Interoperability between archival and bibliographic metadata: an EAD to MODS crosswalk

1. Introduction

The last years, due to the growing number of resources available on the Web, homogeneous access is required to serve users informational needs. However, information sources may differ - even in the same domain or discipline - in various levels, such as system, syntax and in the more complex semantic level. Those diversities provide serious barriers on homogeneous access. To overcome those issues, it is necessary to achieve *interoperability*. Interoperability is succeeded when a set of processes ensures that systems manage their information sources in a way that supports the reuse and exchange of data from various sources inside and outside a particular system; in other words, when it is ensured that information collected by one organization for a particular purpose and possibly encoded using different metadata schemas, can be searched, exchanged, transferred, used and understood by another organization for a different purpose (Hodge, 2005).

Metadata schemas are designed to cover the documentation needs of specific resources and - at the same time - the access needs of particular user communities. Due to those facts, many schemas are defined and used, especially in the collections of *memory institutions* (libraries, archives and museums) where there is an increasing need for documentation of various types of material (e.g. museum objects and collections, archives, electronic material etc). In order to provide homogeneous access for the user to collections that include diverse and multiple metadata descriptions (e.g. digital libraries, repository of harvested metadata etc) *metadata interoperability* becomes an important issue

Two of the most widely implemented metadata schemas for the description of (digital) material in the field of cultural heritage are EAD and MODS. *Encoded Archival Description* (EAD) is the most widely implemented metadata schema for archival description (Library of Congress, 2002a). EAD is a standard for encoding archival finding aids using Extensible Markup Language (XML) (W3C, 2006) so as to make archival resources from many institutions accessible to users. *Metadata Object Description Schema* (MODS) (Library of Congress, 2008a) is an XML bibliographic metadata schema set that may be used for a variety of purposes, and particularly for library applications. It is less complicated than MARCs (such as MARC21, UNIMARC) but at the same time equally powerful, overcoming some shortcomings of MARCs.

In this paper we investigate the problem of semantic interoperability between archival and bibliographic metadata used in digital libraries and repositories. More specifically:

1. We examine the main ideas behind archival description and the EAD metadata standard, in comparison with the bibliographic metadata and the MODS standard. We investigate the semantic relationships between these two metadata standards and discuss the relation between the archival description logic and the documentation logic of a bibliographic metadata schema. We argue that, although EAD is adequate to represent hierarchical structures, this representation is also feasible through MODS.
2. We propose a crosswalk between EAD and MODS consisting of three basic components: a) a semantic mapping from EAD elements/attributes to MODS elements/attributes, b) techniques to map the hierarchical structure of the EAD document to MODS, so as to ensure that the information expressed through the structure of the EAD document is retained in the resulted MODS document, and c) techniques to retain in MODS the information related to the archival components that

is not explicitly assigned to the descriptions of these component in the EAD document, but is inherited from the descriptions of components of higher hierarchy in the EAD document.

Preliminary results of this work have been presented in (Bountouri and Gergatsoulis, 2007).

The paper is structured as follows: In Section 2 we present the basic methods to promote semantic metadata interoperability. In Section 3, we analyze the documentation needs and rules that "penetrate" the archival description, giving also an overview of EAD and MODS. We also discuss the relation of archival metadata to bibliographic metadata used in the library and information science community. In Section 4, we propose an EAD to MODS crosswalk consisting of the semantic mapping between EAD and MODS elements (see Subsection 4.1), the mapping of the EAD structure to MODS (see Subsection 4.2) and a "borrowing" mechanism which aims at achieving completeness of information (see Subsection 4.3). Related research efforts are presented in Section 5 and, finally, conclusions and future research goals are presented in Section 6.

2. Metadata interoperability methods

Managing heterogeneous data is a common issue for organizations that document cultural heritage information, such as archives, libraries and museums. Those organizations dispose and develop various collections with heterogeneous types of material which are described by different metadata schemas, according to their documentation needs. When the case is to manage those data as a homogeneous set of objects - for reasons such as information retrieval and data exchange - metadata interoperability mechanisms must be applied. An analysis of those mechanisms is presented in (Chan, 2005). In this section the most well-known and widely used methods are mentioned:

- *Crosswalks*: A crosswalk defines the *semantic mapping* of the elements of a source metadata schema to the elements of a target metadata schema, in order to semantically translate the description of sources between different metadata. Specific policies and tools have been developed, with the purpose of converting metadata records using crosswalks. As mentioned in (Godby et al., 2004) the main reason that leads to the definition and implementation of crosswalks is the need to locate material in heterogeneous collections. What is more, the last decade crosswalks evolved so as to cope with the need of online information systems to deal with custom-made designed metadata standards developed in reply to the ongoing digital material.
- *Application Profile (AP)*: An Application Profile consists of data elements drawn from one or more metadata schemas, combined together by implementers and used for local applications (Heery and Patel, 2000). An AP is based on *namespaces*, which are used as a stable point of reference to support the definition of particular data element sets or vocabularies.
- *Switching Schema*: In this approach, an existing schema is being used as a switching mechanism among multiple metadata schemas. There is no direct mapping between metadata schemas. Instead each metadata is mapped to the switching model, which is usually general enough in order to incorporate mappings from many and diverse metadata schemas (Godby, 2003). Dublin Core (DCMI, 2008) is often used as a switching schema.
- *Ontology based integration*: Nowadays, ontologies have a vital role in semantic interoperability scenarios, given that they can express semantics in a formal manner. In view of the fact that ontologies can conceptualize a domain, they can be used as an umbrella of terms and meanings expressing the same subject or concept. As said by

(Uschold and Gruninger, 1996; Partridge, 2002), ontologies can fulfill the lack of shared understanding serving as the basis for: a) modeling an information system communication between different communities of users with different needs, b) interoperability between systems that have different modeling methods, languages and software platforms, c) reusability of concepts and relationships among them and reliability by automating the consistency, and d) provision of a common framework across applications for analyzing what entities their data describe.

Crosswalk method is adequate when the case it to define semantic mappings between a limited number of metadata schemas. In a different case, where multiple schemas are involved, defining crosswalks between them requires lot of effort since a growing number of mappings is needed, depended on the number of metadata involved. Using Switching Schema requires only one mapping per metadata schema to the switching mechanism. The weak point of this approach is that due to the generality, simplicity and absence of rich semantics in the switching schema (e.g. unqualified Dublin Core) there is usually loss of information, since many metadata elements cannot be mapped to such a general model or / and the elements of the switching schema have very broad semantics.

APs promote interoperability since their description core is derived from already defined metadata, such as Dublin Core, and in terms of using already known metadata terms from other schemas. An AP is based on one single schema but tailored to different user communities, such as the DC-Library Application Profile that defines the use of the DC metadata schema in library-related applications and projects (Chan and Zeng, 2006). The AP's purpose is to adapt or combine existing schemas into a package that is tailored to the functional requirements of a particular application, while retaining interoperability with the original base schema or schemas. Part of such an adaptation may include the elaboration of

local metadata elements that have importance in a given community or organization, but which are not expected to be usable in a wider documentation context (Duval et al., 2002).

In the ontology based integration approach conceptual representations of the data and of their relationships are used to cope with heterogeneities. Given that ontologies are defined as "*an explicit specification of a conceptualization*" (Gruber, 1993), they can be easily used in a data integration system with the aim of describing and defining the semantics of the data sources and to make their content explicit (Wache et al., 2001). One of the main roles in such a scenario is that they can be used as a *mediated schema*, expressing concepts and the relationships between terms and meanings. When ontologies are implemented as a mediated schema their use is similar to the Switching Schema approach. However, the use of an ontology (usually a domain ontology) instead of a semantically wide metadata schema, such as DC, is preferred when there is a need to integrate various schemas, since ontologies can provide proper alignment of semantics by expressing richer notions and conceptualizing the domains of interest (Bountouri et al., 2008; Kakali et al., 2007; Nussbaumer and Haslhofer, 2007; Stasinopoulou et al., 2007).

3. Encoded Archival Description (EAD)

3.1 Archival Description and EAD

According to International Council on Archives (2000), an *archive* is the whole of the documents, regardless of form or medium, organically created and/or accumulated and used by a particular person, family, or corporate body in the course of the creator's activities and functions. *Archival description* represents an archive (also known as "*fonds*"), which is a complex set of materials that share common provenance. The description involves a complex hierarchical and progressive documentation, beginning with the

description of the whole, and then proceeds to define and describe the sub-components of the whole, the sub-components of sub-components, and so on. This description can even reach the single item level (i.e. an architectural drawing, a map, a photograph).

Encoded Archival Description (EAD) is the most widely used standard for encoding machine readable archival metadata. It is an XML-based descriptive schema, intended to create electronic *finding aids*. Finding aids include the necessary information for the identification, management and interpretation of an archive. EAD has been used so as to provide archival metadata in digital libraries.

The specific standard is based on the philosophy of the *General International Standard Archival Description (ISAD (G))* (International Council on Archives, 2000). ISAD (G) provides general guidance for the preparation of the archival description, which main's scope is to identify and explain the context and content of archival material. For this reason, ISAD (G) proposes seven description areas, with the aim of defining and controlling the structure and the content of a finding aid, and four fundamental rules (known as "*multilevel description rules*") that apply to establish the hierarchy of descriptions. Those rules are listed below:

- *Description from the general to the specific*. In order to represent the context and the hierarchical structure of the archive and its component parts, the resulting descriptions are presented in a hierarchical part-to-whole relationship starting from the broadest (usually the fonds level) to the more specific. In detail, at the fonds level information for the archive as a whole must be provided. At the next and subsequent levels information for the parts being described should be given.
- *Information relevant to the level of description*. For accurately defining the context and content of a description unit, only information appropriate to the level being described must be provided.

- *Linking of descriptions*. A link of each description to its next higher descriptive unit must be given and, if applicable, an identification of the level of description with the intention of explicitly defining the position of the described unit in the hierarchy.
- *Non-repetition of information*. To avoid redundancy of information in hierarchically related archival descriptions, at the highest appropriate level information that is common to the component parts must be present. Any repetition of information at lower levels of description, which has already been given at the higher level, violates this specific rule.

The exploitation/adaptation of the flexible and tree structure based XML encoding language allows EAD to introduce a machine readable form of the archives' multi-level structure, and at the same time, support the inheritance of descriptive information between the hierarchical levels.

3.2 EAD features

An EAD document, starting from the <ead> root element, is composed of three basic elements: a) the *EAD Header* (<eadheader>), which is a mandatory element that includes information about the EAD finding aid (i.e. includes the metadata for the archival description and not the archival description itself), b) the *Front Matter* (<frontmatter>), which is an optional element that contains publication information, such as the title page information of the printed finding aid etc, c) the *Archival Description* (<archdesc>), which is a mandatory element that includes the archival description.

The <archdesc> element is the part of the finding aid that documents the archive itself, providing information about the archive's content and context of creation. In detail, the

subelements of <archdesc> provide the following information for the archival material described:

- *Core identification information*. It is incorporated in the descriptive identification element (<did>), such as the archive's originator (<originator>), its creation date (<unitdate>) etc. The <did> element is mandatory.
- *Administrative and supplemental information*. The specific information group is composed from the use various elements and it facilitates the use of the archival materials. For example, the biography or history (<bioghist>) element that bundles information about the biography/administrative history of the archive's originator, the Conditions Governing Access (<accessrestrict>) element concerning access rules for the material etc.
- *Description of components*. The description of subordinate components is bundled in a wrapper element named “Description of Subordinate Components” (<dsc>), which defines the hierarchical groupings of the materials being described. Having already described the archive as a set within the archival description element, <dsc> contains the hierarchical description of the archival components. As it is mentioned in (Library of Congress, 1999a) an archival component is an easily recognizable archival entity such as series, subseries, file, or item, and it may be in any level within the hierarchical structure of the description. For example, archival series are usually components of archives. Similarly, subseries are parts not only of their parent series but of their “grandparent” unit as well. The description of any single archival component inherits the description of its ancestors. Components are expressed in EAD through nested component elements named either <c> or <c01> to <c012>. In order to provide information for the archival components described inside those

elements, the same elements that define core identification and administrative and supplemental information of the <archdesc> element are used inside the <c> or <c01> to <c012> elements.

3.3 Metadata Object Description Schema (MODS)

Metadata Object Description Schema (MODS) is a bibliographic element set that may be used for a variety of purposes, and particularly for (digital) library applications, hence being currently used in bibliographic catalogues (Library of Congress, 2008c; Copac Academic & National Library Catalogue, 2008). It is a MARC 21 based XML metadata schema, particularly applicable to digital library objects that require explicit descriptions compatible with existing ones in library catalogues. Moreover, MODS is used by current digital library systems, for instance by California Digital Library (University of California, 2005) and supported by retrieval protocols such as SRU (Library of Congress, 2008d). It is simpler than MARC standards and richer than Dublin Core, uses language-based user friendly tags rather than numerical and, although it is less detailed than MARC it is highly compatible with it (Guenther, 2003).

As mentioned in (Music Library Association - Bibliographic Control Committee, 2008), MODS has the following advantages: a) it was developed both for digital objects and MARC format, therefore it is highly compatible with MARC, b) the use of particular cataloging codes is not required, however, it is AACR2 compatible and can accommodate metadata from various sources, and c) it is able to represent constituent parts and express their interrelationships. Additionally, one of the main differences between MARC and MODS is that MODS is recursive, which makes it capable of representing hierarchies (McCallum, 2004). MODS exploit the hierarchical nature of XML to define the structure for related to the primal MODS record resources in a recursive manner.

A MODS record includes twenty top-level elements and many attributes. All of them are optional and they are included under the root elements <mods> (for a single MODS record) or <modsCollection> (for a collection of MODS records).

3.4 EAD and its relation to Bibliographic Metadata

Nowadays, due to the growth of digital repositories and harvested metadata, the description of resources via various metadata schemas and the translation between metadata schemas are frequently observed. Many archival collections - apart from being encoded in EAD documents - are also represented by other metadata schemas, such as bibliographic standards. As part of this phenomenon it is worth discussing the main characteristics of the EAD metadata schema and if those characteristics can be translated into bibliographic metadata schemas.

In this section we mainly analyze the relation of EAD and bibliographic schemas. The mappings are presented in section 4 (semantic mappings between EAD and MODS fields and representation of the hierarchical structure).

As it is mentioned in (Wisser and Roper, 2003), a MARC record can be an excellent companion of archival description. Nevertheless, it is not the adequate solution for documenting finding aids and archives, due to the fact that archival description encompasses several different conceptual levels hierarchically structured, whereas bibliographic description exists on one level (flat structure). While both can describe a variety of materials, they have quite different documentation logic (Imhof, 2008). Thoroughly, EAD encompasses the exact representation of the archival materials and the user can see the material as a whole, and at the same time get an in-depth indication of its arrangement and complexity. On the other hand, a MARC record is less complex and gives a totally horizontal view of the object described. In a MARC record, only one relationships'

level can be documented. For example, in a MARC record the relationship of a monograph to a series can be presented, however, the relationship of the series with its subseries is documented in the MARC record of the series. Moreover, even if typically it is allowed inside a MARC record to include relationships from the general level to lower levels, it is generally implemented in MARCs to define relationships from the lower hierarchical level to the higher level, for example article to journal relationship, due to technical limitations of the record's length (Skvortsov et al., 2005), as defined in ISO2709 (ISO, 1996).

As a result, EAD and bibliographic metadata can supplement but not replace each other. Their co-existence is adequate in various cases. One of the most common cases is the use of EAD to represent the hierarchical relationships (i.e. parent/child relationships), to provide the users with the ability to identify the content and context of collections and to navigate complex hierarchical collections using structural links. On the other hand, bibliographic metadata can be used to present full analytic descriptions of specific (digital) objects, since they provide a satisfactory number of metadata fields for the documentation of various types of material.

At this point it is worth mentioning that MODS overcomes the mentioned MARCs' shortcomings related to represent hierarchical relationships. Thoroughly, MODS is recursive, thereafter it can include multiple hierarchies inside a single MODS record. Furthermore, in a MARC record, due to the limitation of record's length, relationships are defined from the lower level (i.e. article) to the higher level (i.e. journal). MODS has no such limitations, hence it can express the parent/child relationships of EAD.

3.5 EAD to MODS crosswalk: do we need it?

On the whole, there are financial, operational and usability reasons to use bibliographic metadata in order to describe archives. One of the main motives is that library systems -

which implement bibliographic metadata and are regularly used in cultural heritage institutions, such as archives and museums - have been really evolved over the years as an effective documentation means and provide efficient search and retrieval capabilities in a web environment. An indicative motivation is mentioned in (Xu, 1997): "*…For libraries which don't have Z39.50 capabilities, and who need to incorporate records encoded with other metadata schemes into their OPAC databases, a library metadata conversion system is needed as well"*.

MODS can act as a "*crossover metadata format*" for XML applications that use traditional library cataloguing data together with metadata from other origins. MODS purpose is not to express every MARC data element. It mostly filters a record by selecting key elements that can serve a wide variety of metadata needs, retaining at the same time its parallelism to MARC (Music Library Association - Bibliographic Control Committee, 2008).

In addition, it is worth mentioning that EAD, and in general terms, archival description's target is not to provide analytic description for the components described, but mostly to present them as parts of the archive. When there is a need to describe a resource – which may be part of an archive – as part of a separate collection, MODS can act either as a crossover metadata schema, as already mentioned, or as the target schema. For instance, an archive may include single items such as photographs that might also be described as part of another collection (i.e. a digital collection of photographs). It is referable that even in this case, the hierarchical relationships between the archive's components, which are adequately represented by the EAD metadata schema, can be represented in MODS.

In detail, one of the main reasons MODS is chosen as the appropriate bibliographic metadata schema to participate in the specific crosswalk is its recursive ability. In MODS we have both the options to define a hierarchically structured relationship in both directions,

using the <relatedItem> element and its type attribute with its values "constituent" and "host" (see Section 4.2).

Finally, both main entities of an EAD document (the EAD Header and the Archival Description) could be mapped and converted to bibliographic metadata, creating bibliographic records that either document the EAD document itself for the first case either the archive itself or/and its components for the second case. In this paper, we will mainly deal with the Archival Description part of the EAD and its crosswalk to bibliographic metadata, since the specific part incorporates the description of an archive and its components.

4. Proposed crosswalk method

In general terms, there are two important features of EAD that must be translated into another metadata schema in order to accomplish the semantic conformance between them.

a. Semantic mapping the fields of the source schema to target schema's fields. In EAD to MODS crosswalk most of the fields of the first can be mapped to the fields of the second, so as to retain semantic equivalence between them.
b. Mapping the structure of the source metadata schema to the target metadata schema. In the case of EAD, the multi-level hierarchical structure of the description (including the principal of hierarchical inheritance) is one of its features that add semantics to the content, since the description of an archival item is strongly interrelated to the description of the archive itself and its component parts, in view of the fact that they represent the content and context of creation of every archival unit. Those features are vital to define the hierarchical relationship between the components of an archive in MODS.

4.1 Semantic mapping between EAD and MODS

In order to define the semantic mapping from the elements/attributes of EAD to the elements/attributes of MODS the following steps are followed: a) study of the EAD fields (elements and attributes), their semantics and scope notes, according to the EAD Tag Library (Library of Congress, 2002b), b) examination of the MODS fields (elements and attributes), their semantics and scope notes, according to the Detailed Description of MODS Elements (Library of Congress, 2008b), c) definition of the mapping between the EAD fields and their semantic equivalent MODS fields, d) creation of real world examples in order to check the semantic correctness of the mapping.

In Table 1, a fragment of the proposed crosswalk is presented. Due to space reasons, we are presenting a short overview of the crosswalk. For complete reference and analysis of the semantic mappings see (Bountouri and Gergatsoulis, 2008). It is worthy of note that the crosswalk implemented refers to both to the head archival component (<ead>/<archdesc>) and its subordinate components included in <dsc> (such as <ead>/<archdesc>/<dsc>/<c01>, <ead>/<archdesc>/<dsc>/<c01>/<c02> etc), since they implicitly use the same elements for the description of an archival unit. Thus, every EAD path presented in Table 1 can have as a prefix one of the mentioned paths.

INSERT TABLE 1 HERE

Some of the mappings mentioned in Table 1 are analyzed as follows:

In EAD the element <unittitle> encodes the name of the described materials at both the highest unit (<archdesc>, e.g. fonds) and at all the subordinate component levels (<c>, e.g., subseries, file, item). In MODS the element <title> is a word, phrase, character, or group of characters that constitutes the chief title of a resource. In EAD the element <unitdate> is the creation year, month, and/or day of the described materials. The <dateCreated> element of MODS documents the creation date(s) of the material described. In EAD information about the individual responsible for the creation, accumulation, or assembly of the described materials before their incorporation into an archival repository is defined in <origination>/<persname>. In MODS the <name> element includes names associated to the resource. In EAD's path <langmaterial>/<language> the language(s) of the archival materials being described are recorded in coded and/or textual form. In MODS the equivalent data are presented in <language>/<languageTerm> path.

It is considerable – based on the complete reference of the crosswalk (Bountouri and Gergatsoulis, 2008) – that the vast majority of the EAD elements can be semantically mapped to MODS elements. Even if some of them do not have a semantic equivalent field, MODS provides some flexible mechanisms to cover local needs. For instance, it is allowed to implementers to add values to the type attribute inside the <note> element, in order to cover semantics related to the specific element that are not currently included in MODS.

4.2 Mapping the structure

Via MODS it is feasible to thoroughly "carry" the parent/child hierarchical relationship, using the <relatedItem> element and its type attribute with the value "constituent" (for declaring parent to child relationships) or "host" (for declaring child to parent relationships). The specific element provides information to identify other resources related

to the one being described (Library of Congress, 2008b). The <relatedItem type="constituent"> allows to create rich analytics for contained works within a MODS record (Cundiff and Trail, 2007), while the <relatedItem type="host"> allows to define contextual information for works' description. In both cases it is possible to associate descriptive data with the described related resource.

Moreover, it is worth mentioning that via MODS it is possible to also "carry" the inheritance of descriptive information between the hierarchical levels defined in archival description and in particular in EAD, for all the elements of EAD that are semantically mapped to MODS. This is accomplished through the above mentioned use of <relatedItem> element and its type attribute that allows the definition of the parent/child hierarchical relationship; secondly, given that MODS is XML based it adopts the XML characteristics, such as the tree structure that allows the inheritance of information between the nodes and the description they include.

An important issue of policy and system implementation, regardless of the defined attribute type, is whether the <relatedItem> element will include the whole related record(s) or not. MODS also allows the interlinking of related records via the xlink:href attribute of the <relatedItem> element. This attribute is used for an external link, including a link to a related record. Until now, only URI references are allowed via the xlink:href (the MODS schema and implementation guidelines will be revised in order to include more XML linking attributes, such as xlink:role, xlink:title). In addition <identifier> is usually used inside the <relatedItem> for defining a URI for the related resource itself (Library of Congress, 2008b), hence can be used as an alternative of the xlink:href attribute depending on the policy implemented.

To show the two different possibilities consider the following skeleton of an EAD document:

```
1. <ead>
2.    <eadheader>...</eadheader>
3.    <archdesc level="fonds">
4.      <did>
5.        <unitid countrycode="GR" repositorycode="IU">ARC.14</unitid>
6.        <unittitle>Ionian University Archive</unittitle>
7.           ...
8.      </did>
9.       ...
10.     <dsc>
11.       <c01 level="series">
12.         <did>
13.          <unitid countrycode="GR" repositorycode="IU">
   ARC.14/1</unitid>
14.          <unittitle>Research Committee Archives</unittitle>
15.           ...
16.         </did>
17.           ...
18.          <c02 level="item">
19.            <did>
20.               <unittitle>Funding Guidelines - 2006</unittitle>
21.               <unitid countrycode="GR"
   repositorycode="IU">ARC.14/1a</unitid>
22.            </did>
23.              ...
24.          </c02>
25.       </c01>
26.     </dsc>
27.  </archdesc>
28.</ead>
```

The <archdesc> part of the EAD document describes an archive named "Ionian University Archive" and its component (<c01>, line 11-25) is an archival series titled "Research Committee Archives", which includes a single item (<c02>, line 18-24) named "Funding Guidelines - 2006". Based on the first approach of representing, via the <relatedItem> element, the related records of the archival components (in this case the archival series and the single item included), the whole EAD archival description is defined via the nested structure of MODS.

The following MODS skeleton is the result of transforming the above EAD skeleton to MODS using this approach.

```
1.  <mods>
2.    <titleInfo>
3.       <title>Ionian University Archive</title>
4.    </titleInfo>
5.    <identifier>GR-IU-ARC.14</identifier>
6.    <physicalDescription>
7.      ...
```

```
8.       <note type="organization">fonds</note>
9.    </physicalDescription>
10      ...
11.   <relatedItem type="constituent">
12.      <titleInfo>
13.         <title>Research Committee Archives</title>
14.      </titleInfo>
15.      <physicalDescription>
16.          ...
17.         <note type="organization">series</note>
18.      </physicalDescription>
19.      <identifier>GR-IU-ARC.14/1</identifier>
20.      <relatedItem type="constituent">
21.         <titleInfo>
22.            <title>Funding Guidelines - 2006</title>
23.         </titleInfo>
24.         <physicalDescription>
25.            <note type="organization">item</note>
26.         </physicalDescription>
27.         <identifier>GR-IU-ARC.14/1a</identifier>
28.      </relatedItem>
29.   </relatedItem>
30. </mods>
```

The description of the archive itself presented in EAD in lines 3-9 corresponds in the MODS record to lines 2-10. Observe in the above that the <c01> EAD element and its subelements (lines 11-25), representing the archival series, correspond in the MODS record to the <relatedItem type="constituent"> element (lines 11-29), and its <c02> subcomponent (lines 18-24 of the EAD record) correspond to the <relatedItem type="constituent"> element (lines 20-28) of the MODS record, representing the archival item of the series.

More analytically, as part of the semantic mappings, we observe that the attribute level="*value*" in EAD (lines 3, 11, 18) is presented in MODS through the subelement <note> with attribute type="organization" of the element <physicalDescription> (see lines 6-9, 15-18, and 24-26 in MODS). It is important to notice that the <note type="organization"> subelement of <physicalDescription> in MODS denotes the physical arrangement of the material described and it is almost equivalent to the 351 MARC 21 Bibliographic field, which includes "*information about the organization and arrangement of a collection of items*" (Library of

Congress, 1999b). The specific MARC21 field contains a subfield named "$c - Hierarchical level" in which the "*hierarchical position of the described materials relative to other records from the same source*" is defined. The attribute `level` of EAD is mapped to the 351$c hierarchical level field of MARC21 Bibliographic, hence the semantic closeness is obvious.

The alternative method of mapping the structure of the EAD document to MODS promotes the creation of standalone MODS records to represent the archive and each of its components defined in `<dsc>`. According to this approach, in the transformation of the above EAD skeleton, the following MODS record corresponds to the description of the archive (lines 3-9 of the EAD record).

```
f1.  <mods>
f2.    <titleInfo>
f3.       <title>Ionian University Archive</title>
f4.    </titleInfo>
f5.    <identifier>GR-IU-ARC.14</identifier>
f6.    <physicalDescription>
f7.      ...
f8.       <note type="organization">fonds</note>
f9.    </physicalDescription>
f10      ...
f11. <relatedItem xlink:href="URI of the MODS record representing
the archival series" type="constituent"/>
f12. </mods>
```

The following MODS record corresponds to the description of the archival series (lines 11-17 of the EAD record).

```
s1. <mods>
s2.   <titleInfo>
s3.      <title> Research Committee Archives</title>
s4.   </titleInfo>
s5.   <physicalDescription>
s6.       ...
s7.       <note type="organization">series</note>
s8.   </physicalDescription>
s9.   <identifier>GR-IU-ARC.14/1</identifier>
s10.  <relatedItem xlink:href="URI of the MODS record representing
the archival item"  type="constituent"/>
s11.  <relatedItem xlink:href="URI of the MODS record representing
the archive" type="host"/>
s12. </mods>
```

Finally, the following MODS record corresponds to the description of the archival item included in the archival series (lines 18-24 of the EAD record).

```
i1. <mods>
i2.    <titleInfo>
i3.       <title>Funding Guidelines - 2006</title>
i4.    </titleInfo>
i5.    <physicalDescription>
i6.       <note type="organization">item</note>
i7.    </physicalDescription>
i8.    <identifier>GR-IU-ARC.14/1a</identifier>
i9.    <relatedItem xlink:href="URI of the MODS record representing
the archival series" type="host"/>
i10. </mods>
```

All the above records are interconnected through the <relatedItem xlink:href> in order to represent the hierarchical structure of the archival description. The first MODS record corresponds to the archival description part of the EAD record (lines 3-9) and it "calls" through the <relatedItem xlink:href> its constituent parts (line f11 of the MODS record), which in the specific case is the archival series of the EAD record (lines 11-25). The second MODS record via the same mechanism "calls" both its constituent (archival item) and host (archive) parts (lines s10 and s11 of the MODS record), and the third MODS record "calls" its host (archival series) part (line i9). In the specific approach even if the hierarchical structure is not presented identically to the structure in EAD record, and so it is implied through the linking mechanism.

It is clearly an issue of policy, in terms of implementation, which of the two options to follow. Creating independent records is useful in the following cases. Firstly, a standalone record is created once, however, it can be referenced to, if necessary, by other records in various hierarchical levels. What is more, defining standalone records allows its use in other relation scenarios, not necessarily concerning parent/child relationships, and facilitates its use from other records, coming from various metadata schemas. It is important to notice that updating once a standalone record, functions as if this record has been updated in every other record it is included due to the linking mechanism.

Creating records that include nested MODS records inside the `<relatedItem>` is adequate when there is a need to provide users with a complete representation of the resources described. On the other hand, when there is a need to describe a single archival unit, i.e. a photograph, and provide some contextual information about the set it belongs to, i.e. a collection of photographs, it might be more reasonable use the standalone approach and to apply the `<relatedItem type="host">` to relate this record with the record representing the corresponding collection.

It is worthy to note that in case of exchanging records not including information for the related resources might result in providing records with incomplete information. It depends on the system implemented and the options it offers for exporting records to include the necessary information while exporting. This issue is discussed further in the next section.

In order to prove that it is an issue of policy which option to follow, we report part of the Digital Library Federation's Aquifer Implementation Guidelines for Shareable MODS Records (DLF, 2006; Indiana University Digital Library, 2007): "*aggregators, including the Aquifer initiative, are unlikely to follow large numbers of links to related records or parse deeply recursive MODS records effectively. Instead, they are likely to assume one shared record should result in one record presented in a search result set*". Based on that fact the DLF Aquifer MODS Guidelines recommend the use of `<relatedItem>` in specific cases, such as pointing to a full metadata record for a related item record via an `xlink:href` attribute, or including a MODS record inside the `<relatedItem>` when there is a need to provide contextual information useful for full description of the resource or additional information about intellectual constituent units of the resource being described.

A complete reference of the example presented above can be found in the Appendix.

4.3 Borrowing missing nodes based on the inheritance of information

As we have mentioned in earlier sections of this paper, EAD supports the inheritance of information between the hierarchical levels corresponding to the described archival components. In case the inheritance property of EAD is not taken into account during the process of transforming an EAD document into MODS, considerable information may be lost. To cope with that issue we propose two different approaches:

1) **Assuming inheritance of information in the resulted MODS records**. This approach is based on the idea that, due to the hierarchical structure of a MODS record, we may assume that the MODS elements, which correspond (through the proposed crosswalk) to the EAD elements that have the inheritance property, also have the property to be inherited into their descendant MODS elements corresponding to the components of the archive. In other words, target MODS records (in case we choose to represent the archival components as standalone MODS records), or the components of a single MODS record that represent the archive's components (in case we choose the approach where the EAD document is mapped to a single MODS record with nested structure), can be informational complete by "borrowing" the elements missing from its ancestor components. It is worthy of note that, in case of multiple MODS records obtained by transforming an EAD document, the inheritance of information also applies between the distinct MODS records that are connected through the `<relatedItem xlink:href="..." type="constituent">` node. In particular, the missing information of a ("constituent") MODS record can be obtained from its "host" record. Thus, the information is inherited from one descriptive level (e.g. the MODS record representing the archive, with `<note type="organization">`fonds`</note>`) to its included components (e.g. a

<related Item type="constituent"> with <note type="organization">series</note> representing the archival series). It is obvious that, in this approach, it is responsibility of the users of the produced MODS records to take into account the inheritance property when extracting the information they need from those records.

2) **Constructing self-contained MODS records with respect to their information content**. This approach is based on the idea to construct MODS records (by applying either of the two approaches presented in Section 4.2), which are self-contained with respect to their information content. In this case, the information needed in a MODS record representing an archival component should be placed in that record during its construction face. Technically, this construction can be done through one of the following techniques:

   i. Before the conversion phase, the EAD document passes through a pre-processing phase. During that phase, an intermediate EAD document is produced. In this document each of its archival components is self-contained in the sense that it contains all information which is available in the initial EAD document and relates it (directly or indirectly) to that component. In detail, this document is obtained by propagating all elements which have the property of being inherited and that belong to the description of an archival component, to the descriptions of its subcomponents in which the information expressed by these elements is missing. Then, the intermediate EAD document is used to construct the equivalent MODS document (using one of the approaches proposed in Section 4.2). The resulted MODS document incorporates (by construction) self-contained parts that describe the archival components.

ii. During the conversion phase, when a MODS record (or part of a MODS record) corresponding to an archival component is constructed, each time an element is missing from the part of the EAD document corresponding to that component, a "borrowing" mechanism is activated. This mechanism searches for the missing elements in the parent component of the EAD component. In case the element is not present, it continues to its grand-parent component etc. When the mechanism finds the specific element, it passes it to the transformation procedure which treats the element as if they were present in the EAD component being transformed.

Notice that, regardless of which of the two techniques is applied, in this case the produced MODS records are self-contained and there is no need to employ inheritance in MODS to gather information referring to the archival components that they describe.

Some interesting questions arise in relation to the notion of inheritance in archival description: a) is it reasonable to assume that all EAD elements related to an archival component can be inherited to the description of the subcomponents of this component?, b) for those elements that we can assume that can be inherited, are there any constraints that should be fulfilled in order for this inheritance to be meaningful?

We consider that, although these questions are interesting, they are beyond the scope of this paper. We assume that when the transformation from an EAD document to the corresponding MODS records takes places, it is known to the system which EAD elements can/should be inherited.

5. Related work

In Indiana University Digital Library (Billington, 2007) MODS is used to describe image units and collections. Nevertheless, part of this material has its "*master*" metadata in EAD documents. Due to that fact, they convert metadata descriptions from EAD to MODS using crosswalks and adequate mechanisms. Since EAD is XML based, they are using custom XSL document to convert EAD metadata to MODS. The XSL file is applied in specific component levels (e.g. `<c02 level="item">`) and - while converting - it inserts into the MODS record default information, not included in the EAD description of the component. For example, since it is known that the data transformed to MODS are images, the XSL inserts role terms in the creator element (such as "Photographer") and the type of the resource ("still image"). With the aim of converting different types of resources included in an EAD document to MODS, they will have to create separate XSLs for each case, since the XSL they are using is oriented to create MODS metadata only for still images. To support the conversion, they also define basic metadata crosswalks between their existing systems and the new metadata schemas they have developed. In the future, they are planning to design recommended EAD practices to minimize the amount of XSL that has to be written.

In (California Digital Library, 2005) appropriate mappings for several metadata schemas that might be exposed to harvesting are suggested, in addition to the required unqualified DC. Thus, they present a mapping table where MARC21, EAD and MODS are mapped to Unqualified DC and to Generic data fields (fields defined as mandatory). There is no direct mapping from EAD to MODS, although some assumptions can be made, given their simultaneous mapping to DC.

In Brown University (Caldwell, 2005) all the digitized objects of their collections reside in a central database, which in their case uses MODS for describing, cross-searching and retrieving their objects. When they digitize objects documented in EAD they convert the specific descriptions to MODS records.

The pre-mentioned crosswalk approaches deal with local systems' needs. In our effort we try to define a generic crosswalk that could be used in order to cover as much as possible mappings between the two metadata schemas.

A wider effort on metadata crosswalks is presented in (Godby et al., 2004) proposing the creation of a METS (Library of Congress, 2008e) encoded crosswalks repository which could be exploited for searching and harvesting. The proposed EAD to MODS semantic mappings could be part of this repository if encoded according to the specifications proposed in this research attempt. Furthermore, translation services between metadata schemas using crosswalks as part of their infrastructure have been developed, such as the OCLC Crosswalk Web Service (Godby et al., 2008).

Finally, it is worth mentioning that related conversions to bibliographic standards have been implemented with metadata schemas coming from various domains, such as the educational metadata (Kimberly and Ridgway, 2003).

6. Conclusion

In this paper we define semantically accurate crosswalks between archival and bibliographic metadata. We conclude that various parameters must be taken into consideration. In detail, even if a generic crosswalk exists, it is hard enough - while converting - to define disambiguation between the elements of two metadata schemas, given that metadata implementers may often use particular fields with similar but not exactly the same meaning

in order to cope with local description needs. Defining best practices and guidelines based on international and widely implemented cataloguing rules for local description needs could facilitate the implementation of crosswalks as part of a conversion mechanism and promote EAD's interoperability abilities in a wider context. Concerning the implementation of the proposed crosswalk, we are currently working on implementing it through the use of XSL Transformations (XSLT) (W3C, 1999), which is a language for transforming XML documents into other XML documents.

Future work will include an extra characterization of the semantic mappings. In detail, some MODS elements may have broader or narrower semantics in comparison to the EAD elements, such as the `<physicalDescription><note type="organization">`, which is semantically wider than the attribute it is mapped in EAD (`level`). A related issue is the use of the `<relateditem type="constituent">` to represent the hierarchical and inheritance relationships of the archival components. Nevertheless, the specific element-attribute pair is not strictly oriented to define parent/child inheritance relationships, i.e. via the specific element-attribute pair other constituent relationships might be defined, for example a highly structured and rich table of contents.

The use of a mapping language, such as the SKOS Mapping Vocabulary (W3C, 2004), could be employed to define the different kinds of the semantic relationships.

Finally, it is worthy to mention that the proposed crosswalk does not disregard the documentation power of EAD for describing archives. EAD is the only digital libraries metadata standard that maintains the theoretical documentation background of archivists, such as the principle of provenance. However, when there are specific needs our method promotes the creation of semantically equivalent bibliographic metadata.

Acknowledgements

We would like to thank Professor Christos Papatheodorou for his helpful comments in regard to the specific paper. We would also like to acknowledge Michalis Sfakakis and Manolis Peponakis for the knowledge they provided us in relation to the use of bibliographic standards.

References

Billington, L., Moir, S., Wilson, K., and Carpenter, S. (2007). Development of a metadata application profile at the State Library of New South Wales. In *Proceedings of the 13th Information Online Conference, Australian Library and Information Association, Sydney, Australia*. Retrieved December 13, 2008, from http://epubs.scu.edu.au/cgi/viewcontent.cgi?article=1003&context=lib_pubs

Bountouri, L., and Gergatsoulis, M. (2007). Interoperability between archival and bibliographic metadata. In *Proceedings of the First International Workshop on Cultural Heritage on the Semantic Web (in conjunction with the 6th International Semantic Web Conference and the 2nd Asian Semantic Web Conference), 12 November, Busan, Korea, 2007* (pp. 107-108). Retrieved December 13, 2008, from http://www.cs.vu.nl/~laroyo/CH-SW/ISWC-wp9-proceedings.pdf

Bountouri, L., and Gergatsoulis, M. (2008). EAD to MODS crosswalk: Version 1.0. Retrieved December 13, 2008, from http://www.ionio.gr/~boudouri/EAD2MODS/EAD_to_MODS_v1.pdf

Bountouri, L., Papatheodorou, C., Soulikias, V., and Stratis M. (2008). Metadata interoperability in public sector information. (*Accepted for publication by the Journal of Information Science*)

Caldwell, A. (2005). MODS at Brown University. Presented in ALA's meeting: "MODS, MARC, and Metadata Interoperability". Retrieved December 13, 2008, from http://dl.lib.brown.edu/staff/caldwell/MODSatBrown.ppt

California Digital Library (2005). American West project: Baseline metadata requirements. Retrieved December 13, 2008, from http://cdlib-s10.cdlib.org/inside/projects/amwest/AmWest_baseline_metadata.pdf

Chan, L.M. (2005). Metadata interoperability: a methodological study. *Proceedings of the Third China-U.S. Library Conference, Shanghai, China, 23-25 March*, (pp. 110-115).

Chan, L.M. and Zeng, M.L. (2006). Metadata Interoperability and Standardization – A Study of Methodology Part I Achieving Interoperability at the Schema Level. *D-Lib*, 12(6). Retrieved October 31, 2008, from http://www.dlib.org/dlib/june06/chan/06chan.html

Copac Academic & National Library Catalogue (2008). Copac Academic & National Library Catalogue: SRU interface. Retrieved December 13, 2008, from http://copac.ac.uk/interfaces/srw/

Cundiff, M. and Trail, N. (2007). Using METS and MODS to Create XML Standards-based Digital Library Applications. Presented on June 25, 2007, as part of the session "Using Metadata Standards in Digital Libraries: implementing METS, MODS, PREMIS and MIX" in American Library Association Annual Conference 2007. Retrieved December 13, 2008, from http://www.loc.gov/standards/mods/presentations/mets-mods-morgan-ala07/

DCMI (2008). Dublin Core Metadata Element Set, Version 1.1. Retrieved December 13, 2008, from http://dublincore.org/documents/dces/

DLF (2006). Digital Library Federation: Aquifer Implementation Guidelines for Shareable MODS Records, Version 1.0. Retrieved December 13, 2008, from http://www.diglib.org/aquifer/dlfmodsimplementationguidelines_finalnov2006.pdf

Duval, E., Hodgins, W., Sutton, S., and Weibel, S.L. (2002). Metadata Principles and Practicalities. *D-Lib*, 8(4). Retrieved December 13, 2008, from http://www.dlib.org/dlib/april02/weibel/04weibel.html

Godby, C.J., Smith, D., and Childress E. (2003). Two paths to interoperable metadata. In *Proceedings of the International Conference on Dublin Core and Metadata Applications (DC 2003)*, Seattle, Washington, 28 September - 2 October, 2003. Retrieved December 13, 2008, from http://www.oclc.org/research/publications/archive/2003/godby-dc2003.pdf

Godby, C.J., Smith, D., and Childress, E. (2008). Toward element-level interoperability in bibliographic metadata. *The Code4Lib Journal*, 2. Retrieved December 13, 2008, from http://journal.code4lib.org/articles/54

Godby, C.J., Young, J.A., and Childress, E. (2004). A repository of metadata crosswalks. *D-Lib Magazine*, 10(12).

Gruber, T. (1993). A translation approach to portable ontology specifications. *Knowledge Acquisition*, 5, 199-220.

Guenther, R.S. (2003). MODS: The Metadata Object Description Schema. *Libraries and the Academy*, 3(1), 137-150.

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

Hodge, G. (2005). Metadata for electronic information resources: from variety to interoperability. *Information Services and Use*, 25(1), 35-45.

Imhof, A. (2008). Using international standards to develop a union catalogue for archives in Germany: aspects to consider regarding interoperability between libraries and archives. *D-*

Lib Magazine, 14(9-10). Retrieved December 13, 2008, from http://www.dlib.org/dlib/september08/imhof/09imhof.html

Indiana University Digital Library (2007). FAQ for MODS Guidelines Implementers: American Social History Online, Confluence. Retrieved December 13, 2008, from http://wiki.dlib.indiana.edu/confluence/display/DLFAquifer/FAQ+for+MODS+Guidelines+Implementers

International Council on Archives (2000). *ISAD(G): General International Standard Archival Description*. Ottawa: International Council on Archives.

ISO (1996). ISO 2709:1996: Information and documentation -- Format for information exchange.

Kakali, C., Lourdi, I., Stasinopoulou, T., Bountouri, L., Papatheodorou, C., Doerr, M., and Gergatsoulis, M. (2007). Integrating Dublin Core metadata for cultural heritage collections using ontologies. In *Proceedings of the International Conference on Dublin Core and Metadata Applications 2007, August 27-31, Singapore* (pp. 128-139). Retrieved December 13, 2008, from http://www.dcmipubs.org/ojs/index.php/pubs/article/view/16/11/

Kimberly S. L. and Ridgway, J.S. (2003). Generation of XML records across multiple metadata standards. *D-Lib Magazine*, 9(9). Retrieved December 13, 2008, from http://www.dlib.org/dlib/september03/lightle/09lightle.html

Library of Congress (1999a). EAD application guidelines for version 1.0: Creating finding aids in EAD. Retrieved December 13, 2008, from http://www.loc.gov/ead/ag/agcre4.html

Library of Congress (1999b). MARC 21 Format for Bibliographic Data: 351 Organization and Arrangement of Materials. Retrieved December 13, 2008, from http://www.loc.gov/marc/bibliographic/bd351.html

Library of Congress (2002a). EAD version 2002 official site. Retrieved December 13, 2008, from http://www.loc.gov/ead/index.html

Library of Congress (2002b). Encoded Archival Description tag library version 2002 - EAD technical document No. 2. Retrieved December 13, 2008, from http://www.loc.gov/ead/tglib/index.html

Library of Congress (2008a). Metadata Object Description Schema: MODS. Retrieved December 13, 2008, from http://www.loc.gov/standards/mods/

Library of Congress (2008b). MODS user guidelines version 3: Detailed description of MODS elements. Retrieved December 13, 2008, from http://www.loc.gov/standards/mods/v3/mods-userguide-elements.html

Library of Congress (2008c). MODS Implementation Registry. Retrieved December 13, 2008, from http://www.loc.gov/standards/mods/registry.php

Library of Congress (2008d). SRU (Search/Retrieve via URL). Retrieved December 13, 2008, from http://www.loc.gov/standards/sru/

Library of Congress (2008e). Metadata Encoding and Transmission Standard. Retrieved December 13, 2008, from http://www.loc.gov/standards/mets/

McCallum, S.H. (2004). An introduction to the Metadata Object Description Schema (MODS). *Library Hi Tech*, 22(1), 82 – 88.

Music Library Association - Bibliographic Control Committee (2008). Metadata Working Group: Final Report. Retrieved December 13, 2008, from http://www.musiclibraryassoc.org/BCC/BCC-Historical/BCC2008/BCC2008MSWG1.pdf

Nussbaumer, P., and Haslhofer, B (2007). CIDOC CRM in Action – Experiences and Challenges. In L. Kovacs, N. Fuhr, and C. Meghini (eds.) *Proceedings of the ECDL 2007*, *Lecture Notes in Computer Science, Vol. 4675* (pp. 532–533). Heidelberg: Springer.

Partridge, C. (2002). The Role of Ontology in Integrating Semantically Heterogeneous Databases. Technical report, LADSER-CNR.

Skvortsov, V., Zhlobinskaya, O., and Pashkova, A. (2005). UNIMARC XML Slim Schema: living in new environment. In *Proceedings of the World Library and Information Congress: 71th IFLA General Conference and Council, "Libraries - A voyage of discovery", August 14th - 18th 2005, Oslo, Norway.* Retrieved December 13, 2008, from http://www.ifla.org/IV/ifla71/papers/064e-Skvortsov.pdf

Stasinopoulou, T., Bountouri, L., Kakali, C., Lourdi, Irene, Papatheodorou, C., Doerr, M., and Gergatsoulis, M. (2007). Ontology-based metadata integration in the cultural heritage domain. In Dion Hoe-Lian Goh and Tru Hoang Cao and Ingeborg Torvik Solvberg and Edie Rasmussen (eds.), *Proceedings of the Asian Digital Libraries - Looking Back 10 Years and Forging New Frontiers. 10th International Conference on Asian Digital Libraries, ICADL 2007, Lecture Notes in Computer Science, Vol. 4822* (pp. 165-175). Heidelberg: Springer.

University of California (2005). Metadata and encoding. Retrieved December 13, 2008, from http://www.cdlib.org/inside/diglib/resources/onres_dig_metadata.html

Uschold, M. and Gruninger, M. (1996). Ontologies: principles, methods and applications. *Knowledge Engineering Review*, 11(2), 93-155.

W3C (1999). XSL Transformations (XSLT): Version 1.0. Retrieved December 13, 2008, from http://www.w3.org/TR/xslt

W3C (2004). SKOS Mapping Vocabulary Specification. Retrieved December 13, 2008, from http://www.w3.org/2004/02/skos/mapping/spec/

W3C (2006). Extensible Markup Language (XML) 1.0 (Fourth Edition). Retrieved December 13, 2008, from http://www.w3.org/TR/xml/

Wache, H., Vogele, T., Visser, U., Stuckenscmidt, H., Schuster, G., Neumann, H. and Hubner, S. (2001). Ontology-based Integration of Information: a Survey of Existing Approaches. *Proceedings of the International Joint Conference on Artificial Intelligence, Workshop on Ontology Learning*. Retrieved December 13, 2008, from

Wisser, K.M., and Roper, J.O. (2003). Maximizing metadata: exploring the EAD-MARC relationship. *Library Resources and Technical Services*, 47(2), 71-76.

Xu, A. (1997). Metadata Conversion and the Library OPAC. *The Serials Librarian*, 33(1-2), 179-198.

Appendix A

A more elaborated example showing the proposed techniques is presented in this Appendix. The EAD document (Section A1), which has the same structure as the EAD skeleton used in Section 4.2, describes the archives of the Ionian University. Basic identification information for the archive is given, such as its creator (<origination>), the languages it is expressed (<langmaterial>), its creation dates (<unitdate>), its size (<physdesc>/<extent>) etc. Additionally, supplemental information is provided, for example the administrative history of the originator (<bioghist>), access points to the archive (<controlaccess>) and description of the archival subordinate components (<dsc>). One of the basic components of the Ionian University Archive is the Research Committee Archives (archival series represented in the <c01> element), which in turn includes sub-components, such as the Funding Guidelines - 2006 (archival item represented in the <c02> element).

Some of the semantic mappings defined in the crosswalk are the following:

- The titles of the archival components (<did>/<unittitle>) are mapped in MODS to the <titleInfo>/<title> element, due to the fact that both elements represent the main title of the resource described.
- The creators of the archival components (<did>/<origination>/<corpname>) are mapped in MODS to the <name type="corporate">/<namePart> element, given that the specific MODS element represents the name of a person, organization, or event (conference, meeting, etc.) associated with the resource.

Since in the specific EAD document there are three archival components described (<archdesc level="fonds">, <c01 level="series"> and <c02 level="item">), based on the approach of creating a standalone MODS record for every component of the archival description (see Section 4.2), three MODS records are created. These records are related through the <relatedItem type="constituent" xlink:href="..."/> and the <relatedItem type="host" xlink:href="..."/> elements. This transformation is presented in Subsection A.2.

In Section A3, we present the transformation of the EAD document of Section A.1 by employing the approach presented in Section 4.2 in which only one MODS record is created for each EAD document. This specific MODS document contains, in the form of nested MODS records enclosed in the element <relatedItem type="constituent">, the records that correspond to the components of the archive.

It is important to notice that, in both techniques the structure of the archive (encoded through the structure of the EAD document) is fully retained, In the first approach this is defined explicitly through the <relatedItem type="host" xlink:href="..."/> and <relatedItem type="constituent" xlink:href="..."/> elements of the standalone records, while in the second approach this is achieved through the nested structure (element/subelement relation) of the MODS record.

Section A1: the EAD document

In this section we present an EAD document, which describes the Archive of the Ionian University. The specific archival description includes as a sub-component of the archive an archival series, titled "Research Committee Archives". The series consists of several archival

items. In the specific example the item titled “Funding Guidelines - 2006” is presented. The EAD document is as follows:

```
<ead>
  <eadheader>
    <eadid countrycode="GR" mainagencycode="IU">…</eadid>
    <filedesc>…</filedesc>
  </eadheader>
  <archdesc level="fonds">
    <did>
      <unitid countrycode="GR" repositorycode="IU">ARC.14</unitid>
      <unittitle>Ionian University Archive</unittitle>
      <unitdate normal="1984/2007" type="inclusive">1984 - 2007</unitdate>
      <langmaterial>
        <language langcode="gre">Greek</language>
        <language langcode="eng">English</language>
      </langmaterial>
      <origination>
        <corpname>Ionian University</corpname>
      </origination>
      <physdesc>
        <extent>400 files (6 m)</extent>
      </physdesc>
      <repository>
        <corpname>Ionian University Archives Department</corpname>
      </repository>
    </did>
    <bioghist>
    <p>The Ionian University was founded in 1984, the same year as the University of Thessaly and the University of the  Aegean and is located in Corfu. The University consists of the following six Departments (year of establishment in parentheses): Department of History (1985), Department of Foreign Languages, Translation and Interpreting (1986), Department of Music Studies (1992), Department of Archives and Library Science (1993),Department of Computer Science (2004), Department of Audio and Visual Arts (2004)...</p>
    </bioghist>
    <accessrestrict>
     <p>Approval of the user's application by the director of the archive is required for access.</p>
    </accessrestrict>
    <userestrict>
      <p>According to the rules set by the Ionian University Archives Department.</p>
    </userestrict>
    <arrangement>
      <p>The archive was classified thematically.</p>
    </arrangement>
    <acqinfo>
      <p>The archive was collected by the Ionian University Archives Department.</p>
    </acqinfo>
    <controlaccess>
      <controlaccess>
        <subject>University archives</subject>
        <subject>History-Corfu</subject>
      </controlaccess>
      <controlaccess>
        <corpname>Ionian University</corpname>
        <corpname>Department of History</corpname>
        <corpname>Department of Foreign Languages, Translation and Interpreting</corpname>
        <corpname>Department of Music Studies</corpname>
        <corpname>Department of Archives and Library Science</corpname>
        <corpname>Department of Computer Science</corpname>
        <corpname>Department of Audio and Visual Arts</corpname>
      </controlaccess>
```

```
    </controlaccess>
    <dsc>
      <c01 level="series">
        <did>
           <unitid countrycode="GR" repositorycode="IU">ARC.14/1</unitid>
           <container>14.1</container>
           <unittitle>Research Committee Archives</unittitle>
           <unitdate normal="1998/2007">1998 - 2007</unitdate>
           <originator>Research Committee</unitdate>
        </did>
        <bioghist>
       <p>The Special Account for Research Grants (S.A.R.G.) of the Ionian University was established in
1988 in order to meet the need to manage research grants from various sources that are to be invested in
research, educational, training, and developmental projects, as well as other related services or activities
which contribute to the establishment of linkages between education, research and the production
market.</p>
        </bioghist>
        <controlaccess>
          <controlaccess>
          <subject>Research projects</subject>
          <subject>Educational programmes</subject>
        </controlaccess>
        </controlaccess>
          <c02 level="item">
            <did>
             <unittitle>Funding Guidelines - 2006</unittitle>
             <container>ARC.14/1 (1)</container>
             <unitdate label="Project proposal date">01/09/2006</unitdate>
             <daogrp>
               <daoloc href="http://www.ionio.gr/rc/download.php?file=ee
funding2006.pdf"/>
             </daogrp>
           </did>
           <controlaccess>
             <controlaccess>
               <subject>Research Funding</subject>
             </controlaccess>
           </controlaccess>
          </c02>
          . . . Description of other components at the item level . . .
      </c01>
    </dsc>
  </archdesc>
</ead>
```

Section A2: Creating standalone MODS records for each EAD component

In this section we present the transformation of the EAD document of Section A.1 following the approach presented in Section 4.2 in which for every component of the archival description a standalone MODS record is created.

Derived from this specific approach, we notice that the MODS documents created for the description of the archive and its two subordinate components (<archdesc>, <c01 level="series"> and <c02 level="item">, presented in Subsections A2.1,

A2.2 and A2.3 respectively, include the <relatedItem type="host" xlink:href="..."/> and <relatedItem type="constituent" xlink:href="..."/>, which indicate the relationship between the components. For example, in the MODS record for the archive presented in Section A2.1 the <relatedItem type="constituent" xlink:href="..."/> links to the MODS record that represents the child component of the archive, the archival series.

Section A2.1: the MODS record for the archival description

In this section we present the standalone MODS record corresponding to the archival description for the fonds level, including links to its component (the archival series) in order to represent the provenance hierarchy.

```
<mods>
  <titleInfo>
    <title>Ionian University Archive</title>
  </titleInfo>
  <originInfo>
    <dateCreated>1984 - 2007</dateCreated>
  </originInfo>
  <identifier>GR-IU-ARC.14</identifier>
  <language>
    <languageTerm>Greek</languageTerm>
    <languageTerm>English</languageTerm>
  </language>
  <name type="corporate">
    <namePart>Ionian University</namePart>
  </name>
  <physicalDescription>
    <extent>400 files (6 m)</extent>
    <note type="organization">fonds</note>
   </physicalDescription>
   <location>
     <physicalLocation>Ionian University Archives Department</physicalLocation>
   </location>
   <note type="biographical">The Ionian University was founded in 1984, the same year as the University of Thessaly and the University of the Aegean and is located in Corfu. The University consists of the following six Departments (year of establishment in brackets): Department of History (1985), Department of Foreign Languages, Translation and Interpreting (1986), Department of Music Studies (1992), Department of Archives and Library Science (1993), Department of Computer Science (2004), Department of Audio and Visual Arts (2004)... </note>
   <note type="organization">The archive was classified thematically.</note>
   <note type="acquisition">The archive was collected by the Ionian University Archives Department.</note>
   <accessCondition type="restrictionOnAccess">Approval of the user's application by the director of the archive is required for access.</accessCondition>
   <accessCondition type="useAndReproduction">According to the rules set by the Ionian University Archives Department.</accessCondition>
```

```
   <subject>
     <topic>University archives</topic>
     <topic>History-Corfu</topic>
     <name type="corporate">
       <namePart>Ionian University</namePart>
     </name>
     <name type="corporate">
       <namePart>Department of History</namePart>
     </name>
     <name type="corporate">
       <namePart>Department of Foreign Languages, Translation and Interpreting</namePart>
     </name>
     <name type="corporate">
       <namePart>Department of Music Studies</namePart>
     </name>
     <name type="corporate">
       <namePart>Department of Archives and Library Science</namePart>
     </name>
     <name type="corporate">
       <namePart>Department of Computer Science</namePart>
     </name>
     <name type="corporate">
       <namePart>Department of Audio and Visual Arts</namePart>
     </name>
   </subject>
   <relatedItem type="constituent" xlink:href="URI of the MODS record
representing the archival series"/>
</mods>
```

Section A2.2: MODS record for the <c01 level="series">

In this section we present the MODS record corresponding to the archival series "Research Committee Archives", including links to its sub-component (the archival item) and to its parent component (the archive) in order to represent the provenance hierarchy.

```
<mods>
  <titleInfo>
    <title>Research Committee Archives</title>
  </titleInfo>
  <originInfo>
    <dateCreated>1998 - 2007</dateCreated>
  </originInfo>
  <identifier>GR-IU-ARC.14/1</identifier>
  <physicalDescription>
    <note type="organization">series</note>
  </physicalDescription>
  <note type="biographical">The Special Account for Research Grants (S.A.R.G.) of the Ionian
University was established in 1988 in order to meet the need to manage research grants from various sources
that are to be invested in research, educational, training, and developmental projects, as well as other related
services or activities which contribute to the establishment of linkages between education, research and the
production market.</note>
  <subject>
    <topic>Research projects</topic>
    <topic>Educational programmes</topic>
  </subject>
  <relatedItem type="constituent" xlink:href="URI of the MODS record
representing archival item"/>
  <relatedItem type="host" xlink:href="URI of the MODS record representing
the archive"/>
```

```
</mods>
```

Section A2.3: MODS record for the <c02 level="item">

In this section we present the MODS record corresponding to the archival item "Funding Guidelines - 2006", including links to its parent component (the archival series) in order to represent the provenance hierarchy.

```
<mods>
  <titleInfo>
    <title>Funding Guidelines - 2006</title>
  </titleInfo>
  <originInfo>
    <dateCreated>01/09/2006</dateCreated>
  </originInfo>
  <subject>
    <topic>Research Funding</topic>
  </subject>
  <location>
    <url>http://www.ionio.gr/rc/download.php?file=ee funding2006.pdf</url>
  </location>
  <physicalDescription>
    <note type="organization">item</note>
  </physicalDescription>
  <relatedItem type="host" xlink:href="URI of the MODS record representing
the archival series"/>
</mods>
```

Section A3: MODS record for the archive

In this section, we present the transformation of the EAD document of Section A.1 by employing the approach presented in Section 4.2 in which only one MODS record is created for the EAD document. This specific MODS document contains, in the form of nested MODS records enclosed in the element `<relatedItem type="constituent">`, the records that correspond to the components of the archive (series, and item).

```
<mods>
  <titleInfo>
    <title>Ionian University Archive</title>
  </titleInfo>
  <originInfo>
    <dateCreated>1984 - 2007</dateCreated>
  </originInfo>
  <identifier>GR-IU-ARC.14</identifier>
  <language>
    <languageTerm>Greek</languageTerm>
    <languageTerm>English</languageTerm>
  </language>
  <name type="corporate">
    <namePart>Ionian University</namePart>
```

```
  </name>
  <physicalDescription>
    <extent>400 files (6 m)</extent>
    <note type="organization">fonds</note>
  </physicalDescription>
  <location>
    <physicalLocation>Ionian University Archives Department</physicalLocation>
  </location>
  <note type="biographical">The Ionian University was founded in 1984, the same year as the University of Thessaly and the University of the Aegean and is located in Corfu. The University consists of the following six Departments (year of establishment in brackets): Department of History (1985), Department of Foreign Languages, Translation and Interpreting (1986), Department of Music Studies (1992), Department of Archives and Library Science (1993), Department of Computer Science (2004), Department of Audio and Visual Arts (2004)...</note>
  <note type="organization">The archive was classified thematically.</note>
  <note type="acquisition">The archive was collected by the Ionian University Archives Department.</note>
  <accessCondition type="restrictionOnAccess">Approval of the user's application by the director of the archive is required for access.</accessCondition>
  <accessCondition type="useAndReproduction">According to the rules set by the Ionian University Archives Department.</accessCondition>
  <subject>
    <topic>University archives</topic>
    <topic>History-Corfu</topic>
    <name type="corporate">
      <namePart>Ionian University</namePart>
    </name>
    <name type="corporate">
      <namePart>Department of History</namePart>
    </name>
    <name type="corporate">
      <namePart>Department of Foreign Languages, Translation and Interpreting</namePart>
    </name>
    <name type="corporate">
      <namePart>Department of Music Studies</namePart>
    </name>
    <name type="corporate">
      <namePart>Department of Archives and Library Science</namePart>
    </name>
    <name type="corporate">
      <namePart>Department of Computer Science</namePart>
    </name>
    <name type="corporate">
      <namePart>Department of Audio and Visual Arts</namePart>
    </name>
  </subject>
  <relatedItem type="constituent">
    <titleInfo>
      <title>Research Committee Archives</title>
    </titleInfo>
    <identifier>GR-IU-ARC.14/1</identifier>
    <originInfo>
      <dateCreated>1998 - 2007</dateCreated>
    </originInfo>
    <physicalDescription>
      <note type="organization">series</note>
    </physicalDescription>
    <note type="biographical">The Special Account for Research Grants (S.A.R.G.) of the Ionian University was established in 1988 in order to meet the need to manage research grants from various sources that are to be invested in research, educational, training, and developmental projects, as well as other related services or activities which contribute to the establishment of linkages between education, research and the production market.</note>
```

```
    <subject>
      <topic>Research projects</topic>
      <topic>Educational programmes</topic>
    </subject>
    <relatedItem type="constituent">
      <titleInfo>
        <title>Funding Guidelines - 2006</title>
      </titleInfo>
      <originInfo>
        <dateCreated>01/09/2006</dateCreated>
      </originInfo>
      <subject>
        <topic>Research Funding</topic>
      </subject>
      <location>
        <url>http://www.ionio.gr/rc/download.php?file=ee funding2006.pdf</url>
      </location>
      <physicalDescription>
        <note type="organization">item</note>
       </physicalDescription>
    </relatedItem>
  </relatedItem>
</mods>
```