ASSESSMENT AND COMPARISON OF METADATA SCHEMAS FOR ARCHITECTURAL HERITAGE

Paola RONZINO[1], Nicola AMICO[1], Franco NICCOLUCCI[1]

[1]Science and Technology for Archaeology Research Center, The Cyprus Institute, Address

Nicosia, Cyprus

p.ronzino@cyi.ac.cy

n.amico@cyi.ac.cy

f.niccolucci@cyi.ac.cy

Keywords: architectural heritage, cataloguing, metadata schemas, documentation standards, thesauri.

Abstract:

Conservation, protection and study of Architectural Heritage (AH) are based on the complete knowledge and documentation of its complex morphology, its architectonic components and all events related to the building.

A key issue for a full recording of a monument is related to the cataloguing process, which is the action of registration, description and classification of Cultural Heritage (CH) assets.

Many Cultural Institutions and National Bodies have been working towards the development of standards to make documentation uniform, retrieve information about Cultural Heritage, promote data sharing, improve content management and reduce redundant efforts.

To regulate the cataloguing activity, rules, guidelines and indications were defined. In particular, thesauri were adopted to provide, with a controlled vocabulary, structure and guidelines for standardization of terms, avoiding ambiguity and subjectivity of documentation and the loss of important information.

Some of these data standards have been defined within a national framework, such as the ICCD schema (Italy)or the MIDAS standard (England); others aim at guaranteeing data interoperability, such as, among others, LIDO and the CARARE schema.

In this paper we carry out an assessment and comparison of the above mentioned schemas and standards, along with the schema adopted by l'Inventaire général du patrimoine culturel (France).

After an analysis of the formal description of the single metadata schemas, we carried out the mapping of the different schemas to each other. The analysis is performed taking into consideration the RecorDim guidelines for documenting CH places.

As a result, the mappings described in the paper enable the interoperability of data stored according to different metadata schemas. Furthermore the paper highlights properties, equivalencies and shortcomings of the schemas, confirming the idea that a more comprehensive documentation standard for Architectural Heritage is needed.

1. INTRODUCTION

Heritage places (such as archaeological site, single monuments and groups of buildings, historic towns or cultural landscapes) undergo continuous changes (adaptations, renovations, disasters, decay, repairs, and conservation) that affect their status. In order to understand the condition and significance of a Heritage place, it is crucial to identify, locate, and document these changes.

Many Institutions have been working towards the development of standards to make documentation uniform, meaning for documentation the systematic collection and archiving of records, not only for the physical preservation of the AH asset but also for the preservation of its related information for future reference[1,2]. These Institutions have defined guidelines and indications helping in the gathering of information about an asset, adopting thesauri and controlled vocabularies for the standardization of the terms. Some of these data standards have been defined within a national framework, such as the ICCD schema (Italy), the MIDAS standard (England) and the Schéma documentaire appliqué au patrimoine et à l'architecture (SDAPA) adopted by l'Inventaire général du patrimoine culturel (France); others aim at guaranteeing data interoperability, such as LIDO and the CARARE schema, which are CIDOC-CRM compliant.

In order to better understand the ancient buildings, the cataloguing activity can be integrated with digital documentation (historical sources, heritage records, etc.), providing a representation of the CH asset in terms of text, images, drawings and 3D models.

As stated in the *Guiding Principles of Recording, Documentation and Management for the Conservation of Heritage Places* by RecorDIM Initiative [1], Heritage Information is the activity and the product of recording, documenting and managing the information of Cultural Heritage places, providing a way to transmit the acquired knowledge to future generations.

Heritage Information is required to understand meaning and values, promote the interest and involvement of people, permit informed management and ensure long-term maintenance and conservation of Heritage places.

The remainder of this paper is organized as follows: section 2 describes the main international Metadata schemas forms and standards; section 3 reports on the state of the art of existing mappings between metadata schemas. In section 4 we describe the adopted methodology for the metadata schemas analysis and mapping, highlighting equivalencies, difference and shortcomings. In Section 5 we describe the result of the mappings and finally Section 6 concludes the paper.

2. INTERNATIONAL METADATA SCHEMAS

The rapid growth of Internet resources and digital collections has been accompanied by a proliferation of metadata schemas, each of which has been designed according to the requirements of particular user communities, intended users, types of materials, subject domains, project needs, etc.

Several National Bodies and Cultural Institution that deal with cataloguing and documentation of Cultural Heritage developed different forms and standards:

- ICCD form, (Italy): it defines standards and tools for the Cataloguing and Documentation of national archaeological, architectural, art history and ethno-anthropological heritage in agreement with the Regions (http://www.iccd.beniculturali.it/index.php?en/95/institute) [3];
- MIDAS Heritage standard (UK): it is a data standard for information about the historic environment, developed by the English Heritage, for and on behalf of the Forum on Information Standards in Heritage (FISH)[3,4];
- The inventory of the French Cultural Heritage based on the Schéma documentaire appliqué au patrimoine et à l'architecture (SDAPA). (http://www.culture.gouv.fr/culture/dp/schemaDAPA/index.html). [3]
- The VRA Core is a data standard that consists of a metadata element set (units of information such as title, location, date, etc.), as well as an initial blueprint for how those elements can be hierarchically structured. The element set provides a categorical organization for the description of works of visual culture as well as the images that document them (http://www.loc.gov/standards/vracore/).

- The CIDOC CRM [5,6,7] is a formal ontology that defines CH documentation concepts and the relationship between them, used to clarify the documentation process, and to ensure no loss of semantic content when integrating heterogeneous Cultural Heritage data sources.

Metadata schemas:

- The Dublin Core [8], set of metadata elements provides a small and fundamental group of text elements through which most resources can be described and catalogued. Using only 15 base text fields, a Dublin Core metadata record can describe physical resources such as books, digital materials such as video, sound, image, or text files, and composite media like web pages.
- Lightweight Information Describing Objects (LIDO) [9], a metadata harvesting schema developed by the ATHENA Project for harvesting museum data into the service environment of Europeana.
- The POLIS DTD was produced as part of an EU funded Greek national research project to develop an interoperability framework for the Cultural Heritage.
- CARARE (http://www.carare.eu/eng/Resources/CARARE-metadata-schema-outline-v1.0) is a harvesting schema intended for delivering metadata to the CARARE service environment of an organization's online collections, monument inventory database and digital objects.
- The Europeana Data Model (EDM) aims at being integration medium for collecting, connecting and enriching all the elements included on the descriptions provided by Europeana content providers. (http://version1.europeana.eu/c/document_library/get_file?uuid=aff89c92-b6ff-4373-a279-fc47b9af3af2&groupId=10605)

  In the above list only standards employed in Archaeology, AH and museum communities were included. For practical necessity, it focuses on standards adopted in Europe.

3. STATE OF THE ART OF METADATA SCHEMAS MAPPINGS

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 [10]. 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.

Follows a list of metadata schemas mappings already performed.

Mappings between CIDOC- CRM and other schemas is available:

- LIDO v.07-CIDOC CRM v.5.0.1 [11]
- LIDO and CRM_{dig} [12]
- EDM
- Dublin Core (DC)
- MIDAS
- VRA Core 4.0

The ICCD form has been mapped to

- PICO [13]
- VRA Core 4.0 [14]

Mappings between the CARARE schema and Europeana Semantic Elements (ESE), Europeana Data Model (EDM), MIDAS, LIDO are available at CARARE project web site (http://www.carare.eu/eng/Resources).

Mapping between LIDO and ESE is available at Athena project web site (http://www.athenaeurope.org/index.php?en/1/home) and mapping between LIDO and EDM can be automatically achieved since the mapping between ESE and EDM has been made by Europeana.

4. METHODOLODY FOR THE FORMAL ANALYSIS AND MAPPING

The mapping of the models is defined as a sufficient specification to transform each instance of the source model into an instance of the target model with the same meaning.

The methodology adopted for our mapping was the creation of a crosswalk table identifying the correspondences (and non-correspondences) between different metadata element sets of the analyzed metadata schemas. This represents the semantic mapping of fields or data elements in one element set to fields or data elements in another element set.

The ICCD form has been considered as source for our mapping, while the MIDAS standard, LIDO, CARARE and the Schéma documentaire appliqué au patrimoine et à l'architecture (SDAPA) have been considered as targets.

4.1 Formal analysis of metadata schemas

The ICCD catalogue forms are descriptive models that collect the information on the heritage, according to a cognitive 'path' which guides the cataloguer in the documentation process. The ICCD has published different cataloguing models in relation to the different types of heritage and each form is hierarchically structured in paragraphs, fields and sub-fields. In particular we analysed the environmental and architectural heritage form which is composed by 29 paragraphs, 50 structured fields, 176 sub-fields and 14 simple fields. The data described in the form can be grouped as: information about identification of the asset, information about the asset and its components parts and location, data about documentations and source of the asset, information about the structure of the building in all its component parts, information about the related activities at the moment of the survey and restriction of use and administrative information.

The standards used as mapping target model are described below.

The MIDAS standard [4] states what information should be recorded to support effective sharing of the knowledge of the historic environment and the long-term preservation of those records. It covers the individual assets that form the historic environment (buildings, archaeological sites, shipwrecks, areas of interest and artefacts); the work that is undertaken to understand, protect and manage change to those assets; sources of further information. The MIDAS Heritage data standard has a three-level structure. From the broadest to the most specific:

- Themes: the broadest level areas of interest to the historic environment community.
- Information Groups: these set the specific standard for what should be included in an entry covering a particular subject.
- Units of Information: the basic 'facts' or items that make up an entry.

The Inventory of the French Cultural Heritage (L'Inventaire général du patrimoine culturel) is based on the Schéma documentaire appliqué au patrimoine et à l'architecture (SDAPA). The SDAPA XML schema is intended for organizations wishing to provide editorial information and data in different areas of Cultural Heritage. The documentation for this scheme is based on Dublin Core elements.

LIDO is based on CDWA Lite, MuseumDat, the CIDOC CRM and SPECTRUM and is made up of a nested set of wrapper and set elements which structure records and contain data elements holding the information that is being harvested and delivered to the user of the service environment. There are 7 areas in a LIDO record for an object:

- Object identification
- Object classification
- Relations of the object
- Events – in which the object has taken part
- Rights work – information about the rights associated with the object, metadata and digital surrogate
- Record – basic information about the record
- Resource – information about the resource being supplied to the service environment (Europeana)

The CARARE schema is a harvesting schema intended for delivering metadata to the CARARE service environment of an organisation's online collections, monument inventory database and digital objects. The schema focuses on the detailed description of monuments, events in which the monument has been involved and resources which represent and provide sources of information about the monument.

4.2 Mapping

The ICCD form has been used as source model for the mapping between the above mentioned metadata schemas form and standards, while the other schemas were used as target models.

The rationale for choosing ICCD as source was mainly related to the fact that it has a regular structure and completeness of the topics considered in the information sets.

In the particular case of MIDAS standard, since it is a guide for the creation of the inventories and it is not a proper form, much more effort was required in the interpretation of the information set against the ones of the other metadata schemas involved in the analysis.

As first step, in order to facilitate the mapping of the heterogeneous information resources, all the fields of the ICCD form (176 elements) have been listed in a table. Subsequently the elements of the information set from the other schemas where also listed after a comparison. For each element we identified elements with equivalent meaning (represented with a green arrow), similar meaning (orange arrow), and elements that are present just in one schema (red arrow) [15]. A sample of the huge set of mapping correspondences is shown in Figure 1.

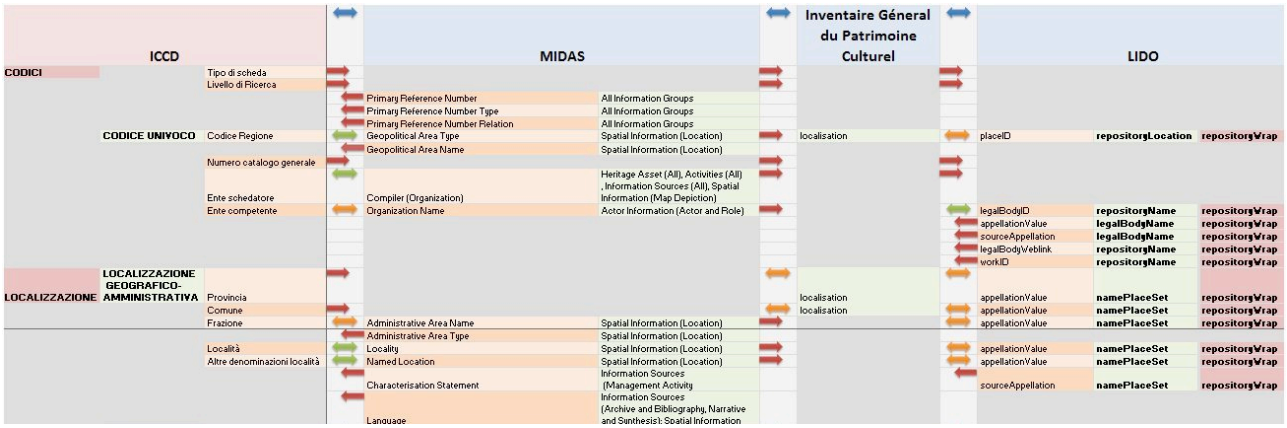

Figure 1: Sample of the 298 records of the mapping crosswalk table

5. ANALYSIS OF THE RESULTS

After the mapping between the above mentioned metadata schemas form and standards has been performed, some considerations about equivalencies, difference and shortcomings between the schemas arose.

In this publication we report in a more detailed way, as sample of the work performed, the mapping between ICCD and MIDAS. What came up from the analysis of the results is that 14% of the elements present in ICCD are equivalent to the ones in MIDAS, while the 75% are completely missing (40% of the element are present just in the ICCD form and the 35% is present just in the MIDAS). Just the 11% can be considered as having similar meaning (Figure 2).

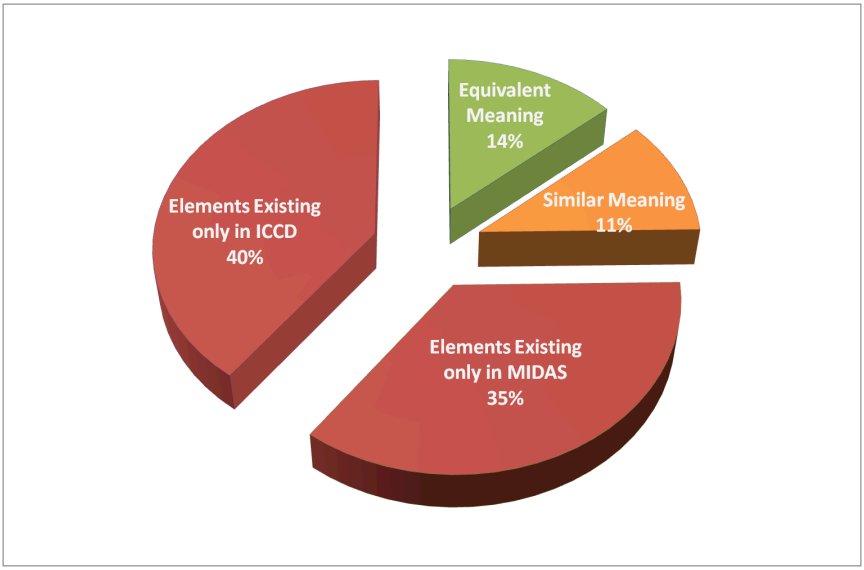

Figure 2: Example of the differences, equivalencies and similarities between ICCD and MIDAS elements.

The major discrepancies between the two forms are related to information about locations of the asset, object description, structural component of the asset, conservation state, activities and intervention on the asset, legislation restriction and information about the compiler.

In details, the ICCD form, in matter of information about location divides the territory in geo-administrative areas attributing codes to regions, provinces, towns, districts, historical centres and urban or extra urban areas, while the MIDAS considers, in addition to these, the general indication of the type of the area in order to understand the environmental context of the asset's location (i.e. moorland, parkland, suburb, conservation area etc.).

In MIDAS there are open fields where it is possible to describe the area. This possibility is offered by the fact that MIDAS is a flexibly designed standard which can be adapted to a range of different situations. The ICCD form has instead a more rigid framework and the only communication between the forms is that they can be related but cannot interchangeable.

In MIDAS users of the standard have control over which information group they choose to comply with. This aspect, which can be considered as giving more freedom in the choice of the forms to fill, can lead to a very subjective interpretation and the absence of a prefixed structure that guides in the documentation of the information can be cause of information loss.

Regarding the structural parts of the asset, they are contemplated in a different way between ICCD and MIDAS (this is the main difference between the two schemas). In fact, ICCD gives the possibility of a very detailed description of the structural plan with the internal division of the spaces, and of each singular component (foundations, vertical and horizontal structures, roofing, staircase, flooring, and decorative elements) describing morphology, materials, masonry etc. MIDAS instead indicates just the general masonry and the basic materials of which a Heritage Asset is composed.

What is missing in the Italian form is the consideration of the dimensional value of the building and the measurement unit, which can be only retrieved from the annexes (maps, sections, plans etc). LIDO, on the contrary, has structured measurement information set about the dimensions, size or scale of the object, indicating type, unit and value.

Another factor of distinction in MIDAS standard is the information about the character of portable items of heritage significance, including, for example, individual artefacts, architectural items and artefact assemblages collected during building repair or demolition. This allows these materials to be studied in their original context.

For what concerns the activities related to the building, the ICCD documents only the restoration and conservation activities, while MIDAS guides in the description of all the possible activities related to the asset and reports on the completeness of the activity, the agent of damage, the vulnerability level, the environmental condition note and the condition date. It foresees also entries record for decision-making

relevant to Heritage assets. This may include for example grants designation management agreement/plans or decision made as part of the town planning process.

Digital documentation is based just on photographs for ICCD and as reference to the use of modern technologies for digital acquisition, without any further specification, for MIDAS, while other schemas such as the CARARE and LIDO contain information about digital resources, including 3D models (CARARE).

LIDO and CARARE, unlike ICCD and MIDAS, geo-reference the objects adopting the GML (Geography Markup Language) specification.

For what concerns the schema SDAPA adopted by L'Inventaire général du patrimoine culturel it is based on Dublin Core and divided into elements, attributes and groups. Its elements are at the basis of LIDO and CARARE and for this particular work were not relevant for the purpose of the analyses.

6. CONCLUSIONS

Understanding a building [16] or a Heritage site means not just understanding its history (changes occurred at the monument over time), its physical configuration and condition, but involves also the understanding of a number of social, political, economic and cultural issues relating to the external environment.

The acquisition of new information deriving from all activities performed on a Heritage place, including Heritage recording, research and investigation, conservation, use and management, maintenance and monitoring, are carried out by a variety of specialists with the common goal of better understanding the asset and the related risks affecting its long-term conservation. The existence of many approaches for ancient building documentation, depending on the purpose of the recording activity (restoration, static analyses, chemical investigation, social studies) and on the target of professionals to whom the documentation process is addressed (restorers, historians, archaeologist, architects, engineers, landscape architects, conservators, managers, planners, sociologist and so on), make difficult the communication and the sharing of the acquired information and knowledge about the asset.

As a consequence of this multi-disciplinary environment the process of recording, documenting, and managing information has to be designed in a way that allows the integration and communication between heterogeneous data in terms of content and format.

The main focus of the paper was to assess, to analyse the formal description and to perform the mapping of the main metadata schema forms and standards, used for cataloguing and information management of Architectural Heritage. As a result of the mapping we highlighted the properties, equivalencies and shortcomings that characterize each form. The results confirm that all the analysed schemas are not designed to fulfil a holistic description of an asset and remark the necessity to avail of a standard of documentation that allows the integration of information coming from different disciplines, avoiding any data loss.

What it is still not well considered in the forms is the documentation of all the information related to the digital provenance of a digital resource (for example 3D models) indicating with details the methodology and the technical steps necessary to achieve the final result (digital data acquisition and data post-processing). These information are of paramount importance for analysis, study and for conservation of the original asset for future generation.

In perspective our future work will concern the creation of a new metadata schema which will allow a complete and holistic approach to the documentation of an Architectural asset, allowing the interoperability of its data information. This schema will take into account the peculiarities of the existing standards with the introduction of new set of information supported by specialized thesauri for the standardization of terms, avoiding ambiguity and subjectivity of documentation and the loss of important information.

7. ACKNOWLEDGEMENTS

This research was partially supported by 3D-COFORM (www.3d-coform.eu). 3D-COFORM is a co-funded Large Scale Integrating Project and has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement n. 231809.

8. REFERENCES

[1] Letellier, R.: *Recording, documentation, and information management for the conservation of heritage places: Guiding Principles*, The Getty Conservation Institute.

[2] Patias, P.: *Cultural Heritage Documentation*, In International Summer School "Digital Recording and 3D Model 2006, Crete, April 2006

[3] Caravale A., *La catalogazione informatica del patrimonio archeologico*, In Archeologia e Calcolatori, 20(2009), 179-187.

[4] Lee E., Turner E. et al: *MIDAS Heritage. The UK Historic Environment Data Stardard*, 2007.

[5] Doerr, M.: *The CIDOC CRM – An Ontological Approach to Semantic Interoperability of Metadata*. AI Magazine, 24(2003)3, 75-92.

[6] Crofts N., Doerr M., et al.: *Definition of the CIDOC Conceptual Reference Model* v. 5.0.2., 2010.

[7] Theodoridou M., Tzitzikas Y., et al.: *Modeling and querying provenance by extending CIDOC CRM*, In Distrib Parallel Databases, 27(2010), 169-210.

[8] D'Andrea A.: *Modelli formali per la rappresentazione della conoscenza in archeologia*, In Documentazione archeologica, standard e trattamento informatico, Budapest, Prime Rate, 2006, 119-137.

[9] Coburn E., Light R., et al.: *LIDO - Lightweight Information Describing Objects v.1.0*, 2010.

[10] Baca, M.: *Practical Issues in Applying Metadata Schemas and Controlled Vocabularies to Cultural Heritage Information*, In Cataloging & Classification Quarterly 2003, 36(2003)3, 47-55.

[11] Koutraki M., Doerr M.: *Mapping LIDO v0.7 to CIDOC-CRM v5.0.1*, ICS-FORTH, 2010.

[12] Pitzalis, D., Niccolucci, F. et al.: *LIDO and CRMdig from a 3D Cultural Heritage Documentation Perspective*, Proceedings VAST 2010, Paris, September 2010, 87-95.

[13] Amaturo M., Berardi E. et al.: *Mapping dei metadati tra le schede ICCD e PICO Application Profile per il Portale della Cultura Italiana*, Istituto Centrale per il Catalogo e la Documentazione, 2007.

[14] Masci M. E., Merlitti D.: *Mapping tra le Schede di catalogo ICCD e VRA Core 4.0. Indicazione dei livelli minimi di catalogazione-v. 2.0*, Istituto Centrale per il Catalogo e la Documentazione, 2008.

[15] Ronzino P., Amico N.: *Mapping between ICCD, MIDAS, LIDO, SDAPA*. Internal report, STARC-The Cyprus Institute, 2011.

[16] King J. (ed.): *Understanding Historic Buildings. A Guide to good recording practice*, English Heritage, 2006