Built Heritage metadata schemas and the integration of architectural datasets using CIDOC-CRM

Paola Ronzino[1]; Franco Niccolucci[1]; Andrea D'Andrea[2]
[1]PIN, VAST-LAB, Prato, Italy; [2]University of Naples "L'Orientale", CISA, Naples, Italy

1. Introduction

The 3D ICONS EU-funded project (www.3dicons-project.eu) has the objective of providing 3D models of European masterpieces of architecture and archaeology to Europeana, the European digital library. Its goal is to enable the public at large to virtually visit such icons of the European cultural heritage, availing of state-of-art technology in data acquisition and visualization. The project implementation must take into account a number of requirements dictated by the limits of general-purpose services such as those provided by Europeana, as regards the access through the Internet limiting the size of the models; and the need to be searchable through the Europeana portal, which poses strict requirements in terms of metadata. Since Europeana hosts only the metadata and not the actual content, users are re-directed to the provider's own repository to access the actual data. This procedure offers an opportunity for developing and offering a richer service to more demanding users. The models may thus be additionally stored in a more detailed - although heavier - version in the project repository, and they may have richer metadata to enable advanced searches.

Therefore the project has established its own metadata schema, which has been mapped to EDM, the schema used by Europeana, to allow integration with its millions of records. The 3D ICONS metadata schema contains much more information and may offer a more advanced search system to technical users and researchers. It may also be integrated with other research infrastructures, for example ARIADNE (www-ariadne-infrastructure.eu), the European research infrastructure integrating archaeological datasets across Europe. For this purpose, it must comply with accepted standards such as CIDOC-CRM, the ISO 21127:2006 standard for cultural heritage documentation. On the other hand, national cultural heritage agencies, like the Italian Ministry of Cultural Heritage and Activities (MIBAC), the French Ministry of Culture, and English Heritage, have established their own documentation standards, and accumulated thousands of monument records according to them. Each of these standards includes a thesaurus for built heritage.

At European level, the project EU-CHIC (Cultural Heritage Identity Card, www.eu-chic.eu) has defined the concept of the CHICEBERG Protocol for the integrated documentation of tangible cultural heritage. The protocol is based on a taxonomy of historic buildings developed within the EU project Perpetuate (www.perpetuate.eu). The scope of EU-CHIC mainly concerns conservation and documentation of environmental changes affecting tangible cultural heritage assets, buildings and monuments, including "natural" deterioration processes and human interventions.

All these initiatives cannot be disregarded. Besides being official initiatives of some national or international authority, they accumulated data that would be

impossible to collect again according to some other scheme. Building techniques, architectural styles and conservation issues, among others, are not bound by national boundaries: they are global in nature but local in features and problems, according to regional diversities and different conditions, for example as far as climate diversity affects conservation.
The acknowledgement of the necessity of a concerted initiative on heritage, especially on built heritage, is clearly expressed in initiatives such the EU Joint Programming Initiative on Cultural Heritage and Global Change (www.jpi-culturalheritage.eu), a framework within which EU Member States jointly address areas where public research programmes can respond to major societal challenges concerning heritage and its preservation. As far as documentation is concerned, this requires the alignment and integration of existing datasets, to overcome their fragmentation in national or sectorial repositories. The present paper aims at providing an initial contribution to this issue. We will start exploring the possibility of aligning some of the current major documentation frameworks, including those in use by some state authorities and the one proposed by 3D-ICONS. Our proposal adopts CIDOC-CRM as common semantic glue. We expect that extensions are likely to be required, taking full advantage of the CRM extensibility, one of the major strengths of the Conceptual Reference Model.

2. Preliminary analysis of existing documentation frameworks

A previous paper [Ronzino, Amico, Niccolucci 2011] on the assessment of metadata schemas and standard adopted for the documentation of Architectural Heritage by Cultural Institutions and National Bodies, demonstrated the need to avail of adequate tools to enable an integrated documentation of the built heritage.
This consideration arose out of a cross-walk mapping among the main schemas and standards in use by National Bodies, notably the schema [ICCD], adopted by the Central Institute for Cataloguing and Documentation (ICCD) of MIBAC, the Italian Ministry of Cultural Heritage; the MIDAS heritage standard adopted by English Heritage [MIDAS]; and the Schéma documentaire appliqué au patrimoine et à l'architecture (SDAPA) used by the Inventaire général du patrimoine culturel of the French Ministry of Culture [SDAPA]. Application profiles developed within European projects related to Europeana, i.e. CARARE [CARARE] and LIDO [LIDO], were also included in the first version of the mapping [Ronzino, Amico, Niccolucci 2011]. The mapping highlighted properties, equivalences, and shortcomings of the various schemas, stressing the divergences among them and their specialized character. The ICCD form has been considered as source for our mapping, while the MIDAS standard, LIDO, CARARE and the SDAPA have been considered as targets. This choice was dictated mainly by the regular structure and completeness of the topics considered in the information sets.
The environmental and architectural heritage ICCD records can be grouped as:
• information about identification of the asset,
• information about the asset and its location,

• data about documentations and source of the asset,
• information about the structure of the building and its component parts,
• information about activities occurred at the moment of the survey
• restrictions of use,
• administrative information.

MIDAS heritage is, instead, a guide for the creation of inventories, a detailed standard intended for full documentation of all aspects of heritage management, not a proper form.

To further extend the analysis on metadata schemas and standards we have included in the latest version of the mapping also the guidelines proposed by the above-mentioned EU-CHIC project, whose main objective was to develop an Identity Card of a heritage asset. The purpose was to define a strategy and select most efficient methods and tools for harmonization of criteria and indicators to be addressed for tracking environmental changes of tangible cultural heritage assets, buildings and monuments, including “natural” deterioration processes and human interventions. The data management scheme proposed by EU-CHIC is based on a Core Data Index for architectural heritage and a Core standard for archaeological heritage. It follows the Council of Europe 2009 Guidance on inventory and documentation of cultural heritage [Council of Europe 2009]. The EU-CHIC scheme is composed by three levels of information, which form the so-called “CHICEBERG”. The first level concerns information describing the asset with free access; the second level concerns the collection of information with restricted or no access. Here detailed data on the history, architecture, previous interventions, risks and so forth can be stored. The third level of information concerns information about activities that contribute to scientific decision-making support: e.g. preventive conservation regarding monitoring of assets and environmental impact controls, and management through the exploitation and planning of regular and extraordinary maintenance activities.

Apart from the administrative information about the heritage asset and the information about location and compilation, which are common in all the analysed schemas, the information set that mostly distinguishes the ICCD schema from MIDAS, EU-CHIC and CARARE is the one concerning the structure of the building (see Table 1 in the next page). ICCD, indeed, allows for a very detailed description of the structural plan, with the internal division of the spaces, and of each singular element of the building structure (i.e. foundations, vertical and horizontal structures, roofing, staircase, flooring, and decorative elements) describing morphology, materials, masonry etc. MIDAS, instead, includes a mere indication of the general masonry and the basic materials that the Heritage Asset is made of, as well as EU-CHIC does. Another important observation concerns the relations. At level 1, the EU-CHIC schema allows for cross-reference to related asset record, to records of fixtures, fittings collections and artifacts; to documentation; and to archaeological records/events and environmental records. Relations are limited in the MIDAS schema, while in the ICCD form is possible to create, e.g. relation of parental type between monument forms and findings forms. In the CARARE schema v.2 the relations can link the heritage asset with digital resources and activities.

Table 1 - Comparison among the features of the most relevant metadata schemas for built heritage documentation

CATEGORY	EU-CHIC	ICCD (form A)	MIDAS	CARARE v.2.0
Asset id.	Unique reference n. of asset	Codice univoco	Primary Reference n.	ID
Heritage Asset	- Type of Heritage asset - Name of heritage asset	- Definizione tipologica - Denominazione	- Monument Type - Material - Heritage Asset Name - Artefact Name Type	- HA/Characters/ heritage asset type - HA/Appellation/ name - HA/Characters/ heritage asset type
Structure	Structural material: - Foundation - Wall/pillars - Interstorey structure - Roof Finishing material: - Foundation - Wall/pillars - Interstorey structure - Roof	- Spazi/suddivisione interna - Impianto strutturale - Pianta - Fondazioni - Strutture verticali - Strutture orizzontamento - Copertura - Scale - Pavimenti e pavimentazioni - Elementi decorativi	- Evidence - Representation Source - Construction Method - Material - Material Component - Note - Material Name - Associated Goods	- HA/description - HA/ construction method - HA/Characters/ materials
Conservation/ restoration	- Current physical condition - General condition - Condition of critical elements - Major Risks - long-term environmental impact - Sudden environmental impact - Anthropogenic impact	- Stato di conservazione - Riferimento alla parte - Indicazioni specifiche	- Modification State - Condition - Condition Statement - Completeness - Condition Date - Agent of Damage - Vulnerability Level - Buffer Zone Width - Enviromental	- HA/Conditions/condition - HA/Conditions/ Condition assessment - HA/Conditions/ Condition date - HA/Conditions/relations
Relations	Cross-reference to: - records of fixures, fittings collections and artefacts - documentations - archaeological records/events - environmental records	Relazioni: - struttura complessa - relazioni dirette		- Has Representation - Is Successor Of - Was Present At - Has Part - Is Part Of - Is Replica Of - Was Digitized By - DC: Relation

Digital documentation includes pictures, movies and other multimedia content for ICCD and as reference to the use of modern technologies for digital acquisition, without any further specification, for MIDAS and EU-CHIC. Other schemas such as the ones used in CARARE and LIDO contain information about digital resources, including 3D models in CARARE.

A detailed analysis of the mapping between the above mentioned metadata schemas revealed the lack in most of them of any element enabling the documentation of the technological aspects related to the production of a 3D

virtual replica.
This problem has been addressed by the CARARE schema that included the 3D models among the digital resources related to the heritage asset, and by CRMdig, an extension of the CIDOC CRM, able to capture the modelling information regarding the provenance of digital objects [Doerr and Theodoridou, 2010]. Both these aspects have been included in the 3D-ICONS schema. Besides various simplifications concerning the reference, rights and record information sections, additional elements were included to meet the needs of the 3D ICONS project, enabling information capture on 3D objects provenance and the purpose of the 3D modelling, called paradata.

3. The 3D-ICONS approach to built heritage documentation: Heritage Asset

3D-ICONS substantially adopted the CARARE schema as far as architectural information is concerned. All the information is contained in a so-called wrapper named Heritage Asset.
The focus of the 3D-ICONS/CARARE schema is on the detailed description of heritage assets (monuments, buildings, landscapes or artefacts) and related digital resources and events in which the heritage asset is represented. The schema is based on MIDAS Heritage with additional elements from LIDO [LIDO] and the so-called Europeana Semantic Elements [ESE] to cover the information needed for the digital resources being made accessible to 3D-ICONS/CARARE and Europeana service environments. This part of the schema is based on the MIDAS Heritage standard, complemented with the POLIS DTD and the CIDOC Core Data Index for Archaeological Sites. The scope of this information set includes archaeological monuments, historic buildings, industrial monuments, and born digital (2D and 3D) models which relate to the archaeological and architectural heritage. The ability to create relations between heritage asset records allows the relationships between individual monuments that form parts of a larger complex to be expressed. The Heritage_Asset information set includes the following elements:
• Record_information: unique ID assigned by the content provider;
• Appellation: the name of the heritage asset and the identifier (ID) and may be repeated if, for example, a monument is known by more than one name or has more than one ID number;
• Description: includes the features of the archaeological monument, historic building, and the born-digital 2D or 3D models;
• General_type: is a broad classification of the general type of the heritage asset or born digital record intended to enable monuments, buildings and landscape areas to be distinguished from other objects.
• Actors: represents the actors involved with this monument; actors include for example creators, builders, sculptors, painters and historical figures who have an association with the monument or building.
• Designations: is an information about any designations for a monument or building which provide it with protection in law;
• Conditions: is about the condition of a monument or building;
• Provenance: (source = DCMI Terms) is a free-text statement of any changes

in ownership and custody of the resource since its creation that are significant for its authenticity, integrity, and interpretation; note that this 'administrative' provenance has nothing to do with the digital provenance concept, nor with archaeological provenance.

• Characters: is a set of information to describe the character of the monument. The information includes:

- o Heritage_asset_type classifies the monument or building with re spect to its function or use, e.g. house;
- o Temporal; about time
- o Materials: is about the basic materials of which a heritage asset is composed, e.g. brick, stone, tile, paper etc. Use of a controlled voca bulary is recommended, and the vocabulary used may be indicated using an attribute;
- o Inscriptions: is a text inscribed on a monument or building;
- o Dimensions (source = MIDAS, LIDO)
  - - Extent is to note the part of the heritage asset being measu red, e.g. Base;
  - - Measurement type is referred to height, length, width, depth, shape;
  - - Units e.g. metres, centimetres;
  - - Scale;
  - - Value is an attribute registering the accuracy of the measu rement to be indicated (exact, approximate);
- o Craft: is a set of information to describe shipwrecks; not relevant in the present context.

• Spatial: is information about the place at which the heritage asset is located, included named places, postal address, the map coordinates and geometry of the heritage asset.

• Repository_location (source = LIDO): identifies the institution which custodies the artefact and possibly the current location;

• Publication_statement;

• Rights (source = DC): is a statement about any rights associated with the heritage asset;

• References: are sources of information about the heritage asset in publications and archival sources (for example, bibliographic references etc.). The information includes:

• Appellation: is the name given to the information source;

• Actors: is the creator, author, contributor, editor, etc.;

• Type includes archive, file, record, book, chapter, article etc.;

• Rights:

• Publication statement:

• Note

Finally, there are two more elements linking the Heritage-asset record to other records concerning the same object.

• Link (Source = LIDO) is the URL where users can find the reference online;

• has_representation: is the relationship between a heritage asset and a digital resource in which it is represented.

4. Conclusions and further work

We have briefly summarized the existing standards for built heritage documentation, in light of the increased use of 3D models for this purpose. It is clear that most of the standards do not take 3D and its characteristics into account, and make Europe-wide integration of archives at least problematic. Further work is requested to establish operational mappings to a common standard such as CIDOC-CRM, a schema that has increased in popularity among heritage professionals. The preconception that it is pointlessly complicated is progressively disappearing as people understand its powerful potential and the necessity of using it to overcome dataset fragmentation in a European framework. For space reasons and to avoid an excessively technical approach we have omitted the complete crosswalk and the mapping of current schemas to the CRM; however, they are available on [Ronzino et al, 2013].

There is still work to do, in order to implement the necessary mereological properties linking the parts of a building or a complex to the whole, a reflection that started with [Doerr et al, 2001] but was then suspended in the CRM team.

Acknowledgements

This paper has been partially supported by the European Commission in the framework of the 3D-ICONS project.

References

CARARE http://www.carare.eu/eng/Resources/CARARE-Documentation/CARARE-metadata-schema.

Council of Europe, 2009, *Guidance on inventory and documentation of the cultural heritage*, Published by the CoE.

Doerr M., Plexousakis P., Bekiari C., 2001, *A Metamodel for Part ñ Whole Relationships for Reasoning on Missing Parts and Reconstruction*, in Kunii H.S., Jajodia S. and Sølvberg A. (eds), *Conceptual Modeling - ER 2001*, Springer - Lecture Notes in Computer Science Volume 2224, pp. 412-425.

Doerr M., Theodoridou M., 2011, *CRMdig: A generic digital provenance model for scientific observation*. http://www.usenix.org/events/tapp11/tech/final_files/Doerr.pdf.

ESE http://www.europeana.eu/schemas/ese/.

ICCD, *Cataloguing Standards* http://www.iccd.beniculturali.it/index.php?en/115/cataloguing-standards.

LIDO, *Lightweight Information Describing Objects* http://www.lido-schema.org/schema/v0.9/lido-v0.9-schema-listing.html.

MIDAS MIDAS *Heritage* http://www.english-heritage.org.uk/publications/midas-heritage/.

Ronzino P., Amico N., Felicetti A., Niccolucci F., 2013, *Built heritage documentation in CIDOC-CRM*, PIN technical report. http://www.vast-lab.org/documents.

Ronzino P., Amico N., Niccolucci F., 2011, *Assessment and comparison of metadata schemas for architectural heritage*, CIPA2011, cipa.icomos.org/fileadmin/template/doc/PRAGUE/127.pdf.

SDAPA Ministère de la Culture et de la Communication, *Schéma documentaire appliqué au patrimoine et à l'architecture* http://www.culture.gouv.fr/culture/dp/schemaDAPA/operationnel/index.html.