

Proceedings of the joint international event 9<sup>th</sup> ARQUEOLÓGICA 2.0 & 3<sup>rd</sup> GEORES, Valencia (Spain). 26–28 April 2021

Received: 26/11/2020 Accepted: 17/03/2021 DOI: https://doi.org/10.4995/Arqueologica9.2021.12090

# VOCABULARY ALIGNMENT FOR SHARING ARCHITECTURAL CONSERVATION DATA

Marta Acierno<sup>a,\*</sup>, Donatella Fiorani<sup>a</sup>, Athanasios Velios<sup>b</sup>

<sup>a</sup> Dipartimento di Storia, Disegno e Restauro dell'Architettura, Sapienza University of Rome, Piazza Borghese 9, 00186 Roma, Italy. marta.acierno@uniroma1.it; donatella.fiorani@uniroma1.it

<sup>b</sup> Ligatus, Chelsea College of Arts, University of the Arts London, 16 John Islip Street London, SW1P 4JU, UK. a.velios@arts.ac.uk

#### Abstract:

In the domain of heritage conservation, much of the decision making relies on assessing risk and planning in advance. This is particularly important for built heritage where environmental conditions cannot be controlled easily and a broader range of risks must be taken into account. Risk assessment requires detailed surveys of built heritage which produce a wealth of data around the vulnerability of a building using domain expert terminology. It also requires information about the environmental condition and the dangerousness of the area that the building is situated. Integrating these data within systems is essential. Further sharing this data with external audiences such as emergency services is also important. In this paper guidelines developed for the Linked Conservation Data project are considered in relation to a risk assessment system: Risk Map of the Italian Ministry of Culture. The guidelines explain how terminology data can be integrated to ensure a common language across systems or system components. Terminology alignment with the Getty Arts & Architecture Thesaurus is considered and specific terms from the built heritage domain are explored as well as the challenges posed by their use in different contexts.

Keywords: integration, risk assessment, cultural heritage, documentation, conservation

### 1. Integration of data

The value of documentation in heritage conservation is emphasised through codes of ethics and international charters. Documentation captures information about the production (materials and techniques), condition and treatment activity of heritage objects and sites. This allows monitoring the development of deterioration processes and assessing past conservation treatments. Built heritage data is a subset of conservation data. It is more complex due to scale and dependency on topology and structure while including mobile objects located in buildings.

Documentation records in conservation are often considered on a per item basis and are produced using free text to be read as part of reviewing a single object or site. Extra value is added to documentation records when they can be summarised at collection level and integrated at cross-collection level. For example, more representative samples of observations on materials and techniques outline the history of technology for objects. Combined observations from the results of conservation treatment allow better assessment of treatment techniques and materials. Monitoring deterioration at scale allows conclusions about deterioration patterns to emerge both in connection to geography and to material type. Systematic sharing of environmental monitoring data combined with condition survey data can assess environmental control and establish causality for damage. These benefits show that integration of conservation records is a pivotal activity for the development of the profession.

This is also true for built heritage with the additional complexity of considering each case individually and therefore more difficult to integrate (i.e., it is difficult to identify categories for a common approach). While complete integration of data for built heritage may be achieved in the future, at the moment the focus is on integrating data for risk assessments to evaluate the possibility of losing historic architecture (Fig. 1) considering buildings and the areas where they are located.

We are introducing a description of risk data here followed by a methodology for integration offered by the Linked Conservation Data (LCD) consortium and its application to Risk Map - a project from the Italian Ministry of Heritage - with particular interest in terminology.

<sup>\*</sup> Corresponding Author: Marta Acierno, marta.acierno@uniroma1.it



**Figure 1**: Risk Map of MIBACT, the historic centre of San Vito Romano (Rome) georeferred with the evidence of landslides.

# 1.1. Risk assessment for historical architecture

Risk is calculated based on territorial dangerousness (area factors) and object vulnerability. We consider three types of territorial dangerousness: a) physical: i.e. landslides, floods, earthquakes, etc., b) environmental/air: i.e. pollution and c) human-related: i.e. depopulation. Vulnerability can be related to structural weakness (for example a design error or a transformation), material decay, bad craftsmanship and inadequate restoration.

Data for risk assessment of built heritage is gathered in different platforms, often in Geographic Information Systems (GIS) where it can be georeferenced, so the vulnerability of a single building and its contents can be related to the dangerousness of the territory that hosts it.

We can collect data about groups of buildings and their components and we can monitor that over time to allow minimal conservation interventions when necessary. We can also compare with observations on similar buildings located nearby and provide a scientific basis for developing conservation strategies.

### 1.2. Linked Conservation Data

Recognising the importance of integrating records, the LCD consortium investigates the steps required for sharing conservation data and for making integration practical for conservators and conservation departments. The consortium currently examines the use of Linked Data technologies to achieve integration. Different working groups are undertaking work into two main areas: modelling and terminology.

Modelling describes the process of aligning local schemas to enable data comparisons. An efficient way of aligning schemas is through a generic ontology. In LCD we are currently examining the suitability of the CIDOC-CRM ontology (CIDOC, 2020) alongside its extensions for modelling conservation data. Work on modelling risk with the CRM ontology is in progress and will be reported separately. This paper is only concerned with the second area of activity in LCD which is the terminology (Velios, & St.John, 2020).

# 2. Terminology for integrating heritage data

The CIDOC-CRM ontology is a generic ontology. As such it does not feature domain-specific classes. The additional granularity needed for describing conservation data is provided through the CRM 'E55 Type' class, i.e. every vocabulary term is matched to 'E55 Type'. Integration at schema level is possible through CRM classes and properties, but if one requires the granularity of domainspecific terminology then vocabulary entries also need to be aligned.

Within the realm of Linked Data there is a dominant standard for encoding and sharing vocabularies known as the Simple Knowledge Organisation System (see SKOS, 2012). SKOS features properties that can assist integration. It allows synonyms for each entry and it features properties that match entries across different vocabularies. These include: skos:exactMatch. skos:closeMatch, skos:broadMatch (and skos:narrowMatch). Using the property skos:exactMatch indicates that two entries from different vocabularies describe the same idea, in practical terms the two entries interchangeably. property used The can be skos:closeMatch indicates semantic overlap useful for retrieval purposes. The property skos:broadMatch (and inverse skos:narrowMatch) indicate that one its vocabulary lacks the detail included in the other but that there is a possible hierarchical connection between terms. There is relatively clear expectation on how skos:exactMatch and skos:closeMatch would be interpreted in implementations for cross-database retrieval systems where a user would enter one word in a search box using their own vocabulary and results with records from matched terms in other vocabularies would Similarly, the skos:broadMatch indicates appear. hierarchical relationships so it would be expected to be utilised when browsing hierarchical vocabularies with terms from one vocabulary appearing as child-terms of another.

The LCD consortium has identified the Getty Arts & Architecture Thesaurus (see Getty Research Institute, 2017) as a vocabulary with enough coverage in conservation to act as a matching hub. Vocabulary maintainers are encouraged to match to the Getty AAT before any other vocabulary to enforce the hub role and minimise the amount of matching needed across vocabularies.

Examples from risk assessment follow after we introduce the Risk Map system in the next section.

# 3. The Risk Map of the Italian Ministry of Heritage: use and data organisation

Risk Map (Fig. 2) is one of the GIS platforms of the Italian Ministry of Culture (Ministero della Cultura - MiC). It is interoperable with other GIS platforms in the Ministry, such as Sigecweb, from the Central Institute of Catalogue and Documentation (ICCD). Risk Map expresses the vulnerability of buildings or building complexes (i.e. a town's historic centre) in relation to territorial dangerousness. The former being data collected of buildings and the latter data from specific thematic maps (Cacace, 2007).

Risk Map is used in case of local disasters – such as earthquakes or floods – because it allows quick identification of heritage at risk. Applications of the system exist for areas of Italy, including Abruzzo, Marche, Sicily and Calabria (Cacace, & Fiorani, 2008; Donatelli, 2010; Acierno, Cacace, & Giovagnoli, 2014).

Risk Map collects data using documentation forms that depend on the type of object or conservation task (i.e.

mobile objects, buildings, historic centres, archaeological remains, archaeological sites, underwater archaeological remains etc). Risk Map extends ICCD to produce these documentation forms. The ICCD provides references, such as a registry and part of the thesaurus for Risk Map which acts as an integration platform. In this paper, we consider how it can further be integrated by sharing records with other systems.



Figure 2: Risk Map of MIBACT, the historic centre of Calcata (Viterbo) georeferred with the info on the selected area.

Like in many other cases, the system of the ICCD is based on historic paper records and therefore inherits legacy data. As such, there is no provision for global identifiers in the system and therefore any linking is done based on internal identifiers or labels. Any attempt to share records from this system would require matching local identifiers and labels with global identifiers possibly provided by well-recognised authority files.

As part of this work, a contribution to Risk Map is made related to the description of historic centres through the creation of six new information forms that correspond to different urban scales. This required the employment of a vocabulary to manage historic centres which reference existing thesauri in ICCD (which are also based on SKOS) (ICCD, 2020) but also adding some new terms.

# 4. Case study: Risk Map for historic centres

The Risk Map vocabularies reflect the nature of the risk assessment records covering generic terms about built heritage and more specific ones deriving from conservation. Comparisons on words only are avoided and instead the use of the word is considered to identify its meaning. The task was to formalise the Risk Map vocabularies and publish them as SKOS Linked Data. This included five steps:

Creation of global identifier and URI. The Risk Map 1) system only allows local identifiers for the vocabulary terms. Therefore, unique identifiers and unambiguous reference points within a global scope were produced outside the system. We use https://w3id.org, whose purpose is to provide persistent re-direction services, namespace is proposed alongside providing a subdirectory 'mibactsapienza', followed by a number (for example; https://w3id.org/mibact-sapienza/1). The w3id consortium requires that the URIs are resolvable to serve content and as Risk Map is not currently offering such functionality changes to the w3id domain yet have not been committed yet.

- Creation of a scope note to describe the vocabulary term. This allows clarifying the meaning of a term based on its use and is required to control synonyms and equivalent terms.
- 3) Establishing the hierarchical position of the term in the vocabulary. To allow browsing the vocabulary, relationships across terms can be built. Associative relationships are between concepts that are relevant using the property skos:related. Hierarchical relationships are between parent and child terms using the properties skos:broader and skos:narrower indicating that the part term is more general than the child term.
- Aligning to target thesauri. To enable integration with 4) datasets built with different vocabularies the AAT is used as a hub thesaurus as expressed by the LCD guidelines. This allows maintaining the Risk Map vocabulary locally while enabling retrieval of data using the AAT. The use of each Risk Map term has been checked against relevant scope notes in the AAT. When semantic relevance was identified, the term was marked as skos:exactMatch. skos:closeMatch was used for terms which diverted slightly from the AAT scope note but were close justify retrieval. enough to More rarely skos:broadMatch or skos:narrowMatch was used when the hierarchical relationship across vocabularies could be established. In other cases, corresponding terms could not be identified and a plan for submitting new terms to the AAT has been made. Note that matching properties are actually described in the official SKOS documentation that specifies that "A skos:closeMatch link indicates that two concepts are sufficiently similar that they can be used interchangeably in some information retrieval applications. A skos:exactMatch link indicates a high degree of confidence that two concepts can be used interchangeably across a wide range of information retrieval applications.
- SKOS publishing. The resulting data was encoded in a spreadsheet and converted to SKOS and validated using the popular SKOS Play tools (Francart, 2020). A sample of that encoding can be seen in the Appendix (see Section 6).

### 4.1. Terms relation and matches

The hierarchy based on Risk Map entries for 'historic centres' is highlighted here. Scope notes were developed after defining the cultural context of built heritage in Italy (Fiorani, 2018, 94-95) based on relevant bibliography which was also the prism under which AAT entries were considered.

The concept 'historic centres' is defined within Risk Map as the "Ensemble of historic buildings within an urban agglomeration formed by World War I (1919). The definition of 'historic centre' is one of the most discussed issue in literature and the choices made for the Risk Map have already been discussed (Fiorani 2018, pp. 88-96). This chronological threshold normally identifies a part - generally central - of the town, but the contemporary city can also include several historic districts that are in fact disjoint. This centre is marked by a perimeter that can sometimes coincide with the boundary of the city walls when they still exist or when they can be identified on the basis of historical records, material evidence, or on natural limits (streams, strong slopes, etc.). The chronological threshold can sometimes result from the evident constructive mutation of the building, verified by the comparison with the available cartographic and historical graphic documentation and direct feedback." Considering the specificity of the term, conceived in a particular geographical context (Italy) and for a particular purpose (Risk Map), an exact match on AAT was not expected. Instead the anticipation was the use of a broader match. The temporal limitation was introduced to identify only those parts of the town constructed with traditional building techniques, since the method for calculating vulnerability within Risk Map is defined upon mathematical models that represent masonry buildings and not reinforced concrete or steel structures. Searching the AAT showed no exact matches.

More thorough investigation indicated three close matches: 'historic district', 'historic quarters' and 'downtown'. The first two disregarded the quality of the buildings and mainly focussed on the administrative role in the designation of the historical value, while the third was mainly focusing on the geographical understanding of the concept. The four scope notes (the three from AAT and the one from our definition of 'historic centres') are different but involve a common core concept which is substantially expressed by the concept 'historic quarters'.

Note that 'historic district' is defined as "Former administrative or electoral districts that have been designated by a modern governing body as being culturally or historically significant, or embodying distinctive characteristics of a period, method of construction, or inhabitants. For historic areas of a city or town, use "historic quarters."; the third as "Core or central areas of cities or towns that are either the oldest areas of the city or the core business areas. The term is thought to derive from these areas, as the oldest areas, having the lowest street numbers; alternatively, term may derive from the development of New York City, where the oldest section is in the south, at the bottom of a map".

The term 'historic centres' declares a whole made of different parts, two of which are defined as 'urban unitsaggregates' and 'urban units-residential or special isolated buildings'. The 'urban units-aggregates' contain further parts which are described with 'building units'. Parts of 'building units' and urban units-residential or special isolated buildings are 'building fronts'. Another part of 'historic centres' is considered as 'urban spaces'. The partitive relationships among these terms do not indicate relevance in terms of characteristics of the things described. For example a facade ('building fronts') is not a building itself ('building unit'), but only part of it. While SKOS does not provide a specific way to express partitive relationships, guidelines indicate that extensions of the SKOS model should be used instead (W3C, 2020). The adoption of a new relationship (lcd:broaderPartitive) within the wider LCD project is considered as shown in our encoding example in the Appendix.

'Urban units' manifest as at least 'urban units-aggregates' and 'urban units-residential or special isolated buildings' and in both cases these are more specific types (specialisations) of 'urban units' where the broader relationship (skos:broader) can be used.

Associative relationships (skos:related) are used to indicate the connection between 'building units' and 'urban units-residential or special isolated buildings' as neither the partitive relationship nor the specialisation is valid.

The term 'urban spaces' is defined within Risk Map as "an open space inside the historic centre bordered by several 'urban fronts' and characterized by the presence of specific elements (gardens, furnishings, flooring, etc.). In theory, all streets and squares contained within the centre are urban spaces." Such definition is broader to AAT 'open space' which is conceived only for pedestrians. Finally, it is interesting to refer to the exact matches that resulted for the concepts 'urban units - aggregates', 'building units' and 'building fronts' whose scope notes exactly match those in AAT 'complexes (buildings and sites)', 'single built works (built environment)' and 'fronts (architectural)'. Conversely, it was impossible to identify an exact match for 'urban units-residential or special isolated buildings' which although scale-wise occupy an urban unit, they are single buildings (not aggregates) and therefore different to AAT 'complexes'.

## 5. Conclusions

The exploration of the subject is ongoing and further experience is gained from matching the local vocabulary of Risk Map to AAT. The advantage of maintaining a local vocabulary is that the local preferences on composite entries and hierarchical arrangements of terms can be preserved thus without requiring modifications to the Risk Map system. To enable interoperability with other datasets alignment of our vocabulary to the AAT is underway. This will allow external users to search our records without having to observe the peculiarities of our local vocabulary. Searching through a well-known vocabulary such as the AAT allows broader exposure of records.

The proposal to encode Risk Map vocabularies intends to amplify the scenario of the digital platform of the Italian Ministry MIC through the alignment with other international thesauri. Collaboration with ICCD experts is initiated to resolve technical issues including those on global identifiers. More generally, the work has to be considered within ongoing research frameworks that aims to integrating conservation data in different institutional contexts in a wider global environment. The work on terminology developed intends to assist the research on conceptual modelling that has focussed on historic centres issue representation (Acierno, & Fiorani, 2019).

The LCD efforts and conceptual modelling open the application of Risk Map to a completely new dimension connecting operational needs of digital instruments for conservation planning and emergency planning with a wider strategy of globally sharing conservation data.

# 6. Appendix

### 6.1. Term relations

Broader relationships (skos:broaderIsA)

urban units

 $\rightarrow$  urban units-aggregates

 $\rightarrow$  urban units-residential or special isolated buildings

Broader partitive relationships (lcd:broaderPartitive)

historic centres

→ urban units-aggregates

 $\rightarrow$   $\rightarrow$  building units

 $\rightarrow$   $\rightarrow$   $\rightarrow$  building fronts

 $\rightarrow$  urban units-residential or special isolated buildings

 $\rightarrow$   $\rightarrow$  building fronts

 $\rightarrow$  urban spaces

Associative relationships (skos:related)

building units ---- urban units-residential or special isolated buildings

### 6.2. Encoding examples

### 6.2.1. SKOS concepts and scope notes

mbs:4 a skos:Concept;

skos:broader mbs:3;

skos:inScheme <https://w3id.org/mibact-sapienza/>;

skos:narrower mbs:6;

skos:prefLabel "Unità urbana-aggregato"@it, "urban units-aggregates"@en;

skos:scopeNote "It is a type of 'urban unit' and is understood as a continuous set of structurally connected but distinguishable building units as organisms that are architecturally coherent. It is bordered by urban spaces with which it communicates through accesses to the ground floor and may contain within it places relevant to one or more building units. It overlooks on urban spaces through fronts that result from the succession of the individual facades of the building units."@en;

Icd:broaderPartitive mbs:1.

mbs:5 a skos:Concept;

skos:broader mbs:3;

skos:inScheme <https://w3id.org/mibact-sapienza/>;

skos:prefLabel "unità urbana-edilizia puntuale residenziale o specialistica"@it, "urban units-residential or special isolated buildings"@en;

skos:related mbs:6;

skos:scopeNote "It is a type of 'urban unit' and is understood as a homogeneous and autonomous unitary element. It is bordered by urban spaces with whom it communicates through accesses to the ground floor and can contain open places inside. It overlooks urban spaces through a variable number of fronts generally figuratively and constructively coherent with each other. 'Residential punctual building units' are buildings such as villas or cottages, while 'Specialist punctual building units' include hospitals or schools."@en;

lcd:broaderPartitive mbs:1.

mbs:6 a skos:Concept;

skos:broader mbs:4;

skos:inScheme <https://w3id.org/mibact-sapienza/>;

skos:prefLabel "building units"@en, "unità edilizia"@it;

skos:related mbs:5;

skos:scopeNote "It is a coherent architectural organism, consisting in a single construction phase or by the merging/recasting of several pre-existing building units. This organism is characterized by the presence of one or more functional units (residential and non-residential) connected to each other through common distribution elements and served by access at the road level; it is bordered by external facades with a generally continuous eaves line and a unitary or composite roof composed of flat and/or inclined parts consistent with all fronts."@en.

mbs:7 a skos:Concept;

skos:inScheme <https://w3id.org/mibact-sapienza/>;

skos:prefLabel "building fronts"@en, "fronte edilizio"@it;

skos:scopeNote "It corresponds to the façade of the single 'building unit' overlooking on an urban space or, more generally, on a street or a town square but also on the natural margin of the town (maritime, river, lake coast, prmonatory limit, etc.)."@en;

skos:topConceptOf <https://w3id.org/mibactsapienza/>;

Icd:broaderPartitive mbs:5, mbs:6.

### 6.2.2. Alignment with AAT

mbs:2 a skos:Concept;

skos:inScheme <https://w3id.org/mibact-sapienza/>;

skos:narrowMatch aat:300008072;

skos:prefLabel "urban spaces"@en;

skos:topConceptOf <https://w3id.org/mibact-sapienza/>
mbs:3 a skos:Concept;

skos:inScheme <https://w3id.org/mibact-sapienza/>;

skos:narrowMatch aat:300264550;

skos:prefLabel "urban units"@en;

skos:topConceptOf <https://w3id.org/mibact-sapienza/>

mbs:4 a skos:Concept;

skos:exactMatch aat:300000202;

skos:inScheme <https://w3id.org/mibact-sapienza/>;

skos:prefLabel "urban units-aggregates"@en;

skos:topConceptOf <https://w3id.org/mibact-sapienza/>

#### References

- Acierno, M., Cacace, C., & Giovagnoli, A. M. (2014). La Carta del Rischio: un approccio possibile alla manutenzione programmata. Il caso di Ancona, Materiali e Strutture. *Problemi di conservazione.* 3(5-6), 81-106.
- Acierno, M., & Fiorani, D. (2019). Innovative Tools for Managing Historical Buildings: the Use of Geographic Information System and Ontologies for Historical Centers. ISPRS. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences., XLII, 2019, 21-27. https://doi.org/10.5194/isprs-archives-XLII-2-W11-21-2019
- Cacace, C. (2007). Risk Map Methodology. In R. Varoli-Piazza (ed.), Sharing Conservation Data, Rome, Italy: Iccrom.
- Cacace, C., & Fiorani, D. (2008). Sistema informativo territoriale della Carta del Rischio, schedatura di chiese dell'Abruzzo e del Molise e calcolo della loro vulnerabilità. In 12a Conferenza Nazionale della Federazione italiana delle Associazioni Scientifiche per le Informazioni Territoriali e Ambientali, (L'Aquila, 21-24 ottobre 2008), L'Aquila, Italy: ASITA, 569-574.
- CIDOC. (2020). CIDOC Conceptual Reference Model (CRM). Retrieved November 12, 2020, from http://www.cidoccrm.org/
- Donatelli, A. (2010). Terremoto e architettura storica. Prevenire l'emergenza. Roma: Gangemi, 199-214.
- Fiorani, D. (2018). Il futuro dei centri storici. Digitalizzazione e strategia conservativa. Roma: Quasar.
- Francart, T. (2020). SKOS Play! Thesaurus & Taxonomies (sparna.fr). Retrieved November 12, 2020, from https://skosplay.sparna.fr/play/
- Getty Research Institute. (2017). Art & Architecture Thesaurus. Retrieved November 12, 2020, from https://www.getty.edu/research/tools/vocabularies/aat/
- ICCD. (2020). Home ICCD Istituto Centrale per il Catalogo e la Documentazione (beniculturali.it). Retrieved November 12, 2020, from http://www.iccd.beniculturali.it/it/percondividere/interoperabilita
- SKOS. (2012). SKOS Simple Knowledge Organization System home page (w3.org). Retrieved November 12, 2020, from https://www.w3.org/2004/02/skos/
- Velios, A., & St.John, K. (2020). Linked Conservation Data terminology guidelines. Retrieved November 12, 2020, from https://www.ligatus.org.uk/lcd/sites/ligatus.org.uk.lcd/files/attachments/193/terminology-working-groupguidelines.pdf
- W3C. (2020). W3C Wiki. Retrieved November 12, 2020, from https://www.w3.org/wiki/SkosCoreGuideToc/SectionExtending /SemanticRelations