ArCo ontology network and LOD on Italian Cultural Heritage

Valentina Anita Carriero
STLab, ISTC
CNR
Rome, Italy
valentina.carriero@istc.cnr.it

Aldo Gangemi
FICLIT
Università di Bologna
Bologna, Italy
aldo.gangemi@unibo.it

Maria Letizia Mancinelli
ICCD
MiBAC
Rome, Italy
marialetizia.mancinelli@beniculturali.it

Ludovica Marinucci
STLab, ISTC
CNR
Rome, Italy
ludovica.marinucci@istc.cnr.it

Andrea Giovanni Nuzzolese
STLab, ISTC
CNR
Rome, Italy
andrea.nuzzolese@cnr.it

Valentina Presutti
STLab, ISTC
CNR
Rome, Italy
valentina.presutti@cnr.it

Chiara Veninata
ICCD
MiBAC
Rome, Italy
chiara.veninata@beniculturali.it

Abstract

ArCo (Architecture of Knowledge) is a collaborative project that involves the institute of the Italian Ministry of Cultural Heritage ICCD (Institute of Catalogue and Documentation) and the Institute of Cognitive Sciences and Technologies of CNR (Italian National Research Council). ArCo aims at modelling the wide domain of Italian cultural heritage for two main purposes: (i) building a network of ontologies, compatible and aligned whenever possible with existing ontologies, that can be used as a de facto standard for representing cultural heritage data; (ii) publishing ICCD data as LOD: about 800.000 publishable files stored in the ICCD General Catalogue database. In this paper, we present ArCo structure, design methods and tools, its growing community, and we delineate its importance, quality, and impact in using semantic technologies in the fruition of Cultural Heritage.

1 Introduction

The increasingly widespread use of semantic technologies and Linked Open Data (LOD) led Digital Humanities to re-think their approach to knowledge management and sharing [1]. These technologies give Digital Humanities a means for representing their knowledge and include it into a network of connected data on the web, thus encouraging its reuse and further enrichment. In this context, ontologies play an essential role, as a technology for organizing knowledge by abstracting data and information of a certain domain.

An increasing number of cultural institutions is choosing ontologies and LOD for modelling and publishing their data, e.g. in Italy the Institute of artistic, cultural and naturalistic heritage of Emilia-Romagna (IBC-ER) [2] and the Fondazione Federico Zeri [3,4], and, in Europe, a lot of institutions within the project Europeana [5].

In this paper we report the results of ArCo (Architecture of Knowledge) [6], a collaborative project that involves the institute of the Italian Ministry of Cultural Heritage ICCD (Institute of Catalogue and Documentation) and the Institute of Cognitive Sciences and Technologies of CNR (Italian National Research Council).

ArCo aims at modelling the wide domain of Italian cultural heritage for two main purposes: (i) building a network of ontologies, compatible and aligned whenever possible with existing ontologies, that can be used as a *de facto* standard for representing cultural heritage data; (ii) publishing ICCD data as LOD: about 800.000 publishable files stored in a database, i.e. the General Catalogue, each describing a specific cultural property from diverse perspectives.

2 Related Work

The cultural heritage domain has an intrinsic complexity, due to the high number of different types of cultural properties that a cataloguer may record, e.g. anthropological material, coin, park, painting, traditional music. They have a lot of

shared information types (e.g. location, bibliography, dating), but also many peculiar characteristics (e.g. staircases and floors in a building). Moreover, their description may be very detailed: for a cataloguer is possible to gather information about measurements, exhibitions, documentation, authorship, inventories, relations between cultural properties, and so forth.

There are many projects and models developed in the context of cultural heritage (CH), to model, publish and connect data on the web: CIDOC-CRM [7,8], EDM [9,10], Cultural-ON [11], Fentry [12] and OAEntry [13] ontologies are some relevant examples. A recent paper [14] discusses the main requirements that a model representing cultural heritage should address, based on an analysis of CIDOC and EDM. Although we build on the good practices of such existing effort, our use case required a level of granularity and a diversity of cultural property types that needed new modeling effort.

To build ArCo, we directly reuse classes and properties from the core (roles, agents, locations) modules of OntoPiA [15], an ontology and controlled vocabulary network for Italian Public Administration, and from Cultural-ON, an ontology that models cultural events and sites [16]. We indirectly reuse patterns from existing ontologies, e.g. CIDOC and Cultural-ON and include explicit alignments to them within ArCo.

3 Methodology

In the development of the project, we followed the principles of eXtreme Design (XD) [17], an ontology engineering methodology based on ontology design patterns [18]. Fig. 1 depicts as XD applied to ArCo.

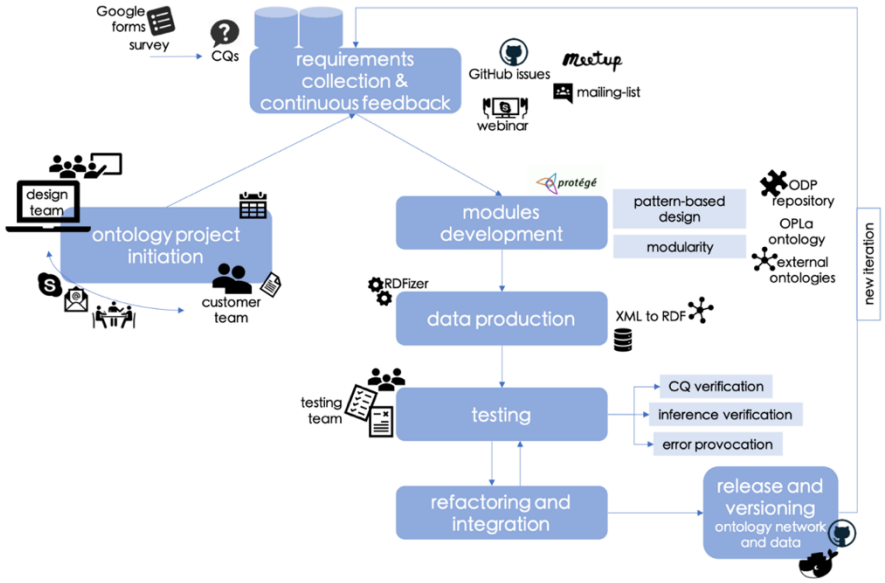

Fig. 1. Implementation of XD methodology in ArCo.

During the project initiation and scoping, domain experts shared with the ontology engineers' team their knowledge of the domain, providing guidelines and data model regulations for interpreting their data. A generic timeline and a release plan with priorities were defined. As recommended, we worked in tight collaboration with our main "customer", i.e. the ICCD. However, given that the ICCD data will be openly published and have high potential for reuse by several other stakeholders, we decided to interact with some representative of them since the very beginning of the process. In addition to domain experts, other agents, such as companies, were involved in the definition of ontology requirements, initially expressed in the form of user stories. The same requirements are reused in the ontology testing phase. Extending XD, four selected companies were also included in an "Early Adoption Program" (EAP) that worked with the incremental unstable releases of ArCo ontologies and data to test them for e.g. publishing their data according to ArCo ontologies, linking their data to ArCo. The EAP members and all the other interested stakeholders created an active community that interacts by means of a dedicated mailing-list [19], GitHub issues tracker [20] and meetups [21].

Pattern-based ontology design plays a central role [22]: by ontology design patterns we mean reusable successful solution to a recurrent modeling problem [23] [24]. XD encourages the reuse of existing ODPs from online repositories [25] as well as the development of new ODPs, when needed. Reused patterns are annotated with OPLa ontology [26], to support users in identification, reuse and ontology mapping.

Since XD is iterative and incremental, ArCo ontology modules and ICCD data are periodically published as unstable releases: this allows us to involve customers and stakeholders in giving us continuous feedback on modeling and testing activities, and to detect new emerging requirements at early stage.

4 ArCo Ontology Network and LOD

4.1 ArCo Release

ArCo release consists of a docker container, available on GitHub [27] and its running instance online [28], which contains:

- the **user guide** accompanying the release, with diagrams and explanations on the content of the release and of each ontology module;
- the **ontologies**, including their source code and a human-readable HTML documentation;
- a **SPARQL endpoint** storing the General Catalogue data in RDF format, generated according to our ontologies;
- examples of **Competency Questions** (CQs), with the corresponding SPARQL queries, for supporting the data query from the community;
- a **RDFizer tool** converting XML data represented according to ICCD cataloguing standards to RDF.

ArCo knowledge graph is also available on the MiBAC official portal [29] with its SPARQL endpoint [30].

4.2 ArCo Ontology Network

ArCo ontology network consists of seven ontology modules connected by *owl:imports* axioms. In Fig. 2, blue circles depict ArCo modules; the green circle indicates directly reused ontologies; the orange circle indicates indirectly reused and aligned ontologies. The network base namespace is https://w3id.org/arco/ontology/, and each module has its own namespace (e.g. https://w3id.org/arco/ontology/core/).

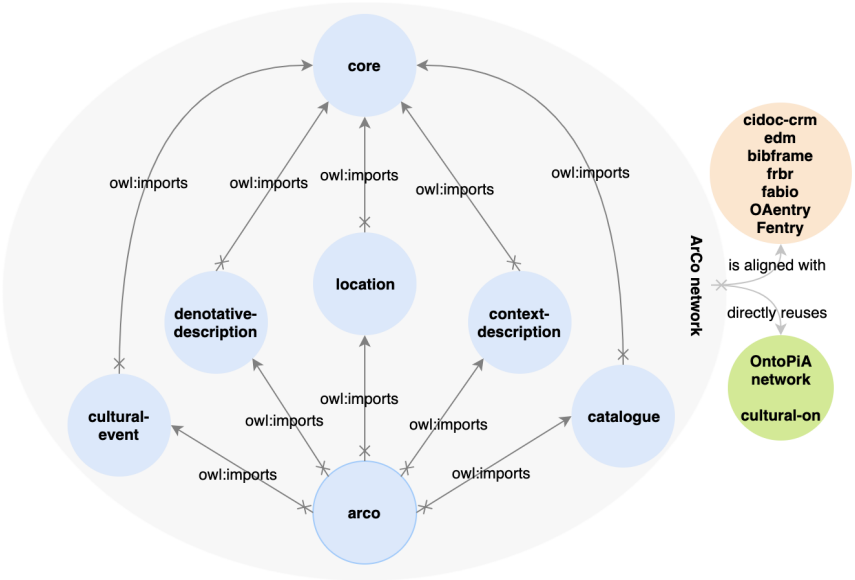

Fig. 2. ArCo ontology network.

The **arco module** [31] represents the network, importing all the other modules. It models top-level concepts from the CH domain, according to the ICCD cataloguing standards [32]. In particular, the hierarchy of the different types of cultural properties is modeled as follows. The top-level class is `:CulturalProperty`, which has two sub-classes `:TangibleCulturalProperty`, and `:IntangibleCulturalProperty`. The first is further specialized in `:MovableCulturalProperty` and `:ImmovableCulturalProperty`.

More specific types of cultural properties are defined as `:DemoEthnoAnthropologicalHeritage`, `:ArchaeologicalProperty`, `:ArchitecturalOrLandscapeHeritage`, `:HistoricOrArtisticProperty`, `:MusicHeritage`, `:NaturalHeritage`, `:NumismaticProperty`, `:PhotographicHeritage`, `:ScientificOrTechnologicalHeritage`, `:HistoricOrArtisticProperty` (see the diagram [33] on Github).

The **core module** [34] represents general concepts orthogonal to the whole network, which are imported by all other ontology modules. This module reuses a number of patterns, such as the *Part-of* [35], the *Classification* [36] and the *Situation* [37] patterns.

The **catalogue module** [38] models concepts related to the ICCD Catalogue, and in particular catalogue records, that is the XML files recording all data gathered by a cataloguer on a particular Italian cultural property. The *Sequence* [39] pattern is reused to model the different versions of the same catalogue record, represented by the class `a-cat:CatalogueRecordVersion`.

The **location module** [40] is intended to cover spatial and geometry information. A cultural property may have multiple locations, represented by the class `a-loc:LocationType`. In addition, the fact that a type of cultural property location holds during a time interval is modeled by the `a-loc:TimeIndexedTypeLocation`, which implements and specialises the *TimeIndexedSituation* [41] pattern.

The **denotative description module** [42] encodes the characteristics of a cultural property observed during the cataloguing process, e.g. measurements, materials, techniques, etc. To represent those characteristics we reused and specialised the *Description&Situation* [43] pattern for modeling both the technical status (`a-dd:CulturalEntityTechnicalStatus`) and the technical description (`a-dd:CulturalEntityTechnicalDescription`) of a cultural property.

The **context description module** [44] represents the context of cultural properties, in a broad sense, including the information related to: authors, collectors, copyright holders, inventories, bibliography, etc. For example, in order to represent the concept of an `a-cd:Archival-RecordSet`, i.e. fonds, series, subseries, etc., we reuse the *Born Digital Archives* [45] pattern.

The **cultural events module** [46] is dedicated to cultural events and exhibitions involving a cultural property. It extends, with some classes and properties (e.g. `a-ce:Exhibition`), the *Cultural-ON* ontology [11].

4.3 ArCo LOD

ArCo knowledge graph currently counts: 7 ontology modules, 327 classes, 379 object properties, 154 datatype properties, 395 restrictions. It counts about 170M triples and provides 24,008 *owl:sameAs* axioms linking to other datasets, such as DBpedia [47], Wikidata [48], the ULAN [49] and TGN [50] Getty Vocabularies, Zeri&LODE [4], YAGO [51], Europeana [52], Geonames [53]. The Entity linking is performed with LIMES [54], and the LIMES configuration files used in the linking process are available on Zenodo [55].

Fig. 3 depicts an example of information of a painting with subject “Madonna con bambino” (tr.en. Madonna with child). On the left side, there is the XML data, expressed as string and stored in the ICCD General Catalogue, and on the right side there is the correspondent data in RDF format generated according to ArCo ontologies.

Fig. 3. An example of XML data from ICCD General Catalogue converted in RDF format according to ArCo ontologies.

5 Impact and Future Work

In order to involve different stakeholders, we have organised a series of meetups associated with the ArCo releases. So far, we had 5 meetups, each attended by about 20 participants, and 1 webinar; we received 35 GitHub issues, and 27 people joined the mailing-list.

ArCo has a potentially very strong impact on both Cultural Heritage and Digital Humanities fields and related domains. At international level, ArCo ontologies allow to represent very detailed information on cultural heritage of many different types and ArCo data can be aligned to other CH data, ensuring a high reliable provenance. These ontology network and dataset will be used by institutions (such as museums, designated for cultural heritage preservation and enhancement), which intend to publish their data as LOD and/or link them to ArCo, as well as by companies and individual consumers (i.e. researchers, students, practitioners, citizens) that own and use CH data for different purposes.

Good examples, among others, of ArCo early adopters are: *Synapta* team [56], which reuses ArCo ontologies for representing musical instruments belonging to *Sound Archives & Musical Instruments Collection* (SAMIC) [57], and *Ricostruzione Trasparente* project [58], which aims at linking its data about areas of Italy damaged by the earthquakes in 2016 to ArCo data.

Currently, an extraordinary amount of data on Italian cultural heritage, in the form of a LOD dataset, is available to anyone interested in querying, consulting and reusing them. ArCo ontologies are released and adopted directly by ICCD, which provides Italian regulations for cataloguing cultural properties. Therefore, ArCo has become, in LOD context, a standard for Italian cultural institutions aiming at creating Linked Data, according to ministerial regulations.

Since the valorization of cultural heritage through LOD enables sharing and reusing of cultural heritage data in an open interconnected and multi-domain knowledge base on the Web, we plan to improve ArCo ontology network and LOD. Future efforts will be directed to: (i) model peculiar information regarding natural heritage and information related to archive and library domains, (ii) improve entity-linking, and (iii) provide tooling support for CH data owners in order to encourage and simplify the adoption of ArCo and other ontologies by domain experts.

References

1. Hyvönen, E.: Semantic Portals for Cultural Heritage. In: Staab S., Studer R. (eds) Handbook on Ontologies. International Handbooks on Information Systems. pp. 757–778, Springer, Berlin, Heidelberg (2009).
2. IBC-ER homepage, https://ibc.regione.emilia-romagna.it/servizi-online/lod, last accessed 2019/05/17.
3. Daquino, M., Mambelli, F., Peroni, S., Tomasi, F., Vitali, F.: Enhancing Semantic Expressivity in the Cultural Heritage Domain: Exposing the Zeri Photo Archive as Linked Open Data. JOCCH 10(4), 1–21 (2017).
4. Fondazione Zeri&LODE homepage, http://data.fondazionezeri.unibo.it/, last accessed 2019/05/17.
5. Europeana Project Homepage, https://pro.europeana.eu/page/linked-open-data, last accessed 2019/05/17.
6. ArCo Project, http://wit.istc.cnr.it/arco, last accessed 2019/05/17.
7. CIDOC-CRM Homepage, http://www.cidoc-crm.org/, last accessed 2019/05/17.
8. Doerr, M.: The CIDOC Conceptual Reference Module: An Ontological Approach to Semantic Interoperability of Metadata. AI Magazine 24(3), 75–92 (2003).
9. Europeana Data Model (EDM) Documentation, https://pro.europeana.eu/resources/standardization-tools/edm-documentation, last accessed 2019/05/17.
10. Charles, V., Isaac, A., Tzouvaras, V. and Hennicke, S.: Mapping Cross-Domain Metadata to the Europeana Data Model (EDM). In: Aalberg T., Papatheodorou C., Dobreva M., Tsakonas G., Farrugia C.J. (eds) Research and Advanced Technology for Digital Libraries. TPDL 2013. Lecture Notes in Computer Science, vol 8092, pp. 484–485, Springer, Berlin, Heidelberg (2013).
11. Cultural-ON Ontology on MiBAC OpenData Website, http://dati.beniculturali.it/lodview/cis/.html, last accessed 2019/05/17.
12. Fentry Ontology, https://essepuntato.github.io/fentry/current/fentry.html, last accessed 2019/05/17.
13. OAEntry Ontology, http://oaentry-ontology.sourceforge.net/index.html, last accessed 2019/05/17.
14. Dijkshoorn, C., Aroyo, L., van Ossenbruggen, J., Schreiber, G.: Modeling cultural heritage data for online publication. Applied Ontology 13(4), 255–271 (2018).
15. OntoPia Ontology Network, https://github.com/italia/daf-ontologie-vocabolari-controllati/tree/master/Ontologie, last accessed 2019/05/17.
16. Lodi, G., Asprino, L., Nuzzolese, A. G., Presutti, V., Gangemi, A., Reforgiato Recupero, D., Veninata, C., Orsini, A.: Semantic Web for Cultural Heritage Valorisation. In: Hai-Jew, S. (eds) Data Analytics in Digital Humanities, Multimedia Systems and Applications, pp. 3–37, Springer, Cham (2017).
17. eXtreme Design, http://extremedesign.sourceforge.net/, last accessed 2019/05/17.
18. Blomqvist E., Presutti V., Daga E., Gangemi A.: Experimenting with eXtreme Design. In: Cimiano, P., Pinto, H.S. (eds) Knowledge Engineering and Management by the Masses, EKAW 2010. LNCS, vol. 6317, pp. 120–134, Springer, Berlin, Heidelberg (2010).
19. ArCo Google Groups, https://groups.google.com/forum/#!forum/arco-project, last accessed 2019/05/17.
20. ArCo Issues Tracker on Github, https://github.com/ICCD-MiBACT/ArCo/issues, last accessed 2019/05/17.
21. Meetup Homepage, https://www.meetup.com/, last accessed 2019/05/17.
22. Presutti, V., Lodi, G., Nuzzolese, A., Gangemi, A., Peroni, S., Asprino, L.: The Role of Ontology Design Patterns in Linked Data Projects. In: Comyn-Wattiau, I., Tanaka, K., Song, IY., Yamamoto, S., Saeki, M. (eds) Conceptual Modeling, ER 2016. LNCS, vol. 9974, pp. 113–121, Springer, Cham (2016).

23. Gangemi, A., Catenacci, C., Ciaramita, M., Lehmann, J.: Modelling Ontology Evaluation and Validation. In: Sure, Y., Domingue, J. (eds) The Semantic Web: Research and Applications, ESWC 2006. LNCS, vol. 4011, pp. 140–154, Springer, Berlin, Heidelberg (2006).
24. Hitzler, P., Gangemi, A., Janowicz, K., Krisnadhi, A.A., Presutti, V.: Ontology Engineering with Ontology Design Patterns: Foundations and Applications, Studies on the Semantic Web, vol. 25, IOS Press (2016).
25. Ontology Design Pattern Homepage, http://ontologydesignpatterns.org/, last accessed 2019/05/17.
26. OPLa ontology on ODP portal, http://ontologydesignpatterns.org/opla/, last accessed 2019/05/17.
27. ArCo release on Github, https://github.com/ICCD-MiBACT/ArCo, last accessed 2019/05/17.
28. ArCo homepage, http://wit.istc.cnr.it/arco, last accessed 2019/05/17.
29. ArCo on MiBAC OpenData, http://dati.beniculturali.it/progetto-arco-architettura-della-conoscenza/, last accessed 2019/05/17.
30. MiBAC Sparql Endpoint, http://dati.beniculturali.it/sparql, last accessed 2019/05/17.
31. ArCo Module Namespace, https://w3id.org/arco/ontology/arco, last accessed 2019/05/17.
32. MiBAC Cataloguing Standards, http://www.iccd.beniculturali.it/it/normative, last accessed 2019/05/17.
33. Cultural Properties Classification on Github, https://github.com/ICCD-MiBACT/ArCo/blob/master/ArCo-release/httpd/public-html/img2/culturalproperty-classification.jpg, last accessed 2019/05/17.
34. Core Module Namespace, https://w3id.org/arco/ontology/core/, last accessed 2019/05/17.
35. Part Of Pattern, http://www.ontologydesignpatterns.org/cp/owl/partof.owl, last accessed 2019/05/17.
36. Classification Pattern, http://www.ontologydesignpatterns.org/cp/owl/classification.owl, last accessed 2019/05/17.
37. Situation Pattern, http://www.ontologydesignpatterns.org/cp/owl/situation.owl, last accessed 2019/05/17.
38. Catalogue Module Namespace, https://w3id.org/arco/ontology/catalogue/, last accessed 2019/05/17.
39. Sequence Pattern, http://www.ontologydesignpatterns.org/cp/owl/sequence.owl, last accessed 2019/05/17.
40. Location Module Namespace, https://w3id.org/arco/ontology/location/, last accessed 2019/05/17.
41. TimeIndexedSituation Pattern, http://www.ontologydesignpatterns.org/cp/owl/timeindexedsituation.owl, last accessed 2019/05/17.
42. Denotative Description Module Namespace, https://w3id.org/arco/ontology/denotative-description/, last accessed 2019/05/17.
43. Description&Situation Pattern, http://www.ontologydesignpatterns.org/cp/owl/descriptionandsituation.owl, last accessed 2019/05/17.
44. Context Description Module Namespace, https://w3id.org/arco/ontology/context-description/, last accessed 2019/05/17.
45. Born Digital Archives Pattern, http://mklab.iti.gr/pericles/BornDigitalArchives_ODP.owl, last accessed 2019/05/17.
46. Cultural Event Module Namespace, https://w3id.org/arco/ontology/cultural-event/, last accessed 2019/05/17.
47. DBpedia Homepage, https://wiki.dbpedia.org/, last accessed 2019/05/17.
48. Wikidata Homepage, https://www.wikidata.org/wiki/Wikidata:Main_Page, last accessed 2019/05/17.
49. ULAN Getty Vocabulary, http://www.getty.edu/research/tools/vocabularies/ulan/, last accessed 2019/05/17.
50. TNG Getty Vocabulary, http://www.getty.edu/research/tools/vocabularies/tgn/, last accessed 2019/05/17.
51. YAGO Knowledge Base, https://www.mpi-inf.mpg.de/departments/databases-and-information-systems/research/yago-naga/yago/, last accessed 2019/05/17.
52. Europeana Linked Open Data, https://pro.europeana.eu/page/linked-open-data, last accessed 2019/05/17.
53. Geonames Homepage, https://www.geonames.org/, last accessed 2019/05/17.
54. LIMES Homepage, http://aksw.org/Projects/LIMES.html, last accessed 2019/05/17.
55. ArCo Entity Linking on Zenodo, https://zenodo.org/record/2630565#.XNhq69MzYUs, last accessed 2019/05/17.
56. Synapta Homepage, https://synapta.it/, last accessed 2019/05/17.
57. Sound Archives & Musical Instruments Collection Homepage, http://museopaesaggiosonoro.org/sound-archives-musical-instruments-collection-samic/, last accessed 2019/05/17.
58. Ricostruzione Trasparente Project, http://ricostruzionetrasparente.it/, last accessed 2019/05/17.