Semantic Web for Cultural Heritage Valorisation

Giorgia Lodi, Luigi Asprino, Andrea Giovanni Nuzzolese, Valentina Presutti, Aldo Gangemi, Diego Reforgiato Recupero, Chiara Veninata, and Annarita Orsini

Abstract Cultural heritage consists of heterogeneous resources: archaeological artefacts, monuments, sites, landscapes, paintings, photos, books and expressions of human creativity, often enjoyed in different forms: tangible, intangible or digital. Each resource is usually documented, conserved and managed by cultural institutes like museums, libraries or holders of archives. These institutes make available a detailed description of the objects as catalog records. In this context, the chapter proposes both a classification of cultural heritage data types and a process for cultural heritage valorisation through the well-known Linked Open Data paradigm. The classification and process have been defined in the context of a collaboration between the Semantic Technology Laboratory of the National Research Council (STLab) and the Italian Ministry of Cultural Heritage and Activities and Tourism (MIBACT) that the chapter describes, although we claim they are sufficiently general to be adopted in every cultural heritage scenario. In particular, the chapter introduces both a suite of ontology modules named Cultural-ON to model the principal elements identified in the cultural heritage data type classification, and the process we employed for data valorisation purposes. To this end, semantic technologies are exploited; that is, technologies that allow us to conceptualise and describe the meaning of data forming the cultural heritage and including such entities as places, institutions, cultural heritage events, availability, etc. These entities have special characteristics and are connected with each other in a profound way. The result is a knowledge base consisting of semantic interconnections with also other data available in the Web to be exploited according to different tasks

G. Lodi (✉) • L. Asprino • A.G. Nuzzolese • D. Reforgiato Recupero
STLab, ISTC-CNR, Via San Martino della Battaglia 44, Rome, Italy
e-mail: giorgia.lodi@istc.cnr.it; giorgia.lodi@gmail.com; luigi.asprino@istc.cnr.it; andrea.nuzzolese@istc.cnr.it; diego.reforgiato@istc.cnr.it

V. Presutti • A. Gangemi
STLab, ISTC-CNR, Via Gaifami 18, Catania, Italy
e-mail: valentina.presutti@cnr.it; aldo.gangemi@cnr.it

C. Veninata • A. Orsini
Ministry of Cultural Heritage and Activities and Tourism, Via del Collegio Romano 27, Rome, Italy
e-mail: chiara.veninata@beniculturali.it; aorsini@beniculturali.it

S. Hai-Jew (ed.), *Data Analytics in Digital Humanities*, Multimedia Systems and Applications, DOI 10.1007/978-3-319-54499-1_1

and users preferences. By navigating the semantic relationships between the various objects of the knowledge base, new semantic paths can be revealed and utilised with the aim to develop innovative services and applications. The process is compliant with Linked Open Data and W3C Semantic Web best practices so that to enable a wider promotion of cultural heritage, and of sharing and reuse of cultural heritage data in the Web. The chapter concludes presenting a number of methodological principles and lessons learnt from the STLab/MIBACT collaboration that are applicable to any cultural heritage context and, in some cases, also to other domains.

Introduction and Motivation

Cultural heritage—both tangible and intangible—is a driving force for economic growth and societal development, an irreplaceable means for building a population's identity and culture, an inspiration for creative industry, and a primary reference for the touristic market. It has been formed by a multitude of local cultures and stories determined by the inhabitants of countries and different territories. This entailed that, over time, cultural heritage has been enriched with a highly heterogeneous set of resources that are diverse in their nature and in the way they are managed by each local culture.

Despite this diversity, cultural heritage's elements are semantically interconnected with each other as the result of contaminations of the different cultures. However, these elements are still mostly managed separately like silos, typically by different organisations (e.g., museums, libraries, holders of archives, etc.) that may also use various classifications, definitions and approaches to describe the same elements. An example is the Italian case where, within the same Ministry of Cultural Heritage and Activities and Tourism (MIBACT), two separate institutes define the concept of "cultural site" in two different ways within the data bases they own, respectively. Overall, this led to a scenario of dispersed data, both at national and international level, that cannot be fully and smartly exploited for such an economic growth and development of the touristic market that cultural heritage can potentially boost.

In the last two decades, web technologies and social networks became pervasively present in our daily life, providing a powerful means for sharing personal knowledge and experiences. Potentially, these technologies are the perfect tools for creating a vast interconnected digital cultural heritage that could be open to anyone so as to promote and improve the social cohesion and economic growth. The advent of new paradigms for data management based on these technologies can also support the involvement of citizens for identifying what is relevant as territorial cultural heritage, enhancing its distinctiveness, and cross-border sharing of local traditions and knowledge. This is key for supporting social integration and recognition, and sense of belonging. In particular, in the cultural domain the Open Data movement and the application of Linked Open Data (LOD) as the main data integration paradigm are emerging as new frontiers to valorise a vast amount of public cultural heritage. However, a diversified scenario where isolated islands of data are available on the Web is still observable. This is the case for instance of the Italian scenario: a number of LOD projects carried out by MIBACT were born but they still struggle to be part of an overall strategy that can guide towards a significant valorisation of one of the largest heritage of the cultural sector.

This chapter describes the work conducted in the context of a collaboration between STLab and MIBACT. The work aims at defining an overall framework for cultural data analysis and valorisation, this latter pursued through the construction of an open knowledge base of Italian cultural data. The knowledge base is meant to be publicly available on the Web, using semantic web standards and technologies, for the reuse by anyone willing to exploit it for the development of innovative and creative applications and services. In computer science, the term "semantic web" refers to W3C's vision of an extension of the Web for dealing with interconnected data that is stored, modelled and handled using so-called semantic web standards and technologies (e.g., Resource Description Framework (RDF), SPARQL query language for RDF (SPARQL), Ontology Web Language (OWL), etc.) The framework we present provides a classification of the cultural data landscape, identified within the collaboration, and a common and standard-based data model that can be used as shared interoperability profile for interconnecting different, even available LOD, datasets managed by various institutes of the Ministry. The model consists of a suite of ontology modules for the cultural domain named Cultural-ON, which are defined following a specific methodology. The methodology has been delineated so as to guarantee sustainability, usability and semantic interoperability requirements, with the application of ontology design patterns. The chapter focuses on the module of the suite that represents "cultural institutes and sites" and "cultural events", which are connected to each other. The process we employed to create LOD datasets of the cultural institutes and of the events, aligned with the defined ontology module, is introduced. In particular, we highlight how to integrate it in a real business process and information system managed by the Ministry, and how to link the datasets to others available in the Web of Data. We claim that our model is sufficiently general to be applied in any organisation working in the cultural sector with such data types as cultural sites and events. We then highlight general methodology principles and best practices that can be replicated in other cultural data management processes and, in some cases, also in other domains.

This chapter is structured as follows. Section "State of the Art" provides a state of the art on projects, ontologies and LOD approaches in the cultural heritage domain. Section "A Cultural Heritage Data Framework: The Case of the Italian Ministry of Cultural Heritage" introduces a cultural data landscape analysis and the objectives of the collaboration we carried out with MIBACT. Section "Cultural-ON: The Cultural ONtologies" describes the methodology we followed to define an ontology for cultural institutes and sites and cultural events, detailing the main modelling choices and the design patterns that we reused for supporting semantic interoperability. Section "The Produced Linked Open Data" provides a detailed description of the process we employed for producing LOD datasets, aligned with the ontology and linked to other data in the Web, and a running example with real data coming from an information system available at MIBACT. Additional methodology principles we learnt from this experience are discussed in section "Lesson Learnt (On Methodological Principles)". Finally, section "Concluding Remark" provides concluding remarks and future directions of the work.

State of the Art

Semantic technologies, and in particular Linked Open Data, have been widely and successfully exploited within the cultural heritage field to improve the access to cultural assets, to enhance the experience of citizens in exploring cultural heritage, to facilitate artworks discoverability, to integrate and enrich data about cultural heritage (Isaac and Haslhofer 2013; de Boer et al. 2013; Szekely et al. 2013; Ruotsalo et al. 2013; Aart et al. 2010). Semantic Web technologies have been used to disclose, in a scalable way, the vast amount of interesting and useful information on cultural heritage. In recent years, several institutions put considerable effort into developing and standardising methods and vocabularies for indexing objects. Semantic Web technologies have given the opportunity of easing these processes and achieving results in a collaborative environment. The collaborative development of ontologies [e.g. CIDOC-CRM (Doerr 2003)] has strengthened the collaboration between organisations so that semantic interoperability requirements could be met within their systems. In addition, the use of common ontologies facilitated the data exchange and the creation of huge digital libraries [e.g. (Isaac and Haslhofer 2013)].

In the context of Semantic Web technologies, the Linked Open Data (LOD) paradigm paved the way to the valorisation of cultural assets. In fact, the LOD paradigm has been used to connect data from different cultural institutions, thus increasing the possibility of reaching cultural data available in the Web of Data. The interlinking of contents of collaborating organisations also contributed to enrich the information in a cost effective manner (Hyvönen 2009). The result has been the creation of a knowledge base that can be reused in different applications.

Semantic annotation of cultural objects has been employed to support indexing and searching within large virtual collections (Schreiber et al. 2008). The indexing of collections has benefited of crowdsourcing techniques where large community of users were invited to assist in the selection, cataloguing, contextualisation, and curation of collections (Oomen and Aroyo 2011; Ridge 2013). Furthermore, semantic technologies allowed going beyond traditional free text search (e.g. Google), providing users with "intelligent" facilities based on ontological reasoning such as semantic search, semantic autocompletion or semantic recommendation of contents (Hyvönen 2009; Ruotsalo et al. 2013; Wang et al. 2007).

Ontologies and Other Knowledge Organisation Systems

In the context of cultural heritage and Semantic Web technologies a lot of ontologies and knowledge organisation systems (KOS) have been developed. We describe here a representative sample of them, which have influenced the design and the methodology we propose in this chapter. In general, we noticed that the most important and used ontologies of the cultural sector are principally focused on modelling cultural heritage objects; that is, movable objects hosted/located in cultural sites.

CIDOC-CRM

CIDOC-CRM (Conceptual Reference Model)[1] (Doerr 2003) is a formal and standard ontology intended to enable information integration for cultural heritage data. Its shared usage can contribute to meet semantic interoperability requirements among heterogeneous sources of cultural heritage information. CIDOC-CRM is the result of a long-term interdisciplinary work and an agreement between the International Committee for Documentation (CIDOC) of the International Council of Museums (ICOM) and a non-governmental organism. Since September 2006 it has been considered for standardisation by the ISO standard body that defined the ISO 21127. CIDOC-CRM has a rich taxonomy of classes and properties that describe Space-Time, Events, Material and Immaterial Things for the cultural domain. This makes the model very useful for building query services (queries can be formulated at various granularities) and eases the extension of the model to other domains, reducing the risk of over-generalisation.

The ontology proposed in this paper has points of convergence and divergence with CIDOC-CRM. On the one hand, the convergences between the models can be found in the way of modelling the material things (such as places and physical objects), collections and catalogues (in CIDOC, *E31 Document*). Both the ontologies model information about places, such as place names, addresses, spatial coordinates, geometries. In fact, the classes representing places in our proposal aligned to CIDOC-CRM's (cf. section "Methodology"). Since the intent of CIDOC is to document objects collected and displayed by museums and related institutions, it provides a finer description of individual items and collections than Cultural-ON. However, the *Cultural Heritage Objects* defined in Cultural-ON can be used as a hook to vocabularies more focusing on the description of artworks (such as CIDOC).

On the other hand, CIDOC lacks of a finer conceptualisation of the access to cultural sites (e.g. opening hours specifications, ticketing information, access conditions) and about organisational structures behind a cultural site, which is addressed by Cultural-ON.

[1] http://cidoc-crm.org/.

Europeana Data Model (EDM)

The Europeana Data Model (EDM)[2] (Isaac and Haslhofer 2013) is a data model used in the Europeana project[3] for integrating, collecting and enriching data on cultural heritage provided by the different and distributed content providers. EDM reuses and extends a set of well-known vocabularies: OAI ORE (Open Archives Initiative Object Reuse and Exchange) for organising objects' metadata and digital representation(s); Dublin Core for descriptive metadata; SKOS (Simple Knowledge Organisation System) for conceptual vocabulary representation and CIDOC-CRM for the representation of events and relationships between objects. The model distinguishes between cultural heritage objects, web resources and aggregations. Aggregations are used to model a data provider's contribution to Europeana on a cultural heritage object. A contribution, in turn, consists of more web resources (i.e., metadata) provided for the object.

As well as CIDOC-CRM, the main focus of the Europeana Data Model is on the description of Cultural Heritage Objects. The information modelled by EDM is orthogonal to the conceptualisation provided by *Cultural-ON*. Therefore, the data modelled with Cultural-ON enriches the information about Cultural Heritage Objects, collected by Europeana.

Getty Vocabularies

The Getty vocabularies[4] is a suite of KOS developed by the Getty Research Institute, consisting in four controlled vocabularies, i.e. The Art and Architecture Thesaurus (AAT) containing generic terms used for describing items of art, architecture and cultural heritage objects; The Getty Thesaurus of Geographic Names (TGN) including names descriptions of historical cities, empires and archaeological sites; The Cultural Objects Name Authority (CONA) containing titles, attributions, depicted subjects about works of several art, architecture, and other elements of the cultural heritage; The Union List of Artist Names (ULAN) including information about artists. AAT, CONA and ULAN are orthogonal to *Cultural-ON*. They can be exploited by data providers for enriching their datasets with information about authoritative research institutes. Users of *Cultural-ON* can take advantage of TGN by reusing its terms to provide a site with its historical nomenclature. Besides providing coordinates of places (also modelled in *Cultural-ON*), CONA provides a finer conceptualisation of the name attribute by modelling different *name types* (e.g. Official, Vernacular, Provisional, etc.). Cultural-ON models three types of names: institutional, preferred and alternative.

---

[2] http://pro.europeana.eu/edm-documentation.

[3] http://pro.europeana.eu/.

[4] http://www.getty.edu/research/tools/vocabularies/index.html.

Projects

The above mentioned ontologies and KOS were successfully used in a number of projects, carried out by cultural public organisations. In this section we report some of them, which we have reused as guidelines for our work.

Europeana

Europeana[3](Isaac and Haslhofer 2013) is a project co-funded by the European Union with the ambition of "unifying" Europe through culture and making cultural heritage available for all. Europeana Foundation promotes and co-founds within it several projects, such as Europeana Vx, Europeana Creative, Europeana Space, Europeana Food and Drink. Among these, *Europeana Vx* (the last project of the series is Europeana V3.0) aims at both creating and coordinating a network of contributing institutions that act as data providers, and managing the collected data through the development of a service infrastructure that enables an ease data discoverability and access. In October 2012, a subset of the Europeana dataset was released in linked open data, under the license Creative Commons CC0. The dataset was modelled using the earlier mentioned Europeana Data Model (EDM) and then made available in data.europeana.eu. The overall repository currently contains over 40 millions of RDF (Resource Description Framework) cultural data. The data that will be modelled through *Cultural-ON* will be able to enrich the Europeana knowledge base. In fact, Europeana mainly focuses on description and aggregation of metadata about cultural objects, whereas *Cultural-ON* is able to model the information for accessing them.

Smartmuseum

SMARTMUSEUM[5] (Ruotsalo et al. 2013) is a research project funded under the European Commission's 7th Framework Programme. The project aims at developing a mobile ubiquitous recommender system that uses the Web of Data to provide tourists with recommendations for cultural heritage. The system delivers the right content to the users by taking as input users' informational needs and contextual information (such as sensors data captured from mobile devices). Semantic technologies are employed to deal with the heterogeneity of data, which may include content descriptions, sensor inputs and user profiles. For modelling users, the project relies on the General User Model Ontology (GUMO). The content are annotated using the Dublin Core metadata standard which is properly extended for the cultural heritage domain including such metadata as material,

[5]http://www.smartmuseum.eu.

object type, and place of creation of the object. The geographical information is modelled using the W3C Geo Vocabulary[6], whereas the objects are indexed with the Getty Vocabularies[7] [i.e., The Art and Architecture Thesaurus (AAT), The Getty Thesaurus of Geographic Names (TGN), The Union List of Artist Names (ULAN)]. The applications developed within *SMARTMUSEUM* can benefit of the data that will be modelled through *Cultural-ON*. In fact, the description and the localisation of the *Cultural Entities* modelled by *Cultural-ON* can help for providing users with more appropriate context information.

Museums and Linked Data

A growing number of museums are relying on semantic technologies for multiple purposes: for providing Internet access to their collections and artworks, for making their assets easily discoverable, for facilitating the integration and enrichment of their data.

In 2011 the *British Museum*[8] released its database of nearly 2 millions of records in linked open data, becoming the first UK art organism to use semantic technologies for making available in the Web its cultural assets. Today the dataset contains over 100 millions of records and is integrated with other data that are present in the Web of Data. The data have been modelled using CIDOC-CRM[1].

In order to make its artworks easily discoverable, in 2014 the *Smithsonian American Art Museum*[9] began to publish authoritative records about American artworks in linked open data. As in the other cases, the records, that have been collected since 1829, are currently modelled using CIDOC-CRM.

To open up and improve the access to its collection of 1 million objects, in 2011 the *Rijksmuseum*[10] in Amsterdam made parts of its collection available online. The Linked Data version of the collection is modelled according to the Europeana Data Model and comprises over 20 millions of RDF data describing almost 600 thousands cultural objects. Furthermore, the collection objects is linked to the *Art and Architecture Thesaurus* of the Getty Vocabularies. The vocabularies provides a structured terminology for art, architecture, decorative arts, archival materials, visual surrogates, conservation, and bibliographic materials.

All these museums can benefit from using *Cultural-ON* for modelling information (such as addresses, contact points, opening hours, ticketing, etc.) for accessing their cultural assets, information about their organisational structure, the services they provide to the public and the events they host.

[6]http://www.w3.org/2003/01/geo/.

[7]http://www.getty.edu/research/conducting_research/vocabularies/.

[8]http://collection.britishmuseum.org/.

[9]http://americanart.si.edu/collections/search/lod/about/.

[10]http://rijksmuseum.sealinc.eculture.labs.vu.nl/home.

A Cultural Heritage Data Framework: The Case of the Italian Ministry of Cultural Heritage

The Italian Cultural Heritage is one of the richest in the world, as it is also evidenced by the fact that Italy includes the greatest number of UNESCO World Heritage Sites (i.e., 50, including 46 cultural sites).

The Ministry of Heritage, Cultural Activities and Tourism (referred to as MIBACT from now on) is responsible for supervising, preserving and promoting this huge patrimony. It also produces and manages a vast amount of information regarding the institutions and organisations operating for the promotion of the cultural heritage. In order to understand the characteristics of this large patrimony, a general assessment of the cultural data landscape becomes crucial.

Analysis of the Cultural Heritage Data Landscape

The cultural heritage domain is characterised by a highly heterogeneous data landscape that embraces a variety of resource types. According to the different objectives and priorities in treating those resources, different classification proposals for them are possible. In the context of this chapter, we propose a general classification of data types we encountered when we dealt with the data owned by MIBACT. In particular, we distinguish between the following principal data categories:

- *Cultural property*—this category includes all types of data regarding the cultural heritage and landscape that belong to the State, territorial government bodies, as well as any other public body and institution, and to private non-profit associations.
- *Cultural events and exhibitions*—this category includes events that are related to cultural properties.
- *Restrictions*—this category embodies all those types of data that are related to legal constraints or restrictions defined on cultural properties for conservation and safeguarding purposes.
- *Statistical data*—this category includes statistical data that can be collected about the exploitation of, and access to, cultural properties.

Each category can be further specialised in additional subcategories. In the context of this work we focus on the cultural property and cultural events and exhibitions categories, only, and we provide in the following a more detailed characterisation of them.

Cultural Property Cultural properties can be classified in (1) cultural institutes and sites and (2) cultural heritage collections and objects. The *cultural institutes and sites* are assets that encompass immovable objects and their organisations, these latter referred to as the set of people and procedures governing the immovable objects. A taxonomy for these objects can be identified. In particular, in the Italian context, the legal framework for the cultural heritage domain defines six main types of immovable objects that must be safeguarded and preserved:

1. *Museum*; that is a permanent structure or institution that acquires, conserves and makes available to the public cultural objects (see below) for educational purposes;
2. *Library*; that is a permanent structure or institution that collects and conserves an organised collection of books and any other published and editorial work available on any media type; a library also ensures the consultation of those works for study purposes;
3. *Holder of Archive*; that is a permanent structure or institution that collects, makes an inventory and preserves documents that own an historical value, ensuring the consultation of those documents for research and study purposes;
4. *Archaeological Area*; that is a site (or a group of physical sites) consisting of natural ruins or historical or prehistoric artefacts and structures;
5. *Archaeological Park*; that is an environment characterised by the presence of important archaeological evidence; it can be considered as an outdoor museum;
6. *Monumental Area or Complex*; that is a set of different buildings built in different eras that acquired over time, and as a whole, a relevant and autonomous artistic, religious, historic and ethno-anthropological value.

In addition to the above typologies, *Cinemas* and *Theatres*, *Cultural Landscape Assets*; i.e., buildings and areas which are the expression of historical, cultural, natural, morphological and aesthetic values of the land (e.g., villas, gardens, vantage points, belvederes, etc.) and *Cultural Research and Promotion Center* fall into the cultural institutes and sites category.

The *cultural heritage collections and objects* can be referred to as the movable cultural objects. The objects are created or collected by people and form an important part of a nation's identity. They can be artistic, historical, technological or natural in origin and located within immovable objects. Examples include for instance paintings, sculptures, amphoras, manuscripts, photographs, books, films, and any other tangible or intangible work with a notable public and cultural interest.

Cultural Events and Exhibitions Cultural events and exhibitions are events that may take place within immovable objects and may be related to specific movable objects. They are typically organised by public bodies or institutions or by private organisations operating for the public interest, with the aim to promote the culture and traditions of a nation. There might exist large scale cultural events that consist in turn of smaller events usually held at local level. An example can be the large scale event created to celebrate the first world war centenary that consists of single dispersed events hosted within cultural institutes or sites available on the territory.

Transversal Data All earlier described data are connected to transversal data types; that is, data that are independent of the specific cultural heritage domain. Examples of these data include, among the others, people, organisations (public or private bodies), location and addresses. Although these types of data may appear less important than the previously defined ones when describing the cultural heritage, they however represent important conjunction points towards other data bases and domains.

Objectives

Traditionally, all these data have been managed separately, thus contributing to the creation of isolated data bases that rarely fully interoperate and that include, in some cases, replication of the same data. In the light of this scenario, the following objectives were identified as crucial for the cultural heritage valorisation process.

- *Cultural heritage unique knowledge*—definition of a general methodology for data integration capable of re-establishing the *unicum* nature of the cultural heritage knowledge. The methodology ensures that technical and semantic interoperability requirements are met when data exchanged among data bases on cultural properties is enabled;
- *Openness*—valorisation of the cultural properties by exploiting novel Open Data paradigms. In particular, the Linked Open Data (LOD) approach, based on the use of standards of the Semantic Web, can be used to create a large open and interconnected knowledge base, This can be offered to anyone, without any legal restrictions, in order to enable the development of innovative services and applications. In the Italian context, there already exist LOD initiatives in the cultural heritage domain[11] promoted by different internal institutes of the main Ministry. These initiatives have the merit to use standard ontologies and thesaurus for data definition and classification (see section State of the Art). However, they all still suffer from the dispersed nature of the cultural properties management. The objective is then to use the LOD paradigm for creating interoperable cultural properties data, linkable among each other and with other data available on the Web. These links can facilitate the integration of data in a distributed manner and can be explored so as to reveal unexpected semantic paths.
- *Common data models*—definition of common and standard data models for cultural heritage domain to be re-used by all stakeholders operating at national and local level. In particular, from the assessment of the data landscape as previously introduced, it turns out that the cultural institutes and sites are recurring data with respect to the other data types. Thus, they can be considered as the basis or nucleus of a unique LOD knowledge base of cultural properties that can be used for starting the data integration and valorisation process.

[11] http://dati.culturaitalia.it/?locale=en, http://dati.acs.beniculturali.it/, http://san.beniculturali.it/web/san/dati-san-lod.

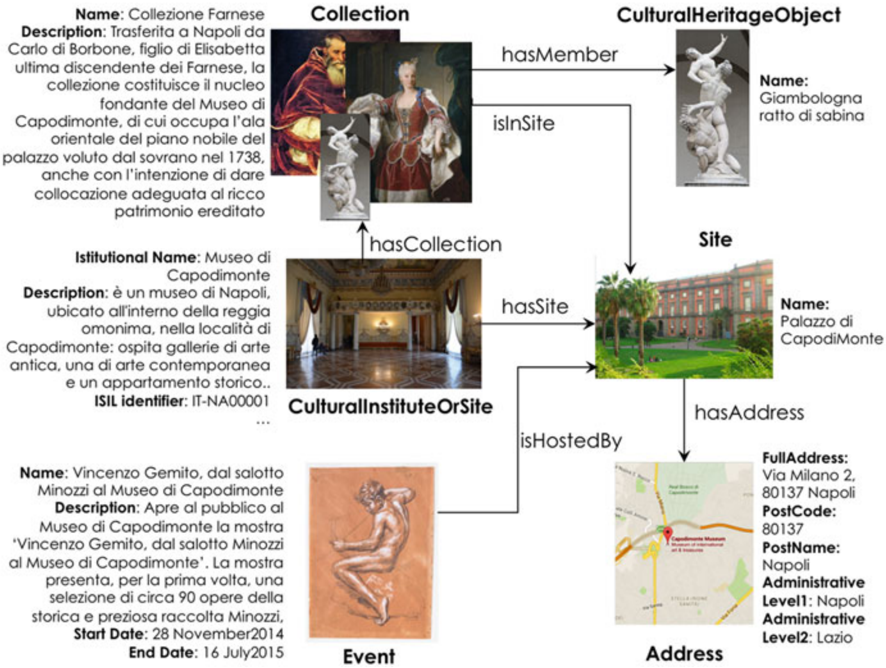

Fig. 1 An example of possible real data on cultural heritage data and their possible relationships

Figure 1 shows an example of real data related to an Italian museum, i.e., "Museo di Capodimonte", which is use as real running example in this chapter, The figure illustrates the real data associated to the principal objects that characterise the museum and the possible relationships among those objects.

Cultural-ON: The Cultural ONtologies

Cultural-ON (Cultural ONtologies)[12] is a suite of ontology modules that provide concepts and relations for modelling knowledge in the cultural heritage domain. At the time of this writing, Cultural-ON embodies one module that allows us to represent cultural institutes and sites and cultural events. In the following sections we provide details about the adopted design methodology (cf. section "Methodology"), the design requirements (cf. section "Design Requirements") and the resulting ontology module (cf. section "Resulting Ontology").

[12]The ontology and its documentation can be retrieved on-line at http://stlab.istc.cnr.it/documents/mibact/cultural-ON_xml.owl and http://goo.gl/2lqj0w.

Methodology

Cultural-ON is designed by following best design practices and pattern-based ontology engineering aimed at extensively re-using Ontology Design Patterns (ODPs)[13] (Gangemi and Presutti 2009) for modeling ontologies. According to Gangemi and Presutti (2009), ODPs are modeling solutions that can be re-used in order to solve recurrent ontology design problems. Hence, they enable design *by-reuse* methodologies. For example, ODPs can be reused by means of their specialisation or composition according to their types and to the scope of the new ontology that is going to be modelled. The design methodology that we followed is based on the *eXtreme Design* (XD) (Blomqvist et al. 2010). XD is an agile approach to ontology engineering that adopts competency questions (CQs) (Grüninger and Fox 1995) as a reference source for the requirement analysis. In fact, XD associates ODPs with generic use cases that are used to match the design requirements. Both the generic use cases and the requirements are expressed by means of competency questions.

Based on this methodology, we modelled Cultural-ON by following these steps:

- **Design requirements elicitation**. The result of this step is a set of requirements expressed in terms of CQs;
- **Key entities extraction**. The result of this step is a set of concepts and relations that provide the basic building blocks for dealing with knowledge within the domain of cultural heritage. The key entities are extracted by analysing the CQs;
- **ODPs identification**. This step is aimed at identifying the most appropriate ODPs. The ODPs are identified by applying the XD methodology and come from the on-line repository ontologydesignpatterns.org;
- **ODPs re-use**. The ODPs resulting from the previous step are modelled in Cultural-ON by means of specialisation. However, we do not directly import any specific implementation of ODPs in our ontology, in order to be completely independent of them (see section "Lesson Learnt (On Methodological Principles)" for a detailed discussion on reuse);
- **Ontology alignment**. Finally, the ontology is aligned to other existing ontologies in the Semantic Web that provide similar conceptualisations.

Design Requirements

The design requirements that we identified are expressed as competency questions. These CQs were extracted by analysing a set of scenarios and real use cases provided us by domain experts, i.e., the Italian Ministry of Cultural Heritage and

[13] http://www.ontologydesignpatterns.org.

Table 1 Design requirements expressed as CQs and identified target ODPs.

ID	Competency question	ODPs
CQ1	What are the cultural institutes or sites of a given type?	Classification
CQ2	What is the name used in a specific period of time for identifying a certain cultural institute or site?	TimeIndexedSituation
CQ3	What are the contacts that can provide information (e.g., opening hours, ticket fares, etc.) about a cultural institute or site?	–
CQ4	What are the opening schedule of a cultural institute or site?	TimeIndexedSituation
CQ5	When can a person contact a reference agent in order ask her information about a cultural institute or site?	TimeIndexedSituation
CQ6	Who is the agent holding a juridical role in a certain period of time for a cultural institute or site?	TimeIndexedPersonRole
CQ7	What is the physical location of a cultural institute or site?	Place
CQ8	Where is a cultural heritage object preserved ?	Place
CQ9	What is the collection of cultural heritage objects preserved by a cultural institute?	Collection
CQ10	What is the description of a certain entity (e.g., a cultural institute)?	Description
CQ11	What are the events organised by a cultural institute or site?	TimeIndexedSituation
CQ12	What are the events associated with cultural heritage objects	TimeIndexedSituation
CQ13	What is the price of the ticket for entering a cultural institute or an event?	Price
CQ14	What is the validity of a ticket for entering a cultural institute or an event?	TimeIndexedSituation
CQ15	What is the cultural institute that another cultural institute belongs to?	PartOf

Activities and Tourism (MIBACT)[14], the italian Central Institute for Cataloguing and Documentation (ICCD)[15] and other internal institutes of MIBACT. The CQs extracted are reported in Table 1 and are associated with the ODPs that address them by providing a re-usable modelling solution. Thus, Table 1 reports the results of the steps 1 and 3 of our ontology design methodology (cf. section "Methodology").

[14] http://www.beniculturali.it/.

[15] http://www.iccd.beniculturali.it.

Resulting Ontology

Cultural-ON is modelled as an OWL ontology. The classes of the ontology that represent the core entities (cf. step 2 of our methodology) are the following:

- `cis:CulturalInstituteOrSite`[16] for representing cultural heritage institutes or sites. An individual of this class is, for example, the Uffizi Gallery or Museum of Capodimonte (see next section);
- `cis:CulturalHeritageObject` for representing cultural heritage movable objects. An individual of this class is any cultural asset (even intangible) of the cultural heritage;
- `cis:Agent` for representing any agentive object, either physical, or social. Individuals of this class can be, for example, people, organisations or any juridical entity;
- `cis:Site` for representing physical places. An individual of this class is, for example, the Uffizi in Florence, Italy. In fact, it is the physical site where the Uffizi Gallery is located in;
- `cis:Collection` for representing collections of cultural heritage objects. For example, the Collections of the Uffizi Gallery are individuals of this class;
- `cis:Event` for representing any kind of event that can be hosted by a cultural institutes and can involve cultural heritage objects;
- `cis:Ticket` for representing tickets. An individual of this class can be any token that grants the entrance to a cultural institute or site or the participation to an event.

These core classes are modelled and organised according to the selected ODPs in order to address the list of requirements presented in Table 1. Figure 2 shows the class diagram of the current version of Cultural-ON represented by using the UML notation.

The requirement CQ1 is addressed by re-using the Classification ODP[17] that in our context allows to represent the category that a certain cultural institute or site belongs to. The category is represented by the class `cis:CISType` whose individuals express actual types of cultural institutes or sites, e.g., museum, library, etc. A `cis:CulturalInstituteOrSite` is related to `cis:CISType` by means of the object property `cis:hasCISType` that in our context plays the role of the object property `classification:isClassifiedBy`[18] as defined in the source Classification ODP.

We did not identify a ODP that could be directly reused for addressing CQ3. However, the class `schema.org:ContactPoint`[19] provides a reference

[16]The prefix `cis:` stands for the namespace http://dati.beniculturali.it/cis/.

[17]http://www.ontologydesignpatterns.org/cp/owl/classification.owl.

[18]The prefix `classification:` stands for the namespace http://www.ontologydesignpatterns.org/cp/owl/classification.owl#.

[19]http://schema.org/ContactPoint.

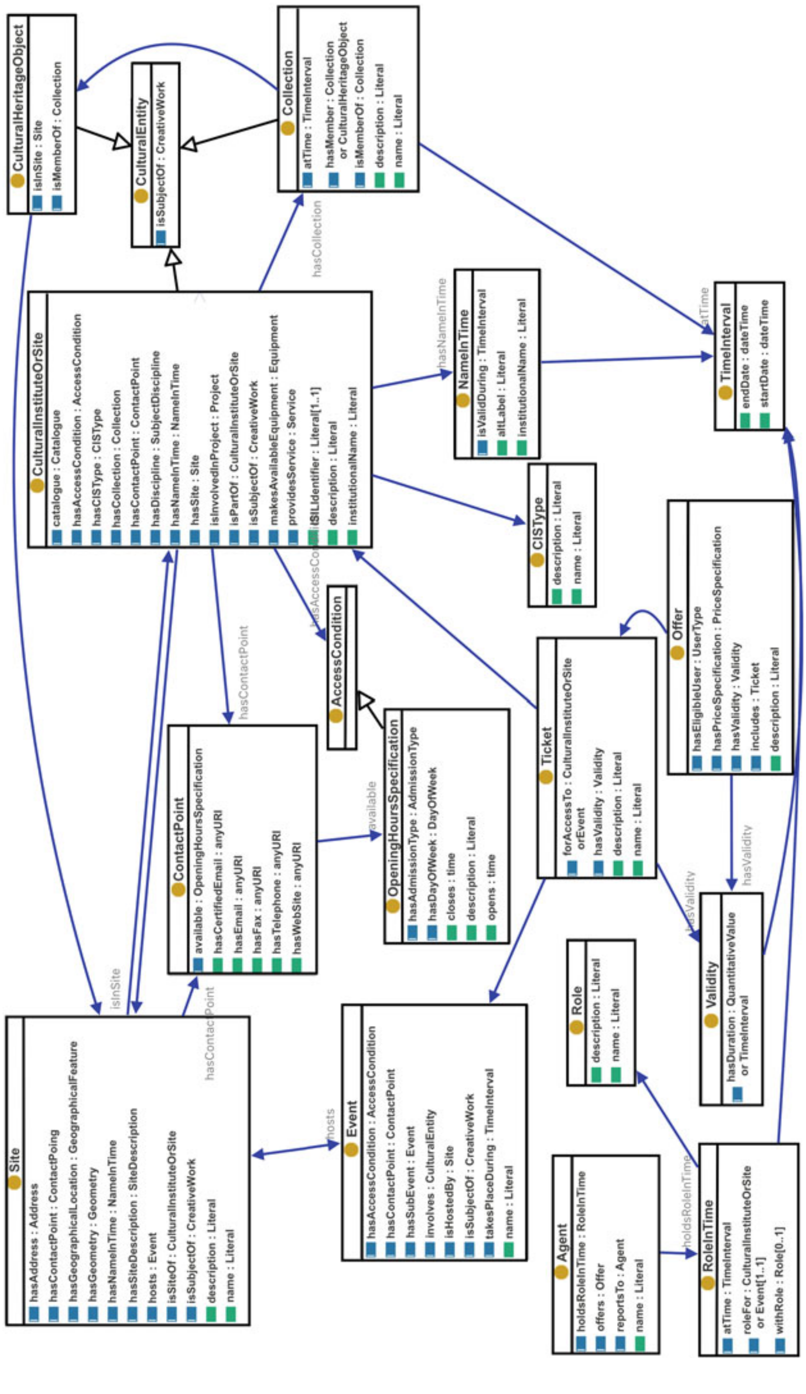

Fig. 2 UML diagram about the core classes and properties defined for modelling Cultural-ON

schema for describing contact points in terms of email addresses, fax numbers, telephone numbers, web sites, opening hours, etc. It is worth mentioning that Schema.org is a collaborative effort aimed at fostering vocabulary sharing over the Web. Hence, Schema.org, similarly to ODP, is a valid source for vocabulary re-use.

The fact the cultural institutes or sites have a name that might change over time (cf. CQ2) is modelled by re-using the TimeIndexedSituation ODP.[20] For example, the Colosseum in Rome was originally known as Amphitheatrum Flavium, while the name further evolved to Coliseum during the Middle Ages. This situations are modelled in Cultural-ON by defining the class `cis:NameInTime` that represents a time indexed situation and allows expressing the temporal validity of the name of a cultural institutes or sites. In fact, a `cis:NameInTime` is associated with (1) a `cis:TimeInterval` by the object property `cis:isValidDuring` and (2) a `cis:CulturalInstituteOrSite` by the object property `cis:hasNameInTime`.[21] The class `cis:TimeInterval` provides the temporal span in terms of a starting and an ending date (i.e., the datatype properties `cis:startDate` and `cis:endDate`, respectively).

The TimeIndexedSituation ODP is also used for addressing the requirements CQ4, CQ5, CQ11, CQ12 and CQ14. For addressing CQ4 and CQ5 we defined the class `cis:OpeningHoursSpecification` that specialises a time indexed situation. Individuals of `cis:OpeningHoursSpecification` can be linked to individuals of `cis:CulturalInstituteOrSite` or `cis:ContactPoint` by the object properties `cis:hasAccessCondtion` and `cis:available`, respectively. In this case the time is represented by an `xsd:time` that is the range of the datatype property `cis:opens`. The property `cis:opens` has the class `cis:OpeningHoursSpecification` as domain. For addressing CQ11 and CQ12 we defined the class `cis:Event` as a time indexed situation. In fact, the temporal dimension is provided by the object property `cis:takesPlaceDuring` whose domain and range are the classes `cis:Event` and `cis:TimeIntervall`, respectively. Instead, the location hosting the event is expressed by the object property `cis:isHostedBy`[22] whose domain and range are the classes `cis:Event` and `cis:Site`, respectively. Finally, in the case of the requirement CQ14, the time indexed situation is represented by the class `cis:Validity`. This class is associated with a `cis:TimeInterval` by means of the object property `cis:hasDuration`. Additionally, it is the range of the object property `cis:hasValidity`. The property `cis:hasValidity` accepts `cis:Ticket` or `cis:Offer` as domain.

[20] http://www.ontologydesignpatterns.org/cp/owl/timeindexedsituation.owl.

[21] The object property `cis:hasNameInTime` has `cis:CulturalInstituteOrSite` and `cis:NameInTime` as domain and range, respectively. Thus, the association of a `cis:NameInTime` with a `cis:CulturalInstituteOrSite` is expressed by the inverse of `cis:hasNameInTime`.

[22] Its inverse property is `cis:Hosts`.

The ticket pricing (cf. CQ13) is modelled by re-using the Price ODP.[23] In fact, we defined a class `cis:Offer` that is associated with (1) a `cis:Ticket` by the object property `cis:includes`; (2) a `cis:PricesSpecification` by the object property `cis:hasPriceSpecification`. The class `cis:PricesSpecification` specialises the original class `price:Price`[24] from the Price ODP and allows to describe a price in terms of an amount and a given currency. The amount can be specified by using the datatype property `cis:amount` whose range is `xsd:double`, while the currency can be specified by using the object property `cis:hasCurrency` whose range are individuals of the class `cis:Currency`.

The approach adopted for modelling the physical site where (1) a cultural institution is located in (cf. CQ7) or (2) a cultural heritage object is preserved in (CQ8), benefits from the re-use of the Place ODP.[25] The physical locations are represented as individuals of `cis:Site`. The object property `cis:hasSite` is designed to associate a `cis:CulturalInstituteOrSite` with a `cis:Site`. In our domain this representation provides a best practice for disambiguating between a cultural institute and its physical place. For example, it allows to represent the "Uffizi" both as an institution (i.e., the "Uffizi Gallery") and a building.

A `cis:Collection` allows modelling collections of cultural heritage objects included in a cultural institute or site and to refer to them as cultural heritage entities. A `cis:Collection` is designed by reusing the Collection Entity ODP.[26] This ODP provides a reference schema for identifying a collection as a whole (i.e., an individual of the class `cis:Collection`) and for associating such a collection with its members by means of the object property `cis:hasMember`.[27] The property `cis:hasMember` has individuals of the class `cis:Collection` as domain while its range can be any individual of `cis:CulturalHeritageObject` or `cis:Collection`. It is worth saying that the class `cis:CulturalHeritageObject` is designed to identify and provide references to cultural heritage objects. Thus, the description of cultural heritage objects with peculiar properties and classes is out of the scope of this ontology. These properties and classes might be defined in other ontologies and, to this respect, the class `cis:CulturalHeritageObject` works as a hook to foster data linking from other linked datasets that provide peculiar knowledge about cultural heritage objects.

---

[23] http://www.ontologydesignpatterns.org/cp/owl/price.owl.

[24] The prefix `price:` stands for the namespace http://www.ontologydesignpatterns.org/cp/owl/price.owl#.

[25] http://www.ontologydesignpatterns.org/cp/owl/place.owl.

[26] http://www.ontologydesignpatterns.org/cp/owl/collectionentity.owl.

[27] The inverse property is `cis:isMemberOf`.

Cultural-ON reuses the PartOf ODP[28] for describing part-of relations among cultural institutes or sites (cf. CQ15). Namely, the object property `cis:partOf` allows to model an individual of `cis:CulturalInstituteOrSite` to be part of another `cis:CulturalInstituteOrSite`, either administratively or juridically.

The requirement CQ6 is addressed by re-using the TimeIndexedPersonRole ODP.[29] This ODP is a specialisation of the AgentRole ODP. In fact, it allows adding the temporal dimension in order to specify when a given role is held by a certain person. In our ontology this ODP is represented by the class `cis:RoleInTime` that represents an *n*-ary relation whose arguments are a `cis:Role`, a `cis:Agent` and a `cis:TimeInterval`.

The requirement CQ10 is addressed by re-using the Description ODP.[30] This ODP is specialised by defining the datatype property `cis:description` whose range is any value of `xsd:Literal` and can be used for providing textual description about `cis:CulturalInstituteOrSite`.

Finally, the ontology alignment step introduced in section "Methodology" allowed us to map our ontology to other ontologies or vocabularies available in the Semantic Web. The alignments are provided in a separate OWL file[31] by means of `owl:equivalentClass`/ `owl:equivalentProperty` and `rdfs:subClassOf`/`rdfs:subPropertyOf` axioms. In particular, Cultural-ON is aligned with Dublin Core,[32] FOAF, [33] GoodRelations, [34] the Organisation ontology, [35] LODE, [36] DOLCE[37] and PRO.[38] For example, the class `cis:Agent` is declared `owl:equivalentClass` to `foaf:Agent` or `cis:Event` is declared `rdfs:subClassOf lode:Event`. Interested readers can refer to section "Lesson Learnt (On Methodological Principles)" for a detailed description of the methodological principles that guided us in the external ontologies alignment process.

[28] http://www.ontologydesignpatterns.org/cp/owl/partof.owl.

[29] http://www.ontologydesignpatterns.org/cp/owl/timeindexedpersonrole.owl.

[30] http://www.ontologydesignpatterns.org/cp/owl/timeindexedpersonrole.owl.

[31] The file containing ontology alignments is available at http://stlab.istc.cnr.it/documents/mibact/alignment_xml.owl.

[32] http://dublincore.org.

[33] http://xmlns.com/foaf/spec/20070524.html.

[34] http://purl.org/goodrelations/v1.

[35] http://www.w3.org/ns/org.

[36] http://linkedevents.org/ontology.

[37] http://www.loa-cnr.it/ontologies/DUL.owl.

[38] http://purl.org/spar/pro.

The Produced Linked Open Data

MiBACT stores records about the Italian cultural institutes or sites (i.e. addresses, opening hours, managing authorities, contacts, etc.) in a database called "DB Unico 2.0".[39] The database also contains information about cultural events[40] (e.g., as exhibitions, seminars, conferences, etc.) hosted in cultural institutes or sites and organised by the MIBACT or by other institutions. A MIBACT's editorial board is responsible for the maintenance of the database.

In the "*Governmental Agenda for the valorisation of the public sector information*", the Italian Digital Agency (Agenzia per l'Italia Digitale—AgID) classified the data included in *DB Unico 2.0* as high-value data to be released to anyone, and for any reuse, through the Linked Open Data paradigm. It was in fact recognised that publishing data on cultural properties using the LOD paradigm can significantly contribute to valorising cultural assets and promoting cultural events, thanks to the possibility granted to anyone to develop new services and applications using those data. Since cultural institutes or sites can be seen as containers of artworks and events, the cultural institutes or sites dataset can serve as the hub of the Linked Open Data cloud of the Italian cultural heritage: users can start discovering the Italian patrimony by exploring the interlinks that this dataset may enable towards other datasets of the cultural domain.

Figure 3 shows the process we employed in order to produce the Linked Open datasets of cultural institutes or sites and cultural events, by means of an activity diagram. The process in this figure is illustrated by highlighting the flow of the information, the actors and the used tools/technologies. Five types of actors participate in the process; namely, *qualified editors*, an *editorial board*, *Linked Open Data engineers*, *ontology designers* and other *domain experts*.

The process consists of a number of phases. The first phase regards the feeding of the database of cultural institutes or sites and events and it is carried out by qualified editors. This phase is currently foreseen in a specific regulation that confers to MIBACT the obligation to manage and maintain such a traditional relational database. Qualified editors are typically members of a variety of public institutions, national and local, operating in the cultural sector. Although the existence of guidelines for database feeding purposes, they can introduce heterogeneity in the data they document within DB Unico. To address this problem, a data cleansing and data harmonisation phase is executed by MIBACT's editorial board. For this activity, several state of the art tools can be used. In the context of our collaboration with MIBACT, the editorial board decided to take advantage of OpenRefine[41] i.e., a tool for cleaning data and for transforming data from a format into another format.

---

[39] http://goo.gl/srh6op.

[40] http://goo.gl/3C5orD.

[41] http://openrefine.org/.

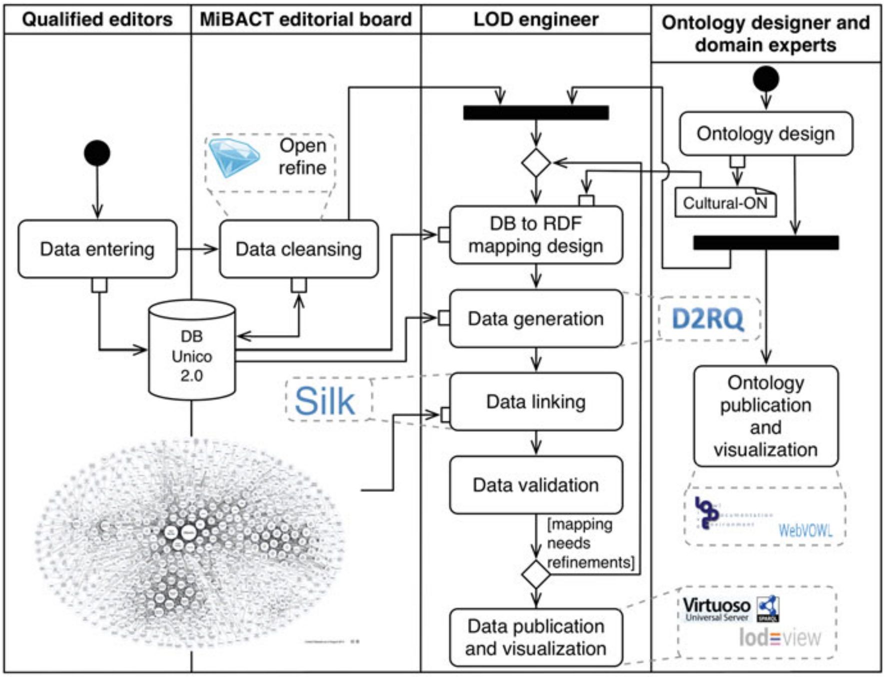

Fig. 3 The process for releasing the information in *DB Unico 2.0* as Linked Open Data

As shown in Fig. 3, in parallel to these activities, the creation of the ontology, following the methodology introduced in the previous section, can be carried out by ontology designers. It is worth noticing that the ontology creation phase requires a strict interaction and collaboration between ontology and domain experts. These latter are the only ones capable of clearly defining the requirements as previously represented, since they know precisely the peculiarities of the data, the legislation that applies, and their everyday working experience.

Once the ontology is defined, it can be published on the Web. To this end, a stable URI for it is selected, human-readable documentation and basic metadata for the ontology are specified, labels and comments are defined in more languages (currently in Italian and English), and the content negotiation is implemented. In this phase, we leveraged LODE (Peroni et al. 2013), i.e., a tool that, used in conjunction with content negotiation, allows us to make available the ontology as a human-readable and browsable HTML page with embedded links. Additional tools such as WebVOWL[42] can be used to graphically visualise on the Web the elements of the ontology.

[42]http://vowl.visualdataweb.org/webvowl.html.

Once the data cleansing and ontology creation activities are completed, LOD engineers can be involved in the process phases devoted to the generation, linking and publication of Linked Open datasets. For these phases we used the D2RQ[43] platform; that is, a system for accessing relational databases as RDF graphs. The platform provides two ways of working: *d2r-serv* and *dump-rdf*. The former queries a relational database as a virtual RDF graph; the latter dumps the content of the whole database into a single RDF file. Both ways need a customisable mapping file as input where rules specifying how to transform the relational data in RDF are defined (the activity *DB to RDF mapping design* in Fig. 3). The mapping file can be defined using a declarative mapping language through which describing the relation between an ontology and a relational data model. In this case, LOD engineers need the ontology and database schema as inputs in order to execute this phase of the process.

Once the RDF dump is produced, the data it includes can be linked (data linking in Fig. 3) to other well-known datasets available on the Web of Data (e.g., DBpedia, GeoNames, and any other of the cultural sector). In order to carry out this activity, we used Silk,[44] a linked data integration framework that enables the integration of heterogenous data sources. As D2RQ, Silk requires a mapping configuration for specifying how to discover links between datasets.

A data validation phase is then required in order to verify whether the produced datasets match the desired data quality. If the validation is positive, the datasets can be published: otherwise, the mapping phase has to be re-iterated applying proper mapping refinements.

The final phase of the process involves the Linked Open datasets publication. Data are uploaded in a triple store; that is, a data base capable of storing RDF graphs and enabling SPARQL interrogations on the data. SPARQL is the W3C standard protocol, similar to SQL, that allows to query RDF triples. In our work, we used Virtuoso SPARQL endpoint with its open source version.[45] Finally, in order to favour a better end-users experience, we deployed LodView[46]; that is a Java Web application based on Spring and Jena that implements URI dereferencing and allows users to browse RDF resources in a user-friendly way, as HTML pages.

The described process ensures that the following principles are met: (1) the published information is created and managed within a real business process of a public body; (2) database updates are reflected in the Linked Open datasets by configuring the above process in such a way to be run periodically when changes to the main authoritative sources are applied; (3) the dataset is linked to broadly-known datasets available on the Web, e.g. DBpedia, GeoNames; (4) the data and the ontology are easily accessible through user-friendly interfaces (e.g., LodLive, LODE, WebVOWL) that allow non-expert users to explore the model and the datasets.

[43] http://d2rq.org/.

[44] http://silkframework.org/.

[45] https://github.com/openlink/virtuoso-opensource.

[46] http://lodview.it/.

Example of Linked Open Data Aligned to Cultural-ON

In the following we provide an example of usage of the ontology *Cultural-ON* for exporting in RDF the data stored in *DB Unico 2.0*. The example is based on data regarding the National Museum of Capodimonte and an event it hosted. We provide some crucial code snippets; interested readers can refer to the complete example available in TURTLE and in RDF/XML formats at.[47] The example can be explored through LODView by accessing any URI defined in the RDF file, e.g.[48]

Frame 1 provides the definition of the museum as "*Cultural Institute or Site*". The resource `cissite:Museo_di_Capodimonte` represents the National Museum of Capodimonte. The resource is of type `cis:Museum`, i.e. a Cultural Institute or Site of type Museum. A Cultural Institute or Site has a name `cis:institutionalName` and a description `cis:description`. The values of these properties are also specified in the RDF schema properties `rdfs:label` and `rdfs:comment`, respectively, in order to improve the readability of the data when using semantic web browsers and visualizers. Every Cultural Institute or Site is associated with a unique identifier named ISIL, i.e., International Standard Identifier for Libraries and Related Organizations. In the example of Frame 1 the ISIL identifier is fictive since in Italy it is currently available only for libraries and holders of archives, not for museums (the Ministry is planning to assign an ISIL identifier to all cultural institutes or sites in the near future).

```
cissite:Museo_di_Capodimonte a cis:Museum ;
  rdfs:label "Museo di Capodimonte"@it ;
  rdfs:comment "Nel 1738 Carlo di Borbone ... "@it ;
  cis:institutionalName "Museo di Capodimonte"@it ;
  cis:description "Nel 1738 Carlo di Borbone ... "@it ;
  cis:ISILIdentifier "IT-NA00001" ;
  cis:hasSite site:Sede_del_Museo_di_Capodimonte ;
  cis:hasNameInTime nit:Museo_di_Capodimonte ;
  cis:hasCollection collection:Collezione_Farnese ;
  cis:catalogue cat:Catalogo_Collezione_Farnese ;
  ...
```

Frame 1 The code snippet for the definition of the *Cultural Institute or Site* "Museo di Capodimonte"

The 'National Museum of Capodimonte', intended as Cultural Institute, is located in a building commonly known with the name 'National Museum of Capodimonte' represented by the resource `site:Sede_del_Museo_di_Capodimonte`. The definition of the Site is provided by Frame 2.

The separation of `cis:CulturalInstituteOrSite`, i.e. a complex entity that involves an organisation of people and a physical site, from `cis:Site`, i.e. a

[47] https://w3id.org/cis/example/.

[48] https://w3id.org/cis/resource/CulturalInstituteOrSite/Museo_di_Capodimonte.

place where a cultural institute or site is physically located, is required to describe situations where multiple institutes or sites are located in the same building or where a single cultural institute or site is dislocated in several buildings. Furthermore, Frame 2 shows an example of usage of the ODPs "*Description*" and "*Place*". The `sd:Sede_del_Museo_di_Capodimonte` is the description of the building regarding the attribute "Area".

```
site:Sede_del_Museo_di_Capodimonte a cis:Site ;
  cis:name "Sede del 'Museo di Capodimonte'"@it ;
  cis:hasSiteDescription sd:Sede_del_Museo_di_Capodimonte ;
  cis:hosts event:Vincenzo_Gemito_dal_salotto_Minozzi_
    al_Museo_di_Capodimonte ;
  cis:isSiteOf cissite:Museo_di_Capodimonte ;
  ...
sd:Sede_del_Museo_di_Capodimonte a cis:SiteDescription ;
  cis:description "The Museum has an area of 1000
   square meter"@en ;
  cis:hasAttribute attribute:Area .
attribute:Area a cis:Attribute ;
  rdfs:label "Area"@en ;
  cis:description "The extent in square meter of a Site"@en .
```

Frame 2 A code snippet defining the site of the National Museum of Capodimonte

Frame 3 provides an example of usage of the class `cis:NameInTime`. `cis:NameInTime` (as well as `cis:RoleInTime`) implements a specialisation of the ODP called *Time Indexed Situation*. The situation defined by `nit:Museo_di_Capodimonte` describes the names assigned to the museum since it was founded. Since 1957 the museum has been officially called "*Museo di Capodimonte*"; it also has "*Museo Nazionale di Capodimonte*" as alternative name.

```
nit:Museo_di_Capodimonte a cis:NameInTime ;
  cis:institutionalName "Museo di Capodimonte"@it ;
  cis:altLabel "Museo Nazionale di Capodimonte"@it ;
  cis:isValidDuring ti:Intervallo_di_tempo_per_denominazione_
   nel_tempo_Museo_di_Capodimonte .

ti:Intervallo_di_tempo_per_denominazione_nel_
   tempo_Museo_di_Capodimonte a cis:TimeInterval ;
  cis:startDate "1957" .
```

Frame 3 A code snippet showing an example of usage of the class `cis:NameInTime`

Frame 4 shows an example of the classes `cis:Collection` and `cis:Catalogue`. The resource `collection:Collezione_Farnese` represents the "Farnese" collection that has been managed by Museum of Capodimonte during a specific time interval `ti:Collezione_Farnese_a_Museo_di_Capodimonte`. In this example, the collection contains only one *Cultural Heritage Object*; `cho:Giambologna_Ratto_di_una_sabina`. It is worth noting that the property `cis:hasMember` of the class Collection allows including in a collection any resource defined in other datasets (e.g. Cultura Italia). In other words, this property allows us to link the dataset generated from the data contained in *DB Unico 2.0* to other datasets related to the descriptions of artworks. The resource `cat:Catologo_Collezione_Farnese` of type `cis:Catalogue` is used to describe collections.

```
collection:Collezione_Farnese a cis:Collection ;
  cis:name "Collezione Farnese"@it ;
  cis:description "Trasferita a Napoli... "@it ;
  cis:atTime ti:Collezione_Farnese_a_Museo_di_Capodimonte ;
  cis:hasMember cho:Giambologna_Ratto_di_una_sabina .

cho:Giambologna_Ratto_di_una_sabina
  a cis:CulturalHeritageObject ;
  cis:name "Giambologna - Ratto di una Sabina" ;
  cis:isMemberOf collection:Collezione_Farnese ;
  cis:isInSite site:Sede_del_Museo_di_Capodimonte .

cat:Catologo_Collezione_Farnese a cis:Catalogue ;
  cis:describes collection:Collezione_Farnese ;
  cis:isCatalogueOf cissite:Museo_di_Capodimonte ;
  ...
```

Frame 4 An example of usage of the classes `cis:Collection` and `cis:Catalogue`

The resource `event:Vincenzo_Gemito_dal_salotto_Minozzi_-al_Museo_di_Capodimonte`, illustrated in Frame 5, defines an event hosted by Museum of Capodimonte. The event was an exhibition of Vincenzo Gemito's artworks. In particular, the event involved the *Cultural Heritage Object* "Il giocatore di carte" represented by the resource `cho:Il_giocatore_di_carte`.

```
event:Vincenzo_Gemito_dal_salotto_Minozzi_al_Museo_di_Capodimonte
  a cis:Event;
  cis:name "Vincenzo Gemito, dal salotto Minozzi
    al Museo di Capodimonte" ;
  cis:description "Apre al pubblico al Museo di ... " ;
  cis:isHostedBy site:Sede_del_Museo_di_Capodimonte;
  cis:involves cho:Il_giocatore_di_carte ;
  ...
```

Frame 5 An example of event hosted by the museum

Lesson Learnt (On Methodological Principles)

Besides the results described in the previous sections, this experience allowed us to reflect on how to improve our ontology development methodology, in particular as far as collaboration with domain experts and ontology reuse are concerned.

Communicating with Domain Experts

Domain experts are always present as one of the main actors in ontology engineering methodologies (Simperl et al. 2010). Their contribution to the process is crucial, especially in defining the domain and task requirements, which drive the ontology design and testing phases. Similarly, in software engineering methodologies, domain experts are the main actor and input source of software and database requirements definition, which determine the corresponding software modelling choices. Nevertheless, there is a difference in how software and databases are perceived by domain experts, as compared to how they perceive ontologies. Software and databases are expected to be black boxes able, respectively, to do what was expressed in form of requirements and to provide, when queried, the data they are expected to store.

Ontologies provide a data model for building knowledge bases, however they are also expected to reflect a shared understanding of a knowledge domain and, in some cases, to be evolved and maintained by domain experts without the help of ontology experts. This important commitment of ontologies is often wrongly interpreted as if the communication between domain experts and ontology experts should be mediated by sharing the ontology implementation (e.g. its OWL representation) and by discussing the modelling choices, applied during its implementation. In other words, the ontology itself is often used as the reference document for discussing domain requirements. Based on our experience, this is a wrong practice.

The good practice, as suggested in most methodological guidelines (Simperl et al. 2010; Suárez-Figueroa et al. 2015; Blomqvist et al. 2010), is to focus the communication on ontological requirements. It is important though to identify a common communication code. In the case of the Cultural-ON project, we used user stories (extracted from normative documents and discussions with domain experts) and competency questions, as shown in section "Cultural-ON: The Cultural ONtologies".

Although we acknowledge that keeping the domain experts focused on requirements is a good practice, we also have to recognise the importance of sharing with them the ontology, at some level of detail. It is important that they have an understanding of the ontology main concepts and usage, in order to facilitate reuse, and favour technological transfer to the domain experts, who may want to maintain and evolve the ontology. This need is not properly addressed, currently: the only way to share ontology details with domain experts is to explain them the OWL representation. Considering that most of them do not have a knowledge representation background, this practice slows down the process significantly and forces ontology and domain experts to perform training sessions in logics and ontology design.

Our experience (not limited to the Cultural-ON project) suggests the need to identify an intermediate ontology representation, able to provide, at the same time, both a clear understanding and a simplified view of the ontology, which shows the essential modelling components, and hides such technical details as complex axiomatisation and *logical sugar*. Ontology design patterns (ODP) may be a practical means for sharing a view of the ontology with domain experts, addressing this idea. A proposal, which we are currently investigating, is to use UML component diagrams as intermediate ontology representation (by defining an appropriate UML profile), where each component represents an ODP, or more in general an ontology module, and its interfaces show the ODP's most important concepts. The interfaces of an ontology module can be either classes or properties. Their role is twofold: (1) to show the points of interaction with the ontology component, and (2) to provide details on the most important concepts covered by the ontology module. Interacting with an ontology component means both communication between two ontology modules, and communication between an ontology module and an external client (for example, a user that wants to instantiate that component).

As a proof of concept, we show in Fig. 4 an example UML component diagram describing Cultural-ON as a set of interconnected ODPs. Grey components represent more general ODPs, which could be omitted depending on the level of detail that one wants to have. The white components show the core parts of the ontology and highlight their main concepts. For example, the component `CulturalInstituteOrSite` shows the class CIS[49] and the fact that one of its main characteristics is the composition (i.e. the `partOf` relation). The component `RoleInTime` shows its three main concepts, providing the intuition of a *n*-ary relation which models the role of an agent during a certain time period. Such component is linked to the component `CulturalInstituteOrSite` by means of the concept `roleAt`, indicating that an agent has a role at a certain CIS. For example, this notation allows to hide the OWL-specific modelling of an *n*-ary relation, which requires reification, while still conveying the conceptual semantics of this part of the ontology.

The choice of what concepts to show as interfaces and which ones to visualise as links between components can change, depending on the intention of the designer. The aim of this example is to show a possible direction of research, more than providing a definitive solution, which is still under analysis.

Practices and Methods for Ontology Reuse

An important reflection that is worth sharing is related to the practice of ontology reuse. The Cultural-ON project was meant to favour the principle of reusing existing resources as much as possible, in order to facilitate interoperability. At the same time the Ministry had the need of having complete control on the evolution and maintenance of Cultural-ON, for ensuring its sustainability. In this section, we

[49] Short for `CulturalInstituteOrSite`.

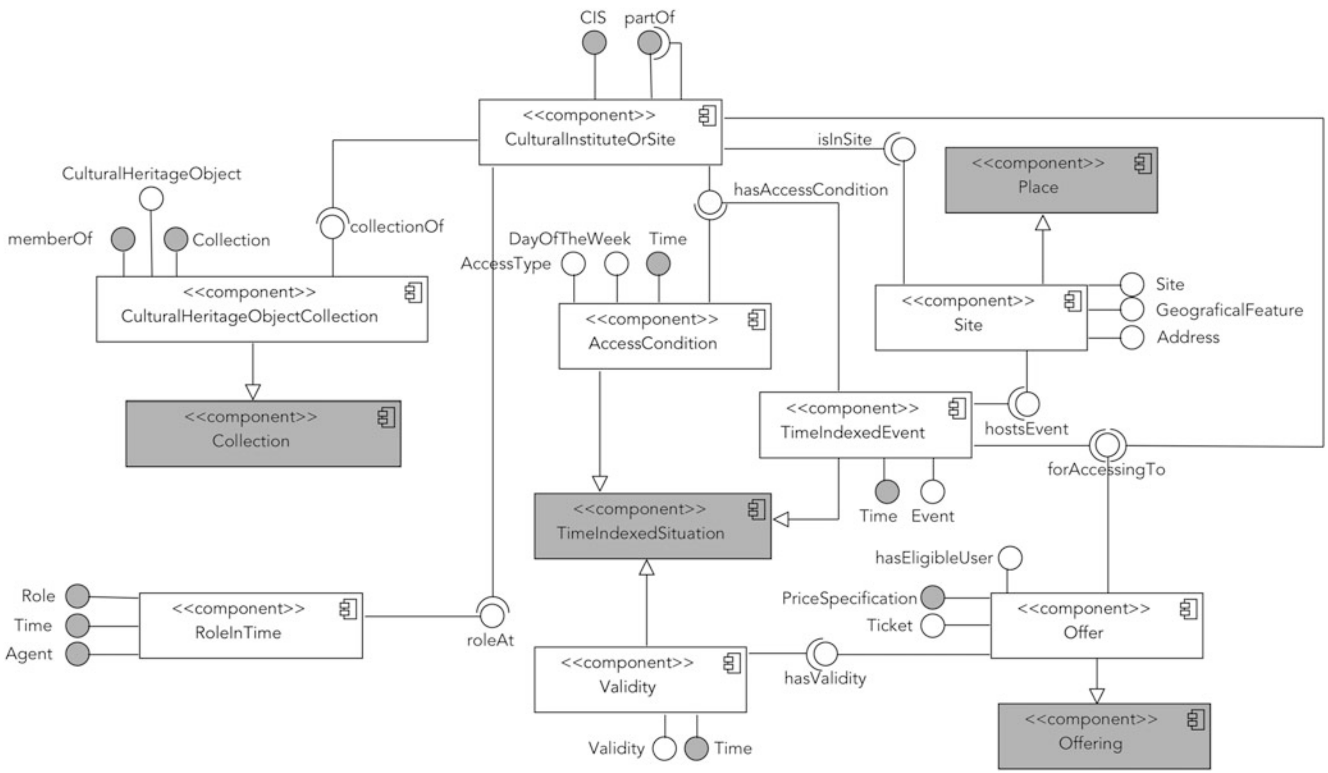

Fig. 4 Example of UML Component Diagram used for sharing the main concept modules of Cultural-ON with domain experts. This example provides the intuition of an ongoing work on defining an intermediate notation for sharing ontology modelling details with domain experts without exposing them to language-specific modelling details

discuss our choices in this respect, by generalising their principles, after a brief survey on the different practices in ontology reuse and their impact on the resulting ontologies.

Ontology reuse is a recommended practice both in scientific (Simperl et al. 2010) and in standardisation (W3C, ISO, ANSI, OASIS) literature. It is considered a good practice as it favours semantic interoperability among diverse knowledge bases, especially in the Semantic Web and Linked Data context. Nevertheless, a standardisation of ontology reuse practices is still missing because ontology design had a non-linear evolution and because modelling requirements may vary significantly depending on the domain, the availability of existing ontologies, the sustainability constraints within a given organisation, the specific requirements of an ontology project, the trends, and finally the personal taste of the ontology developer.

In the light of these considerations, it is possible to highlight a number of ontology reuse models, emerging from the current practices in the Semantic Web and Linked Data community, that can be classified based on the type of ontology that is reused:

- Reuse of foundational ontologies or top-level ontologies, for example when DOLCE[50] or the DBpedia ontology[51] are specialised for modelling events and participation to events, such as the case of the institutes that are involved in an exhibition;
- Reuse of ontology patterns,[52] for example when the "participation" pattern[53] is applied for modelling the participation to events;
- Reuse of domain ontologies, for example when the Music Ontology[54] is reused in another ontology project for modelling musical events.

Reuse models can also be classified based on the type of reused ontology fragments:

- Reuse of individual entities (classes, relations, individuals), for example when an element such as `dolce:hasParticipant` or `dbpedia-owl:Event` is reused as a specific entity;
- Reuse of *groups* of entities (modules, patterns, arbitrary fragments), for example when the pattern "participation", or `dolce:hasParticipant` together with all relevant related entities (`dolce:Object`, `dolce:hasParticipant`, `dolce:Event`, etc.), or the Event Ontology[55] are reused.

---

50 http://www.ontologydesignpatterns.org/ont/dul/DUL.owl.

51 http://dbpedia-org/ontology.

52 http://www.ontologydesignpatterns.org.

53 http://www.ontologydesignpatterns.org/cp/owl/participation.owl.

54 http://musicontology.com.

55 Event Ontology http://motools.sourceforge.net/event/event.html.

or they can be classified based on the amount of reused axioms (an axiom in an ontology is a rule that restricts or extends the possible inferences that can be done on the data):

- Reuse of ontologies by importing all their axioms, for example the whole Music Ontology;
- Reuse of ontologies by importing only the axioms within a given “neighbourhood” of certain entities or fragments, for example `dolce:hasParticipant` with all relevant entities and axioms that link them up to a certain “graph distance” (such neighbourhood is named “module”);
- reuse of ontologies without importing axioms, that is by importing only entities (the URI of those entities), for example `dolce:Event`.

Finally, reuse can be classified based on the alignment policy:

- Reuse of ontologies as the source of entities and axioms for the new ontology project, for example when an entity such as `dolce:Event` is directly reused as the type of cultural events;
- Reuse of ontologies through alignments of entities of the ontology under development to the existing ones, for example when the axiom `rdfs:subClassOf` is used for aligning a class of Cultural-On to a class of DOLCE (e.g. `dolce:Event`).

The only characteristic that all reuse models share is to reuse entities with the same logical type as they were defined. For example, an entity defined as `owl:Class` in an ontology, is commonly reused as such, and not as, e.g. `owl:DatatypeProperty`. Besides this shared characteristic, reuse practices can be as diverse as 36 different alternatives (as described above). Possibly, even more than one adopted at the same time in an ontology project. Independently of the personal taste and current trends, a certain choice of reuse practice impacts significantly on the semantics of an ontology, on its sustainability, and on its potential interoperability.

Semantics of Reuse

It spans from *maximum* reuse, when a whole ontology is reused, by importing all axioms and by specialising them for concepts that are not already addressed; to *minimum* reuse, when only some entities are reused, by aligning them to entities defined in the *local* ontology, and without importing their related axioms. The *safest* and most complete semantics is obtained with the maximum reuse approach.

Sustainability of Reuse

In the case of maximum reuse, the usability of the ontology decreases, because the imported ontologies introduce entities and axioms that may be unnecessary (or sometimes undesired) in the local ontology. Furthermore, maximum reuse creates a strong dependency on the reused ontologies; for example, if the local ontology imports the whole Event Ontology, and the latter has some axioms changed, the local ontology might become incoherent with respect to its original requirements.

Interoperability and Reuse

In the case of minimum reuse, interoperability is simplified as there is no constraint caused by the semantics of the reused ontology (its axioms).

In summary, aspects related to the quality of semantics are better addressed by a maximum reuse, while sustainability and interoperability are better satisfied by a minimum reuse. For this reason, the attitude of designers, the scope of the project and the nature of the data to be modelled influence the decision on the reuse approach.

Direct Versus Indirect Reuse of Ontology Elements

Based on the general discussion of the previous sections, we summarise three possible approaches to ontology reuse that have been considered during the development of the Cultural-ON project. Afterwords, we report our choice and the principles and requirements that motivated it.

Minimal Reuse of Individual Entities

This approach seems to be advised by the Linked Data community, however it is a routine, not a good practice, at all. In fact, this practice of reuse is essentially driven by the intuition of the semantics of concepts (which can be relations or classes) based on their names, instead of their defining axioms. In addition to this problem related to the formal semantics of the reused entities, that may be incompatible with the intended semantics to be represented, there is the problem of creating a dependency of the local ontology with all ontologies whose parts are reused. This dependency may put at risk the sustainability and stability of the ontology and its associated knowledge bases.

PROS: Sharing of Linked Data community praxis.

CONS: Semantic ambiguity, difficulty in verifying the consistency among the diverse reused concepts, dependency on external ontologies, risk of instability and unsustainability of the ontology.

Direct Reuse of Ontology Design Patterns and Alignments

Main concepts and relations of the ontology are defined in the local namespace (e.g. `cis:CulturalInsituteOrSite`). They are aligned, when possible and appropriate, to existing ontologies by means of logical relations (e.g. `owl:equivalentTo rdfs:subClassOf`). The modelling of some concepts and relations, which are relevant for the domain but applicable to more general scopes, is delegated to external ontologies, by means of ontology design pattern (ODP) reuse. For example, an instance of the class `cis:CulturalInsituteOrSite` can be involved in an event. Such event can be modelled as an instance of the class `ex:Event`, defined in an external ontology. The class `ex:Event` can be directly reused as the range of a property that expresses such a relation. In this example, the pattern "Event" associated with the class `ex:Event` in the external ontology, is *directly* reused. When third parties will reuse the Cultural-ON ontology for modelling their data, they will represent information about events by using the relations defined in the "Event" pattern, according to their needs. With this approach, the ontology shows a modular architecture, with a *core* component that defines the main domain entities and other components that identify relevant ODPs that model other concepts, needed for addressing the ontology requirements.

PROS: Stability and sustainability of domain relations and concepts, modularity, interoperability.

CONS: Possible heterogeneity in the ODPs usage, risk of instability and unsustainability limited to external modules.

Indirect Reuse of Ontology Design Patterns and Alignments

ODPs are used in this case as templates. This approach is an extension of the previous one. It keeps all PROS and decreases the CONS. The ontology is designed by identifying relevant ODPs from external ontologies to be reused, and by reproducing them in the local ontology. At the same time, the ontology guarantees interoperability by keeping the appropriate alignments to the external ODPs, and provides extensions that satisfy more specific requirements. For example, following this approach the class `cis:Event` and the relation `cis:isInvolvedInEvent` are locally defined and then aligned to `ex:Event` and `ex:isInvolvedInEvent` by means of equivalent (or other logical) relations. The alignment axioms may be published separately from the core of the ontology. If the ontology will need further extensions, the same approach must be followed, in order to avoid the introduction of dependencies.

PROS: Stability and sustainability of the ontology, modularity, interoperability, minimisation of dependencies from external ontologies, identification of external ontologies that can be used for more specific and extended requirements.

In all the three approaches described, the alignments must be maintained following the evolution of the reused ontologies. However, in the situation of incoherence raised by a change in an external reused ontology, the third approach guarantees the easiest maintenance, as it will be enough to revise the affected alignment relations.

In the context of the Cultural-ON project, we decided to adopt the third approach, i.e. indirect reuse of ODPs and alignments. This choice is motivated by requirements of the project that can be summarised as follows:

- to favour interoperability;
- to remove and minimise the dependency on external ontologies, in favour of the sustainability of the ontology;
- to provide to potential users of Cultural-ON a “self-contained” ontology, which reuses external ontologies by means of alignments;
- to define a sufficient number of axioms to address both competency questions and the legacy data (i.e. DB Unico);
- to find the best tradeoff in terms of number of reused entities, ODPs and ontology fragments that guarantee coherence among the axioms of the reused ontologies and of the local one. The coherence is tested and checked at design time, by using query engines and reasoners, according to Blomqvist et al. (2012).

Concluding Remark

This Chapter described a work conducted in the context of a collaboration established with the Italian Ministry of Cultural Heritage and Activities and Tourism. The work consisted in two important phases: (1) a cultural data landscape analysis, which allowed us to understand the peculiarities of part of the large Italian cultural patrimony; and (2) the definition of a common and standard-based data model capable of representing cultural property resources, The model has been defined so as to pave the way to the construction of a large open interconnected knowledge base which can, among others, guarantee that different organisations operating in the cultural sector could agree on the precise meaning of common and exchanged information on cultural properties, preserving that meaning throughout the data exchange.

The data model is actually a suite of ontology modules we named Cultural-ON (Cultural ONtologies), where the module related to “cultural institutes and sites” and “cultural events”, we presented in this Chapter, is expressed using OWL (Ontology Web Language), a semantic markup language for publishing and sharing ontologies on the Web. The module has been designed by applying a methodology that, starting from a clear definition of requirements to meet, makes extensively re-use of ontology design patterns. Together with ontology design patterns, CulturalON has been structured so as to be linked to other ontologies available on the Web. In doing so, sustainability and semantic interoperability requirements were considered.

Based on the data model, the chapter proposes a process that allows us to produce linked open datasets on cultural institutes and sites and cultural events that can be maintained up-to-date over time. This was realised by integrating the proposed process into the real business process of data management employed by the Ministry.

The chapter also discusses a number of lessons learnt and recommendations that can be sufficiently general to be applied in analogous cultural projects and in other domains.

Future Works

We are planning to extend the Cultural-ON suite with a module for the representation of cultural objects and collections. Currently, the cultural institutes and sites ontology offers a preliminary representation of objects and collections as "hooks" for future extensions, which will capture the peculiarities of those objects and collections. In this sense, standard ontologies such as CIDOC-CRM (as described in the state of the art) can be investigated.

Finally, we are evaluating the reuse of the data model and produced linked open datasets to support the valorisation of less known cultural properties, thus fostering both their fruition by a wide audience of users and the creation of a richer shared culture and a wealthy touristic market.

References

C. Aart, B. Wielinga, W. Robert Hage, Mobile cultural heritage guide: location-aware semantic search, in *Proceedings of the 17th International Conference of Knowledge Engineering and Management by the Masses (EKAW-2010)*. ed. By P. Cimiano, H. Sofia Pinto (Springer, Lisbon, 2010), pp. 257–271

E. Blomqvist et al., Experimenting with eXtreme Design, in *the Proceeding of 17th International Conference of (EKAW-2010)*, ed. By P. Cimiano, H. Sofia Pinto. Lisbon. Lecture Notes in Computer Science, vol. 6317 (Springer Berlin Heidelberg, Lisbon, Portugal, 2010), pp. 120–134. Doi:10.1007/978-3-642-16438-5_9

E. Blomqvist, A. Seil Sepour, V. Presutti, Ontology testing - methodology and tool, in *Knowledge Engineering and Knowledge Management - the Proceeding of 18th International Conference (EKAW-2012)*, ed. By A. ten Teije et al., Galway City. Lecture Notes in Computer Science, vol. 7603 (Springer Berlin Heidelberg, Galway City, Ireland, 2012), pp. 216–226. Doi:10.1007/978-3-642-33876-2_20

V. de Boer et al., Amsterdam museum linked open data. Semantic Web **4**(3), 237–243 (2013)

M. Doerr, The CIDOC conceptual reference module: an ontological approach to semantic interoperability of metadata. AI Magazine **24**(3), 75–92 (2003)

A. Gangemi, V. Presutti, Ontology design patterns, in *Handbook on Ontologies*, 2nd edn., ed. By S. Staab, R. Studer. International Handbooks on Information Systems (Springer, Berlin, 2009), pp. 221–243. Doi:10.1007/978-3-540-92673-3_10

M. Grüninger, M.S. Fox, The role of competency questions in enterprise engineering, in *Benchmarking – Theory and Practice*, ed. by A. Rolstadås. IFIP Advances in Information and Communication Technology (Springer, Boston, MA, 1995), pp. 22–31. Doi:10.1007/978-0-387-34847-6_3

E. Hyvönen, Semantic portals for cultural heritage, in *Handbook on Ontologies*, 2nd edn., ed. By S. Staab, R. Studer (Springer, Berlin, Heidelberg, 2009), pp. 757–778

A. Isaac, B. Haslhofer, Europeana Linked Open Data - data.europeana.eu. Semantic Web **4**(3), 291–297 (2013)

J. Oomen, L. Aroyo, Crowdsourcing in the cultural heritage domain: opportunities and challenges, in *Proceedings of 5th International Conference on Communities & Technologies (C&T-2011)* (ACM, Brisbane, 2011), pp. 138–149

S. Peroni, D. Shotton, F. Vitali, Tools for the automatic generation of ontology documentation: a task-based evaluation. Int. J. Semantic Web Inf. Syst. **9**(1), 21–44 (2013)

M. Ridge, From tagging to theorizing: deepening engagement with cultural heritage through crowdsourcing. Curator: Museum J. **56**(4), 435–450 (2013)

T. Ruotsalo et al., SMARTMUSEUM: a mobile recommender system for the Web of Data. Web Semantics: Sci. Serv. Agents World Wide Web **20**, 50–67 (2013)

G. Schreiber et al., Semantic annotation and search of cultural-heritage collections: the MultimediaN E-Culture demonstrator. Web Semantics: Sci. Serv. Agents World Wide Web **6**(4), 243–249 (2008)

E.P.B. Simperl, M. Mochol, T. Bürger, Achieving maturity: the state of practice in ontology engineering in 2009. Int. J. Comput. Sci. Appl. (IJCSA) **7**(1), 45–65 (2010). ISSN: 0972-9038

M.C. Suárez-Figueroa, A. Gómez-Pérez, M. Fernández-López, The NeOn methodology framework: a scenario-based methodology for ontology development. Appl. Ontol. **10**(2), 107–145 (2015). Doi:10.3233/AO-150145

P. Szekely et al., Connecting the Smithsonian American Art Museum to the linked data cloud, in *Proceedings of the 10th Extended Semantic Web Conference (ESWC-2013)*, ed. By P. Cimiano et al. (Springer, Montpellier, 2013), pp. 593–607

Y. Wang et al., Interactive user modeling for personalized access to museum collections: the Rijksmuseum case study, in *Proceedings of the 11th International Conference of User Modeling (UM-2007)*, ed. By C. Conati, K. McCoy, G. Paliouras (Springer, Corfu, 2007), pp. 385–389