



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Semantic metadata enrichment and data augmentation of small museum collections following the FAIR principles

Citation for published version:

Vlachidis, A, Antoniou, A, Bikakis, A & Terras, M 2021, Semantic metadata enrichment and data augmentation of small museum collections following the FAIR principles. in K Golub & Y-H Liu (eds), *Information and Knowledge Organisation in Digital Humanities: Global Perspectives*. 1 edn, Digital Research in the Arts and Humanities, Routledge, pp. 106-129. <https://doi.org/10.4324/9781003131816-11>

Digital Object Identifier (DOI):

[10.4324/9781003131816-11](https://doi.org/10.4324/9781003131816-11)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Publisher's PDF, also known as Version of record

Published In:

Information and Knowledge Organisation in Digital Humanities

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



6 Semantic metadata enrichment and data augmentation of small museum collections following the FAIR principles

Andreas Vlachidis, Angeliki Antoniou, Antonis Bikakis, and Melissa Terras

Introduction

Over a decade has passed since the European Agenda for Culture (European Commission 2007) recognised digitisation as a fundamental driver for fostering cultural diversity and intercultural dialogue. National galleries, libraries, archives, museums (GLAM) institutions across Europe have undertaken initiatives for digitising major parts of their collections but the same promise of digitisation is yet to be realised by many smaller and regional organisations. In particular, south European cultural heritage institutions appear to be slower adopters of ICT (Information and Communication Technologies) compared to their northern counterparts and the reasons for such lack of uptake are complex (Peacock 2008; Gombault et al. 2016), despite the fact that heritage in such regions is rich, diverse and forms a key strategic resource for economic development (see, e.g., Bonaccorsi et al. 2007). Semantic Web technologies can significantly benefit digitised collections by disclosing information in a scalable and interoperable way, aggregating previously heterogeneous and siloed data (Benjamins et al. 2004). Particularly, in the domain of cultural heritage, such aggregations can be extremely useful for providing context over relations between persons, artefacts, works, locations etc. while also supporting information seeking activities through semantic linking, recommendation and visualisation techniques (Clough et al. 2008).

This chapter explores the role and application of semantic models on smaller cultural heritage collections for facilitating data dissemination, interlinking and augmentation and for making their collection data FAIR, namely Findable, Accessible, Interoperable and Reusable (Go-Fair.org 2020), presenting a case study on the Archaeological Museum of Tripoli to demonstrate the applicability of this approach. The task of modelling and enrichment with semantic metadata has been achieved to deliver descriptions, references and structures of artefacts within the collection, withdrawing the silo barriers of museum items and enabling interoperable, multi layered representations that can be used to deliver a variety of user experiences. We reflect on the benefits of using the Conceptual Reference Model

of the International Committee for Documentation of the International Council of Museums (CIDOC-CRM; Doerr 2003) aligned to the FAIR principles and present considerations of how best the sector can support such work.

The remainder of the chapter is structured as follows. In Section Background, the background and relation of the semantic technologies with FAIR data principles is discussed. In Section The Archaeological Museum of Tripoli case study, the main case study is presented, following the methodological choices of the metadata model design. The chapter concludes with observations regarding the adoption of the FAIR data principles in small heritage organisations and the benefits for semantic linking and interoperability across cultural heritage collections.

Background

The FAIR data principles

The FAIR data principles are a set of guidelines and best practices for the management of scholarly data aiming at facilitating their discovery, integration and reuse both by humans and computer agents (Wilkinson et al. 2016; Go-Fair.org 2020). They refer to different aspects of data and the processes used for its production and communication, such as format, communication protocol, usage license and provenance. They were swiftly adopted by data producers in several scientific domains, but more widely in the life and natural sciences (see (van Reisen et al. 2020) for an analysis). This is not surprising considering the volume, volatility and diversity of the data used in these domains. However, other domains, such as the humanities, also suffer from similar problems that hinder the discovery, integration and reuse of data, particularly around data-complexity such as the “ambiguity of symbols, too many persistent identifiers for the same concept, semantic drift and linguistic barriers” (Mons 2018, 3).

The FAIR data principles have also recently gained the attention of Digital Humanities scholars. For example, the PARTHENOS (2019) project (Pooling Activities, Resources and Tools for Heritage E-research Networking, Optimization and Synergies) aims at developing a “trans-humanities research infrastructure” by integrating existing infrastructures and initiatives from different disciplines such as linguistic studies, humanities, cultural heritage, history and archaeology. Project activities were designed around FAIR data principles. They include the definition of data curation and intellectual property policies and the development of guidelines, standards, methods, services and tools that enhance the discoverability, interoperability and reuse of Digital Humanities resources. Some other initiatives and studies contribute to the specification and expansion of the FAIR data principles for Digital Humanities researchers given the specific characteristics of the data in its related disciplines.

For example, several European organisations, such as the Archives Portal Europe Foundation (APEF 2019), the European Research Infrastructure for Language Resources and Technology (CLARIN 2021), the Digital Research Infrastructure for the Arts and Humanities (DARIAH 2021), the Europeana and the European Research Infrastructure for Heritage Science (E-RIHS 2020) are working together on the conformation of a heritage data reuse charter. In the mission statement of this initiative (Heritage Data Reuse Charter 2020), they recommend a new set of principles: *Reciprocity* (agreement between institutions and researchers to share content and knowledge equally with each other); *Interoperability* (data is made available in formats that facilitates its reuse); *Citability* (data and any data-driven research should be fully citable); *Openness* (data should be shared under an open license whenever possible); *Stewardship* (attention should be paid to the long-term preservation, accessibility and legibility of cultural heritage data) and *Trustworthiness* (the provenance of data should be clear and openly available).

The Dutch knowledge centre for digital heritage and culture (DEN), has also published a similar set of minimal requirements along with associated guidelines, principles, policies, references and roadmaps for digitisation of cultural heritage, focusing on six areas of attention: Rights Management, Findability, Creation, Presentation, Digital Sustainability/Preservation and Description. These map onto the FAIR principles, although expand certain details and implementation, showing how they must be discussed in relation to particular domains. In a recent study, Koster et al. (2018) reviewed the FAIR data principles and other similar initiatives focusing on the reuse of Digital Humanities data, making a distinction between objects, object metadata and metadata records and proposing a roadmap for implementing the principles in libraries, archives, museums (LAM) collections. More recently, Barbuti (2020) argued that more emphasis should be put on the reusability of data in digital cultural heritage, and suggested an extension of the R element with four more requirements: Readability, Relevance, Reliability and Resilience. He also demonstrated how these principles were applied to the design of three digital libraries in the context of the Terra delle Gravine (2017) project. The adoption, specification and implementation of the FAIR data principles in Digital Humanities have also been the focus of some recent new conferences such as FAIR Heritage or among the main topics of interest of some bigger conferences such as the CIDOC conference (CIDOC 2018).

Using semantic technologies to implement the FAIR data principles

With regard to the application of the FAIR data principles in practice, most current approaches rely on the use of semantic technologies such as the Resource Description Framework (RDF) data model, the SPARQL Protocol and RDF Query Language (SPARQL), ontologies encoded in the

Web Ontology Language (OWL) and Linked Data vocabularies (for an overview of these technologies, see Vassalo and Felicetti (2020)). RDF is a W3C standard specification (W3C 2015) for the conceptual description of data as triples that express relations of data in the form of object-predicate-subject statements (e.g., statue – is made of – marble). Rich data structures expressed in form of triples can be stored in specialised databases known as triple-stores that enable data interrogation using SPARQL, which is the standard language for retrieving and manipulating RDF data. OWL, another W3C standard, is a family of knowledge representation languages for authoring ontologies, namely “explicit formal specifications of the terms in a domain and the relations among them” (Gruber 1993, 199) that aim at providing a shared conceptualisation and understanding of common domains between different communities. OWL ontologies can be processed by computer programs (called reasoners) to verify their consistency and to compute inferred knowledge. Linked Data is a collection of open interrelated structured datasets on the Web, which are formatted in the RDF and follow some standard principles (Berners-Lee 2020).

The adoption of semantic technologies to implement the FAIR data principles is not a surprise, as the design principles of Linked Data are to a large extent in line with those of FAIR data. The first Linked Data principle is to use URIs (Universal Resource Identifiers) to name web resources. This can be seen as a way to implement the first principle of FAIR data, according to which “(meta)data are assigned a globally unique and persistent identifier” (Go-Fair.org 2020). The second Linked Data principle, which is to use HTTP URIs so that people can look up those names, is in line with the first accessibility principle of FAIR data, according to which data and metadata “should be retrievable via a standardised communication protocol” (Go-Fair.org 2020). The third Linked Data principle is that when someone looks up a URI, some relevant useful information should be provided. This can be seen as a way to implement the FAIR data principles, according to which data should be described with rich metadata, using a formal, accessible, shared and broadly applicable language for knowledge representation. Formal models such as ontologies can guarantee the use of well-defined and interoperable knowledge representations that carry definitions and conceptual arrangements of entities and relationships to describe a domain or a resource. The fourth Linked Data principle is that data should include links to other URIs so that users can discover more things (by following such links). This can be seen as a way to implement the third interoperability principle, according to which (meta) data should “include qualified references to other (meta)data” (Go-Fair.org 2020).

In contrast with FAIR principles, the Linked Data principles refer to specific technologies with which the principles can be implemented (URIs, RDF, SPARQL). They can, therefore, serve as a good starting point for

making data FAIR. The implementation of the remaining FAIR principles requires some further steps. Most of them, however, point to the use of ontologies such as the second principle of interoperability I2 where “(meta) data use vocabularies that follow FAIR principles” and the first principle of reusability R1 where “meta(data) have a plurality of accurate and relevant attributes”.

Following the last reusability principle, which recommends that “(meta)data meet domain relevant community standards” (Go-Fair.org 2020), most Digital Humanities projects using semantic technologies to implement the FAIR data principles rely on the use of standard ontologies, such as CIDOC-CRM and the Simple Knowledge Organization System (SKOS). CIDOC-CRM a well-established ISO standard (ISO 21127 2006) for the modelling of cultural heritage information (Doerr 2003). The model provides an extensible semantic framework of real-world entities such as Events, Types, Appellations, Actors, Places, Physical and Conceptual objects and others that any cultural heritage information can be mapped to. The model can be applied to traditional relational database implementations as well as to contemporary semantic web frameworks such as RDF and OWL. SKOS is a W3C standard for representing the semantics of structured vocabularies, such as thesauri, classification schemes, taxonomies, subject-heading systems, enabling publication of such vocabularies as Linked Data, defining classes and properties to represent common features like preferred labels, synonym, broader, narrower terms relations and others (Miles and Bechhofer 2009). This additional layer of concepts enables the typological specification of individual items, which cannot be achieved solely by the semantics of CIDOC-CRM and other ontologies.

For example, the Virtual Research Environment in Southeast Europe and the Eastern Mediterranean (VI-SEEM 2018; Vassallo and Felicetti, 2020) project develops an integrated research infrastructure for scientific communities in life sciences, climate science and digital cultural heritage following the FAIR data principles, based on CIDOC-CRM and its extensions, CRMsci and CRMdig. The project identifies CIDOC-CRM as the ideal knowledge representation and communication framework to address the requirements of the proposed infrastructure. It aims to: enable access and retrieval of various digital resources from different domains; enable interoperability among heterogeneous data; trace the provenance of data and to facilitate its interpretation and reuse. In the same paper, the approach is demonstrated using two case studies, in which datasets from different research domains, described according to different metadata formats, were mapped to CIDOC-CRM. This has enabled both a richer description of the available resources and interoperability between the different datasets.

The Advanced Research Infrastructure for Archaeological Dataset Networking in Europe (ARIADNE 2017) has a similar aim: to develop an

integrated research infrastructure for archaeology by integrating existing infrastructures, services and datasets. In order to achieve this, the project developed the ARIADNE Catalogue Data Model (ADCM), which is used to describe all available services, language and data resources from different project partners. Furthermore, the mapping of ADCM to CIDOC-CRM allowed the reuse of such resources in other fields of Digital Humanities, improving their discoverability and interoperability (Aloia et al. 2017). In the same domain, De Haas and van Leusen (2020) recently proposed the adoption of CIDOC-CRM as the standard ontology for archaeological research, as a means for implementing the reusability principles. They also proposed a domain-specific extension of CIDOC-CRM, which meets the specific data modelling needs of archaeology and described their work towards standardising the proposed extension. In the field of historical research, Beretta (2020) argues that the application of the FAIR data principles requires the development and use of a standard ontology and proposes adopting CIDOC-CRM as the core ontology for this domain, in combination with two other foundational ontologies, C.DnS (*Constructive Descriptions and Situation*; Gangemi et al. 2008) and DOLCE (*Descriptive Ontology for Linguistic and Cognitive Engineering*; Gangemi et al. 2002). Recently, the Science Museum Group's Heritage Connector project, a foundational project of the UK's *Towards a National Collection* initiative, aims to semantically link collections metadata using a variety of approaches to automatically establish cross-references between major heritage collections, looking also at how the outputs can be best shared (Lewis and Stack 2020).

Adoption of the FAIR data principles by cultural institutions

Having identified the potential benefits of FAIR data and their role in promoting and facilitating research, several GLAM institutions are working on making their collections FAIR. Some notable examples are the Rijksmuseum, British Library, Yale Center for British Art and the Wellcome Collection. They have all made part or all of their collections and the associated metadata publicly available, either as direct downloads or via APIs, relying in most cases on semantic technologies and standard ontologies. A comprehensive list of institutions that make their digital resources and metadata openly available using APIs is available at the "Museums and the machine-processable web" online forum (Museum-API 2015). However, there are very few examples of small institutions represented there. An obvious reason for this is that most small institutions have not digitised their collections, so do not have anything that they can share online.

According to a Europeana white paper (Nauta et al. 2017), 78% of European heritage has not been digitised and only 58% of the digitised content is available online, while a very small percentage of it (3%) can be accessed for reuse. In recent years, the GLAM sector has received support

from national and regional initiatives to develop digital capacity. In the UK, programmes such as the Digital Skill for Heritage (DSfH) and the Digital Heritage Lab have been designed to support digital capabilities and to raise digital skill sets of small and medium size cultural heritage organisations. However, the COVID-19 pandemic has revealed the clear disparity across GLAM organisations with 40% of the UK museums during lockdown lacking any digital access to their collections (Finnis and Kennedy 2020). The main challenges that cultural institutions face for making their collections open remain and include: extra time, effort and costs required for the digitisation of their collections, their proper documentation and rights clearance; technical challenges; lack of metadata for their collections and lack of relevant skills among their staff (Estermann 2015).

These challenges are more common in small and medium-sized institutions. The same study also identified some risks for these museums, the most common being the reuse of their content without proper attribution to the institution or creator and the misuse or mis-representation of content and copyright infringements by third parties. On the other hand, according to the same study, the majority of cultural institutions identify that opening up their collections will provide benefits including raising the visibility or perceived relevance of the institution, improving the discoverability of their holdings, extending their audiences, facilitating networking with other cultural institutions, encouraging interactions with their audiences and more generally, it will allow them to better fulfil their mission.

Thus, despite the potential risks, small and medium-sized cultural institutions seem to identify the benefits of adopting the FAIR data principles and want to follow the larger ones in making their collections and metadata FAIR. Semantic technologies and Linked Data design guidelines seem to be the most promising way for achieving this goal, as they open up the potential for interoperability and discoverability of the objects within these collections. In the next part of this chapter, we demonstrate an approach for making the collection of a small museum FAIR using semantic technologies. We focus on the Archaeological Museum of Tripoli, an example of a small and not well-known, peripheral archaeological museum with low visitor numbers and limited digital presence, but with a unique collection of regional antiquities.

The archaeological museum of Tripoli case study

Overview of the case study

The Archaeological Museum of Tripoli houses around 7,000 items, covering a large period of Peloponnese regional history. Its holdings provide a rich collection from various local excavations of Ancient Greek sanctuaries, cemeteries, residential areas and Roman villas. Despite this important collection, this small museum has very low numbers of physical visitors and

is not popular with locals and tourists. The museum participated as a pilot case in the EU Horizon 2020 CrossCult project which aimed at facilitating interconnections between different pieces of cultural heritage information, public viewpoints and physical venues by taking advantage of the advances of digital technologies, particularly focused on the aspects of interactivity, recollection and reflection (Lykourantzou et al. 2016). The pilot case offered non-typical, crosscutting and transversal viewings of the museum items, in order to allow the visitors to go beyond the typical level of history presentation (e.g., type of a statue or its construction date), into deeper levels of reflection, over social aspects of life in antiquity and its power structures. Prior to joining CrossCult, the museum did not use any technologies to assist its visitors and its collection was not exposed in any digital media. In addition, inside the museum, all object related information was presented to the visitors with old-fashioned labels.

The museum of Tripoli is not unique in its traditional, non-digital practices. There are thousands of small- and medium-sized museums across Europe that house important collections, but have a very limited or even a non-existent digital presence, with a variety of factors that can cause this. Lack of budget, limited human resources, slow technology acceptance, organisational attitudes and bureaucracy associated with decision-making in public institutions (Gombault et al. 2016), but also the geographical location of the museum, the museum type and reasons internal to the organisation such as leadership (Bonaccorsi, et al., 2007; Peacock, 2008) can provide hindrances for fulfilling the potential of digitisation and receiving the benefits of an effective adoption of ICT. Similar obstacles in the provision of interoperable online access to archival holdings and metadata have been reported by a comparative survey of the views of archivists from Croatian, Finnish and Swedish archive (Tanacković Faletar et al. 2017). Following discussions with the Tripoli museum, we discovered there were a number of hindrances for fulfilling their potential via digitisation, including general negative attitudes towards the use of technology within an archaeological museum, believing that the visitors' attention will shift to the technology rather than the exhibits and general fear of the unknown.

In addition, numerous bureaucracy procedures, e.g., the requirement of multiple licences, seem to delay processes. Furthermore, Greek IPR law does not easily permit the use of images of archaeological items and time-consuming procedures need to be followed to obtain licences. Finally, there is the wider recognised issue of a lack of understanding and communication between humanities and computing experts, which can be traced to their training: specialists from both domains seem to face difficulties in communicating, creating obstacles (Terras 2012). However, technology can help such institutions emerge from isolation and attract many new visitors. Innovative and engaging apps to present museum content, social media activity and games could be all employed to revive such a traditional space (Antoniou et al. 2019; Kontiza et al. 2020).

We selected a total of 36 Archaeological Museum of Tripoli items, following consultation with the museum director. These items were representative of different art styles, eras, functions and all had a very high historical value. Seven topics were also identified common to these 36 items, all related to women in antiquity: “social status”, “appearance”, “mortality and immortality”, “daily life”, “education”, “names-animals-myths” and “religion and rituals”. Reflection upon history was at the heart of the project, and the identified topics aimed at increasing engagement and historical re-interpretation.

The 36 museum items were digitised and a simple set of metadata was created for each of them by museum personnel who photographed the items. The set contained elements about the title, type, period and other relevant information about the items. This early set constituted the foundation upon which the FAIR metadata model was designed as discussed in sections below. In addition, the information related to the items was significantly enhanced, since each item was associated with a number of narratives. Narratives were also accompanied with numerous digital objects, like related images (usually from similar objects in other museum collections), videos (e.g., documentaries, clips from European cinema etc.), games (specially created games to enhance the museum objects and engage visitors further) etc. Thus, not only were the museum items digitised, but the digital content of the museum was augmented with the addition of multimedia items and narratives. This allowed a broad range of digital activities to be built upon the digitised collection.

Pre-processing the collection

When dealing with a traditional small museum collection which has undertaken limited digitisation, it is important to assess the state of available data in terms of existing format, structure and volume. This will help understanding the current state and coverage of data, the availability of any associated metadata, and to identify appropriate methods and tools for data cleaning required prior to metadata assignment. In the case of the Tripoli museum, available digital data and metadata was extremely limited and sparse. The majority of museum artefacts enjoyed unstructured data in the form of short descriptions, e.g.;

2279: Marble pediment tombstone with a representation of a family (enface). The female figure bears a chiton and a cloak. The male figure and the boy bear a short chiton. On the architrave there is the inscription ANTIOXIC ΦΟΡΤΟΥΝΑΤΟΥ

ΘΥΓΑΤΗΡ ΚΑΛΛΙΣΤΗ. Found in Herod Atticus villa in Loukou, Kynouria. Roman era work (middle Antonine era, 161 A.D - 180 A.D.). Dimensions: Height 1.60m, Width 0.82m. Location: Room 15, 1st floor.

The descriptions contained concise but rich information about the object in terms of: a physical and structural properties (size, material, type etc.); b descriptive attributes and symbols (inscriptions, representations); c related facts and events (place of excavation, period) and d curation related (location, category). The size and level of description was consistent across museum items. This, then, could be used to generate metadata descriptions in line with the CIDOC-CRM semantics, including e.g., elements of type, identifier, location, dimension and other as discussed in the Metadata Model section below.

Delivering structured metadata from unstructured descriptions is a challenging task that requires human intellectual effort or employment of sophisticated techniques for automatic metadata generation. The size of the collection (36 artefacts) did not justify the development of a Natural Language Processing (NLP) application for automatic extraction of information and generation of metadata. Instead, a manual process was followed to identify and extract relevant information from the descriptions in order to produce semantic metadata. Arguably, it would be inefficient to attempt a manual process on a larger collection when automated information extraction approaches have been used with relative success in the semantic indexing and automatic metadata generation with respect to CIDOC-CRM semantics (Vlachidis and Tudhope 2016).

The manual extraction task focused on identifying entities of interest that would support information retrieval, cross-searching and discovery across the collections participating in the CrossCult infrastructure (Lykourantzou et al. 2016). It included mentions of exhibit type, related material, temporal, spatial information, dimensions and other features such as inscriptions or visual representations. The major benefit of the manual process was the accuracy of metadata, which was ensured through a peer-review process which involved humanities experts and a research associate. The process involved review of annotated documents for identifying the individual metadata elements and a subsequent process that delivered structured metadata from the unstructured annotated text. The structured output organised the extracted information (metadata) into a machine-processable format (i.e., XML).

The intermediate (pre-processing) step is necessary before the final delivery of metadata in a semantic web serialisation for the following reasons. Firstly, it enables an agile development of a database for holding and reviewing the extracted values by initially hiding the finer semantics of the conceptual model. This is particularly beneficial for communicating the focus of the task to non-technical experts with no background of semantic metadata and conceptual models. Secondly, it decouples the process of storing data from aligning the data to the conceptual model. This provides flexibility to explore the finer semantics of the model and to implement diverse mappings and relationships from the original extracts as illustrated in [Figure 6.1](#). And finally, it supports automation of data cleaning tasks that correct and

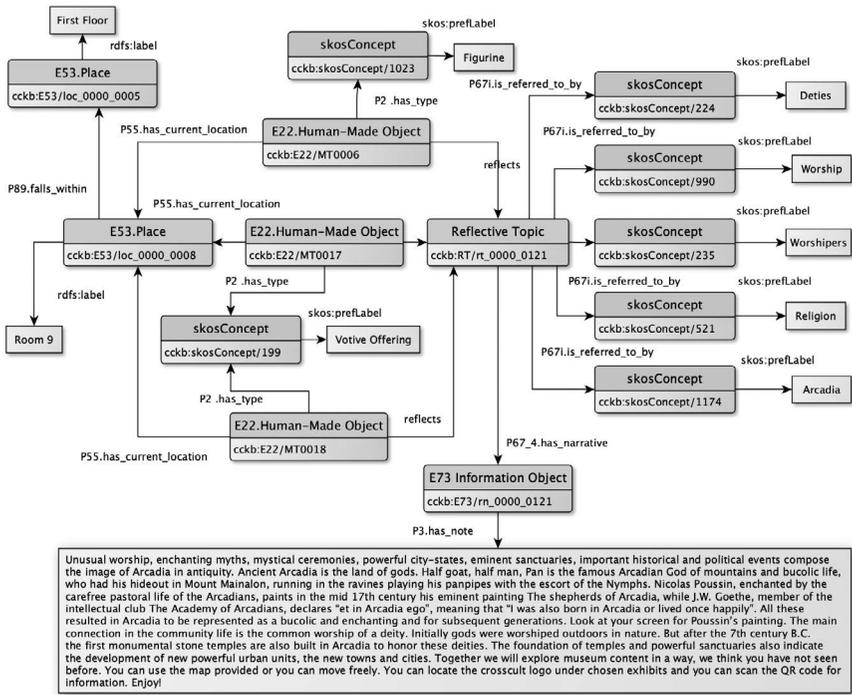


Figure 6.1 Semantic relationships of co-located museum exhibits in room 9 (Archaeological Museum of Tripoli).

standardise variations of data (e.g., labels of periods and units), data wrangling tasks that transform data to different serialisations (e.g., from XML to RDF), and enrichment techniques such as those discussed in section *Semantic Enrichment and Augmentation*.

Modelling the collection

The data modelling process had its own challenges. The first was the selection of the underlying ontology. Despite its growing popularity in the Cultural Heritage domain and its rich expressive capabilities, CIDOC-CRM was not the obvious choice to everyone. Researchers from CrossCult with an Information Science background preferred solutions based on taxonomies, classification systems or simpler metadata schemas (e.g., Dublin Core), while software developers found CIDOC-CRM unnecessary complicated and verbose for the needs of the services and applications that they wanted to develop. The adoption of the model came after careful consideration of the importance of modelling the relationships between the different cultural heritage resources of the museum, the need for semantically linking such resources with the collections of other cultural

heritage venues, and the FAIR principle for the use of “domain-relevant community standards”.

In order to facilitate semantic data interoperability and to enable discovery, integration and reuse, we aligned the data of the museum collection to the upper-level ontology of the CrossCult project (Vlachidis et al. 2017). The upper-level ontology implements a core-subset of CIDOC-CRM semantics, which is supplemented with project-specific extensions that handle entities and properties on reflections and narratives. The upper-level ontology maintains full compatibility with CIDOC-CRM containing the least minimum set of CRM concepts, ensuring reusable and interoperable qualities of the data. The data of the 36 museum items were aligned to the classes of the upper-level ontology.

At the core of the model resides the CIDOC-CRM entity *E18 Physical Item*, which comprises all persistent physical items with a relatively stable form, human made or natural. The entity enables the representation of a vast range of items of interest, such as museum exhibits, gallery paintings, artefacts and monuments while providing extensions to specialised definitions for human made objects, physical objects and physical features. The well-defined semantics enable the description of static parameters of a museum item, such as dimension, unique identifier, title and type and allow for rendering rich relationships between the physical item and entities describing the item in terms of ownership, production, location and other conceptual associations.

Another critical challenge is related to the population of the ontology with appropriate individuals and statements describing the available cultural heritage resources. Mapping the terms used by historians in the original descriptions of the resources and the elements of the ontology was not straightforward. Reaching a common understanding of the precise meaning of the original descriptions and determining their mappings to the ontology required extensive communication between ontology experts and historians. By focusing on a representative sample from the museum collection and the collections of the other venues that participated in CrossCult, we developed semi-automatic processes, which could then be re-used for all the available data. The different backgrounds of the experts who were involved in the development of the CrossCult Classification Scheme (information scientists, historians and museum experts) brought two more challenges to the project: how to determine the scope of the vocabulary and how to come up with a commonly agreed structure. Two decisions that helped us address such challenges were: (i) to rely as much as possible to standard external vocabularies such as the Arts and Architecture thesaurus; (ii) to setup and use an online environment for collaborative development and management of vocabularies, thesauri and taxonomies. Among others, the environment enables discussions on the terms and structure of the ontology, linking the vocabulary to external terms and creating RDF descriptions of the vocabulary.

The entity *E22.Human-made Object*, which is a specialisation of the *E18.Physical Item* was selected as the key component for holding together the range of properties that specify a museum item. In detail, the information of the museum item “2279”, as shown in the respective description, was aligned to the following properties of an *E22.Human-made Object*:

- *P1_is_identified_by* > *E42.Identifier* “2979”
- *P2_has_type* > *E55.Type* “Tombstone”
- *P3_has_note* “the original textual description”
- *2x P43_has_dimension* -> *E54.Dimension* “height 1.60m” and “width 0.82m”
- *P45_consists_of* > *E55.Type* “Marble”
- *P50_has_current_keeper* > *E40.Legal Body* “The Archaeological Museum of Tripoli”
- *P55.has_current_location* > *E53 Place* “Room 15”
- *P102.has_title* > “Antiohis pediment tombstone”
- *P138i.has_representation* > *E38.Image* “antiohis.jpg”.

The CIDOC-CRM is a high-level model of concepts and relationships not tied to any particular vocabulary to describe types or ontology individuals. This level of abstraction, albeit very useful for the applicability of the model across the semantic requirements of the broader cultural heritage domain, does not cover the need for finer type specifications. For example, a *E22.Human-made Object* can have the particular type “Tombstone” which is a concept originating from a thesaurus structure. The relationships of the thesaurus structure can be further explored for revealing connections between items (i.e., CIDOC-CRM instances) as illustrated in [Figure 6.2](#) and discussed by the scenario 2 “Connections across museums” in section *Interoperable Output and Retrieval*. Vlachidis et al. (2018b) discuss the details of the creation of the CrossCult knowledge base which aggregates a set of thesauri resources and the CIDOC Conceptual Reference Model.

The majority of the vocabulary terms originate from three external resources: the Arts and Architecture Thesaurus of Getty (AAT), the EuroVoc multilingual thesaurus and the Library of Congress Subject Authorities (LC) vocabulary. Vocabulary entries originating from these resources were arranged in a hierarchical terminological structure, called the CrossCult Classification Scheme (CCCS), providing a controlled vocabulary for specifying types, subjects and appellations. For example, the museum item titled “Antiohis pediment tombstone” has the type “tombstone”, which is identified by the internal CrossCult reference “943” and links to the AAT term “tombstones (sepulchral monuments)”. The model was also compensated by Dublin Core elements to support references to digital resources such as images and audio-visual media. For example, the *E38.Image* “antiohis.jpg” is assigned a *dc:source* that resolves to the actual web location of the image.

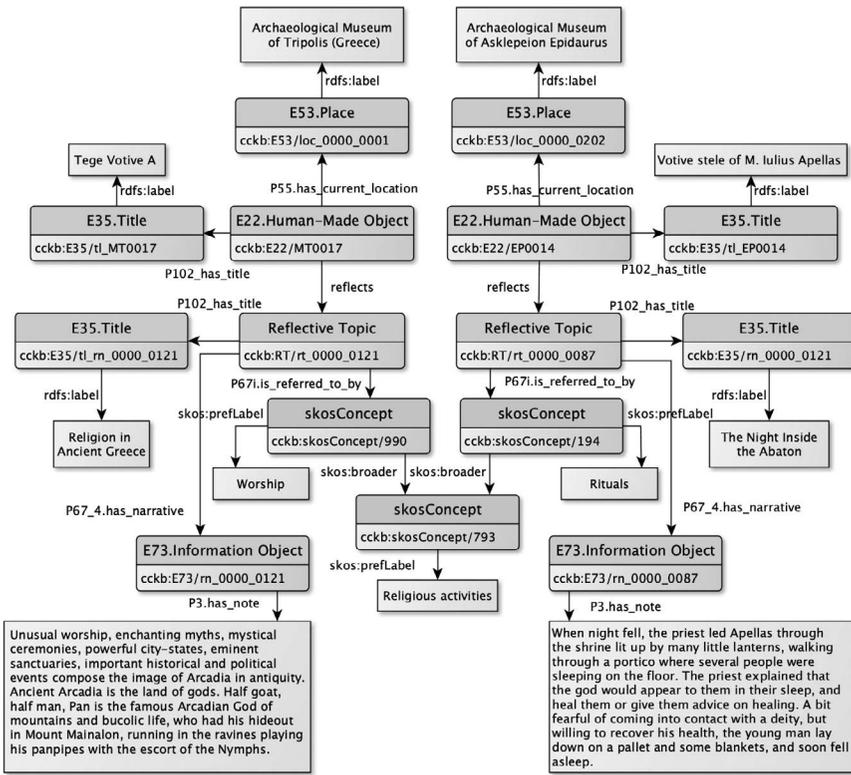


Figure 6.2 Connections across museums based on the concept of religious activities.

It is worth noting that the model supports chains of relationships to describe finer semantics beyond simple subject–predicate–object expressions. The metadata information about the dimensions of a museum item can be elaborated in expressions that define units, actual values and type. For example, an E54.Dimension P2_has_type E55.Type (height), P91_has_unit (cm) and P90_has_value (160). Such relationships can be explored at the discovery layer to enable sophisticated retrieval and complex data aggregation as discussed below. In addition, the museum item ingests subject metadata from the CCCS, which add to its description, enhancing its retrieval capacity. For example, the subject “Roman (culture or period)” is used to describe the item in addition to “tombstone” and “marble” that were used to describe its type and material, respectively.

Semantic enrichment and augmentation

A semantic enrichment and data augmentation phase succeeded the data modelling. A key contribution of the CrossCult data-modelling phase, beyond the data alignment to CIDOC-CRM, was the definition and

specification of the classes and properties responsible for handling the semantics around historic reflection and interpretation. The CrossCult project aimed at spurring a change in the way European citizens appraise history by facilitating interconnections among pieces of cultural heritage information that have cross-cultural, cross border, cross-religion and cross-gender qualities. Central to this endeavour, from a data modelling perspective, was the definition of the *Reflective Topic* entity, which enables the creation of a network of points of view, aiding reflection and prospective interpretation, enabling interconnection between physical or conceptual objects of hand-made or natural origin (Vlachidis et al. 2018b). The *Reflective Topic* can be understood as an extension of the CIDOC CRM *E89.Propositional Object* entity, extended with the project-specific property *reflects* (and its inverse property, *is reflected by*). The Tripoli museum item “2979 (Antiohis pediment tombstone)” was set as the primary subject of reflection for the topics “social status”, “daily life”, “name-animals-myths” and “appearance”.

Reflective topics are abstract propositions that take the form of a keyword or a short phrase. Necessary for their contextualisation is the creation of associated narratives: short stories authored by historians and social scientists. These contextualise a reflection topic with inspiring viewpoints and facts around a particular artefact or a broader collection of exhibits, aiding reflection and re-interpretation by storytelling. For example, the following reflective narrative was used to interweave the museum item “2979 (Antiohis pediment tombstone)” with the reflective topic instance “Daily Life: Middle Class Family”.

A family is represented in this Roman time tombstone found in the amazing Herod Atticus Villa. You can access a video about this unique villa by the Arcadian coast and the excavations [there], as well as a map to take you [there]. Look at the family here. This is the basis of ancient society. Marriage is a very important institution in this ancient society. Only children born inside a marriage can have citizenship. It is unthinkable for a woman to be pregnant outside marriage. There are strict controls for the behaviour of women, since respectable women should stay indoors, should not have contact with males other than her family members and they are usually married at a young age. This is a reassurance to a man that the children he raises are his. This is the time before DNA testing and men worry about the paternity of their children. This is what Telemachus answers to the disguised Athena, when she asks about his parents, in Homer’s *Odyssey*: “My mother certainly says I am Odysseus’ son; but for myself I cannot tell. It’s a wise child that knows its own father”.

The reflective topic instance “Daily Life: Middle Class Family” was related via the CIDOC CRM property *P67i is referred* (inverse property of *P67 refers to*) to a number of subject keywords from the CCCS, such as

“Married Man”, “Married Woman”, “Marriage” and “Family”. Moreover, the reflective narrative is further enriched with links to DBpedia concepts (crowd sourced structured content from the information created in various Wikimedia projects) following a Named Entity Recognition and Linking process (NERL). The process entails the automatic recognition of entities such as persons, places, historic periods etc. and their linking to definitions in the Web (i.e., Linked Data resources). This was achieved using the DBpedia Spotlight tool (Mendes et al. 2011), which automatically recognises mentions of DBpedia resources in natural language text of the target document, following a match and disambiguate process that links unstructured information sources to the Linked Data cloud. The process increased the semantic interoperable properties of the narratives by adding an additional layer of subject heading semantics. For example, the above reflective narrative was linked to the following DBpedia concepts and topics: “Arcadia”, “Athena”, “Family”, “Roman Empire”, “Citizenship”, “Herodes Atticus”, “Ancient History”, “Headstone”, “Homer”, “Odyssey”, “Telemachus”, “Paternity (law)” and “Marriage in ancient Rome”. The NERL process managed to automatically expand the original metadata coverage and to provide a new set of concepts for facilitating retrieval and reuse of the reflective narrative resource and the associated museum item. This benefited the discoverability and cross-linking quality of the item enabling retrieval and interconnection across museum collections, as discussed below.

Interoperable output and retrieval

The semantic metadata was accommodated by the CrossCult Knowledge Base (hereafter CCKB), which provided the framework for facilitating storage, reasoning and retrieval across the disparate museum collections that participated in CrossCult (Vlachidis et al. 2017). Based on a data ingestion process, the metadata was imported in the knowledge base for immediate use or storage. The ingestion covered the whole range of metadata of the 36 museum items, including structure, reference, description, relation to reflective topics and linking to DBpedia entities. DBpedia contains structured information extracted from Wikipedia articles which it makes freely available using Semantic Web and Linked Data standards. The creation of XML Schema Definition (XSD) was necessary for providing consistency to the automated data ingestion process by providing formal description of the elements and structure of the XML documents, which were used as the intermediate data format for mediating the transition of semi-structured data formats to the final OWL output. The resulting OWL statements consisted of class assertions, property assertions and named individual declarations. The final version of the unified OWL structure for 36 items of the Tripoli museum collection, including relevant reflective topics and narratives, contained in total 18,184 axioms and 3,491 ontology individuals of museum items, CCCS vocabulary entries and DBpedia references. Our final

data is made available (following FAIR guidelines) for others to access and reuse (Vlachidis et al. 2018c).

The semantic metadata can facilitate information retrieval and association discovery between museum items stored in the CCKB. The CCKB has been deployed as a triple store and the retrieval scenarios were executed using SPARQL queries. The scenarios demonstrate the capacity of the metadata model to support complex associative queries. The effectiveness of the model has been explored through the project pilot “One venue, non-typical transversal connections” but it has not been formally evaluated using standard information retrieval evaluation metrics. The details of the queries and their results are available in the CrossCult project deliverable D2.4 (Vlachidis et al. 2018a). The following sections discuss particular scenarios that demonstrate the advantages of using metadata semantics for retrieving information and revealing connections, which can leverage serendipity, stimulate curiosity and foster reflection. The examples unfold two separate information seeking and association discovery scenarios. The first scenario promotes the discovery of museum exhibits that are co-located in the same room and have a common reflective topic. The second scenario expands from the first to reveal cross-collection connections by discovering museum items that belong to different museums and share a common reflective topic.

In the first scenario, a user walks into the Archaeological Museum of Tripoli and wishes to find objects relating to a common reflective topic. In total, 15 separate reflective topics are associated with the items in the room, linking to concepts such as “Life events”, “Hair styles”, “Public education”, “Deities”, “Social class”, “Worship”, “Immortality” etc. The user makes the choice to retrieve items and narratives relating to the topic of “Worship”. Three museum items are returned (we use here their item IDs): *MT0006* (a figurine of a woman wearing a veil), *MT0017* (bronze object – votive offering, probable earring) and *MT0018* (bronze object – votive offering, bracelet). The items *reflect* the reflective topic *rt_0000_0121*, which is about “Religion” and “Rituals”. As shown in the diagram in [Figure 6.1](#) below, all three items are located in the Room 9 and reflect the same reflective topic, which is referred by the CCCS concepts “Deities”, “Worship”, “Worshippers”, “Religion” and “Arcadia”. The reflective topic has a narrative, which is about worship and community life of Arcadia in antiquity.

The reflective topics in the CCKB can be composed by others, e.g., books are composed by chapters, which in turn can be composed by sections. The scenario presents the relation of three museum items to a reflective topic, which can be unfolded to a further composition of reflective topics. The reflective topic *rt_0000_0121* is composed of the reflective topics *rt_0000_0136*, *rt_0000_0137*, and *rt_0000_0146*, which are reflected in the museum items *MT0017*, *MT0018* and *MT0006*, respectively. Each reflective topic carries (has) a narrative, which presents a story of an object around a particular topic, in this case “Worship”. For example, the museum item

MT0017, is a bronze object, possibly personal jewellery, which was offered as a votive and was found in Tegea. The museum item *MT0018* is also a votive offering found in Tegea, which according to the respective narrative *rn_0000_0137* “was a very important Arcadian city, known for the temple of Alea Athena”. The item *MT0006* is a clay figurine found in Mantinea, the place of prehistoric Ptolis founded by “Mantineas, the mythical grandson of Pelasgos, the first parent of Arcadians”.

This rich network of narratives and items fosters a rich user experience impossible to deliver without employing the CCKB metadata semantics. The example clearly demonstrates the added value that can be brought by the knowledge base, where three museum items, which in other cases might have gone unnoticed, deliver a rich interlinked narrative that enables users to find out more about worship and ritual in antiquity, particularly linked to the area of Arcadia, Greece.

The second scenario (Figure 6.2) demonstrates the retrieval of information and related narratives of artefacts across museums connected through a common subject or topic of interest. This extends the previous scenario by retrieving items located in different venues, which relate to common or similar reflective topics. For example, a visitor of the archaeological museum of Tripoli, having experienced its narratives and items, wishes to find more items that reflect topics relating to worship, which might be located elsewhere. The item *MT0017* is a votive offering reflecting the reflective topic *rt_0000_0121*, which is about religion and rituals in ancient Greece, and relates to the topic of “Worship”. By semantically expanding on the topic through its broader concept “Religious Activities”, the query retrieves the item *EP0014*, which is located in the Archaeological Museum of Asklepeion in Epidaurus. The item is a votive stele of M. Iulius Apellas, reflecting the Reflective Topic *rt_0000_0087* entitled “The night inside Abaton” and relating to the topic of “Rituals”. The associated narrative is about Apellas experiencing the healing ritual of spending a night in the Abaton, a dormitory for those awaiting Asklepios’ advice on healing. The second scenario demonstrates how meaningful and serendipitous connections between museum items of separate venues can be achieved through reflective topic associations. In this case, a votive offering located in the Archaeological Museum of Tripoli and a votive stele located in the Archaeological Museum of Asklepieion Epidaurus, can trigger reflections around the topics of ritual and worship in the relationships between religion and healing practices in antiquity (as well as providing a means by which to compare museum holdings and collection practices).

Conclusion

The Archaeological Museum of Tripoli case study demonstrated useful lessons regarding the adoption of the FAIR data principles in small heritage organisations. Practical challenges existed, such as the unwillingness and

scepticism towards the use of technology, the cumbersome bureaucracy procedures that delay the processes of digitisation and making the data open using appropriate licences and the difficult communication between humanities and computing experts. Other challenges included selecting the data modelling approach to take, setting up standardised procedures to the population of the ontology given differences in common understandings regarding original descriptions and terminology and balancing the needs of and relationships between the cultural heritage organisation staff, digital humanities researchers and computer scientists.

Despite these initial obstacles, through the CrossCult project the Archaeological Museum of Tripoli has finally received a wealth of technological tools, such as storytelling applications, games and augmented reality. Additionally, its new social media presence has allowed digital promotion, seeing it included in museum networks, becoming more widely known to the public and attracting new visitors. Furthermore, the network of CrossCult museums and sites allowed the museum to be connected to different sites around Europe, through dedicated narratives and the data-led infrastructure we describe here. For example, the Archaeological Museum of Tripoli content was connected to the content of the National Gallery in London and also to the content of the Epidaurus Archaeological site. Both these places attract vast numbers of visitors every year. The content of the Archaeological museum of Tripoli was digitally enhanced with objects from these venues and was also advertised to the visitors of these popular places. In doing so, we demonstrate how preparing collection data to align with standardised ontologies and FAIR principles can help smaller GLAM organisations reuse and repurpose their collection data, allowing their holdings to become accessible to a wider heritage audience.

This has ramifications for other organisations in the cultural heritage sector. Moving from static digitisation of collections, and the work of aggregators of digital collections such as Europeana, we now need to move forward in discovering connections and associations between objects, collections, venues and narratives. In doing so, the semantic data approaches described here (including data cleaning, preparation, aligning with standardised ontologies and thesauri and sharing this data widely) are necessary to structure data, reveal patterns and allow rich interoperability as well as analyses. Additionally, developing robust user testing will allow us to reflect upon where these approaches can be best deployed (user testing of the outcomes of this work is described elsewhere (Dahroug et al. 2019)). Allowing collections data to be as reusable as possible will also be beneficial to organisations at a time of continued austerity in the heritage sector, ensuring that the resources put into digitisation and cataloguing of their collections can be redeployed elsewhere. This is also, then a question of efficiency, expanding ideas regarding why collections are digitised, allowing the data to become more usable and therefore sustainable. The collections effectively become

“collections as data”, allowing others to manipulate, analyse and reframe them (Padilla 2018).

Following the FAIR principles ourselves, we welcome others accessing and reusing the dataset described here (Vlachidis et al. 2018c), which is available under a Creative Commons 4.0 Attribution International license. However, we also identify the need to develop discovery mechanisms for other such datasets, allowing the work undertaken in creating these rich semantic structures to be easily available for others to build upon: the fractured nature of LAM sector work in this area means there is no one clear mechanism for sharing and disseminating semantically enriched collections data. Providing such a mechanism will allow others to see the benefits, and examples, of semantic cross walking of collections, but also avoid both waste and duplication of effort, while providing opportunities for data-led innovation between and across collections.

The processes described in the present work also require resources that are not always available especially for small- and medium-sized museums when the entire collection needs to be digitised. However, the present approach allows the digitisation of a few, representative and carefully selected items that will allow the museum to enter the networked world by also keeping the cost low. The present work, with the digitisation of only 36 items, gave the museum the opportunity to enter a community of connected venues, to promote its content and attract new audiences. We do not underestimate the work and skillsets necessary to successfully adopt semantic linking, recommendation and visualisation techniques, which requires interdisciplinary working groups to successfully navigate overlapping areas of expertise. We recommend that those working in semantic technologies look to heritage collections as a rich use case for development, and an area of deployment that can lead to social good. If those in collections and museums management can understand the benefits of the creation of open, rich, shareable datasets of collections, shareable under FAIR principles, then this should lead to added support for this semantically enriched approach across the sector. At a time of increased social distancing, encouraging open sharing of detailed, structured, collections data can support many across the sector in outreach, engagement and understanding.

Bibliography

- Aloia, Nicola, Franca Debole, Achille Felicetti, Ilenia Galluccio, and Maria Theodoridou. 2017. “Mapping the ARIADNE catalogue data model to CIDOC CRM: Bridging resource discovery and item level access”. *SCIRES-IT-SCientific RESearch and Information Technology* 7 (1): 1–8.
- Antoniou, Angeliki, et al. 2019. “Bringing a Peripheral, Traditional Venue to the Digital Era with Targeted Narratives”. *Digital Applications in Archaeology and Cultural Heritage* 14: e00111.

- APEF. 2019. “Archives Portal Europe Foundation.” <http://www.archivesportaleuropefoundation.eu>
- ARIADNE 2017. “Advanced Research Infrastructure for Archaeological Dataset Networking in Europe”. <https://ariadne-infrastructure.eu/> (Date accessed 27/09/2021)
- Barbuti, Nicola. 2020. “Creating Digital Cultural Heritage with Open Data: From FAIR to FAIR Principles”. In *Digital Libraries: The Era of Big Data and Data Science. IRCDL 2020. Communications in Computer and Information Science*, vol. 1177, edited by Michelangelo Ceci, Stefano Ferilli, Antonella Poggi, 173–181. Cham: Springer.
- Benjamins, V. R., J. Contreras, M. Blázquez, J. M. Doderó, Garcia, E. Navas, F. Hernández, and C. Wert, 2004. “Cultural Heritage and the Semantic Web”. In *European Semantic Web Symposium*, 433–444. Berlin: Springer.
- Beretta, F., 2020. A Challenge for Historical Research: Making Data FAIR Using a Collaborative Ontology Management Environment (OntoME). *Semantic Web* 12 (2): 279–294.
- Berners-Lee, Tim. 2020. “Linked Data – Design Issues”. <https://www.w3.org/DesignIssues/LinkedData.html>
- Bonaccorsi, Andrea, Lucia Piscitello, and Cristina Rossi. 2007. “Explaining the Territorial adoption of new technologies. A spatial econometric approach.” *Applied Evolutionary Economics and Economic Geography*, 256.
- CIDOC. 2018. “Special sessions of CIDOC CRM”. http://www.cidoc2018.com/special_sessions (Date accessed 27/09/2021)
- CLARIN. 2021 “European Research Infrastructure for Language Resources and Technology”. <https://www.clarin.eu/> (Date accessed 27/09/2021)
- Clough, Paul, Jennifer Marlow, and Neil Ireson. 2008. “Enabling Semantic Access to Cultural Heritage: A Case Study of Tate Online”. In *Proceedings of the ECDL Workshop on Information Access to Cultural Heritage*, 978–990.
- CrossCult. 2019. “Empowering Reuse of Digital Cultural Heritage in Context-aware Crosscuts of European History”. <https://cordis.europa.eu/project/id/693150> (Date accessed 27/09/2021)
- Dahroug, Ahmed, et al. 2019. “Using Dates as Contextual Information for Personalised Cultural Heritage Experiences”. *Journal of Information Science* 47 (1): 82–100. <https://doi.org/10.1177%2F0165551519871823>
- DARIAH. 2021. “Digital Research Infrastructure for the Arts and Humanities”. <https://www.dariah.eu/> (Date accessed 27/09/2021)
- de Haas, Tymon, and Martijn van Leusen. 2020. “FAIR Survey: Improving Documentation and Archiving Practices in Archaeological Field Survey Through CIDOC CRM”. *Fasti On/line Documents & Research*.
- Doerr, Martin. 2003. “The CIDOC Conceptual Reference Module: An Ontological Approach to Semantic Interoperability of Metadata”. *AI Magazine* 24 (3): 75–92.
- E-RIHS. 2020. “Europeana and the European Research Infrastructure for Heritage Science”. <http://www.e-rihs.eu/> (Date accessed 27/09/2021)
- Estermann, Beat. 2015. “Diffusion of Open Data and Crowdsourcing among Heritage Institutions. Based on data from Finland, Poland, Switzerland, and The Netherlands”. *Presented at the EGPA 2015 Conference, held on 26–28 August 2015 in Toulouse, France*. http://survey.openglam.ch/publications/EGPA2015_Estermann_Diffusion_of_Open_Data_and_Crowdsourcing_among_Heritage_Institutions_20150901.pdf

- European Commission. 2007. "Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions on a European agenda for culture in a globalizing world {SEC(2007) 570}": <https://eur-lex.europa.eu/legal>
- Tanacković Faletar, Sanjica, Koraljka Golub, and Isto Huvila. 2017. "The Meaning of Interoperability and its Implications for Archival Institutions: Challenges and Opportunities in Croatia, Finland and Sweden." *Information Research* 22 (1): 463–482.
- Finnis, Jane, and Anra Kennedy. 2020. "The Digital Transformation Agenda and GLAMs – Culture24 Findings and Outcomes. Europeana". <https://pro.europeana.eu/post/the-digital-transformation-agenda-and-glams-culture24-findings-and-outcomes>
- Gangemi, Aldo, Nicola Guarino, Claudio Masolo, Alessandro Oltramari, and Luc Schneider. 2002. "Sweetening Ontologies with DOLCE". In *International Conference on Knowledge Engineering and Knowledge Management*, 166–181. Springer: Berlin.
- Gangemi, Aldo, Lehmann Jos, and Carola Catenacci. 2008. "Norms and Plans as Unification Criteria for Social Collectives". *Autonomous Agents and Multi-Agent Systems* 17: 70–112.
- Go-Fair.org. 2020. "FAIR Principles". <https://www.go-fair.org/fair-principles/>
- Gombault, Anne, Allal-Chérif Oihab, Décamps Aurélien. 2016. "ICT Adoption in Heritage Organizations: Crossing the Chasm". *Journal of Business Research* 69 (11): 5135–5140.
- Gruber, Thomas R. 1993. "A Translation Approach to Portable Ontologies". *Knowledge Acquisition* 5 (2): 199–220.
- Heritage Data Reuse Charter. 2020. "Mission Statement of the Heritage Data Reuse Charter". <https://datacharter.hypotheses.org/mission-statement>
- Hyvönen, Eero. 2012. "Publishing and using cultural heritage linked data on the semantic web". *Synthesis Lectures on the Semantic Web: Theory and Technology* 2 (1): 1–159.
- Kontiza, Kalliopi, et al. 2020. "On How Technology-Powered Storytelling Can Contribute to Cultural Heritage Sustainability across Multiple Venues – Evidence from the CrossCult H2020 Project" *Sustainability* 12 (4): 1666. <https://doi.org/10.3390/su12041666>
- Koster, Lukas, and Saskia Woutersen-Windhouwer. 2018. "FAIR Principles for Library, Archive and Museum Collections: A Proposal for Standards for Reusable Collections". *Code4Lib Journal* 40. Available from: <https://journal.code4lib.org/articles/13427>
- Lewis, Rhiannon, and John Stack. 2020. "Sidestepping the Limitations of Collection Catalogues with Machine Learning and Wikidata. Heritage Connector Blog". *Science Museum Group*, accessed March 3 2021. <https://thesciencemuseum.github.io/heritageconnector/post/2020/09/23/sidestepping-the-limitations-of-collections-catalogues-with-machine-learning-and-wikidata/>
- Lykourantzou, Ioanna, Yannick Naudet, and Luc Vandennebeele. 2016. "Reflecting on European History with the Help of Technology: The CrossCult Project". *GCH*, 67–70. <http://dx.doi.org/10.2312/gch.20161384>
- Mendes, Pablo N., Max Jakob, Andrés García-Silva, and Christian Bizer. 2011. "DBpedia Spotlight: Shedding Light on the Web of Documents". In *Proceedings of the 7th International Conference on Semantic Systems (I-Semantics '11)*, 1–8. New York, NY: Association for Computing Machinery. <https://doi.org/10.1145/2063518.2063519>

- Miles, Alistair, and Sean Bechhofer. 2009 “SKOS Simple Knowledge Organization System Reference”. *W3C Recommendation*. <https://www.w3.org/TR/skos-reference/>
- Mons, Barend. 2018. *Data Stewardship for Open Science. Implementing FAIR Principles*. Boca Raton, FL: CRC Press.
- Museum-API. 2015. “Helping Cultural Heritage Organisations Make Content Re-Usable; Helping Programmers Access Cultural and Historic Content Through Open Cultural Cata”. <http://museum-api.pbworks.com>
- Nauta Jan Nauta, Wietske van den Heuvel, and Stephanie Teunisse. 2017. “D4.4 Report on ENUMERATE Core Survey 4”. *Europeana DSI 2- Access to Digital Resources of European Heritage*. Europeana. https://pro.europeana.eu/files/Europeana_Professional/Projects/Project_list/ENUMERATE/deliverables/DSI-2_Deliverable%20D4.4_Europeana_Report%20on%20ENUMERATE%20Core%20Survey%204.pdf
- Padilla, Thomas G. 2018. “Collections as data: Implications for enclosure”. *College and Research Libraries News* 79 (6): 296–300.
- PARTHENOS. 2019. “Pooling Activities, Resources and Tools for Heritage E-research Networking, Optimization and Synergies”. <https://www.parthenos-project.eu/> (Date accessed 27/09/2021)
- Peacock, Darren. 2008. “Making Ways for Change: Museums, Disruptive Technologies and Organisational Change”. *Museum Management and Curatorship* 23 (4): 333–51. <https://doi.org/10.1080/09647770802517324>
- Sauermann, L., R. Cyganiak, and Völkel, M. 2008. “Cool URIs for the Semantic Web”. <https://www.w3.org/TR/cooluris/>
- Terra delle Gravine. 2017. “Percorso Partecipato per la Progettazione Culturale.” <https://terradellegravineprogettazioneperlacultura.wordpress.com/il-progetto/> (Date accessed 27/09/2021)
- Terras, Melissa. 2012. “Being the Other: Interdisciplinary Work in Computational Science and the Humanities”. In *Collaborative Research in the Digital Humanities: A Volume in Honour of Harold Short, on the Occasion of his 65th Birthday and his Retirement*, edited by Willard McCarty, and Marilyn Deegan, 213–230. London: Routledge.
- van Reisen, Mirjam, Mia Stokmans, Mariam Basajja, Antony Otieno Ong, Christine Kirkpatrick, and Barend Monshidden. 2020. “Towards the Tipping Point of FAIR Implementation”. *Data Intelligence* 2 (1–2): 264–275.
- Vassallo, Valentina, and Achille Felicetti. 2020. “Towards an Ontological Cross-Disciplinary Solution for Multidisciplinary Data: VI-SEEM Data Management and the FAIR principles”. *International Journal of Digital Libraries*. <https://doi.org/10.1007/s00799-020-00285-5>
- VI-SEEM. 2018. “Virtual Research Environment in Southeast Europe and the Eastern Mediterranean”. <https://vi-seem.eu> (Date accessed 27/09/2021)
- Vlachidis, Andreas, and Douglas Tudhope. 2016. “A Knowledge-based Approach to Information Extraction for Semantic Interoperability in the Archaeology Domain.” *Journal of the Association for Information Science and Technology* 67 (5): 1138–52.
- Vlachidis, Andreas et al. 2017. “The CrossCult Knowledge Base: A Co-inhabitant of Cultural Heritage Ontology and Vocabulary Classification”. In *New Trends in Databases and Information Systems. ADBIS 2017. Communications in Computer and Information Science*, vol. 767, edited by Mārīte Kirikova et al., 353–362. Cham: Springer. https://doi.org/10.1007/978-3-319-67162-8_35

- Vlachidis, Andreas et al. 2018a. “CrossCult D2. 4 Refined Digital Cultural Resource Data & Data Structure”. <https://www.crosscult.lu/en/resources/deliverables/>
- Vlachidis, Andreas et al. 2018b. “Semantic Representation and Enrichment of Cultural Heritage Information for Fostering reinterpretation and Reflection on the European history”. In *Digital Cultural Heritage. Lecture Notes in Computer Science*, vol. 10605 edited by Marinos Ioannides, 91–103. Cham: Springer. https://doi.org/10.1007/978-3-319-75826-8_8
- Vlachidis, Andreas, et al. 2018c. “The CrossCult Knowledge Base (Version Final, v3.8 owl) [Data set]”. Zenodo. <http://doi.org/10.5281/zenodo.1209325>
- W3C. 2015. “World Wide Web Consortium – Semantic Web Standards.” <https://www.w3.org/standards/semanticweb/>
- Wilkinson, M. Dumontier, et al. 2016. “The FAIR Guiding Principles for Scientific Data Management and Stewardship”. *Scientific data* 3: 160018. <https://doi.org/10.1038/sdata.2016.18>