Cultural Heritage Information Retrieval: Data Modelling and Applications

Babak Ranjgar[a], Abolghasem Sadeghi-Niaraki[a,b*] and Maryam Shakeri[a], Soo-Mi Choi[b]

[a] *Geoinformation Technology Center of Excellence, Faculty of Geodesy & Geomatics Eng., K.N. Toosi University of Technology, Tehran, Iran*
[b] *Department of Computer Science and Engineering, Sejong University, Seoul, Republic of Korea.*

*Corresponding author. E-mail: a.sadeghi.ni@gmail.com

Abstract. The Cultural Heritage (CH) community is one of the domains to adopt Semantic Web recommendations and technologies, which can provide interoperability between various organizations by creating a shared understanding in the community. The CH employed Semantic Web technologies step by step along its evolution process for better knowledge management and a uniform understanding among the community. To identify this evolution process, there is a need to review CH knowledge engineering and the process to improve information retrieval, which new researchers could follow the newest developments in the area. This paper presents this process from its initial steps and the various challenges faced to the latest developments in the CH information retrieval. CH has the goal of preserving and dissemination of the historical information to people and society. Therefore, by making data machine-readable and achieving data interoperability thus a better information retrieval, there is a wide set of opportunities to develop smart applications based on rich CH information as a form of interactive, user-friendly, and context-aware dissemination of information to users. We also reviewed intelligent applications and services developed in the CH domain after establishing semantic data models and Knowledge Organization Systems. Finally, challenges and possible future research directions are discussed.

Keywords: Cultural heritage, data modelling, Semantic Web, information retrieval, ontology, Knowledge Organization Systems

1. Introduction

If the problem was shortage and unavailability of information in some 30 years ago, today it is information overload with the advent of digitization and more importantly, the web. With the revolution of the web, information accessibility became easier and faster. More effort was put in to digitizing information in papers and creating central databases to store the data produced and also to find and reuse them efficiently by taking advantage of the technological advancement of computers. There are two important issues here as discussed in [1]. The first is technical interoperability, which is solved by the decentralized architecture of the web and its platform independent protocols for data sharing and exchange. The very web itself lead to the second problem, which is semantic interoperability. By connecting vast amounts of databases with unstructured data with no or little standardization, it caused a trouble that is called “the digital dark age” [2]. In the current web of documents, one can only search for words and their co-occurrences [3]. However, this is not a suitable way to search and retrieve information, since users do not always know the name of the thing they are searching for or basically their question is a semantic one. For example, artists who lived in a desired city during a special period of time. The current web cannot handle these types of queries, and it has certain limitations. There is a need for data integration and understanding to reach short-term accessibility and long-term preservation, or the data produced with great deal of effort and high cost will fade into disuse, or even worse, be unusable [4]. In the late 20th century, the Semantic Web was proposed to solve this problem [5] and since then it has been an active research field. The Main aim of the Semantic Web is to transform the current web of documents into a web of data and information by making the available data machine-readable [6]. With machines understanding the data, information retrieval can be easier, better, and faster. The Semantic Web has developed standards and technologies to structure and harmonize heterogeneous data, and its latest recommendation is design and usage of formal ontologies to achieve that goal. Of course, knowledge organization and information integration is not a new idea [4]. This idea of formal ontology is based on the valuable past efforts and the traditional knowledge engineering methods.

The Cultural Heritage was one of the first domains to adopt Semantic Web methods, tools, and recommendations [7], [8], [9] for modeling collections of memory organizations, which are also known as GLAMs. This is because of its needs and the importance of its goal, which is to record and preserve heritage knowledge that is a society's identity and also disseminate it in a way to be reusable and accessible to their people. Establishing data models and information integration standards and knowledge management in the Cultural Heritage domain is of great importance, because its data has different formats and types. Also, scientists and specialists from many communities and expertise contribute to this multidisciplinary field [10]. The data in this domain has different types of forms, such as texts, audios, videos, images, 3D models, and spatial data. This data is also related to various types of subjects, such as art, literature, archaeology, spatial science and geometry, physics, and architecture. Additionally, data acquisition and curation techniques differ from archive to archive and from country to country. Alongside these issues, lack of standards and shared understanding has had a substantial effect on data heterogeneity in the Cultural Heritage domain.

In this paper, we are going to present the efforts invested in the cultural heritage knowledge engineering and the process in the way to develop better information retrieval systems, mature top level schemas and data models. Furthermore, there will be discussion about the work to publish the structured knowledge for it to be used and reused. Then, the intelligent and context-aware services and applications developed in the CH domain will be reviewed, which were only possible based on the mature data integration and harmonizing systems. The paper concludes with the discussion of the challenges ahead and possible future research needs.

2. Methodology

Developing better information retrieval methods lies within information science and knowledge management areas of expertise. Therefore, in order to take a survey of efforts in this manner, we had to search for work about knowledge management and knowledge organization in the CH domain. With our method, we came across to several famous projects, such as Europeana, CultureSampo, ARIADNE, and EEXCESS. We also found successful data models that were developed, such as CIDOC CRM, and EDM. Following these models lead to a better and more complete understanding of the progress made.

However, it seemed a little incomplete and partial just to focus on data models and techniques developed for information retrieval. After all, these models were not developed for their own sake and there were definitely some higher level goals behind them. Especially in the CH domain with such vital information that their preservation, organization, management, manipulation, and dissemination are of great importance for the memory conservation of a society and the world. We decided to divide this paper in two main parts. The first part focuses on information modelling efforts in the CH domain with the goal of dealing with the heterogeneity of CH data and achieving interoperability, and the second on taking advantage of the interoperable information to develop interactive, user-friendly information retrieval system. Therefore in second part the focus is on smart publishing systems and intelligent applications and services developed based on models and structures in the first part. Figure 1 illustrates the general approach taken in this research with each part consisting their subparts, which are discussed in detail in the paper. These two steps are substantial to take the CH knowledge and prepare them for presentation to users in a convenient and efficient way.

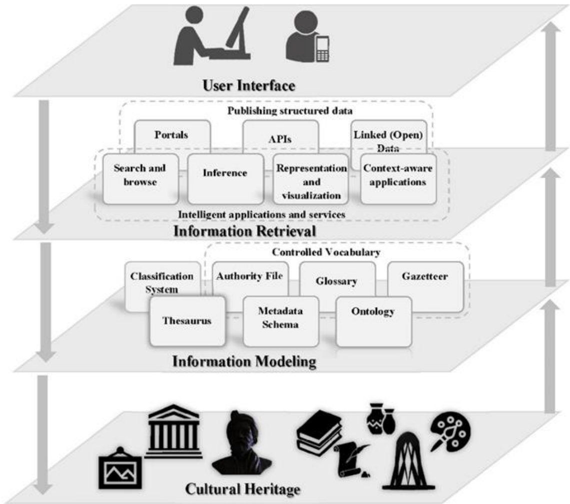

Fig. 1. Overall methodology of the research.

3. Information modelling

3.1. Preliminary knowledge organization systems (KOSs): early solutions

From the very beginning, human beings were interested in classifying and categorizing different branches of knowledge in a hierarchical, so-called "tree-like" method [4]. As a result, the simplest form of knowledge organization systems were classification systems. After that there were controlled vocabularies and thesauri. These systems were created and used before the web era in libraries, museums, and archives. With the advent of computers and the web, there were computerized versions of them to search and find the information in central database systems, but after going online there were problems that were discussed in the previous section. These types of KOSs were not enough to address the heterogeneity of the data and semantic interoperability. The Semantic Web and its technologies were started to handle the previously mentioned issues. The initial recommendation of the Semantic Web was to use metadata schemas to describe the resources on the web in a machine-readable form to better structure and thus retrieve information. Although, metadata schemas were a breakthrough solution, it was not yet enough and had some drawbacks, which lead to ontological data models. In this section, we discuss traditional knowledge organization systems and the steps taken towards metadata schemas. The formal ontologies and conceptual models are based on the past KOSs. Without understanding them and the challenges and issues that were faced, it would be difficult to understand what ontological data models are. The evolution mentioned is shown in Figure 1, which will be discussed in detail in the following sections.

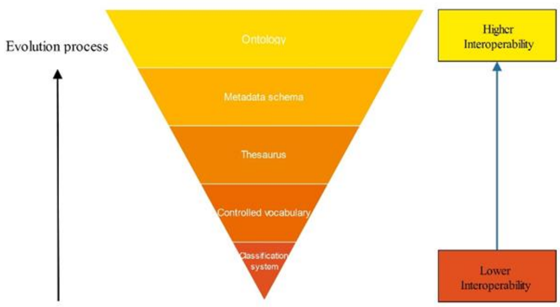

Fig. 2. Process of Knowledge Organization Systems evolution

3.1.1. Classification systems

Classification systems intend to organize the knowledge for information storage and retrieval purposes mostly in libraries [11]. With these types of systems, users are able to browse through the collection for their content of interest without prior knowledge of its existence [12]. Later, they were converted to computer formats that created digital libraries, which provided the search and find service online. One of the first classification systems that gained widespread attention was the Dewey Decimal Classification (DDC), which is a system of 10 numeric sections with decimal extensions. Later, it was combined with bibliographic classification and punctuation marks and symbols to link and relate different areas of knowledge. This system is named the Universal Decimal Classification (UDC), which is now used in 150000 libraries in 130 countries and is published in over 40 languages. Its web service is available on webdewey[1].

The Library of Congress Classification (LCC)[2] is another classification system which was initially developed for the Library of Congress in the late 19th century. It uses letters for classes and each class has a subclass that is identified by two letters.

Iconclass[3] is a classification system designed for art and iconography. It is a well-known tool used for the description and retrieval of subjects represented in images. Ten main divisions of Iconclass are coded by digits 0 to 9. The Classes have subdivisions both in digits and letters.

3.1.2. Controlled vocabulary

A controlled vocabulary or term list is an ordered set of limited words and phrases, which are used to index content [13]. Vocabulary control is used to standardize the naming and provide uniformity, which improves indexing, browsing, and the retrieval of data [14]. There are four types of controlled vocabularies, which include authority files or lists, glossaries, dictionaries, and gazetteers [12]. Dictionaries are lists of words in an alphabetic order that include their definitions. They have no special structure within them and are not of much interest in the CH domain.

Authority files are lists of terms, names, and phrases that are used to control the variant names for an entity. This type of controlled vocabulary is used mostly in the library domain, where the bibliographic records are arranged through a procedure called authority control. Changes in a person's name can occur due to a variety of reasons, such as artistic nicknames, and personal reasons. In these cases, the use of an authoritative controlled vocabulary maintains a consistent method of referring to the same entity with the same name within the bibliographic catalogue. It also accounts for alternatives that should refer back to the standardized designated name [4]. Examples of such lists are the LCNAF[4] (Library of Congress Name Authority File) and the INIS's Authority List for Journal Titles. There are many lists of this kind in different countries, which encouraged libraries to aggregate their data to form a complete reference list. The United States Library of Congress, the OCLC (Online Computer Library Center), and the German National Library began a proof of concept project to link their authority records in 1998. After four years of testing this method, this group formed the VIAF[5] (Virtual International Authority File) consortium at the 69th IFLA (International Federation of Library Associations and Institutions) General Conference. Later, many libraries from various countries contributed to the VIAF, which became an OCLC service.

A glossary is a list of words and terms from a specific subject field or from a particular work, and it usually contains their definitions. It is used mostly within the archive domain to help with research in archives collection and records. The Glossary of Archives and Records Terminology[6] of the Society of American Archivists (SAA) and the Glossary of the Rules for Archival Description[7] are examples of these type of lists.

[1] http://dewey.org/webdewey/
[2] https://www.loc.gov/catdir/cpso/lcc.html
[3] http://www.iconclass.nl/
[4] http://id.loc.gov/authorities/names
[5] https://viaf.org/
[6] https://www2.archivists.org/glossary
[7] https://www2.archivists.org/glossary/terms/r/rules-for-archival-description

A gazetteer is a list of place names. Traditional gazetteers were some sort of a geographic dictionary that were published as a book or in conjunction with maps or atlases. The contents of a gazetteer can include a subject's location, the feature types (e.g. river, town, etc.) country, state, and other descriptive information. The Gazetteer of British Place Names[1] and the World-Historical Gazetteer[2] are two examples of many of their kinds.

3.1.3. Thesaurus

A thesaurus is a type of controlled vocabulary that establishes relationships among its terms using taxonomies and a variety of semantic relations, such as hierarchy, equivalence, and association. These relations are clearly displayed by standardized relationship indicators which are employed reciprocally [15]. Thesauri are much more functional when it comes to retrieval of information from a system [16]. Relationships are usually indicated by the notation BT (Broader Term), NT (Narrower Term), SY (Synonym), and RT (Associative or Related Term). However, relations can exceed the ones mentioned above in some thesauri. These types of relationships and structures makes a thesaurus resemble an ontology, but they are an exploration of terms rather than formalized conceptual entities. Furthermore, the lack of a definition of relating functions, has resulted in less or no ontological commitment [4].

One of the top level thesauri in the CH domain is the UNESCO Thesaurus[3], which covers a wide range of subject fields, such as education, culture, natural sciences, social and human sciences, communication, and information. It is compliant with the ISO 25964 standard that includes all aspects of developing a monolingual or multilingual thesaurus. Many thesauri have been developed based on it and it serves as a top level thesaurus. For example, the UKAT[4] (United Kingdom Archival Thesaurus) is a thesaurus that was developed on the basis of the UNESCO Thesaurus for archives in the UK to help with indexing their collections and catalogues. The LCSH[5] (Library of Congress Subject Headings), which is now in its 40^{th} version, is a complete thesaurus of subject headings used for bibliographic records and it is maintained by the Library of Congress. Another thesaurus from the Library of Congress is the TGM (Thesaurus for Graphic Materials) which is a tool for indexing visual materials by subject and by genre/format. The thesaurus includes more than 7,000 subject terms and 650 genre/format terms to index the types of photographs, prints, design drawings, ephemera, and other pictures. In fact this is a merged form of the previously separated two thesauri of the TGM I (Thesaurus for Graphic Materials I: Subject Terms) and the TGM II (Thesaurus for Graphic Materials II: Genre and Physical Characteristic Terms) since 2007. The most used thesauri in the CH domain are possibly those developed by the Getty Institute. The Getty vocabularies[6] (AAT, TGN, ULAN, and CONA) contain structured terminology for art, architecture, decorative arts, material culture, archival materials, visual surrogates, conservation, geographic names, the names of artists, and bibliographic materials. Compliant with international standards of ISO and NISO, they provide authoritative information for catalogers, researchers, and data providers. They were and continue to be critical contributions to cultural heritage information management and documentation. The AAT (Art and Architecture Thesaurus) is for generic concepts related to art, architecture, conservation, archaeology, and other cultural heritage. It includes work types, styles, materials, and techniques. The CONA (The Cultural Objects Name Authority) is composed of titles, attributions, depicted subjects, and other metadata about works of art, architecture, and other cultural heritage, which are both extant and historical, physical and conceptual, linked to museum collections, special collections, archives, libraries, and other resources. The ULAN (The Union List of Artist Names) is a structured vocabulary, that includes names, biographies, related people, and other metadata about artists, architects, firms, studios, museums, patrons, sitters, and other people and groups involved in the creation and study of art and architecture. The TGN (The Getty Thesaurus of Geographic Names) is a structured vocabulary that includes names, and descriptions for extant and historical cities, empires, archaeological sites, and physical features important to the research of art and architecture.

---

[1] https://www.gazetteer.org.uk/
[2] http://whgazetteer.org/
[3] http://vocabularies.unesco.org/
[4] https://ukat.aim25.com/
[5] http://id.loc.gov/authorities/subjects.html
[6] www.getty.edu/research/tools/vocabularies/

3.1.4. Metadata Schemas

As previously mentioned, the Semantic Web has the goal to convert the current web of documents to a web of data by providing machine-readable formats for information. Metadata schemas are actually machine-readable data about data and according to NISO, they are intended to increase data exchange with minimal loss of content and functionality through platform independent approaches. A metadata consists of a set of elements that are usually structured in a form of textual information [4], which describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information resource [17]. Generally, metadata schemas are classified in two categories [18]:

1. Descriptive Metadata describes an information resource which can also be broken down into two subcategories [17]:
   1.1. Content-based metadata that describes the content of a resource through tags, such as genre for movies and books or material type for an artifact.
   1.2. Content-independent metadata that is not about content of the resource, but it is associated with it, such as an author of a book or the last modification of a multimedia object.
2. Administrative metadata is used for managing collections and resources, and it stores information, such as the acquisition state and the location of information.

Two key necessary technologies to create and use metadata schemas are XML and RDF, which are W3C recommendations. XML[1] (eXtensible Markup Language) is a markup language that is similar to HTML, but its tags are not predefined. It can be extended to any field of interest, and it is both human-readable and machine-readable. XML is independent from platforms and languages, and it has a fundamental role toward interoperability. However, as XML is only at syntactic level, machines cannot clearly determine the meaning of XML tags. As a result, W3C has developed RDF with the goal of addressing the XML problems by adding semantics on top of the XML [19]. RDF[2] or Resource Description Framework is a data model similar to classical conceptual modeling (entity-relationship) for representing and modelling information about web resources. These description are in the form of subject-predicate-object called a triple. Predicate indicates a relationship between the object and the subject that are unique web resources and have a stable web identifier called a URL (Uniform Resource Locator). This is an important issue, since it helps to resolve the uniformity of an identity problem in the harmonization of different information sources. The triples of RDF are actually called statements, and the subject and a set triples can form a linked graph with subjects and objects as nodes and predicates (or properties) as edges.

Metadata schemas are only possible with the aforementioned technologies and basically relational databases. The Cultural Heritage domain has developed many metadata in its various fields, such as libraries, archives, and museums. One of the most prominent metadata schemas is the Dublin Core[3]. It originates from libraries and now is used in many other organizations. DC contains 15 core elements that is called the Dublin Core Metadata Element Set (DCMES), which includes the title, creator, and the date. These elements were later extended to 55 elements called the DCMI (Dublin Core Metadata Initiative) for a broader range of purposes and of business models. From the very start of the RDF model by W3C, DC adapted it, and it became a popular metadata for use with RDF [9].

Before the web era, the Library of Congress started an initiative in the 1960s to create MARC (MAchine-Readable Cataloging), which later became an international standard. In 1999, MARC21 was designed by combining United States and Canadian MARC formats (USMARC and CAN/MARC). It was name MARC21, because it was refined for the 21st century and to make it more accessible to the international community. Later, it adapted an XML markup language and developed MARCXML[4] in order to facilitate the sharing of and the networked access to bibliographic information. The Library of Congress' Network Development and the MARC Standard Office developed MODS[5] (Metadata Object Description Schema) which is much easier to understand for humans compared to MARCXML, as it uses language-based tags rather than three-digit numeric tags. Moreover, it is compati-

[1] https://www.w3.org/XML/
[2] https://www.w3.org/RDF/
[3] http://dublincore.org/
[4] http://www.loc.gov/standards/marcxml/
[5] http://www.loc.gov/standards/mods/

ble to outside metadata, such as DC and its mapping is more convenient. MADS[1] (Metadata Authority Description Schema) is an XML schema developed by the same organization to provide authority element sets and complement existing object descriptive MODS metadata.

VRA (Visual Resource Association) Core Categories[2] are developed based on DC to describe the work of visual culture as well as the images that document them. This standard is hosted by the Network Development and the MARC Standard Office of the Library of Congress in partnership with the Visual Resource Association. After a series of revisions, it is now called Core4, which is the only metadata standard devised especially for the description of images and the cultural heritage objects they represent.

The CDWA[3] (Categories for the Description of Works of Art) is a set of guidelines and also a metadata schema for the description and cataloging works of art, architecture, groups and collections of works, and related images. The CDWA includes 532 categories and is more expressive than the VRA Core. Also, it is maintained by the Getty Institute.

Table 1. Summary of traditional KOSs

Type		Name	Usage area	References
Classification System		UDC	Classification of library sources of knowledge	[11], [12]
		LCC	Classification of library sources of knowledge	
		Iconclass	Classification of images of art and iconography by the subject depicted	
Controlled Vocabulary	Authority File	LCNAF	Authority control of names of persons for bibliographic reason	[4], [12], [13], [14]
		Authority List of Journal Titles	Authority control of names of journals	
	Glossary	Glossary of Archives and Records Terminology	Glossary of terms used in archives and its records in US & Canada	
		Glossary of the Rules for Archival Description	Glossary of rules used in description of archival records	
	Gazetteer	Gazetteer of British Place Names	Indexing place names and their historic and administrative county in Great Britain	
		World-Historical Gazetteer	Indexing various places such as historic, natural, residential, and etc. all over the world	
Thesaurus		LCTGM	Indexing visual materials by subject and by genre/format	[4], [15], [16],
		LCSH	Thesaurus of subject headings used for bibliographic records in libraries	
		ULAN	A structured vocabulary, including names, biographies, related people, and etc. about artists.	
		TGN	A thesaurus, including names, and descriptions for extant and historical cities, empires, archaeological sites.	
		CONA	Titles, attributions, depicted subjects, and other metadata about works of art, architecture, and other cultural heritage.	
		AAT	A thesaurus about cultural heritage, including work types, styles, materials, techniques, and etc.	
		UKAT	Indexing collections and catalogues within archives in UK	

[1] https://www.loc.gov/standards/mads/
[2] https://www.loc.gov/standards/vracore/
[3] http://www.getty.edu/research/publications/electronic_publications/cdwa/

Metadata Schema	DC	Metadata for web resources	[4], [9], [17], [18], [19], [20]
	MARCXML	Metadata for bibliographic information of libraries	
	MODS	User-friendly form of MARCXML	
	MADS	Provides authority element sets to complement MODS metadata.	
	VRA Core	Describing work of visual culture	
	CDWA	For description and cataloging works of art, architecture, and related images	
	MARC AMC	Exchanging information about archival data	
	EAD	Encoding standard for archival finding aids	

In 1977, the SAA (Society of American Archivists) initiated a working group (NISTF) to develop a method for exchanging information about archival data. MARC AMC[1] (MARC for Archives and Manuscripts Control) metadata was created as a result of efforts of the task force. Since the MARC standard only supports one level of description, it was not a substitute for the more detailed finding aids that were produced by the archivists [20]. This problem encouraged the Berkeley Finding Aid Project to create a platform-independent, machine-readable encoding standard for archival finding aids. The EAD[2] (Encoded Archival Description) is an XML standard developed for this matter. It is based on the notion that archives are hierarchical in nature, and its descriptions are based on inheritance which enables it to provide information on different levels of detail [20]. The EAD standard is jointly administered and maintained by the United States Library of Congress and the Society of American Archivists. The following table summarizes the traditional KOSs discussed in this section.

3.2. Data integration at the metadata level

Before we discuss metadata integration approaches, it is better to discuss one last important concept related to traditional KOSs. W3C developed the SKOS[3] (Simple Knowledge Organization System) to support the use of the traditional KOSs, such as classification systems, controlled vocabularies, thesauri, and others that are called concept schemes within the framework of the Semantic Web. These systems were developed with a lot of effort and are incorporated in many organizations and it is not possible to stop using them easily. With SKOS, they can be used in the Semantic Web space and they can be interoperable, so various organizations can exchange their data and data integration can be easier. SKOS provides specifications and standards to represent knowledge organization systems using the Resource Description Framework (RDF). Encoding this information in RDF allows it to be interchangeable between computer applications in an interoperable way. It also enables the population of elements of metadata schemas with them that adds to uniformity of description and accessibility of data over the web. SKOS has three main features to help represent a concept in simple and understandable way [21]:

Labeling properties that are used to connect a concept to the terms that represent it in natural languages so the concept can be represented seamlessly in multilingual environments. For example, skos:altLabel is used to show alternative terms for the concept, such as synonyms and its name in other languages.

Semantic properties that are employed to represent the semantic relationships between terms in a concept such as a thesaurus. For example, skos:broader indicates generalization BT (Broader Term).

---

[1] https://www2.archivists.org/glossary/terms/m/marc-amc-format
[2] https://www.loc.gov/ead/
[3] https://www.w3.org/2004/02/skos/

Documentation properties that are used to encompass the important notes and documentations of a KOS. Notes in a documentation have different roles. SKOS has notations like skos:scopeNote and skos:definition. For explanatory notes and notations, such as skos:historyNote for management notes.

So after the fact that every part of CH data providers and memory organizations developed their own specific metadata schemas describing their own data, efforts began to integrate the data from various institutions to create a virtual large-scale memory organization for seamless access to various and different aspects of the cultural heritage, such as Europeana[1], Netherlands E-culture[2], and CultureSampo[3] in Finland. This would also make information of small-scale organizations richer, and the users could be able to find more information in an interested area by aligning and integrating similar data from different sources. Of course this level of integration is a step behind ontology-based information integration, which is going to be discussed in next section. By understanding this method and its limitation, we can have better insight towards ontological data models in reaching semantic interoperability. There are many approaches for achieving metadata interoperability [22], but two main methods are employed for this matter in the CH domain.

- First, a single metadata schema is chosen and the contents of the databases are transformed into that metadata schema. This method is applied in project MuseumFinland. In [23], the authors stated that this approach guarantees a level of consistency and interoperability, but the enforcement of data into one metadata would cause damage to the rich original data. It loses its own metadata that has specific elements describing itself, and there is no one-size-fits-all metadata for heterogeneous data in the CH domain [24].
- Second, the original metadata schemas and relative KOSs are kept, and a series of alignments and mapping is applied between the metadata to integrate the data and create interoperability between the different schemas and concepts. In this process, the similar and correspondent elements and also the non-correspondent ones are identified. A mapping occurs between similar elements to connect them which is called "crosswalk" [24]. In [21], the authors integrated data from two Dutch CH institutions by aligning their KOSs. First, they SKOSified the two institutions' KOSs, and then they used Falcon and S-Match tools to carry out mapping between them. Finally, they implemented a faceted browser to provide seamless access to collections of both institutions. In [25], a massive amount of cultural heritage objects were chosen from six collections and a series of mapping and alignment was done, since the number of metadata schemas and vocabularies used in various collections were high and a visualization system called "demonstrator" was developed at the end, which offered results for semantic queries of the users. Although, compared to the first one, the second method is better, it still has some drawbacks. In this approach, some of the elements are put aside in the mapping process due to lack of correspondent elements in other metadata, which brings about a loss of information [26].

At the end of the day, metadata is a useful tool for providing content and administration description of data that can help with its retrieval, though it seems not enough for the Cultural Heritage field. In the next section, limitations of pre-ontological knowledge management are discussed in detail as well as the need for another data model.

3.3. Why formal ontologies? (Limitations of traditional KOSs)

Traditional knowledge organization systems are limited in terms of semantic expressivity. Users of such systems are forced to choose from the available limited list of options to search for the information of their interest. These types of standardized frameworks may have satisfactory results in a small-scale and local organization for data entry and retrieval, but it is not a fundamental long-term solution for large-scale data integration in the complicated CH domain [4]. The reasons that indicate that such systems cannot provide interoperability at a large-scale are numerous. First of all, the linguistic limitations that are imposed by language are a major drawback of these systems. There are two kinds of lexical ambiguity, homonymy and polysemy. "The bark of a dog versus the bark of a tree is an example of homonymy; review as a noun and as a verb is an example of polysemy [27]." Inability in distinguishing the meanings of the words is a classical information retrieval problem. The performance of such systems can be improved by incorporating a hierarchical structure, which allocates categories for words. This solution can be useful to disambiguate some terms but not all of them, because some terms cannot be classified in

[1] https://www.europeana.eu/
[2] http://multimedian.project.cwi.nl/
[3] http://www.kulttuurisampo.fi/

a special category [4]. Moreover, the classical hierarchical systems cannot represent fundamental relationships, such as parts and wholes theories, for example, mereology and mereotopology. Another problem with these systems is that they are very committed to their structure and its correctness. They take it as a one to one correspondence to the real world, and this is evident with metadata schemas [4]. Metadata is constructed with a human processing point of view and is not appropriate for automated tools to infer and drive new knowledge from existing information. A reason is this type of information is implicit in metadata and the relation between the entities are not considered as it is in the real world [1]. Therefore, the right solution is one that respects every party involved in the community to reach a consensual conceptualization of the domain independent from linguistic defects and other problems.

Ontologies are of special interest in AI (Artificial Intelligence) and its subfields, such as knowledge engineering and knowledge representation, since they allow for the exchange and reuse of knowledge in computational form [28]. This notion also gained widespread attention in fields of information integration and information retrieval. This is due to what ontologies promise which is to provide a shared understanding of a domain that can facilitate communication between different parties of the community and also computers [28], [29]. This method tries to deal with the information integration problem of heterogeneity with a new approach by avoiding the aforementioned issues of former knowledge organization methods such as linguistic ambiguities or commitment to a single structure that is set to model the real world perfectly. Ontologies do not intend to be in a one-to-one correspondence with the universe, and they have a functional purpose and concentrate on the particular viewpoints of domain users to provide an adequate model to their aims and are consistent with reality [30]. There are many definitions for ontology, but the widely accepted one is given in [31]: An ontology is a formal, explicit specification of a shared conceptualization. A 'conceptualization' refers to an abstract model of some phenomenon in the world by having identified the relevant concepts of that phenomenon. 'Explicit' means that the type of concepts used, and the constraints on their use are explicitly defined. 'Formal' refers to the fact that the ontology should be machine readable, which excludes natural language. 'Shared' reflects the notion that an ontology captures consensual knowledge, and it is not private to a particular individual, but accepted by a group [28]. There is an attempt to understand the concepts not "in general" but with regards to their functionality within the defined domain of use [32].

A formal ontology usually is comprised of a scope declaration and a series of classes and properties extracted from the discourse between the users involved. A class is "a category of items that share one or more common traits serving as criteria to identify the items belonging to the class [33]." which is described by a scope note that indicates the intension of that class by a text. The intension of a class is a description of that category such that a human being can read it and identify instances of it. The clarity of such descriptions is of the highest importance for the effectiveness of an ontology and research presently continues in this area [34]. "A property serves to define a relationship of a specific kind between two classes [33]." Properties are generalizations of types of relations that can be possible among classes. Their formalization results from research into how users actually conduct reasoning and relate objects in the domain [4]. There are two additional concepts that should be defined for a property to form a well-defined ontology. The first is the domain, which is the class that a property is defined for, and every property must have exactly one class as its domain. The other one is the range, which is the class that comprises of all potential values of a property [33]. The specification of these relations is the premise of the possibility of reasoning and inferring over the data at later stages [4]. However, the primary tool to gain expressive power within the ontology is the use of an is-a relation over the classes and properties. Formal ontologies make use of a function of inheritance provided by the is-a relation in order to structure classes from more general to more specific. An ontology is left for open discussion and it never defines all possible classes. Whenever there is no class appropriate for particular data, a revision process starts between knowledge engineers who design the model and the domain community to develop a new (sub) class/ (sub) property within the model to support the new phenomenon [4].

As previously mentioned, formal ontologies should be encoded in a machine-readable formal language to avoid natural language pitfalls. Typical AI languages that can be used for implementing ontologies are description logics for reaching the KR (Knowledge Representing) community needs of representing declarative knowledge. Examples of such description logics include KL-ONE, KIF, LOOM, KRYPTON, and CYCL [28]. It was right after the Semantic Web initiative that substantial progress occurred in this field with development of RDF. RDF is a neutral description tool for web resources that does not define its meaning. RDFS (RDF Schema), which is an extension of RDF, provides small but useful vocabulary including simple taxonomical relationships to declare classes and properties, which makes it a basic tool for implementing ontologies. W3C identified some applications and used cases where the RDFS showed poor expressivity, and its limitations are discussed here [35]. W3C's Web Ontology Working Group developed OWL (Ontology Web Language), which is built upon RDF and RDFS. It is the mostly used ontology language and has gained widespread acceptance, since it covers RDFS limitations and shortcomings. Three versions were developed for OWL (OWL full, OWL DL, and OWL lite) due to a set of different and incompatible needs, such as full RDFS compatibility, efficiency in computation, and the high expressivity power with the combination of RDFS and a full logic. As we go from OWL full to OWL lite, expressivity power and RDF(S) support decreases as a trade-off for higher use convenience and computation efficiency. More details about this matter can be found in [35].

3.4. Ontologies in CH domain

Cultural heritage that has been keeping itself up-to-date with knowledge representation techniques, embraced the ontological modeling of data, as it promised to be a useful tool for information integration and providing interoperability between various parts of the community. As stated earlier, ontologies are functional and intended to model the interactions in a domain with respect to its aims. Functions in this domain as said in [36] are: Collection management that involves tasks such as acquisition, registration, and compiling inventories of objects and their description, hosting exhibitions, providing insurance, rights, and protection zones. Conservation, which is comprised of tasks, such as the diagnosis of deterioration, establishing preventive measures, planning interventions and applying treatments and chemical agents when needed. Research includes investigation, description, interpretation of cultural objects and works. Presentation of retrospective knowledge is simply the most important function of all. Besides the functions mentioned above, the information in the CH domain have some special characteristics that further affects the model, and they should be developed for this domain. Information is usually discrete and lacks consistency that exists in other disciplines, such as geology, and it also has an event-centric meaning, so people, and things are connected via events. Finally, its descriptions are retrospective and about past, which is contrary to information in fields that deal with phenomenon in the future and involve tasks like planning and predicting. Since there is not a unified true assumption of the past, information cannot be integrated and normalized on the basis of an assumed past [36].

Perhaps the most widely known and accepted ontology in the CH domain is the CIDOC Conceptual Reference Model, which provides definitions and a formal structure for describing the implicit and explicit concepts and relationships used in cultural heritage documentation. The CIDOC CRM [37] is a formal ontology intended to promote a shared understanding of cultural heritage information by providing a common and extensible semantic framework that facilitates the integration, mediation, and exchange of heterogeneous cultural heritage information. It can provide the "semantic glue" necessary to mediate between different sources of CH information, such as items published by galleries, libraries, archives and museums (also called GLAMs). The CIDOC CRM is the result of a series of work and discourse by interdisciplinary domain experts and specialist such as computer science, archaeology, museum documentation, history, library science, physics, and philosophy over the years. The International Committee for Documentation (CIDOC) of the International Council of Museums (ICOM) initiated the work to solve knowledge engineering and representation that museums were faced with in late 20th century. The first result of the activities was the CIDOC Relational Data Model, a relational database model with more than 400 tables, which was actually difficult to implement in a wide range [36]. Therefore, the CIDOC Documentation Standards Working Group (DSWG) decided to change to the object-oriented method for its benefits over the relational approach, and this work resulted in the first edition of CRM. This model with 50 classes and 60 properties reduced the complexity of the relational model dramatically, and it encouraged the official creation of the CIDOC CRM Special Interest Group (SIG). This group was responsible for the development of CRM as an international standard for the museum community.

The task was achieved in 2006 since the CIDOC CRM was accepted as an official standard (ISO 21127:2006). Initially, it was released in textual form to stress its independence from specific knowledge representation formats [36]. Later, valid formal definitions for CRM were developed in TELLOS, KIF, RDFS, and OWL. One of the trusted OWL formats of CIDOC CRM, which began from its 4.2.4 version is called Erlangen CRM [47]. It was developed by scientists from Erlangen-Nuremberg University in Germany. Currently, CRM is the only data model that is an ISO standard in the CH domain, and it has gained many attention and acceptance. It has been used in various projects and lots of development is taking place around it. CRM is a bottom-up model based on empirical CH data, and it is open ended, which means it can be extended for new phenomenon observed and specialized for user needs. At the moment, CIDOC CRM is in version 6.2.3 containing 99 classes and 188 properties.

It has established an event-centric approach for modeling data, in which objects, persons, and concepts are connected via events. On its way to becoming a formal ontology for the CH domain with such wide aspects, CRM was harmonized with different top level ontologies to become a core ontology. First, it was harmonized with ABC ontology, which is a data model for integrating multimedia information in digital libraries [48] during the years 2001 and 2003, which both models affected each other. For further reading the technical issues you can refer to [49]. Figure 3 shows major concepts and modelling notion of this ontology. As it can be seen, temporal entities that include events is in the focus of the model and other entities, such as objects, actors, places, and time-spans are connected to it. The classes, type and appellation, can be applied to any class in the model for deeper specializations [37].

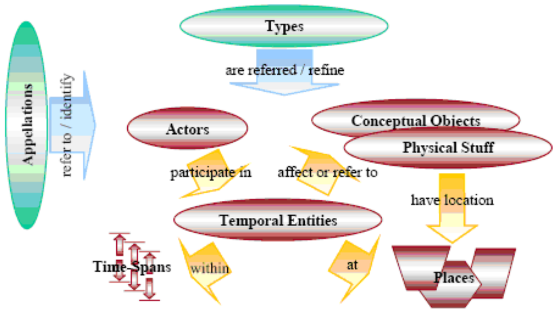

Fig. 3. Major concepts of the event-centric CIDOC CRM model [37].

Another important mediation task that was conducted is the harmonization of FRBR (Functional Requirements for Bibliographic Records) and CIDOC CRM. FRBR is an entity-relationship model for bibliographic information in the library domain developed by IFLA to overcome difficulties in the Dublin Core metadata for the integration and retrieval of information in libraries. It is a data model similar to the object-oriented format of CRM and with this harmonization both models benefited from each other. On one hand, CRM extended its coverage of the CH domain by adopting library information field, and on the other hand, FRBRoo [51] ontology was developed for IFLA, which benefits from the event-centricity of CRM. Recently, an extension for FRBRoo is developed called PRESSoo, which handles documents published continuously and are long lasting serials [52]. For different purposes, several extensions are developed for CRM. CRMdig [38] is an extension to record the description information related to the processes and approaches of production of digital models and representations whether 2D, 3D, animations, and other types created by various technologies. This model actually documents and integrates provenance information which is and essential factor in data evaluation assessment and trustworthiness [39]. CRMsci [40] is about general provenance data in various descriptive and empirical studies related to cultural objects and also scientific observations and measurements carried out. It considers relevant standards, such as INSPIRE (earth science), OBOE (life science), SEEK (ecology), Darwin Core (biodiversity), national archeological standards for excavation, digital provenance models and others. CRMinf [41] is an argumentation model and extends CIDOC CRM formal ontology to integrate metadata about argumentation and inference making in empirical and descriptive sciences. It proposes classes to document states of belief made in the observation phase. This model is not yet completed, and it is under development but a validation process was done in the British Museum Discovering Sloan project. CRMarchaeo [42] is another extension developed in compliance with CIDOC CRM to model the metadata about the archaeological excavation process. The reason for this kind of model was to maximize the interpretation capability and evaluation of the procedure carried out, since archaeological excavation activities are destructive themselves. CRMba [43] is an extension for CRM that was developed to model archaeological information of standing heritage buildings. It is harmonized with CRMarchaeo [44], because it uses archaeological information, such as stratigraphic units from that model and attaches them to relative parts of buildings. It also uses mereology and mereotopology theories between various parts of buildings tailored to their architecture to model such information for heritage buildings. Due to its characteristics, it incorporates classes and properties from other extensions in particular, CRMarchaeo, CRMsci, and CRMgeo which will be discussed later [43]. Lastly, the CRMgeo extension was created to support spatio-temporal reasoning over heritage information that will be discussed in next section. Its major achievements were to harmonize spatial standards of the OGC (Open Geospatial Consertium) with the CIDOC CRM standard. Also, it made some changes to the core entities of the CRM, such as introducing the Space Time Volume (SPV) concept. Details can be found in [45] and [46]. A schematic view of CIDOC CRM core concepts and its extension is shown in Figure 4.

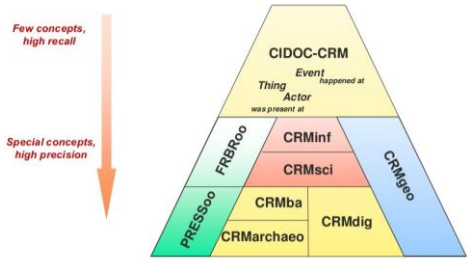

Figure 4. CIDOC CRM core concepts and its extensions [69].

Due to high activities and wide-spread engagements, new extensions are still under development and proposal to CRM SIG, such as extensions A and B, proposed to harmonize MIDM (Multiple Interpretation Data Model) with CRM [50]. CRM is implemented from large-scale projects to small-scale ones. In the ARIADNE project, CRM was used as the backbone ontology model for heterogeneous data integration. ARIADNE (Advanced Research Infrastructure for Archaeological Dataset Networking in Europe) is an e-infrastructure with the purpose of creating a place for archaeological data providers across Europe to register and connect their resources, and it is also a portal with services, such as search, and access. Some of the extensions above (CRMba, CRMarchaeo, CRMgeo, and CRMsci) were developed within this project due to heterogeneity of data involved [53]. The Research Space project developed an infrastructure for integrating the British Museum data and for this matter it used a simplified form of CRM [54]. The WissKI project aims to provide a Virtual Research Environment (VRE) for managing scholarly data in memory organizations that is completely open source and free to use. Also, it will enable researchers to work and collaborate from different places. It has developed a semi-automatic text annotator, which uses semantic web technologies, and Erlangen CRM (ECRM) was employed as its top ontology [55]. Arches is a project supported by the Getty Conservation Institute and the World Monuments Fund. It is a WebGIS tool for management, monitoring, risk mapping, and conservation planning of built heritages. It employs OGC standards for spatial data and analysis and also uses CIDOC CRM to model its database, which ease its use for organizations already compatible with CRM [56].

There are data models that were developed based on CRM in some countries. CRM-EH (English Heritage) was developed by the English Heritage for the specific excavation events data of the Center for Archaeology with series of work with CRM authorities and experts. It was designed with the intention to capture the detailed excavation/analysis procedures [57]. In a project named STAR, a semi-automated tool was developed for extracting data from five archaeological databases and mapping them to the CRM-EH model to achieve interoperability and a better search and retrieval of the information [58].

In Korea, KCHDM (Korean Cultural Heritage Data Model) was developed mainly based on CIDOC CRM. It is an ontological model for integrating heterogeneous heritage data from different institutions in Korea and serve as a mediating means for collecting and connecting various database systems [59].

For the CultureSampo (Finnish culture on the semantic web) project, Hyvönen et al. developed national ontology based on the thesauri of their own country in the FinnONTO [60] project. They just employed content independent recommendations of W3C, such as RDF, SKOS, and OWL, but they converted their national ISO abiding thesauri into light weight ontologies and created the national KOKO ontology infrastructure, which consists of one high level and mediating ontology called YSO and 14 other field specific ontologies [61].

In the Europeana project which aimed to collect, enrich, and provide access to cultural heritage information of institutes all over the Europe, a data model was developed that is called EDM (European Data Model). This top level ontological model was created to replace the older flat ESE (Europeana Semantic Elements) metadata due to general shortcomings metadata schemas, which were elaborated on in section 2.3. The model reuses constructs from other standards, such as Dublin Core and FOAF, to which institutions can map their data (even CIDOC CRM can be mapped to EDM) [62].

MONDIS (MONument Damage Information System) is an ontological framework developed to capture and reason over the built heritage documentation of damages, interventions, changes, and natural disaster occurrences, for diagnosing current condition of the buildings that can be helpful for their conservation. The intention for this ontology was to complete the existing ontologies with the possibility to describe monument damage and its causes and consequences [63].

Recently, HERACLES (HEritage Resilience Against CLimate Events on Site) ontology is being developed in the course of a project with the same name. It aims for better management and monitoring of built heritage health by modeling climate change effects and different types of damage it can cause for various type of materials through specific mechanisms. It is still in the early stages, going through tests and awaiting acceptance of experts and stakeholders [64].

CH from the very beginning embraced Semantic Web technologies, so it evolved as it did. These were some of the data models developed in the CH domain for different purposes (summarized in Table 2). By undergoing various evaluations and experiments, they were updated and modified to cover the problems reported. As a result of these types of activities, the data models became more and more mature during this time. In [65], an evaluation was done on three prominent data models, and its results depicted that they acted well and were appropriate for CH needs based on 6 main criteria that consisted overall 10 sub criteria, especially CIDOC CRM, which is a well-established standard ontology, showing excellent performance in 6 out of 10 and ok in remaining 4.

Table 2. Summary of ontologies in CH domain

Ontology	**Owner/authority**	**Licensing**	**Language(s)**	**Purpose**	**References**
CIDOC CRM	ICOM	?	RDF, RDFS, OWL	An standard general upper ontology for heterogeneous information integration in cultural heritage domain with a wide range application from collections of applied and fine arts to archaeology, built heritage and GLAMs	[37]
EDM	Europeana Foundation	CC0 1.0[1]	RDF(S) & OWL	For aggregation and representation of objects information in Europeana portal coming from GLAMs over the Europe with various metadata schemas	[62]
KOKO	National Library of Finland	CC3.0[2]	RDF	Heterogeneous cultural heritage information integration and representation on semantic web for Finland	[61]
CRM-EH	English Heritage	?	RDF	Integration of archaeological databases containing excavation and finds information	[57]
KCHDM	?	?	OWL	Cultural heritage data integration based on contextual information for South Korea	[59]
MONDIS	?	?	OWL2	Built heritage damage documentation, risk prediction, and intervention recommendation	[63]
HERACLES	HERACLES project	CC BY-SA 4.0[3]	OWL	for effective resilience of built heritage against climate change effects	[64]

3.4.1. CRM vs. EDM

[1] https://creativecommons.org/publicdomain/zero/1.0/
[2] http://creativecommons.org/licenses/by/3.0/
[3] https://creativecommons.org/licenses/by-sa/4.0/

Arguably, EDM and CRM are two of the most dominant and widely used ontologies in the CH domain, which were introduces in the previous section. In this section, we elaborate on the differences between them, and through the process the two different modelling methods are discussed. Both EDM and CRM are top level ontologies developed to model CH data with a set of classes and properties that result in some level of abstraction and interoperability, but they have certain dissimilarities due to their specific intentions.

The outstanding distinction between the models are their structures and the way they organize descriptive information. There are two approaches for modeling information in the CH, which are object-centric and event-centric approaches. In the former, the object is in the center and all other descriptions and information are connected to it. For example an object has a creator, creation date, and a location. In the latter, the information related to the object is connected to the object through different events. For example, an actor's involvement in a production event at a specific period of time and in a particular place leads to the creation of an object. CIDOC CRM uses the event-centric method to model cultural heritage data [37]. In contrast, EDM employs an object-centric approach [66]. Both of the methods have advantages and disadvantages, and it is not our intention to say one is better than the other. EDM is used in the Europeana portal, which gathers CH information from institutions in Europe. Since the object-centric approach is widely used and its constructs are already available, it is employed in EDM. Memory institutions have stored their data mostly in an object-centric way, and its conversion to the event-centric type needs a great deal of effort. However, the event-centric method can store more detailed information for a CH object. For example, if an object has more than one contributor in its creation or has gone changes during time with multiple acquisition events in different times and locations, the object-centric method will definitely not be able to store all the information, but in an event-centric approach different events of various type can be defined with specific spatio-temporal properties to store the abovementioned information without losing data [67]. This leads to the second difference between the models which is storage of changes and provenance data over time. In EDM, this kind of information is stored in a textual format, which is difficult for machines to understand and reason over. In CRM, detailed information relating to creation, evolution, transition of objects, and other changes to it are stored through a chain of events, which is fully machine understandable, and it is possible to query and infer from them. We can conclude that the object-centric approach stores only one state of the world, since it connects the object to one creator, one location, and one time period [66]. One aspect in ontologies is the domain and range of the classes and properties that we discussed before in section 2.3. For every single class and property in CIDOC CRM, its range and domain is defined, but in EDM some of the classes do not have a specified domain and range. This shows that CRM has higher ontological commitment than EDM. An unspecified domain and range would result in inconsistencies and therefore makes it difficult for inference engines to automatically deduce the types of instances used as domains and ranges of such properties in EDM [67]. Finally, there is a difference between the two models in providing different views and representations for a single object. An object can have various representations, for example, images, post cards, and 3D models. Also, it may have different views, which means different institution provide various types of descriptions for the same object. EDM, due to its purpose, handles and supports different representations and views for an object. It provides a construct called aggregation, which connects different digital representations to its object, which allows it to represent it in various forms. It also provides different views for an object with the construct called proxy [61]. Different descriptions for the same entity are gathered to provide multiple views for a single object. CRM does not have any special class dedicated to provide different representations and views. However, with some of its general properties, this result can be somewhat achieved [67].

Table 3. Comparison of CIDOC CRM and EDM ontologies

	Modeling structure	**Ontological commitment**	**Changes and provenance data**	**Multiple views and representations**
CRM	Event-centric	High	Chain of events – machine understandable	Difficult to achieve
EDM	Object-centric	Low	Textual – difficult to infer	Fully supports

3.4.2. Ontology-based metadata interoperability

Although metadata schemas have certain shortcomings discussed before in detail, they are widely used in museums and other memory institutions and cannot be disposed that easily. Metadata interoperability approaches based on metadata are not suitable and have some downsides (fully discussed in section 2.2). However, ontologies provide an efficient approach for metadata interoperability, in which no metadata element is omitted and it keeps the original richness of data. On the other hand, ontologies act as a mediating medium and convert data between different metadata formats by defining mapping paths from metadata schemas to a core ontology and back to them. For example, in [67] this functionality is discussed and a mapping is developed for converting data in DC format to the CIDOC CRM ontology. Another advantage of ontology-based metadata interoperability is making implicit information in metadata become explicit. This is achieved if the ontology uses an event-centric structure since the events can bring more details and also enables them to be reasoned over by machines [23], [68].

4. Information retrieval

Until now we have discussed various aspects of CH data and the road taken from preliminary knowledge organization methods to the latest developments to structure this data for better retrieval. Now that the information is mapped to the structures created, it is time to provide services based on this structured data to search, browse, and retrieve them. These types of services could be an important achievement to replace the currently used text-based search engines. Also, there is the opportunity to develop smart and intelligent applications, since the information is in a machine-understandable form. Data is in interoperable formats which brings about the ability to develop inter-institutional systems to create a shared understanding of the issues and goals. In this section, we discuss the efforts put in this way.

4.1. Publishing structured data for its use and reuse

4.1.1. Portals

The first thing to do after structuring data is to create possibilities for its use and reuse. One of the Semantic Web promises is to prevent a digital dark age and the loss of data that is generated with a lot of efforts and costs. Semantic portals are great tools for aggregating the heterogeneous data from various publishers and institutions. They can act as a single publishing channel for local and small institutions [70]. As stated in [71], there are three types of portals: service portals accommodating a set of services e.g. Yahoo!, community portals serving as virtual meeting venues, and finally the kind of portal that we focus in CH case, information portals acting as hubs of data. When the content of such portals are Semantic Web content, they are called semantic information portals. These portals are based on Semantic Web technologies, and they can be useful for CH information both for the users and data publishers [71]. End-users can enjoy a global view of the data gathered from multiple sources in a seamless homogenous repository thus reducing time and effort needed for finding them. Users can also take advantage of semantic searching, browsing, recommendation, and other intelligent services and applications developed in the context of the portal. On the other hand, semantic portals can be beneficial for content hosts. Creating portals for distributed data provided by various memory organizations in a central manner is costly and not feasible but Semantic Web technologies are promising tools for collecting and integrating distributed heterogeneous data from various sources (semi-) automatically into a global portal. This kind of portal can be a shared, cost-effective publication channel for participating organizations with the common goal of promoting cultural knowledge among society and experts. As Semantic Web technologies like metadata and ontologies link the related information with each other, they in fact enrich the content of every organization involved for free [71]. MuseumFinland [72] and its successor CultureSampo [73] are well-known examples of semantic information portals in the CH domain.

4.1.2. APIs

Despite all the positive aspects, portals cannot guaranty the reuse of data. Portals are appropriate tools for the usage of data through various applications to provide different services but their data is static and the possibility for reuse of the data is very low. However, APIs are suitable for this matter. An API (Application Programming Interface) can lower the technical barriers and required effort and time for reusing data and services provided for developing another applications and services [74]. The Europeana project is one of the outstanding and large-scale examples of and API based CH data aggregator. The aim of this project as discussed before is to harvest, aggregate, and integrate heterogeneous CH data from different data providers across the Europe with the help of Semantic Web technologies and standards. Europeana provides an API enabling third parties and other communities to reuse the rich data collected for their own needs [62], [75]. Various applications and intelligent services from portals to location-based applications can be developed by consuming the data provided by APIs. The Europeana portal is developed based on its own API created within the project [75]. A noteworthy point here is that some metadata schemas are created for accessing and searching information from APIs. These metadata are called harvesting and searching metadata [9]. They were not developed for structuring data but rather querying the APIs and harvesting information. LIDO (Light Weight Information Describing Objects)[1] is an XML schema developed collaboratively by CDWA Lite, museumdat, SPECTRUM and CIDOC CRM communities intended for delivering metadata for use in a wide range of online services. It covers a variety of descriptive information about museum objects. It is mostly based on CIDOC CRM and borrowed its event-centric concept [76]. There are also numerous protocols developed for federated search. Z39.50[2] is amongst the first protocols developed by the Library of congress for searching and information retrieval from a database. It is a client-server protocol which is a NISO/ANSI standard but it dates before the web era and HTTP protocol. Z39.50 has been updated into the SRU[3] protocol (Search/Retrieval via URL), which uses the HTTP protocol and REST. SRU has a twin protocol SRW[4] (Search/Retrieve Web Service) that is based on Web Service SOAP messages. Queries in SRU and SRW are expressed using the simple Contextual Query Language (CQL), which is a standard based on Z39.50. The result set is returned as an XML document. A widely used system targeted for only harvesting metadata is the OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting)[5] .The OAI-PMH protocol is based on HTTP where request arguments are issued as GET or POST parameters of a URL. Data providers are repositories that expose structured metadata via service providers. Then they make OAI-PMH service requests to harvest that metadata. OAI-PMH responses are encoded in XML syntax and it supports harvesting records in any metadata format encoded in XML.

4.1.3. Linked (Open) Data

As we mentioned in the early stages of the paper, recently there has been a global tendency to move from the web of documents to the web of data, in which data is machine-readable and structured and information retrieval can be improved dramatically. In 2006, Tim Berners Lee introduced the concept of Linked Data and its principles [77]. Linked Data is concerned about data on the web and providing connections and links between them as web of data. We can follow the links between the pages in web of document, humans and machines follow links between data to find other related data [78]. Linked Data employs to main technologies RDF and HTTP to connect structured data on the web to each other and to real world entities such as persons, places, books, films, music, and companies, which are given unique identifiers URI. The web of data can be accessed through Linked Data browsers which navigate users between connected data by the RDF links provided. Also, its search engines can provide complex queries that were just possible in relational databases [79]. This can turn the web to a single global database, which is sometimes referred to as the *global data space*. In [77], Tim Berners Lee outlines four basic rules for publishing data on the web to become a part of the Linked Data:

1. Use URIs as names for things
2. Use HTTP URIs so that people can look up those names.
3. When someone looks up a URI, provide useful information, using the standards (RDF, SPARQL)
4. Include links to other URIs, so that they can discover more things.

The web of data started with an initial project, Linking Open Data in 2007 supported by the W3C Semantic Web Education and Outreach Working Group (SWEO). The intention behind the project was to identify the data sets that were available under open licenses and re-publish them on the web in RDF format with links and connections between them. Through this time, the cloud of Linked Open Data grew bigger and bigger. The central parts of the cloud are DBpedia and Geonames which act as linking hubs [79]. Most of the things we refer are within

---

1 http://network.icom.museum/cidoc/working-groups/lido/
2 http://www.loc.gov/z3950/
3 http://www.loc.gov/standards/sru/
4 http://www.loc.gov/standards/sru/companionSpecs/srw.html
5 https://www.openarchives.org/pmh/

these two. DBpedia[1] extracts Wikipedia information in RDF and includes URIs for a wide range of entities that can be referred to, while Geonames provide URIs for names of places and the spatial relationships between places in RDF format. With the growth of the Linked Open Data cloud, there was a need to create a means to validate this information. In an update of his notes in 2010, Tim Berners Lee stated a five-star system for evaluating data put on web.

- 1 Star: Data is available on the web (whatever format), but with an open license.
- 2 Stars: Data is available as machine-readable structured data (e.g., excel instead of a scanned image of a table).
- 3 Stars: Data is available as (2) but in a non-proprietary format (e.g., CSV instead of excel).
- 4 Stars: All the above, plus use open standards from the W3C (RDF and SPARQL) to identify things, so that people can link to it.
- 5 Stars: Data is available according to all the above, plus outgoing links to other people's data to provide context.

LOD publishers provide different methods to access the data published. Linked Data browsers such as Tabulator[2], Disco[3], OpenLink[4], Ontology-browser[5], and Zitgist[6], make data on the web browse able based on URI erencing. Another way is to provide SPARQL endpoints for querying data in a standard approach to be used in mash-up applications. SPARQL endpoints enable machine and human users to make SPARQL queries to an RDF repository conveniently using HTTP. The data is also available to download in RDF dumps, which sometimes can be used for offline purposes. Lastly, there are human user interfaces that can search RDF data on the web. Examples of such application interfaces are as Falcons[7], Sindice[8], Swoogle[9] and Watson[10].

Linked Open Data can be great opportunity for the CH community, because its data is distributed in various formats. Adoption of LOD would have definite effects to improve reusability and interoperability of CH information [68], [80]. With the integration of the data with other data on the web, it can increase the richness of CH data and possibility of generating new knowledge. One of the early adopters of LOD is the Library of Congress publishing its authority files and thesauri. Later other organizations in library domain joined LOD, such as the German National Library and the British National Library publishing entities that can be referred on the CH data network. Recently, the Getty institute published its thesauri (AAT, ULAN, CONA, IA, and TGN) as LOD under the Open Data Commons Attribution License (ODC-By) 1.0., which can be used in many applications of the CH domain. Europeana started a pilot project in 2011 to move its data to LOD [81]. They provided a part of the data aggregated in the portal in EDM format. It is available in three ways of URI dereferencing, SPARQL endpoint, and bulk download [82]. There are two strategies for memory organizations when it comes to publish LOD. The first is to invest in infrastructure and publish your data as LOD, which small institutions cannot afford and it is not feasible for them. This approaches requires the organization to choose or develop a domain ontology to map their data and extract it as RDF. The second is to provide their data in a special structure and format, depending on the host, to large-scale aggregators like Europeana so it gets published as LOD [68], [80]. In this approach, the institute does not have to map its data to a specific ontology, but it has to provide the data in a special format and structure designated by the aggregator. In the STELLAR project [83], the goal was to develop an automatic tool for mapping the archaeological data to CIDOC CRM. The archaeological extension they developed is called CRM-EH and they extracted them in RDF/XML with the intention of publishing it as Linked Data. Since automatic ingestion and mapping by a large-scale aggregator may cause damages to the original richness of data, authors in [80] proposed a methodology for small institutions to map their data to EDM on their own to keep the richness of data and link their metadata to Linked Data. The tool developed in this paper is called Amalgame which is a part of the ClioPatria semantic web toolkit. To evaluate the approach, they converted the Amsterdam

[1] https://wiki.dbpedia.org/
[2] https://www.w3.org/2005/ajar/tab
[3] https://www.w3.org/2001/sw/wiki/Disco
[4] http://ode.openlinksw.com/
[5] https://www.w3.org/2001/sw/wiki/OntologyBrowser
[6] https://www.w3.org/2001/sw/wiki/Zitgist
[7] https://www.w3.org/2001/sw/wiki/Falcons
[8] https://www.w3.org/2001/sw/wiki/Sindice
[9] http://swoogle.umbc.edu/
[10] https://www.ibm.com/watson/services/discovery/

Museum metadata to Linked Data, which made the museum to be the first small scale memory organization to join the Linked Data cloud. By following the idea of this project, the Amsterdam Museum data of the Smithsonian American Art Museum (SAAM) was published as linked data with slight differences [84]. In this project EDM ontology was chosen and it was tailored to the SAAM data. Also, a tool named Karma was developed to do the mapping automatically and it also had the capability of visualizing the links it made so that the authorities could check the accuracy of linking to LOD sources.

4.2. Intelligent applications and services

4.2.1. Search and browse

Searching and browsing are the basic services that an institution portal can offer. After all, data structuring and information management approaches are for better and more accurate retrieval of data. When the data is structured with Semantic Web technologies like ontologies and schemas ordered in hierarchies with the help of classes and properties, some services can be provided for the users to improve searching. For example, semantic auto completion, which completes the words that a user is typing in the search box by the annotations of data that it has. This can help the user when they cannot fully remember the name of the thing they are searching for. Also, after retrieving the results, the system can order and group them based on their semantic categories, which further guides users to their interests. Moreover, semantic recommendations can be provided for users while they view a piece of information through the links and properties that connect it to other things. The CultureSampo portal [73] offers the abovementioned services for the end-user. The concept of facet browsing interface is one of the most popular and widely used approaches for browsing large collections of data. This concept was used in early MuseumFinland and the CultureSampo portal and in the Europeana portal. In a facet browsing system, there a number of facets and each of them highlights one aspect and the dimension of the underlying data. The user can select the desired values in the facets and in this way narrow down the collection data to reach to the interested information. Traditional facet browsers assume a fixed set of facets to select and navigate through relatively homogeneous data. However, data in the CH domain is heterogeneous, and this causes problems in employing facet. In [85], the authors developed /facet, which is a browser for Semantic Web repositories that covers the problems mentioned. It has the capability to dynamically generate facets based on the type of resources chosen by the user in the GUI developed and also incorporates a cross-type selection. /facet does not require manual prior software configuration in contrast to traditional facet browsers. It was tested on a diverse dataset gathered from three institution with multiple varying thesauri in contrast to other projects, such as MuseumFinland, which mapped the entire data to a single schema. /facet was developed as a part of the Netherlands MultimediaN e-culture project.

Recently there’s a growing interest towards inspirational information retrieval in CH. It is argued that a substantial number of users and researchers visit these data repositories to stimulate their creativity, so there should be mechanism that balance retrieval between expected results and surprising answers yet relevant to users queries [86]. The PATHS project [87] is amongst few examples that have studies possibilities of serendipitous search result in CH domain. However, this paradigm in searching and browsing collections needs more attention.

4.2.2. Inference

As discussed previously, implementing ontologies in description logics like OWL DL can increase the expressiveness of it, so that helps computers to reason over them easily. There are number of inference engines that have been developed to reason upon OWL DL such as RACER [88], FaCT++ [89], pellet [90], and HermiT [91]. They can perform the following basic logical deductions [47]:

- *Concept satisfiability* which is to check whether a newly defined concept is consistent with the knowledge base as well as satisfiability of the knowledge base as a whole.
- *Subsumption* that is to compute the proper place for a newly defined concept in the concept hierarchy.
- *Proper instantiation* that is to check whether a given individual belongs to the class it is designated to.
- *Realization* which is to compute the class a given individual belongs to and retrieval of the instances of a given class.

For the first time, the authors in [92] developed an OWL version of CIDOC CRM (Before Erlangen CRM [47]) to do some reasoning on it. They developed a knowledge discovery interface based on RACER inference engine that carried out some simple reasoning. However, OWL suffers from some limitations that lower its capability of reasoning. Although OWL provides a variety of constructors for classes, it has a limited set of constructors for properties. The concept-based modelling of OWL prevents it from performing inferences based on the properties. For example, OWL lacks composition constructors for properties that makes it unable to capture the relationship between concepts associated with a combination of properties. The typical example here is the “uncle rule” [93]. To infer the uncle relationship, there is a need to reason over the composition of parent and brother properties, which is not possible with OWL alone. The consensus way to cope with this problem is to extend OWL with “rules languages” to increase its expressivity. Rules define specific conditions and operations to infer and extract new knowledge from a knowledge base. SWRL (Semantic Web Rule Language) [95] is a rule language developed based on OWL DL and OWL Lite sublanguages of OWL and Unary/Binary Datalog RuleML sublanguages of the Rule Markup Language. SWRL is a crucial step towards a standardized and interoperable inference framework in the Semantic Web infrastructure. In [94], for their purpose of extracting new knowledge from data with a combination of facts distributed over different sources, They used information from three art related databases and modeled them with CIDOC CRM ontology in OWL language using Protégé[1] software. Since, CRM is event-centric and OWL is insufficiently expressive for property based ontologies, a set of rules were defined in SWRL. Using Jess inference engine they showed the applicability of the method to derive new knowledge that was not contained in a single database.

4.2.3. Data representation and visualization

Creative smart representations and 3D visualization of thematic data can give users both experts and non-experts, a holistic overview in an interactive and easy-to-grasp manner. For example, in the CultureSampo portal [96], cultural information is represented on a map based on their locations, and they are also categorized in different types. Each of them is visualized with a specific color that gives the user the opportunity to browse a city or a place on map and find out about various cultural heritage resources there. In this portal, there is the possibility to overlay historic boundaries of cities and historic maps on the google map. There are also other intuitive ways of representing and visualizing data there, such as displaying visual items of a chosen type on a time line to realize the changes through time and visualizing the social network of famous historic figures and persons, in which user can search if there is a connecting path between two persons and how they are related. 3D visualization can be beneficial both for understanding and analyzing thematic data. In [97], after successfully enriching BIM model of a historic building with cultural heritage documentation by integrating IFC model with Semantic Web technologies, the authors visualized the 3D model of the building in unity game engine and the linked information to the different parts of the building. In [98], archaeologists and computer science researchers collaborated towards a connection between 3D spatial representation and archaeological knowledge, by integrating observable (material) and non-graphic (interpretive) data. After acquisition of 3D model of a built heritage with photogrammetry techniques, they integrated the model with geometric, topological, and temporal semantics to model Units of Stratigraphication. They visualized the integrated information on the 3D model with different colors that provided a convenient way for the user to capture information of different parts. In [99], the ADE developed based on CityGML for cultural heritage architecture (CHADE) was used for 3D data of a historic church. Different parts of the building in 3D model is linked to its own class and enriched with its relating information. The resulting model is rendered in a 3D GIS environment that provided different geometric measurements and visualization of thematic information. Expert users with such systems can do basic measurements and information retrieval like material of a part of the built heritage without the need to on-site experiments which are destructive. Also, they can visualize the thematic data like year of construction with colors to see the status of different parts.

However, visualization is quiet young in CH domain. InfoVis techniques can be very helpful for presenting CH data. The amount of information stored in collections is enormous and in some cases like British museum it takes years for users to visit all items in the institution. InfoVis methods can provide various holistic and generous

[1] https://protege.stanford.edu/

viewpoints on data and very importantly serendipitous information retrieval [100]. Nonetheless, due to the vast amount of data and its heterogeneity, this fairly new promising field of research is still very much challenging.

4.2.4. AR/VR applications

CH employed advanced multimedia technologies from an early time, as these systems could attract more audience especially younger people. ARCHEOGUIDE, an AR location-based user guide application, in 2001 was the start in CH [101]. Visual multimedia systems are easier for people to follow therefore, its usage can help CH to reach a wider dissemination of cultural knowledge. These methods can make learning process in museum more interactive than just reading labels and descriptions. Also they have the potential to increase user engagement by enhancing the sense of place so that they could bring an added value to CH objects [102]. However, it is argued that the use of such technologies remove the focus from CH objects and user pay more attention to the virtual graphics [103]. For this problem, combination of multimedia technologies with semantics seems to be a kind of solution. As in this way the information used can be more organized and machine-readable so that users can select what they want to learn about the CH site or an special item, thus increasing user interaction compared to the situation where the developer of AR/VR apps choose what to see for users.

Structured and rich Linked Data can provide the possibility to create context-aware AR/VR applications that can improve retrieving personalized data, which could further contribute to increasing the CH experiences for the user. In [104], a mobile AR application was designed and implemented for user guide. The application uses LOD published through the MultimeadiaN e-culture project and is based on the location context of the user capture via GPS sensor harvest cultural heritage data for user's nearby POIs. It is based on the user view extent and heading acquired by mobile sensors data for the specific POI that the user is facing and is displayed. Also, to display the retrieved data facet-based approach [85], which was developed within the same project is utilized. In [105], a mobile AR web-based application was developed called LOD4AR. It harvests and integrates LOD from three separate sources DBpedia, LinkedGeoData, and Data.Gov.ro which is Romanian museum data. Kim et al. [106] first mapped and integrated data from five Korean heritage databases to a data model that was a previously developed data model for Korean cultural heritage data KCHDM [59]. Then they developed a mobile AR application for three POIs in a palace. The location detection is based on vision-based methods that match the camera image and the POIs image databases. The data is displayed based on the 5 super classes of the KCHDM, and the user can select the category they are interested in and browse various types of information and multimedia content.

4.2.5. Context-aware applications

To this end, we have discussed this matter until now and we can see that the evolution of the process in the CH domain is extraordinary and lots of effort has been put in to the work, but still if users want to access specific information, they first should find a repository, and then they have to learn how to search within that system.

Context-aware ubiquitous computing (ubicomp) can be a solution to this problem. Mark Weiser in 1991 proposed the term ubiquitous computing and defined its features in some papers, most importantly in [107]. He defined ubiquitous computing as invisible, non-disturbing, and calm that "weaves itself into the fabric of everyday life" [107]. This vision of Mark Weiser means that cyberspace should be brought to the real world, so people could interact with almost every object in the environment for computing rather than a single access point to the cyberspace, which is the monitors of personal computers. In this way, users would pay less attention to the computing technologies and thus focus on their actions in the real world. Also, ubicomp is not passive like PCs, yet it is active and sometimes proactive. This means that while the user is acting with the object in their focal point of attention ubicomp technologies act around it without the user defining their needs. Ubicomp technologies provide the services that they are designed for. In this way, they help the user in their everyday life without intruding their attention. However, a prerequisite step for ubicomp to reach this level is for it to be context-aware, which means that it should capture and understand the context of the environment that the user is in. Context and its definition was first introduced by Bill Schilit [108]. Context-aware computing tries to make assumptions about the current situation of the user. Dey also defined context as "any information that can be used to characterize the situation of an entity" [109]. So context can be a variety of information from the location of the user, time, and to even size of the interface of the user's device. While context-aware applications can provide many services including location-aware user guides and recommender systems, they can increase the precision of personalized information retrieval. In the EEXCESS project [110], an application was developed with the purpose of providing ubiquitous access to the cultural heritage information for users. After acquiring the user's context, which is the main topics of the text of the web page that the user is reading, this application carries out a federated search on content providers using LIDO schema and aggregates the collected data based on the EDM data model complemented by W3C PROV ontology as EDM lacks provenance metadata. Then it ranks the list of data gathered based on the user's information need and shows previews of recommended items at the bottom of the page for the user. The integrated and structured data through Semantic Web technologies can increase efficiency of recommender systems. This is assessed and verified in the SMARTMUSEUM project [111]. The SMARTMUSEUM application is a mobile context-aware recommender system for users interested in cultural heritage, which utilized the web of data. This application is designed for three outdoor, indoor, and web-based scenarios. In the outdoor phase, the user moves around the city and based on their context of location acquired by GPS sensor, the visit time of the sites, and the interested type of the information that is manually inputted by the user, cultural heritage sites are recommended to the user. In the indoor phase, the user enters a recommended place and information about various objects are provided for the user on-site. The location context of the user is acquired by RFID sensors and based on their interested information context, personalized content is retrieved for the user. In the desktop scenario, all the context is edited by the user on the web interface. Also, users can rate the content recommended that is used by the system to refine its next recommendation in a personalized way. Authors reflect that ontology-based data structuring is effective in better matching user context and retrieved information thus increasing recommendation accuracy.

5. Critical discussion

We presented the process that CH community has taken to stablish data interoperability through Semantic Web technologies from creating domain specific vocabularies and metadata schemas to top-level and application specific ontologies, also, the various opportunities that this transfer of data and information into knowledge can provide. However, there are still challenges need to be tackled in future studies, which we will mention here.

5.1. Sustainability

First, sustainability is an important issue. CH is one of the four primary pillars of sustainable development [112], therefore sustainability within the CH and its data is vital. The data silos and repositories mainly are project-led in CH and the presumed data persistence is at stake [113]. The challenge arising here is data stewardship and the responsibility to maintain the data [114]. This problem first came up in the course of Archaeology Data Service (ADS) development. This archive holds the archeological data based in the UK and it has been funded and maintained from 1996 up to now in collaboration with researchers, heritage agencies, and funders [113], pointing out the significance of data responsibility and conservation. Other regional or large-scale portals must ensure researchers and stakeholders of its data conservation and preservation. Another issue in this regard is that small institutes and countries that lack technical and financial ability to provide their data for the integration in aggregator portals due to high standards and protocols. In this regard, during the continental-scale ARIADNE infrastructure project [115], for example, weaker GLAMs were given help to prepare their data according to the standards [113] to address the hurdles in the way of this collaborative innovation of creating a digital heritage data discovery infrastructure. In addition, these small institutes have another problem of low publicity as they attract low number of visitors and their information is left unexplored [116]. Semantic Web solutions can integrate data of these institutes to the Linked Open Data cloud for more visibility [117] and thus reaching sustainability in whole CH ecosystem.

5.2. Data reuse and dissemination

Infrastructures and portals are not the ultimate goal for digital data [118]. These digital archives are created for linking and integrating data to make its retrieval and reuse more convenient. Despite the unanimous agreement on this process of collecting, documenting, modeling, and packaging the data, the actual reuse of it is the missing part

[118]. It is not clearly shown how and to what extent these efforts have been effective in terms of reusability [119]. Digital repositories and researchers need to save their projects' lessons learnt on the best practices of the methods applied and their impact on the dissemination and reusability of data. Interoperability might have been the challenge of the last decade and it has been solved largely with various kinds of data models and metadata developed. However, todays' challenge could be reusing of the semantically linked data [113].

5.3. Spatio-temporal aspect of CH data

There is a quiet famous assertion among geoscientists, which claims that almost up to 80 percent of all data in the world has some spatial or geographic reference. This has been proved to some extend of reliability in [120] by evaluating the Linked Open Data of Semantic Web (LOD is going to be discussed later in detail). CH data is no exception and a large proportion of cultural resources has some sort of connection to space. Therefore, they can be retrieved by search terms that refer to locations [124]. As in many knowledge management systems ontologies and data models, heritage objects are linked to their coordinates and temporal periods, indicating that location and time are important factors in cultural events. In addition, the TGN (Thesaurus of Geographic Names) vocabulary, which was introduced, is a well-known thesaurus in the CH domain that is a structured list of place names and their previous historical names. It is used to link CH data to their location, which helps in semantically annotating and inferencing information. Geospatial science deals with phenomenon that relates to space, and it has an information system for analyzing and visualizing data called GIS (Geographic/Geospatial Information System) [121]. With the emergence of web and web services, GIS systems also went on the web for numerous reasons, intending to provide services over the web, which introduced another paradigm called WebGIS. As a matter of fact, geospatial science also have adapted semantic web technologies [122], [123] since spatial information is in different formats, such as vector and raster and also spatial features have various feature types, such as point, line, polygon, and etc. and also to model the different topological relations.

GISicence and geospatial semantic web can benefit CH in two general aspects of 3D semantics of heritage building's architecture and spatio-temporal reasoning. In recent years lots of research was done in the GIS field for the 3D recording of cultural heritage, the important and difficult part of the work is to create structures to handle this data and integrate them while building semantic models and heritage documentation and standards. This could be helpful in making geometric measurements, management and monitoring health of heritage building, preservation and protection planning alongside sustainable smart city visions. Work and effort in 3D and architectural aspect of spatial science for the benefit of CH community is more than the other aspect. For example, [128] Recommends a method to digitally record cultural heritage buildings, enrich them with topological relations and semantics, and transfer it to a 3D GIS environment for further analysis and management. In this approach, authors use their previously developed HBIM (Historic Building Information Modeling) [129], which is a model for capturing and modeling historic building structures from 3D models generated using BIM. After completing the 3D model with its parts that were semantically defined, it was transferred to a 3D GIS environment and CityGML was chosen for this purpose. CityGML [130] is an OGC (Open Geospatial Consortium) standard for storage and exchange of 3D models in an interoperable way that allows the same data to be reused in different applications. The purpose for its development is to provide a standard and common definition of basic entities, attributes, and relations of a 3D city model. After moving to the CityGML environment, different segments of the model is recognized such as rooftops, windows, and roads. Then the model is ready for further analysis regarding geometry, topology, and semantics. Thematic views and analyses are also possible, but for this matter an ADE (application domain extension) is needed to be developed for CityGML [128]. In [131], an ADE is proposed for CityGML by extending it through XML that is capable of modeling thematic information of parts of architectural heritage buildings in multilevel views from LoD (Level of Detail) 1 to LoD 5 increasing details of parts and related information. In [131], another ADE is developed for CityGML called CHADE (Cultural Heritage Application Domain Extension). In this spatial ontology, some classes are proposed for CityGML and incorporated Getty's AAT (Art and Architecture Thesaurus) vocabulary, so that the model is capable of providing geometric measurements also thematic information representation in different levels of detail. There are a lot of applications and data models developed for spatial reasoning and manipulation of 3D models for management of cultural heritage resources. An overview of them can be found in [133].

There has been a great tendency to develop location-based and beyond that, location-aware applications for numerous purposes like recommendation systems for users and tourist guide systems [125]. These applications can assist users to query surroundings spatially and find desired locations and POIs. However, there are some problems with CH information, mentioned in [125], that hinder achieving the applications discussed. First, there is a problem in annotations content that have georeferenced locations with varying granularity. For example, a heritage object may refer to a country name while another refers to a city name in that country, and the missing semantic relationship between them would cause an error in accuracy of information retrieval when processing a spatial query. Second, place names and extents of places have change during time and based on cataloging time of objects their location name may differ, which would again cause certain problems. Finally, nearby POIs are not just the one with small distances. Their accessibility and the time needed to reach the place should be analyzed. The Geospatio-temporal Semantic Web can play an important role in solving the problems mentioned above. Although the majority of CH resources are georeferenced, the concept of place is poorly defined in these data. Efforts for integrating spatio-temporal reasoning into CIDOC CRM began in 2013, which tried to harmonize the two OGC and CRM standards [126]. The efforts resulted in CRMgeo extension, which integrates geoinformation and CRM ontology through conceptualizations, formal definitions, encoding standards, and topological relationships defined by OGC's GeoSPARQL. Unlike other CRM extensions, CRMgeo brought changes to the core classes of the data model by introducing Spacetime Volume. It also defined some new subclasses and properties, such as phenomenal and declarative space, and integrated geospatial featuretypes and relationships from GeoSPARQL into CIDOC CRM ontology [46]. Although CRMgeo provides links to spatial standard GeoSPARQL, it suffers in temporal aspect, as it lacks links to any time ontology [127]. With this being said, this extension has not been used widely and overall the aspect of spatio-temporal semantics is left unexplored, which could be very helpful in spatial reasoning. This can make it possible to infer new knowledge and links that were not known before. It can reveal for example, which type of art started in which place, influence of different kinds of art from one place to another, or types of artifacts in a special place and so forth.

5.4. Tourist engagement and social intelligence

There is a need to engage the tourists, whether native or foreigner, more with the tangible and intangible heritage for a better dissemination and education of cultural heritage. While a great number of studies agree on the capacity of CH to create attachment, entertainment, and social bonding, these aspects have remained less explored [134]. Storytelling is one of the unique features of museums and galleries [135]. CH professionals began to provide digital storytelling tools to enable engagement and interactivity between users and heritage objects [136]. Various technological possibilities have been applied in digital storytelling and narrative authoring tools such as multimedia presentation, VR/AR interfaces [137], and indoor navigation [112] to connect several object with a specific narrative. However, there are a number of suggestions that can be helpful in making digital storytelling easier and more effective. Usage of ontologies can be of great assistance in devising the plot of stories as it relates objects to each other and historical event based on their relationships and also modelling the sequence of the conceptual map of the story [138]. Ontologies can reveal unseen connections between objects and events in real world, which can bring up interesting stories. Museum experts can benefit from ontologies based systems that collect information from variety of sources related to objects to build narratives [137].Another aspect is that heritage sites are mostly visited by groups of people rather than individually, however most applications and services are developed for individuals [134]. Social and shared digital experience is a necessity in CH domain, which has not gained much attention. The CHESS [139] and Emotive [140] project are amongst few that have focused on storytelling for groups of people, increasing collective participation and engagement.Context-awareness can also be beneficial in digital storytelling by tailoring the stories to the preferences of different users. Users of CH vary from professional expert, scientists, students, and regular people and they have different ages. It can also relate the story to the day of visit or the trending topics in social media [141]. CH can benefit from social intelligence to enable the community to develop brand-new strategies for engaging and attracting more visitors. Nature of both the social media and CH data are big, heterogeneous, highly unstructured, and involves wide range of collaborators and stakeholders [142]. Therefore, a semantic approach seems very profitable in linking the social media to CH and raising awareness, interest, and engagement in a wider scale.

6. Conclusion

In this paper, the evolution of information engineering techniques and Knowledge Organization Systems (KOSs) for a more precise and personalized information retrieval in Cultural Heritage domain was discussed in detail. Knowledge management is crucial in CH due to obvious reasons such as dealing with rich heterogeneous data and involving different organizations and people from various fields of expertise. GLAMs (Galleries, Libraries, Archives, and Museums) are great and rich sources of CH information. This information is of vital importance in memory preservation, education of new generations, tourism, and other possible areas. A better knowledge management would lead to more achievement in the aforementioned aims of CH both in the local and global scale.

Now that CH has mature data models, more and more memory organizations should adopt them and contribute to Linked Open Data (LOD) where machine-understandable data is linked together, which creates a wide set of opportunities to use and re-use the data. Furthermore, intelligent and personalized applications and services can be developed for people utilizing the web of data to achieve a better user interaction, engagement, and heritage information dissemination. The CH community has invested a lot of effort and time to develop data models and means for knowledge organization, now it is time to take advantage of this great source to provide smart applications, which are still in the initial stages. This could further help the CH industry.

References

[1] Signore, O. (2009). Representing knowledge in archaeology: from cataloguing cards to semantic web. Archeologia e calcolatori, 20, 111-128.

[2] Ghosh, P. (2015). Google's Vint Cerf warns of 'digital Dark Age'. BBC News, 13, 2015.

[3] Benjamins, V. R., Contreras, J., Blázquez, M., Dodero, J. M., Garcia, A., Navas, E., Hernandez, F., & Wert, C. (2004, May). Cultural heritage and the semantic web. In European Semantic Web Symposium (pp. 433-444). Springer, Berlin, Heidelberg.

[4] Bruseker, G., Carboni, N., & Guillem, A. (2017). Cultural Heritage Data Management: The Role of Formal Ontology and CIDOC CRM. In Heritage and Archaeology in the DigitalAge (pp. 93-131). Springer, Cham.

[5] Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The semantic web. Scientific american, 284(5), 34-43.

[6] Shadbolt, N., Berners-Lee, T., & Hall, W. (2006). The semantic web revisited. IEEE intelligent systems, 21(3), 96-101.

[7] Noor, S., Shah, L., Adil, M., Gohar, N., Saman, G. E., Jamil, S., & Qayum, F. (2018). Modeling and representation of built cultural heritage data using semantic web technologies and building information model. Computational and Mathematical Organization Theory, 1-24.

[8] Liu, D., Bikakis, A., & Vlachidis, A. (2017, September). Evaluation of semantic web ontologies for modelling art collections. In Advances in Databases and Information Systems (pp. 343-352). Springer, Cham.

[9] Hyvönen, E. (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.

[10] Vavliakis, K. N., Karagiannis, G. T., & Mitkas, P. A. (2012). Semantic web in cultural heritage after 2020. SW2022, Boston, Nov, 11.

[11] Beghtol, C. (1998). Knowledge domains: multidisciplinarity and bibliographic classification systems. KO KNOWLEDGE ORGANIZATION, 25(1-2), 1-12.

[12] Hodge, G. (2000). Systems of Knowledge Organization for Digital Libraries: Beyond Traditional Authority Files. Digital Library Federation, Council on Library and Information Resources, 1755 Massachusetts Ave., NW, Suite 500, Washington, DC 20036.

[13] Harpring, P., Lanzi, E., Whiteside, A., & McRae, L. (2006). Cataloging cultural objects: A guide to describing cultural works and their images. American Library Association.

[14] Vállez, M., Pedraza-Jiménez, R., Codina, L., Blanco, S., & Rovira, C. (2015). Updating controlled vocabularies by analysing query logs. Online Information Review, 39(7), 870-884.

[15] National Information Standards Organization (US). (2005). Guidelines for the construction, format, and management of monolingual controlled vocabularies. NISO Press.

[16] Alvarenga, L., & de Paiva Oliveira, A. (2004). "Thesaurus" and "Ontology:" A Study of the Definitions Found in the Computer and Information Science Literature, by Means of an Analytical Synthetic Method. KO KNOWLEDGE ORGANIZATION, 31(4), 231-244.

[17] de Carvalho Moura, A. M., Campos, M. L. M., & Barreto, C. M. (1998). A survey on metadata for describing and retrieving Internet resources. World Wide Web, 1(4), 221-240.

[18] NISO, N. (2004). Understanding metadata. National Information Standards.

[19] Shah, U., Finin, T., Joshi, A., Cost, R. S., & Matfield, J. (2002, November). Information retrieval on the semantic web. In Proceedings of the eleventh international conference on Information and knowledge management (pp. 461-468). ACM.

[20] van Ballegooie, M., & Duff, W. (2006). DCC Digital Curation Manual: Instalment on Archival Metadata. HATII, University of Glasgow; University of Edinburgh; UKOLN, University of Bath; Council for the Central Laboratory of the Research Councils.

[21] Isaac, A., Schlobach, S., Matthezing, H., & Zinn, C. (2008). Integrated access to cultural heritage resources through representation and alignment of controlled vocabularies. Library Review, 57(3), 187-199.

[22] Haslhofer, B., & Klas, W. (2010). A survey of techniques for achieving metadata interoperability. ACM Computing Surveys (CSUR), 42(2), 7.
[23] Ruotsalo, T., & Hyvönen, E. (2007). An event-based approach for semantic metadata interoperability. In The Semantic Web (pp. 409-422). Springer, Berlin, Heidelberg.
[24] Baca, M. (2003). Practical issues in applying metadata schemas and controlled vocabularies to cultural heritage information. Cataloging & classification quarterly, 36(3-4), 47-55.
[25] Schreiber, G., Amin, A., Aroyo, L., van Assem, M., de Boer, V., Hardman, L., Hildebrand, M., Omelayenko, B., van Osenbruggen, J., Tordai, A. and Wielemaker, J. (2008). Semantic annotation and search of cultural-heritage collections: The MultimediaN E-Culture demonstrator. Web Semantics: Science, Services and Agents on the World Wide Web, 6(4), 243-249.
[26] Lim, S., & Li Liew, C. (2011, September). Metadata quality and interoperability of GLAM digital images. In Aslib Proceedings (Vol. 63, No. 5, pp. 484-498). Emerald Group Publishing Limited.
[27] Krovetz, R. (1997, July). Homonymy and polysemy in information retrieval. In Proceedings of the eighth conference on European chapter of the Association for Computational Linguistics (pp. 72-79). Association for Computational Linguistics.
[28] Studer, R., Benjamins, V. R., & Fensel, D. (1998). Knowledge engineering: principles and methods. Data and knowledge engineering, 25(1), 161-198.
[29] Uschold, M., & Gruninger, M. (1996). Ontologies: Principles, methods and applications. The knowledge engineering review, 11(2), 93-136.
[30] Zúñiga, G. L. (2001, October). Ontology: its transformation from philosophy to information systems. In Proceedings of the international conference on Formal Ontology in Information Systems-Volume 2001 (pp. 187-197). ACM.
[31] Gruber, T. R. (1993). A translation approach to portable ontology specifications. Knowledge acquisition, 5(2), 199-220.
[32] Davis, R., Shrobe, H., & Szolovits, P. (1993). What is a knowledge representation?. AI magazine, 14(1), 17.
[33] Le Boeuf, P., Doerr, M., Ore, C. E., Stead, S. (May 2018). Definition of the CIDOC Conceptual Reference Model. Technical Report version 6.2.3
[34] [34] Guarino, N., & Welty, C. (2000, October). A formal ontology of properties. In International Conference on Knowledge Engineering and Knowledge Management (pp. 97-112). Springer, Berlin, Heidelberg.
[35] Antoniou, G., & Van Harmelen, F. (2004). Web ontology language: Owl. In Handbook on ontologies (pp. 67-92). Springer, Berlin, Heidelberg.
[36] Doerr, M. (2009). Ontologies for cultural heritage. In Handbook on Ontologies (pp. 463-486). Springer, Berlin, Heidelberg.
[37] Doerr, M. (2003). The CIDOC conceptual reference module: an ontological approach to semantic interoperability of metadata. AI magazine, 24(3), 75-75.
[38] Doerr, M., & Theodoridou, M. (2011). CRMdig: A Generic Digital Provenance Model for Scientific Observation. TaPP, 11, 20-21.
[39] Hartig, O. (2009). Provenance Information in the Web of Data. LDOW, 538.
[40] Doerr, M., Kritsotaki, A., Rousakis, Y., Hiebel, G., & Theodoridou, M. (2014). CRMsci: the Scientific Observation Model.
[41] Stead, S., & Doerr, M. (2015). CRMinf: The argumentation model. An extension of CIDOC-CRM to support argumentation.
[42] Doerr, M., Felicetti, A., Hermon, S., Hiebel, G., Kritsotaki, A., Masur, A., May, K., Ronzino, P., Schmidle, W., Theodoridou, M. and Tsiafaki, D. (2016). Definition of the CRMarchaeo, an extension of CIDOC CRM to support the archaeological excavation process.
[43] Ronzino, P., Niccolucci, F., Felicetti, A., & Doerr, M. (2016). CRM ba a CRM extension for the documentation of standing buildings. International Journal on Digital Libraries, 17(1), 71-78.
[44] Ronzino, P. (2017). Harmonizing the CRMba and CRMarchaeo models. International Journal on Digital Libraries, 18(4), 253-261.
[45] Doerr, M., Hiebel, G., & Eide, Ø. (2013). CRMgeo: Linking the CIDOC CRM to GeoSPARQL through a spatiotemporal refinement. Institute of Computer Science, Tech. Rep. GR70013.
[46] Hiebel, G., Doerr, M., & Eide, Ø. (2017). CRMgeo: A spatiotemporal extension of CIDOC-CRM. International Journal on Digital Libraries, 18(4), 271-279.
[47] Goerz, G., Oischinger, M., & Schiemann, B. (2008). An implementation of the CIDOC conceptual reference model (4.2. 4) in OWL-DL. In Proceedings of the 2008 Annual Conference of CIDOC-The Digital Curation of Cultural Heritage.
[48] Lagoze, C., & Hunter, J. (2001, October). The ABC ontology and model. In International Conference on Dublin Core and Metadata Applications (pp. 160-176).
[49] Doerr, M., Hunter, J., & Lagoze, C. (2003). Towards a core ontology for information integration. Journal of Digital information, 4(1).
[50] Van Ruymbeke, M., Hallot, P., Nys, G. A., & Billen, R. (2018). Implementation of multiple interpretation data model concepts in CIDOC CRM and compatible models. Virtual Archaeology Review, 9(19), 50-65.
[51] Bekiari, C., Doerr, M., Le Bœuf, P., & Riva, P. (2017). FRBR object-oriented definition and mapping from FRBRER, FRAD and FRSAD (version 3). International working group on FRBR and CIDOC CRM harmonisation.
[52] Le Boeuf, P., & Pelegrin, F. X. (2014). FRBR and serials: the PRESSoo model.
[53] Meghini, C., Scopigno, R., Richards, J., Wright, H., Geser, G., Cuy, S., Fihn, J., Fanini, B., Hollander, H., Niccolucci, F. and Felicetti, A. (2017). ARIADNE: a research infrastructure for archaeology. Journal on Computing and Cultural Heritage (JOCCH), 10(3), 18.
[54] Oldman, D., & Tanase, D. (2018, October). Reshaping the Knowledge Graph by Connecting Researchers, Data and Practices in ResearchSpace. In International Semantic Web Conference (pp. 325-340). Springer, Cham.
[55] Scholz, M., & Goerz, G. (2012, August). WissKI: A Virtual Research Environment for Cultural Heritage. In ECAI (Vol. 242, pp. 1017-1018).
[56] Myers, D., Dalgity, A., Avramides, I., & Wuthrich, D. (2012, October). Arches: an open source GIS for the inventory and management of immovable cultural heritage. In Euro-Mediterranean Conference (pp. 817-824). Springer, Berlin, Heidelberg.
[57] Cripps, P., Greenhalgh, A., Fellows, D., May, K., & Robinson, D. (2004). Ontological Modelling of the work of the Centre for Archaeology. CIDOC CRM technical paper.

[58] Binding, C., May, K., & Tudhope, D. (2008, September). Semantic interoperability in archaeological datasets: Data mapping and extraction via the CIDOC CRM. In International Conference on Theory and Practice of Digital Libraries (pp. 280-290). Springer, Berlin, Heidelberg.
[59] Kim, S., Ahn, J., Suh, J., Kim, H., & Kim, J. (2015, September). Towards a semantic data infrastructure for heterogeneous Cultural Heritage data-challenges of Korean Cultural Heritage Data Model (KCHDM). In 2015 Digital Heritage (Vol. 2, pp. 275-282). IEEE.
[60] Hyvönen, E., Viljanen, K., Tuominen, J., & Seppälä, K. (2008, June). Building a national semantic web ontology and ontology service infrastructure–the FinnONTO approach. In European Semantic Web Conference (pp. 95-109). Springer, Berlin, Heidelberg.
[61] Mäkelä, E., Hyvönen, E., & Ruotsalo, T. (2012). How to deal with massively heterogeneous cultural heritage data–lessons learned in CultureSampo. Semantic Web, 3(1), 85-109.
[62] Doerr, M., Gradmann, S., Hennicke, S., Isaac, A., Meghini, C., & Van de Sompel, H. (2010, August). The europeana data model (edm). In World Library and Information Congress: 76th IFLA general conference and assembly (pp. 10-15).
[63] Cacciotti, R., Valach, J., Kuneš, P., Čerňanský, M., Blaško, M., & Křemen, P. (2013). Monument damage information system (MONDIS): An ontological approach to cultural heritage documentation. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 5, W1.
[64] Hellmund, T., Hertweck, P., Hilbring, D., Mossgraber, J., Alexandrakis, G., Pouli, P., ... & Padeletti, G. (2018). Introducing the HERACLES Ontology—Semantics for Cultural Heritage Management. Heritage, 1(2), 377-391.
[65] Liu, D., Bikakis, A., & Vlachidis, A. (2017, September). Evaluation of semantic web ontologies for modelling art collections. In European Conference on Advances in Databases and Information Systems (pp. 343-352). Springer, Cham.
[66] Isaac, A. (2013). Europeana data model primer.
[67] Dijkshoorn, C., Aroyo, L., Van Ossenbruggen, J., & Schreiber, G. (2018). Modeling cultural heritage data for online publication. Applied Ontology, (Preprint), 1-17.
[68] Kakali, C., Lourdi, I., Stasinopoulou, T., Bountouri, L., Papatheodorou, C., Doerr, M., & Gergatsoulis, M. (2007, August). Integrating Dublin Core metadata for cultural heritage collections using ontologies. In International conference on Dublin core and metadata applications (pp. 128-139).
[69] Theodoridou, M., Daskalaki, M., Doerr, M., & Bruseker, G. (2016). Methodological tips for mappings to CIDOC CRM. In *44th Computer Applications and Quantitative Methods in Archaeology Conference (CAA 2016)" Exploring Oceans of Data* (pp. 04-29).
[70] Sidoroff, T., & Hyvönen, E. (2005, November). Semantic e-goverment portals-a case study. In Proceedings of the ISWC-2005 Workshop Semantic Web Case Studies and Best Practices for eBusiness SWCASE05 (Vol. 7).
[71] Hyvönen, E. (2009). Semantic portals for cultural heritage. In Handbook on ontologies (pp. 757-778). Springer, Berlin, Heidelberg.
[72] Hyvönen, E., Mäkelä, E., Salminen, M., Valo, A., Viljanen, K., Saarela, S., Junnila, M. and Kettula, S. (2005). MuseumFinland—Finnish museums on the semantic web. Web Semantics: Science, Services and Agents on the World Wide Web, 3(2-3), 224-241.
[73] Hyvönen, E., Mäkelä, E., Kauppinen, T., Alm, O., Kurki, J., Ruotsalo, T., Seppälä, K., Takala, J., Puputti, K., Kuittinen, H. and Viljanen, K. (2009, May). CultureSampo: A national publication system of cultural heritage on the semantic Web 2.0. In European Semantic Web Conference (pp. 851-856). Springer, Berlin, Heidelberg.
[74] Foulonneau, M., Martin, S., & Turki, S. (2014, February). How open data are turned into services?. In International Conference on Exploring services science (pp. 31-39). Springer, Cham.
[75] Concordia, C., Gradmann, S., & Siebinga, S. (2010). Not just another portal, not just another digital library: A portrait of Europeana as an application program interface. IFLA journal, 36(1), 61-69.
[76] Coburn, E., Light, R., McKenna, G., Stein, R., & Vitzthum, A. (2010). LIDO-lightweight information describing objects version 1.0. ICOM International Committee of Museums.
[77] Berners-Lee, T. (2006). Linked Data - Design Issues. Retrieved March 11, http://www.w3.org/DesignIssues/LinkedData.html
[78] Bizer, C., Heath, T., & Berners-Lee, T. (2011). Linked data: The story so far. In Semantic services, interoperability and web applications: emerging concepts (pp. 205-227). IGI Global.
[79] Bizer, C., Heath, T., Idehen, K., & Berners-Lee, T. (2008, April). Linked data on the web (LDOW2008). In Proceedings of the 17th international conference on World Wide Web (pp. 1265-1266). ACM.
[80] De Boer, V., Wielemaker, J., Van Gent, J., Hildebrand, M., Isaac, A., Van Ossenbruggen, J., & Schreiber, G. (2012, May). Supporting linked data production for cultural heritage institutes: the amsterdam museum case study. In Extended Semantic Web Conference (pp. 733-747). Springer, Berlin, Heidelberg.
[81] Haslhofer, B., & Isaac, A. (2011). data. europeana. eu-The Europeana Linked Open Data Pilot.
[82] Isaac, A., & Haslhofer, B. (2013). Europeana linked open data–data. europeana. eu. Semantic Web, 4(3), 291-297.
[83] May, K., Binding, C., Tudhope, D., & Jeffrey, S. (2011). Semantic Technologies Enhancing Links and Linked Data for Archaeological Resources. Proceedings Computer Applications and Quantitative Methods in Archaeology (CAA2011).
[84] Szekely, P., Knoblock, C. A., Yang, F., Zhu, X., Fink, E. E., Allen, R., & Goodlander, G. (2013, May). Connecting the smithsonian american art museum to the linked data cloud. In Extended Semantic Web Conference (pp. 593-607). Springer, Berlin, Heidelberg.
[85] Hildebrand, M., Van Ossenbruggen, J., & Hardman, L. (2006, November). /facet: A browser for heterogeneous semantic web repositories. In International Semantic Web Conference (pp. 272-285). Springer, Berlin, Heidelberg.
[86] Petras, V., Hill, T., Stiller, J., & Gäde, M. (2017). Europeana–a Search Engine for Digitised Cultural Heritage Material. *Datenbank-Spektrum*, *17*(1), 41-46.
[87] Fernie, K., Griffiths, J., Stevenson, M., Clough, P., Goodale, P., Hall, M., Archer, P., Chandrinos, K., Agirre, E., de Lacalle, O.L., & de Polo, A. (2012, September). PATHS: Personalising access to cultural heritage spaces. In *2012 18th International Conference on Virtual Systems and Multimedia* (pp. 469-474). IEEE.
[88] Haarslev, V., & Möller, R. (2003, October). Racer: A Core Inference Engine for the Semantic Web. In EON (Vol. 87).
[89] Tsarkov, D., & Horrocks, I. (2006, August). FaCT++ description logic reasoner: System description. In International joint conference on automated reasoning (pp. 292-297). Springer, Berlin, Heidelberg.
[90] Sirin, E., Parsia, B., Grau, B. C., Kalyanpur, A., & Katz, Y. (2007). Pellet: A practical owl-dl reasoner. Web Semantics: science, services and agents on the World Wide Web, 5(2), 51-53.

[91] Shearer, R., Motik, B., & Horrocks, I. (2008, October). HermiT: A Highly-Efficient OWL Reasoner. In OWLED (Vol. 432, p. 91).
[92] Koutsomitropoulos, D. A., & Papatheodorou, T. S. (2007). Expressive Reasoning about Cultural Heritage Knowledge using Web Ontologies. In WEBIST (2) (pp. 276-281).
[93] Horrocks, I. (2005). OWL rules, ok?. Rule Languages for Interoperability, 34.
[94] Lin, C. H., Hong, J. S., & Doerr, M. (2008). Issues in an inference platform for generating deductive knowledge: a case study in cultural heritage digital libraries using the CIDOC CRM. International Journal on Digital Libraries, 8(2), 115-132.
[95] Horrocks, I., Patel-Schneider, P. F., Boley, H., Tabet, S., Grosof, B., & Dean, M. (2004). SWRL: A semantic web rule language combining OWL and RuleML. W3C Member submission, 21(79).
[96] Hyvönen, E., Mäkelä, E., Kauppinen, T., Alm, O., Kurki, J., Ruotsalo, T., Seppälä, K., Takala, J., Puputti, K., Kuittinen, H. and Viljanen, K. (2009, April). CultureSampo–Finnish culture on the semantic web 2.0. Thematic perspectives for the end-user. In Proceedings, Museums and the Web (pp. 15-18).
[97] Pauwels, P., Bod, R., Di Mascio, D., & De Meyer, R. (2013, October). Integrating building information modelling and semantic web technologies for the management of built heritage information. In 2013 Digital Heritage International Congress (DigitalHeritage) (Vol. 1, pp. 481-488). IEEE.
[98] Drap, P., Papini, O., Pruno, E., Nucciotti, M., & Vannini, G. (2017). Ontology-based photogrammetry survey for medieval archaeology: Toward a 3d geographic information system (gis). Geosciences, 7(4), 93.
[99] Noardo, F. (2018). Architectural heritage semantic 3D documentation in multi-scale standard maps. Journal of Cultural Heritage, 32, 156-165.
[100] Windhager, F., Federico, P., Schreder, G., Glinka, K., Dörk, M., Miksch, S., & Mayr, E. (2018). Visualization of cultural heritage collection data: State of the art and future challenges. *IEEE transactions on visualization and computer graphics*, *25*(6), 2311-2330.
[101] Vlahakis, V., Ioannidis, M., Karigiannis, J., Tsotros, M., Gounaris, M., Stricker, D., ... & Almeida, L. (2002). Archeoguide: an augmented reality guide for archaeological sites. *IEEE Computer Graphics and Applications*, *22*(5), 52-60.
[102] Bekele, M. K., Pierdicca, R., Frontoni, E., Malinverni, E. S., & Gain, J. (2018). A survey of augmented, virtual, and mixed reality for cultural heritage. *Journal on Computing and Cultural Heritage (JOCCH)*, *11*(2), 1-36.
[103] Cameron, F., & Kenderdine, S. (2007). *Theorizing digital cultural heritage: A critical discourse*.
[104] Van Aart, C., Wielinga, B., & Van Hage, W. R. (2010, October). Mobile cultural heritage guide: location-aware semantic search. In International Conference on Knowledge Engineering and Knowledge Management (pp. 257-271). Springer, Berlin, Heidelberg.
[105] Vert, S., Dragulescu, B., & Vasiu, R. (2014, October). LOD4AR: Exploring Linked Open Data with a Mobile Augmented Reality Web Application. In International Semantic Web Conference (Posters & Demos) (pp. 185-188).
[106] Kim, H., Matuszka, T., Kim, J. I., Kim, J., & Woo, W. (2017). Ontology-based mobile augmented reality in cultural heritage sites: information modeling and user study. Multimedia Tools and Applications, 76(24), 26001-26029.
[107] Weiser, M. (1991). The Computer for the 21 st Century. Scientific american, 265(3), 94-105.
[108] Schilit, B., Adams, N., & Want, R. (1994, December). Context-aware computing applications. In wmcsa (pp. 85-90). IEEE.
[109] Dey, A. K. (2001). Understanding and using context. Personal and ubiquitous computing, 5(1), 4-7.
[110] Seifert, C., Bailer, W., Orgel, T., Gantner, L., Kern, R., Ziak, H., Petit, A., Schlötterer, J., Zwicklbauer, S. and Granitzer, M. (2017). Ubiquitous access to digital cultural heritage. Journal on Computing and Cultural Heritage (JOCCH), 10(1), 4.
[111] Ruotsalo, T., Haav, K., Stoyanov, A., Roche, S., Fani, E., Deliai, R., Mäkelä, E., Kauppinen, T. and Hyvönen, E. (2013). SMARTMUSEUM: A mobile recommender system for the Web of Data. Web semantics: Science, services and agents on the world wide web, 20, 50-67.
[112] Kontiza, K., Antoniou, A., Daif, A., Reboreda-Morillo, S., Bassani, M., González-Soutelo, S., Lykourentzou, I., Jones, C.E., Padfield, J., & López-Nores, M. (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.
[113] Wright, H., & Richards, J. D. (2018). Reflections on collaborative archaeology and large-scale online research infrastructures. Journal of Field Archaeology, 43(sup1), S60-S67.
[114] Law, M., & Morgan, C. L. (2014). The archaeology of digital abandonment: online sustainability and archaeological sites. Present Pasts, 1-9.
[115] Meghini, C., Scopigno, R., Richards, J., Wright, H., Geser, G., Cuy, S., Fihn, J., Fanini, B., Hollander, H., Niccolucci, F., & Felicetti, A. (2017). ARIADNE: a research infrastructure for archaeology. Journal on Computing and Cultural Heritage (JOCCH), 10(3), 1-27.
[116] Lykourentzou, I., Naudet, Y., & Vandenabeele, L. (2016, October). Reflecting on European History with the Help of Technology: The CrossCult Project. In GCH (pp. 67-70).
[117] De Boer, V., Wielemaker, J., Van Gent, J., Hildebrand, M., Isaac, A., Van Ossenbruggen, J., & Schreiber, G. (2012, May). Supporting linked data production for cultural heritage institutes: the amsterdam museum case study. In Extended Semantic Web Conference (pp. 733-747). Springer, Berlin, Heidelberg.
[118] Huggett, J. (2018). Reuse remix recycle: repurposing archaeological digital data. Advances in Archaeological Practice, 6(2), 93-104.
[119] Webster, P. (2018). Book review: cultural heritage infrastructures in digital humanities edited by Agiatis Benardou, Erik Champion, Costis Dallas and Lorna M. Hughes. LSE Review of Books.
[120] Hahmann, S., Burghardt, D., & Weber, B. (2011). "80% of All Information is Geospatially Referenced"??? Towards a Research Framework: Using the Semantic Web for (In) Validating this Famous Geo Assertion. In Proceedings of the 14th AGILE Conference on Geographic Information Science.
[121] Maguire, D. J. (1991). An overview and definition of GIS. Geographical information systems: Principles and applications, 1, 9-20.
[122] Egenhofer, M. J. (2002, November). Toward the semantic geospatial web. In Proceedings of the 10th ACM international symposium on Advances in geographic information systems (pp. 1-4). ACM.
[123] Kuhn, W. (2005). Geospatial semantics: why, of what, and how?. In Journal on data semantics III (pp. 1-24). Springer, Berlin, Heidelberg.

[124] Jones, C. B., Alani, H., & Tudhope, D. (2001, September). Geographical information retrieval with ontologies of place. In International Conference on Spatial Information Theory (pp. 322-335). Springer, Berlin, Heidelberg.
[125] Kauppinen, T., Paakkarinen, P., Mäkela, E., Kuittinen, H., Väätäinen, J., & Hyvönen, E. (2011). Geospatio-temporal semantic web for cultural heritage. In Digital Culture and E-Tourism: Technologies, Applications and Management Approaches (pp. 48-64). IGI Global.
[126] Doerr, M., Hiebel, G., & Eide, Ø. (2013). CRMgeo: Linking the CIDOC CRM to GeoSPARQL through a spatiotemporal refinement. Institute of Computer Science, Tech. Rep. GR70013.
[127] Nys, G. A., Van Ruymbeke, M., & Billen, R. (2018, August). Spatio-temporal reasoning in CIDOC CRM: an hybrid ontology with GeoSPARQL and OWL-Time. In CEUR Workshop Proceedings. RWTH Aachen University.
[128] Dore, C., & Murphy, M. (2012, September). Integration of Historic Building Information Modeling (HBIM) and 3D GIS for recording and managing cultural heritage sites. In 2012 18th International Conference on Virtual Systems and Multimedia (pp. 369-376). IEEE.
[129] Murphy, M., McGovern, E., & Pavia, S. (2009). Historic building information modelling (HBIM). Structural Survey, 27(4), 311-327.
[130] Kolbe, T. H., Gröger, G., & Plümer, L. (2005). CityGML: Interoperable access to 3D city models. In Geo-information for disaster management (pp. 883-899). Springer, Berlin, Heidelberg.
[131] Costamagna, E., & Spanò, A. (2012, October). Semantic models for architectural heritage documentation. In Euro-Mediterranean Conference (pp. 241-250). Springer, Berlin, Heidelberg.
[132] Noardo, F. (2016, April). A spatial ontology for architectural heritage information. In International Conference on Geographical Information Systems Theory, Applications and Management (pp. 143-163). Springer, Cham.
[133] Jancsó, A. L., Jonlet, B., Hoffsummer, P., Delye, E., & Billen, R. (2017, September). An Analytical Framework for Classifying Software Tools and Systems Dealing with Cultural Heritage Spatio-Temporal Information. In International Conference on Spatial Information Theory (pp. 325-337). Springer, Cham.
[134] Perry, S. E., Roussou, M., Mirashrafi, S., Katifori, A., & McKinney, S. (2019). Shared digital experiences supporting collaborative meaning-making at heritage sites.
[135] Bedford, L. (2001). Storytelling: The real work of museums. Curator: the museum journal, 44(1), 27-34.
[136] Katifori, A., Karvounis, M., Kourtis, V., Perry, S., Roussou, M., & Ioanidis, Y. (2018, December). Applying Interactive Storytelling in Cultural Heritage: Opportunities, Challenges and Lessons Learned. In International Conference on Interactive Digital Storytelling (pp. 603-612). Springer, Cham.
[137] Valtolina, S. (2016). A storytelling-driven framework for cultural heritage dissemination. Data Science and Engineering, 1(2), 114-123.
[138] Winer, D. (2014). Review of ontology based storytelling devices. In Language, Culture, Computation. Computing of the Humanities, Law, and Narratives (pp. 394-405). Springer, Berlin, Heidelberg.
[139] Katifori, A., Karvounis, M., Kourtis, V., Kyriakidi, M., Roussou, M., Tsangaris, M., Vayanou, M., Ioannidis, Y., Balet, O., Prados, T., & Keil, J. (2014, November). CHESS: personalized storytelling experiences in museums. In International Conference on Interactive Digital Storytelling (pp. 232-235). Springer, Cham.
[140] Katifori, A., Roussou, M., Perry, S., Drettakis, G., Vizcay, S., & Philip, J. (2018, November). The EMOTIVE Project-Emotive Virtual Cultural Experiences through Personalized Storytelling. In CIRA@ EuroMed (pp. 11-20).
[141] Bampatzia, S., Bravo-Quezada, O. G., Antoniou, A., Nores, M. L., Wallace, M., Lepouras, G., & Vassilakis, C. (2016, September). The use of semantics in the CrossCult H2020 project. In Semanitic Keyword-based Search on Structured Data Sources (pp. 190-195). Springer, Cham.
[142] Chianese, A., Marulli, F., & Piccialli, F. (2016, February). Cultural heritage and social pulse: a semantic approach for CH sensitivity discovery in social media data. In 2016 IEEE Tenth International Conference on Semantic Computing (ICSC) (pp. 459-464). IEEE.