

# Representing and Linking Dunhuang Cultural Heritage Information Resources Using Knowledge Graph

Xiaoguang Wang\*, Wanli Chang\*\* and Xu Tan\*\*\*

Wuhan University, School of Information Management, Wuhan, Hubei, 430072, CN,

\*<wxguang@whu.edu.cn>, \*\*<wanlichang@whu.edu.cn>, \*\*\*<txishuo@whu.edu.cn>



Xiaoguang Wang is the vice dean and a professor in the School of Information Management at Wuhan University, China. His research interests include digital asset management, knowledge organization, semantic publishing, and digital humanities.

Wanli Chang is a postgraduate student in the School of Information Management at Wuhan University, China. His research interests focus on data management of cultural heritage and knowledge discovery.

Xu Tan is a doctoral student in the School of Information Management at Wuhan University, China. Her research interests include digital humanities, semantic web technology, digital libraries, and cultural informatics.



Wang, Xiaoguang, Wanli Chang and Xu Tan. 2020. "Representing and Linking Dunhuang Cultural Heritage Information Resources Using Knowledge Graph." *Knowledge Organization* 47(7): 604-615. 19 references. DOI:10.5771/0943-7444-2020-7-604.

**Abstract:** This study employs a knowledge graph approach to realize the representation and association of information resources, promote the research, teaching, and dissemination of Dunhuang cultural heritage (CH). The Dunhuang Mogao Grottoes is a UNESCO world CH site, and digitization of Dunhuang CH has produced a large amount of information resources. However, these digitized resources continue to lack the systematic granular semantic representation required to correlate Dunhuang cultural heritage information (CHI) in order to facilitate efficient research and appreciation. To respond to this need, new approaches for representing CHI are being developed. This study identifies five facets and their semantic relationship to Dunhuang CH, constructs an ontology model to regulate the entities, attributes, and relationships of Dunhuang CH knowledge, and subsequently processes the resulting data using various techniques (such as semantic annotation and entity association) to facilitate rendering the data in a knowledge graph construction. Finally, we constructed a DH-oriented knowledge graph service platform in order to provide a user friendly visual display and semantic retrieval service.

Received: 1 June 2020; Revised: 30 September 2020, 16 October 2020, 23 October 2020; Accepted: 23 October 2020

Keywords: knowledge, Dunhuang, data grottoes semantic

## 1.0 Introduction

Cultural heritage (CH) resources are numerous and widely distributed, containing rich and diverse cultural knowledge. Effective protection and utilization of CH resources is of great significance to the sustainable development of human civilization. Due to the growth of digital humanities (DH), the acquisition, organization, management, and dissemination of CH resources using digital technology has aroused great interest from libraries, archives, and museums (LAMs) in recent years. In the field of Dunhuang CH, the

development of digital technology has produced fruitful results over the last forty years as evidenced by the proliferation of digitized Dunhuang cultural heritage information (CHI) resources.

Structuring and storing information resources in databases is the basic method currently used to facilitate the management and organization of CHI. Through the implementation of knowledge organization system (KOS) principles and best practices, fine-grained and semantic description of CHI can be used to facilitate knowledge management and mining. KOS is a universal term for the various

norms and methods used to organize knowledge. It is an important means of acquiring and utilizing knowledge and is a generic term for semantic tools used to express and effectively organize the elaboration of various knowledge structures. KOS is divided into four sub-levels based on the complexity of the structure and main functions, namely: term lists, metadata-like models, classifications and categorizations, and relationship models (Zeng 2010). In the field of CH, the semantic representation and linking of CHI can be achieved by using different levels of KOS.

A series of thematic databases have already been established in the field of Dunhuang CH. Such as, the International Dunhuang Project (IDP), Mellon International Dunhuang Archive (MIDA), E-Dunhuang, Dunhuang Academic Resources Database, and so on. These projects have organized a plethora of DH resources and improved public access to them. However, many impediments to efficient scholarship and appreciation of Dunhuang CH still persist today. From a KOS perspective, these shortcomings include:

1. Metadata and taxonomy are used by these discrete databases to provide a basic description of their information resources. However, this basic flat description is no longer adequate in today's multi-source, heterogeneous, multimodal, big data environment, where data can also be semantically described and linked through more advanced knowledge representation and organization models (such as ontologies and knowledge graphs) for the purpose of deep revelation of resource content. In addition, problems also persist in databases management and interoperability posed by redundancy in the construction of resource repositories employing different elements and properties in these flat resource descriptions.
2. Although there are a lot of images and 3D resources residing within in these databases, the cave structure of Dunhuang CH might not be accurately described in the absence of related data models.
3. In the context of DH, data service providers need not only provide digital information resources but must also integrate the information needs of users to provide associated resources and context. However, most existing Dunhuang related databases lack this necessary connection to external resources and other databases are document-centric rather than entity-centric. This approach has limited the value and usability of these discrete information silos.

Therefore, this study establishes the Dunhuang Grottoes Knowledge Graph and user-oriented application service platform covering multimodal resources with functions such as semantic retrieval and knowledge navigation to provide better support services for the protection, research,

teaching, and dissemination of Dunhuang CH. Such a model would be capable of accurately and comprehensively revealing intrinsic features and external information regarding Dunhuang CHI and finally meet the needs of knowledge mining and analysis in order to revolutionize the efficiency of knowledge utilization.

## 2.0 Related work

Semantic representation and organization of CHI is the basis for open sharing and knowledge mining. At present, communities of LAMs have developed many general and special KOSs (knowledge organization system). For example, the metadata models include DCMI (Dublin Core Metadata Initiative), VRA Core (the core categories for visual resources), CDWA (Categories for the Description of Works of Art), RDA (Resource Description and Access). The thesaurus includes AAT (Art and Architecture Thesaurus), ULAN (Union List of Artist Names). The ontologies include *CIDOC CRM* (*CIDOC Conceptual Reference Model*), ABC, FOAF (Friend-of-a-Friend), and so on.

In accordance with the growing trend towards big data in CH, massive multi-source heterogeneous data must be organized and linked, and a knowledge graph approach provides the possibility for this. A new knowledge representation technology of KOS- knowledge graph has ontology technology at its core. As a result, knowledge graph provides a path and method for extracting structured knowledge from multimodal digital resources and representing and storing this structured knowledge in a graphical data model (Singhal 2012). With the help of knowledge graph, advanced forms of knowledge services such as knowledge visualization, semantic retrieval, and intelligent question answering can be better realized. The core technologies of knowledge graph mainly include entity relationship identification and extraction, knowledge fusion, entity linking, and knowledge reasoning (Moura et al. 2014). Moreover, knowledge graphs are widely used in intelligent search at well-known internet companies, such as Google (based on Wikidata and freebase database), Microsoft (Satori Knowledge Base), and Baidu. With the development of knowledge graph technology, knowledge graphs have also been successfully applied in CH projects, such as European (Purday 2009), Getty Digital Museum (Alexiev et al. 2015), Venice Time Machine Project (Abbott 2017), and Finland's CultureSampo project (Mäkelä et al. 2012).

This study contends that knowledge graphs are also particularly useful in linking and organizing the rich and inherently complex CHI of three-dimensional structures. Historical buildings are valuable and a non-renewable CH resource. In order to preserve, restore, and recreate historic buildings, an increasing number of research projects are employing digital technologies to describe, manage, and ana-

lyze architectural heritage information. 3D modeling is an important technology in CH conservation, which primarily uses computers to describe spatial relationships between objects in a mathematical way. Allen et al. (2003) performed a comprehensive 3D modeling of St. Pierre's Cathedral using a 3D laser scanner, and Foni et al. (2007) of the University of Geneva use the Cathedral of Saint Sophie as an example in order to perform a virtual reconstruction and simulation of a complex and endangered historical building.

In the field of architectural engineering and construction (AEC), building information modelling (BIM) is widely studied and applied with regard to the development of digital cities and smart cities. BIM is the integration of the physical and functional information of the whole life cycle of the building, which contains rich and detailed structural information. Due to developments in interdisciplinary research, some DH scholars have begun to use architectural information models to represent and organize the structural information of historical buildings in the field of CH. For example, Simeone integrates building information systems and ontology-based systems to enhance knowledge representation and management in the process of building heritage (Simeone et al. 2019). Similarly, Noor et al. (2018) believes that the traditional architectural heritage metadata model standard lacks a description of the key aspects of 3D modeling and proposes a model that uses semantic web technology and building information models to represent building data and historical data. Pauwels et al. (2013) also achieves the integration of architectural elements and building-related CHI, including historical documents, design drawings, etc., by combining building information models and semantic web technologies to demonstrate the feasibility of this approach using the Ghent Book Tower in Belgium as an example.

### 3.0 The semantic contents of Dunhuang CH

The triadic world refers to the spatial form made up of the physical world, human society, and information space, the connectivity and integration of which is an important feature and trend of the new generation of computing technologies (Ma et al. 2019). In the context of DH, with the continuous development of digital technology, a large number of digital resources have been generated on the basis of CH. Traditional humanities research is gradually shifting to DH research, and this evolution in research paradigm has gradually transformed the structure of knowledge space in the field of CH from the original binary space (i.e., physical space and cultural space) to a ternary space. According to the triadic world theory, Dunhuang CH can be divided into physical space, cultural space, and information space: 1) physical space refers to the physical space constituted by the

internal architecture of the Dunhuang grottoes and their external rock walls, which is an external manifestation of Dunhuang CH; 2) cultural space refers to the mythological world and human social world depicted by the Dunhuang grottoes. (the mythological world generally includes Buddhist figures narratives and time, as well as other religious figures, non-religious mythological figures, and so on. The human social world covers elements, such as administrative institutions, religious organizations, general social groups, ancient ancestors, and folklore); and, 3) the information space is a virtual space that realizes the digital twin of Dunhuang CH physical space and cultural space through the flow of information using the Dunhuang CHI element as its carrier.

In consultation with Dunhuang Academy research scholars, this paper divides the physical space, cultural space, and information space of the Dunhuang grottoes into five facets. These facets include: cave structure knowledge, academic works knowledge, digital information resource knowledge, mural content knowledge, and, finally, the facet of historical and cultural knowledge. Each of these facets is discussed in further detail in the following section.

#### 3.1 The facet of cave structure knowledge

A religious product, grottoes provided a mechanism for transmitting Buddhist ideas to subsequent generations. Therefore, they contain rich historical and cultural information. The different forms of grotto architecture express not only the development of Buddhism but also the shift from one cultural form to another. Therefore, they are an integral part of the process of organizing CHI.

The Dunhuang grottoes are a type of grotto temple building constructed on the cliffs of the Gobi Desert in northwest China, where the design of the interior space is more important due to the influence of the external environment. The interior of Dunhuang grottoes is a closed structural space. There is a commonality between different grottoes. No matter how the internal structure of the main room changes, the grotto always maintains the same basic spatial pattern composed of three parts: the front room, the main room, and the hallway. Moreover, in addition to the east wall with doors or windows, there are many beautiful murals painted on the north, south, and west walls as well as on top of the caves. For example, Cave No. 285 displays a classic grotto temple structure (Figure 1). The architectural form of the main room of the grotto may be subdivided into four categories. These categories include: the zen grotto, the central pillar grotto, the overlying grotto, the back-screen grotto, as well as a small number of big Buddha grottoes, nirvana grotto, and altar grotto (Ning 1986). The variations in the forms of each of the Dunhuang grottoes

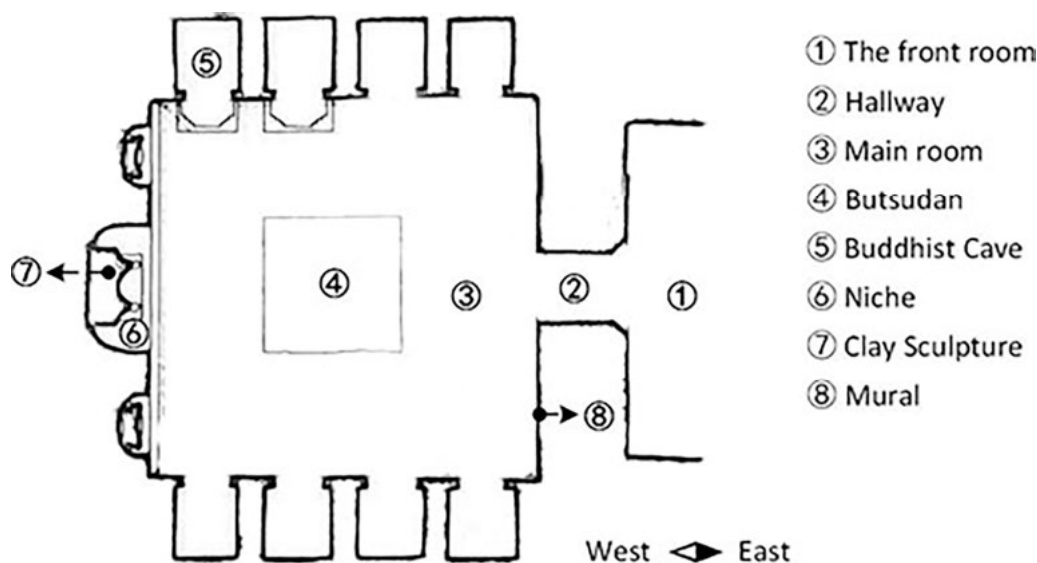


Figure 1. Floor Plan of Cave No.285.

indicates the architectural structures of different grotto forms were determined by their historical and cultural functions.

### 3.2 The facet of academic works knowledge

In the process of studying the history and culture of Dunhuang, *Dunhuangology* was formed. *Dunhuangology* is practiced by many academic fields, including: religion, art, history, archaeology, language, literature, anthropology, geography, philosophy, science, technology, architecture, ancient book collation, cultural relics protection, sports, folk custom, and so on. Over the past century, *Dunhuangology* has been continuously developed and deepened. The study of Dunhuang grottoes has been very fruitful, as a result, a large amount of academic research resources, including research papers, books, archaeological reports, and picture albums have been created. These academic works and documents have provided important value for the excavation of CHI in the Dunhuang grottoes. Fan (2000) divides the stages of research into the Dunhuang grottoes into three periods: beginning, development, and full development. From the beginning of data collection, picture interpretation, content verification, to topic discussion, comprehensive research, etc., a great deal of research and analysis has been published. These studies have laid a strong foundation for further in-depth research in the future.

### 3.3 The facet of digital information resource knowledge

Digital information resources are an important manifestation of CHI. They can reshape the physical space of the

grotto through digital photography, three-dimensional modeling and other digital technologies, allowing the grotto to break through its physical and geographical limitations. The characteristics of digital information resources can be divided into data characteristics, structural characteristics, and content characteristics. Data characteristics that describe that CH digital resources are massive, multi-source, heterogeneous, multi-modal, and cross-cultural. Structural characteristics describe the medium or format of CH digital resources, including audio, video, photographic, and so forth. Finally, content characteristics include digital murals, digital archives, 3D models, laser point clouds, multi-spectral images, etc. At present, the Dunhuang Academy has cooperated extensively with renowned scientific research institutions and universities to form a set of technical specifications for the digitization of cultural relics suitable for the characteristics of immovable cultural relics such as grottoes. It has accumulated several digital resources of different data types, including multimedia resources such as pictures, audio, video, and 3D models.

### 3.4 The facet of mural content knowledge

The artistic composition of the Dunhuang grottoes has four main elements: grotto temple architecture, statues, frescoes, and Dunhuang MS. The most important part of Dunhuang CH, Dunhuang murals are cultural treasures in the history of mankind. Considered to be magnificent “libraries on the walls,” they have extremely high artistic and appreciation value. Because the murals display the creativity and integration of Dunhuang artists of all dynasties, they formed their own unique artistic characteristics and techniques for expression. Compared with other elements, the

murals are numerous and rich in semantics, with a total area of nearly 45,000 square meters preserved in the Mogao Cave in Dunhuang (Whitfield and Otsuka 1995).

Because Dunhuang murals are so rich in semantic information and are dependent on contextual revelation and artificial understanding of labeling, it is difficult to symbolize, formalize, and semanticize. Traditional methods require professional domain experts to interpret the mural content, but domain experts may not be able to adapt to this complicated form of annotation. As a result, semantic representation of the mural content has gradually become a hotspot in Dunhuang grotto research. With deep semantic annotation based on image theory, researchers can work with domain experts to realize the semantic description of mural contents. This approach facilitates knowledge mining and discovery and is convenient for users to understand the mural content in depth (Wang et al. 2020a).

### 3.5 The facet of historical and cultural knowledge

From the fourth century AD to the fourteenth century AD, the Dunhuang grottoes were continuously excavated and rebuilt over the course of more than ten dynasties. Dunhuang's ruling groups, eunuchs, monks, noblemen, folk societies, common people's families, passers-by, and other people from all walks of life were involved in the construction of the grottoes. The creation of the Dunhuang Mogao Grottoes and its Buddhist activities became an important part of social life in Dunhuang during the middle ages. Records of construction activities, included recordings of the era of construction and related conditions of construction activities in Cave No.55, and a portrait painted of the donor who built the cave Cao Yuanzhong. According to records, Buddhist activities in the Dunhuang grottoes mainly included fasting, scripture writing, and worship. For example, there is an inscription in Cave No.469 that records a large-scale 2,000-person fasting party hosted by Cao Yuanzhong at Dunhuang in 953 AD. The construction activities and Buddhist activities of the Dunhuang grottoes were carried out throughout almost the entire major historical period of the Dunhuang grottoes, so the description of these activities or events can help to shed light on the historical and cultural context of the Dunhuang grottoes.

### 4.0 Domain ontology building

Ontology is a clear formal specification of the conceptual model to improve its interoperability and accessibility. The Dunhuang Grotto Ontology Model was extended and customized to the characteristics of Dunhuang grottoes CH by referring to the widely used and recognized ontology models in the world.

The Dunhuang Grotto Ontology Model displays a high level of abstraction and generalization of Dunhuang grottoes domain knowledge. It expresses informal knowledge clearly as various concepts and their relationships. It deeply describes and reveals the semantic characteristics of Dunhuang grottoes' research resources, and the semantic connections between them. According to the division of Dunhuang CH by the ternary world theory, the knowledge of Dunhuang CH may be divided into five facets. Moreover, knowledge of Dunhuang's CH is organized around the Dunhuang grottoes (the structure of grottoes and the contents of its murals) on the basis of "people-things-time-land-things."

The Dunhuang Grotto Ontology Model references and reuses *CIDOC CRM*, *FOAF*, *FRBR* and other internationally widely used ontology models and data standards. In addition to this achievement by the Dunhuang Grotto Ontology, this paper formulates corresponding data standards and data mapping specifications for the literature, multimedia resources, databases, metadata information, and other data in the field of Dunhuang grottoes in order to finally provide a unified and standardized data model for the construction of a knowledge base for the Dunhuang grottoes. In accordance with the core constituent elements of Dunhuang grottoes CH, the data model is divided into the following classes: grotto, indoor space, component, clay sculpture, mural, subject, academic work, digital resource, event, location, time, and agent. Each class is associated and described by object attributes, and each class has its own defined data attributes.

#### 4.1 Modeling of the spatial structure

When expressing the spatial structure in the Dunhuang grottoes, the existing international standards such as CityGML and IFC of BuildingSmart provide the geometric expression, attributes, and visual data model of the building composition. However, these standards do not provide for detailed description of the orientation relationship inside the indoor space structure. This inability results in the loss of some information in the process of knowledge organization (Kim et al. 2014). Moreover, although these model standards can describe modern buildings well, they fail to meet the need for fine-grained semantic representation in the field of CH architecture. In particular, different CH buildings have very different architectural characteristics, which needs to be taken into consideration when organizing structural knowledge of CH buildings.

The spatial structure of the Dunhuang grottoes can be semantically divided into indoor space, components, and indoor objects according to the characteristics of the spatial structure of the grottoes (Figure 2).

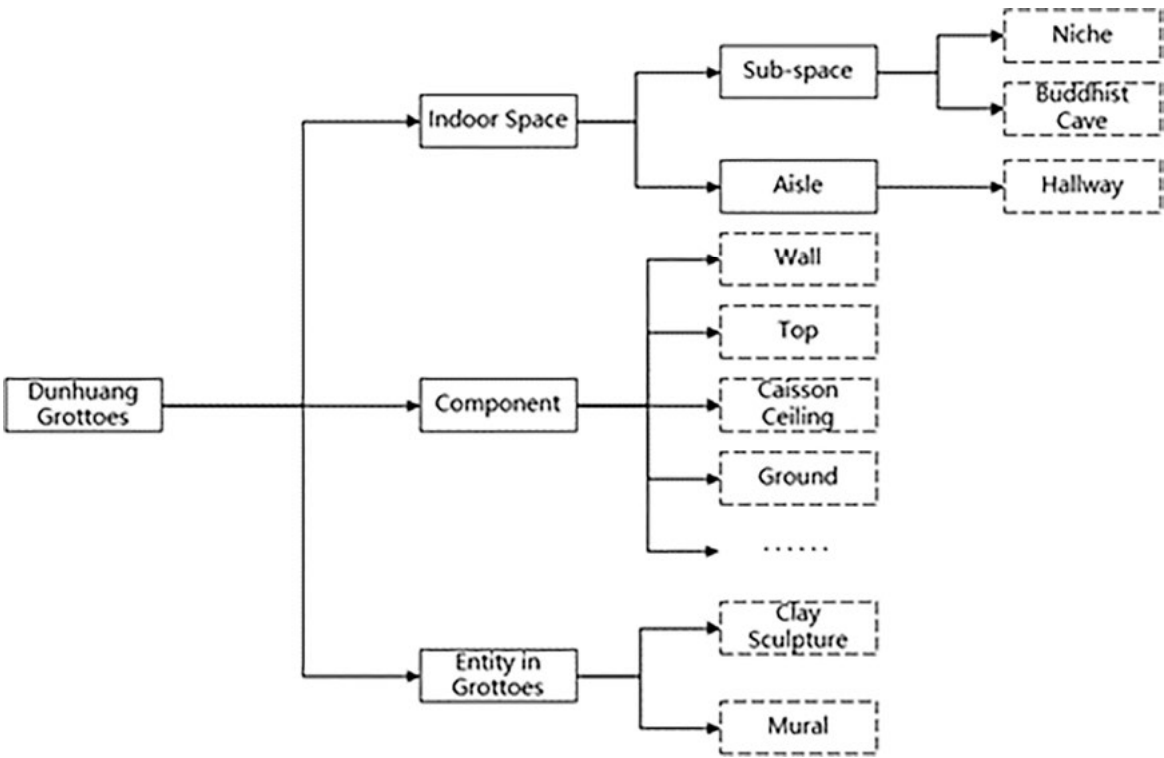


Figure 2. Semantic division of spatial structure in grotto.

Semantic location is an important means of linking semantic objects in space and can be direct or indirect. Direct semantic positional relationships include: coordinates, place names, addresses, etc., while indirect semantic positional relationship may include a combination of distance, orientation, and relative topological relationships. In this paper, the semantic objects are mainly related through indirect positional relationships. Reference is made to the division of spatial positional relationships in MPEG-7, which are spatial orientation relationships, topological relationships and spatial semantic relationships. These spatial position relationships are discussed further in Table 1 below.

Type	Definition	Example
Spatial direction relationship	Represents regional location information of semantic objects in space	top of, bottom of, middle of
Topological relationship	Spatial orientation location relative to other components of or in the space.	nearby, within, contain, adjacent to
Spatial semantic relationship	Represents the logical relationship between semantic objects in space, such as subordination, dependency, etc.	belongs to, part of, consist of

Table 1. Spatial position relationship.

According to the semantic division of the spatial structure and the classification of spatial orientation relationships, this paper designs a spatial structure ontology representation model suitable for accurately representing the Dunhuang grottoes. This new spatial structure ontology representation model includes five categories: grottoes, space, components, statues, and murals. The data attributes of spatial orientation are defined for each class, signifying the spatial orientation relationship, topological relationship, and spatial semantics between the cave structures.

4.2 Modeling of the academic work

Academic research in the area of Dunhuang grottoes has accumulated a large amount of academic research resources, including research papers, books, and paintings. These academic works and documents provide important documentation and analysis of the historical and cultural knowledge contained in the Dunhuang grottoes. The facet of academic work knowledge provides a correlation between research literature and the content of Dunhuang grottoes in order to reveal and link the academic knowledge contained in the field of Dunhuang grottoes. The attributes of the facet academic work knowledge mainly include the titles, authors, introductions, editions, publication times, and so on.

### 4.3 Modeling of the digital resource

Tremendous enduring interest in the Dunhuang grottoes has resulted in the proliferation of many digital resources of different data types, including multimedia resources such as pictures, audio, video, and 3D models. These digital resources provide important documentation and analysis of the historical and cultural knowledge contained in the Dunhuang grottoes. It mainly includes names, types, description formats, sources, and so on.

### 4.4 Modeling of the mural content

The knowledge of the mural content is a representation of the content and themes depicted in the mural based on the grotto's structure knowledge. The attributes of the mural content category include the mural content description, mural type, painting technique, and the theme expressed in the mural's content. The theme is a semantic understanding of the content of the mural. The semantic content of the mural can, therefore, be identified as the mural's theme.

### 4.5 Modeling of the historical cultural knowledge

Dunhuang grottoes related historical data and research literature constitute the main data sources of its historical and cultural knowledge. The ontology model expresses and reveals the historical and cultural knowledge of Dunhuang grottoes from different angles. It is based on people, things, times, places, and events. From the perspective of the ontology model, the historical and cultural knowledge is contained in multiple categories of rich semantic relations, and the knowledge facts and logical relationships of the historical culture are expressed from multiple angles. The facet of Dunhuang CHI mainly includes four categories: people, things, times, and places. The time category contains two sub-categories: dynasty and the epoch year. The epoch year can represent the beginning and end years of the relevant dynasty.

### 5.0 Knowledge graph construction of Dunhuang

The knowledge representation and organization method based on the knowledge graph should include the complete process of data collection and processing, data semantic extraction, entity association, and knowledge graph visualization and retrieval. The process of constructing the knowledge graph is shown in Figure 3. The knowledge resource layer is the foundation of the project. In this layer, structured, semi-structured, and unstructured data are collected and organized to build the original corpus based on the five previously described ontological facets. There are two main steps in the knowledge representation and organ-

ization layer. The first step is to clean the unstructured original corpus, filter out effective information resources, and annotate and extract the data resources according to the Dunhuang Grotto Ontology Model. The second step is to associate structured data, semi-structured data, and open external data through data mapping. The next layer, the knowledge service layer, consists of building a knowledge service-oriented KOS based on the completed knowledge database. This paper builds a knowledge service platform for Dunhuang CH by means of our knowledge graph. This knowledge service platform provides functions such as knowledge navigation, visual analysis, and semantic retrieval.

### 5.1 Data acquisition and preprocessing

The resources related to Dunhuang CH are numerous and widely available. According to the degree of structure, they can be divided into structured data, semi-structured data, and unstructured data. These data categories range from multimedia resources such as pictures and videos, Dunhuang research documents, to books and other resources. Structured data specifically includes Dunhuang related bibliographies, citation indexes, and metadata registries. Semi-structured data includes unstructured metadata descriptions and annotated resources. Finally, unstructured data is mainly multimedia resources and descriptive text of grotto structure content.

In recent years, the Dunhuang Academy has established a large number of databases, including the full-text database of the Dunhuang Conference, and the Dunhuang Studies Book Catalog Database. At the same time, through digital technology, the Dunhuang Academy has preserved a large number of high-definition images of Dunhuang murals, 3D models, and other visual resources. We were, therefore, able to obtain structured, semi-structured, and unstructured data from these data sources through cooperation with Dunhuang Academy. We also cooperated with CNKI (China National Knowledge Infrastructure)—the largest literature database in China. CNKI includes more than twenty-nine million records. Humanities, social science, natural science, and application science researchers use CNKI as a network publishing platform to supplement their academic works. This data was harvested, cleaned, and transformed to obtain valid Dunhuang CH data for this study.

### 5.2 Knowledge representation and organization

In order to accurately express the spatial structure of the Dunhuang grottoes according to the ontology, we extracted the description of the Dunhuang grottoes' structural knowledge in the General Contents of Dunhuang Grot-

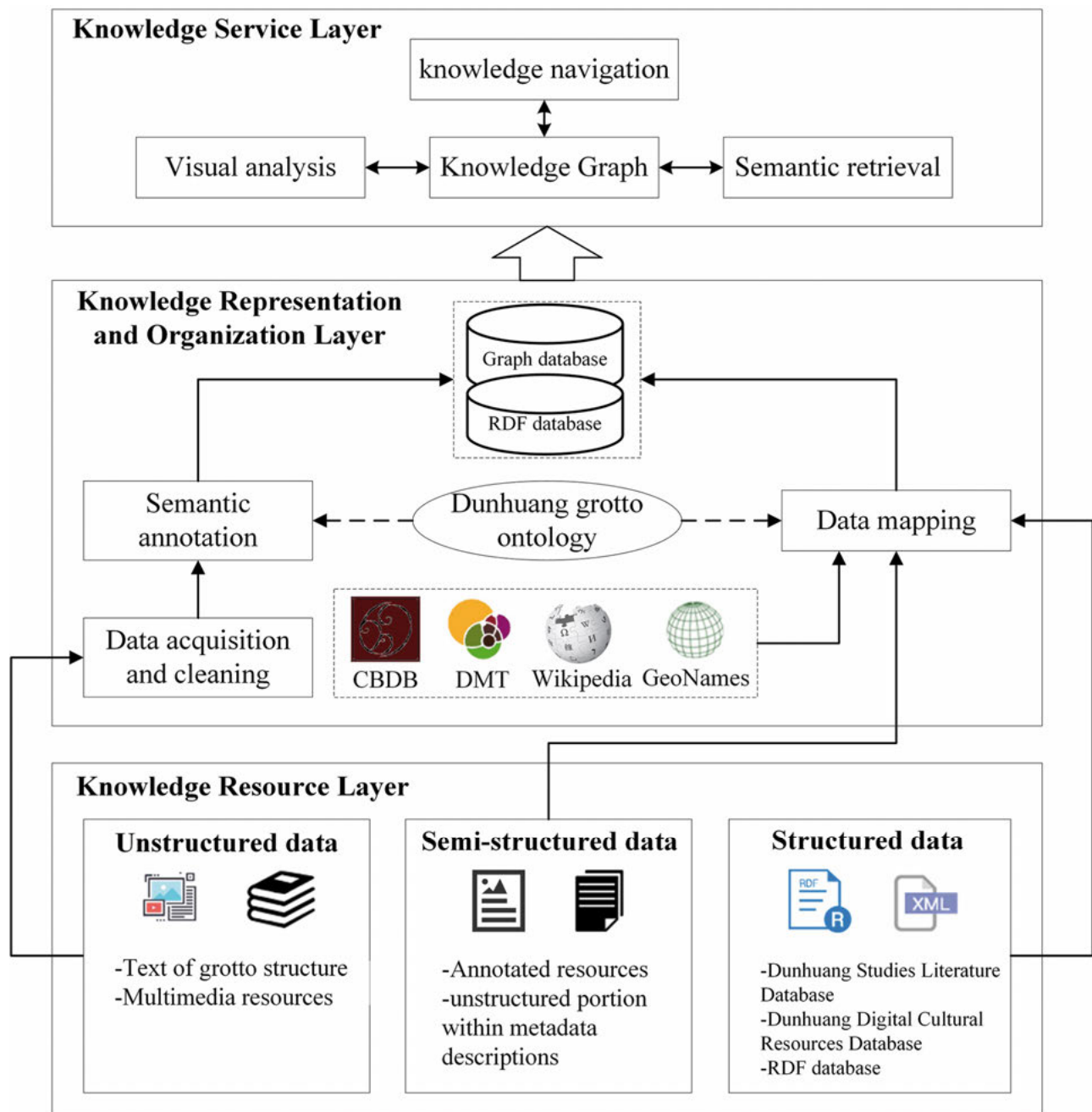


Figure 3. Knowledge organization model.

toes. In this step, the unstructured text data composed of the detailed description of the structure and content of each cave was converted into structured knowledge. To assist researchers in their semantic annotation of Dunhuang domain knowledge, we developed a fine-grained text semantic annotation tool that used predefined ontologies.

The structure space of the grotto contains a lot of information. There is a strong semantic relationship between the different composition structures and the objects in the grotto. The semantic division of the spatial structure can be

represented by a three-dimensional space structure, which can then be transformed into structured knowledge of spatial structures according to the spatial structure ontology (Figure 4).

Once the spatial structure of the cave has been semantically extracted, it is possible to integrate the other four facets of knowledge into this ontology, including the facets of academic works knowledge, multimedia resources knowledge, history and culture knowledge, and the facets of mural content knowledge.



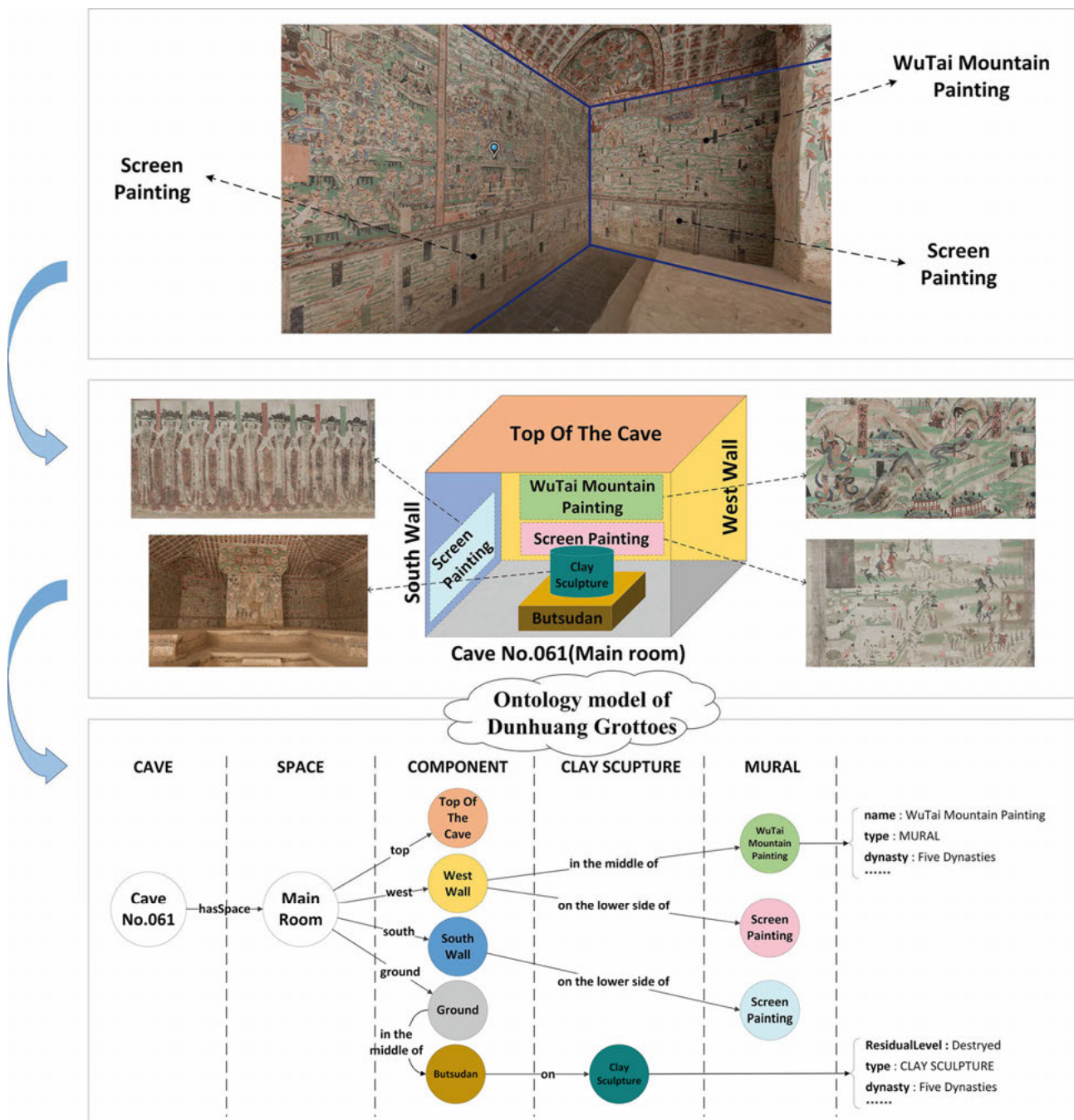


Figure 4. The process of abstracting the spatial structure (Cave No.61).

When implementing the description of the semantic relationship between different entities in Dunhuang CH, it is also necessary to consider establishing association with external data as much as possible, so as to make the content of the knowledge base more abundant and accurate. In this paper, CBDB, Dunhuang Mural Thesaurus (DMT) (Wang et al. 2020b), AAT, GeoNames, and Wikipedia were selected for entity association. DMT is a thesaurus based on the Dunhuang Dictionary and other literature related to Dun-

huang grottoes, which is constructed through the top-down and bottom-up methods, with AAT as the reference. The thesaurus provides a detailed division and interpretation of Dunhuang-related vocabulary.

### 5.3 Knowledge graph service platform

After organizing the CHI and storing it in the knowledge base, we constructed a platform for displaying our know-

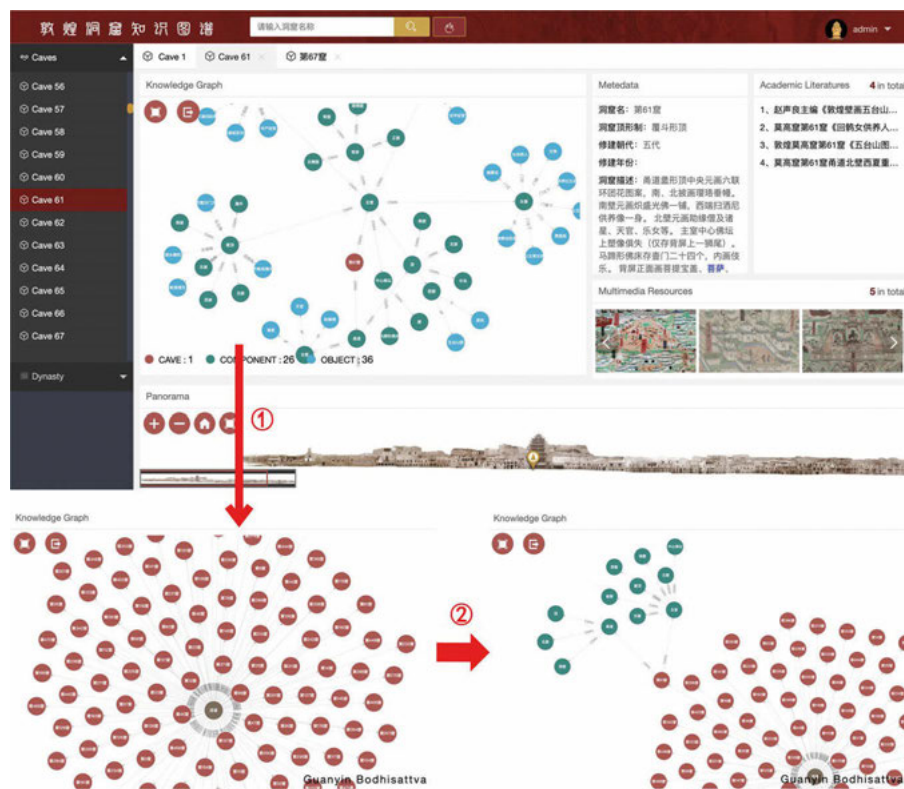


Figure 5. Knowledge graph of Dunhuang cultural heritage.

ledge graph of Dunhuang grottoes. This platform provides functions such as knowledge navigation, semantic retrieval, and visualization. This platform will be opened and provided to scholars as well as members of the general public who are interested in Dunhuang CH. Under the guidance of standardized knowledge organizations and domain experts, this platform can provide standardized and accurate knowledge services for Dunhuang CH.

Within the knowledge graph, nodes represent instances, and the connection between two nodes represents the semantic relationship between entities. Through the semantic relationship between nodes, knowledge inference and mining can be achieved. The navigation bar supports retrieval operations by cave number and dynasty. For example, Figure 4 illustrates how the information linked to Cave No.61 is retrieved, as well as how the spatial structure information of the grotto is displayed in the system, allowing a clear visualization of the structural distribution of the grotto. By clicking on different entity nodes, the metadata information, research literature information, and related multimedia resources related to the node are displayed.

Semantic retrieval is an important aspect of providing knowledge services, which can help users quickly locate desired knowledge. This knowledge graph platform also supports semantic search functions. For example, if you search for “Guanyin Bodhisattva,” all caves containing the search

term will be displayed on the interface, and knowledge discovery can be realized by clicking on the corresponding cave. This search is illustrated in Figure 5, steps one and two.

Because the knowledge graph is associated with external data, such as the Dunhuang Mural Thesaurus, users can click on the corresponding entity term to view externally associated data if they need to go deeper or expand on some content when using it. When you click “Bodhisattva,” you can jump to the Dunhuang fresco thesaurus to view the interpretation of the word. Likewise, when you click on a research resource, you will be redirected to the detailed page of the document in CNKI. Thus, external linkages enrich the knowledge graph and provide searchers with additional context and avenues for further exploration and clarification.

## 6.0 Discussion

In order to realize the knowledge organization and management of Dunhuang CH resources, this paper analyzes the characteristics of Dunhuang CH and proposes a knowledge ontology model for Dunhuang grottoes. The Dunhuang Grotto Knowledge Graph was constructed to realize the semantic representation and association of Dunhuang grottoes CHI and can provide high-quality knowledge services with distinct advantages. The following section summarizes the ideas presented in this study and discusses how they

could be further expanded and optimized by future projects.

- 1) In CH, building structure information is an indispensable part of architecture, but the current research on the organization of Dunhuang CHI has mainly focused on Dunhuang's murals and manuscripts while ignoring the rich semantic information of the cave structure itself. Therefore, a spatial structure representation ontology based on the semantic features of Dunhuang grottoes was constructed in this paper.
- 2) When describing the semantic contents of CH, the composition of the resources can be clearly recognized by dividing the knowledge of different topics or types, which provides a reference for accurate and reasonable ontology construction. For example, this article describes the contents of Dunhuang CH through five knowledge facets, which provided guidance for the subsequent construction of a Dunhuang Grotto Knowledge Graph.
- 3) The content of the murals in the Dunhuang grottoes is rich in semantic information. This information needs to be represented and mined through a more refined semantic annotation model. The focus of this paper is to organize and correlate the multi-source heterogeneous and multimodal resources of Dunhuang CH. The representation of mural contents was performed at a medium granularity level. The system could be integrated with a digital interactive display system in the future to facilitate deeper understanding of the plot, action, and context of mural content.
- 4) The evaluation of the knowledge graph is an indispensable step of its construction. In this paper, to ensure the accuracy, completeness, and usability of the knowledge graph, we invited experts in the field of Dunhuang CH to conduct a preliminary assessment of the Dunhuang CH knowledge graph. In order to expand the scale of the knowledge graph, enhance the degree of relevance of the knowledge in the graph, and provide richer and more personalized functions, further systematic evaluation of the knowledge graph will be carried out through questionnaire surveys, user interview and observation, and eye-tracking experiments. The results of these evaluations will be examined in future research.

## 7.0 Conclusion

This paper analyzes the current state of Dunhuang DHI from the organization of CHI, representation of building structure, to existing CH knowledge services. A Dunhuang Grotto Ontology (covering cave structure, academic work, multimedia resources, mural content, and historical and cultural) was also designed to address current challenges in Dunhuang CHI. A multi-source, heterogeneous, multi-

modal Dunhuang Grotto Knowledge Graph was subsequently constructed for the purpose of representation and association. In addition, a knowledge graph platform was built to provide knowledge service.

In this paper, the Dunhuang Grotto Knowledge Graph was realized via constructing a high-level conceptual ontology, semantic annotation, extracting entities and relationships of Dunhuang CHI, linking the knowledge of five facets, and storing them into Neo4j. By constructing a visualization and retrieval service platform, this study may promote the research, teaching, and dissemination of Dunhuang CH and provide a methodological reference for related CH projects. In the future, a larger dataset and richer knowledge network system of Dunhuang CHI will be constructed. Varied data resources will be fused and re-constructed into smart big data to meet the demand for multiple knowledge services under different retrieval tasks and usage scenarios.

## References

- Abbott, Alison. 2017. "The 'Time Machine' Reconstructing Ancient Venice's Social Networks." *Nature* 546: 341-4.
- Alexiev, Vladimir, Joan Cobb, Gregg Garcia and Patricia Harpring. 2014. "Getty Vocabularies Linked Open Data: Semantic Representation." <http://vocab.getty.edu/doc/>
- Allen, Peter K. and Alejandro Troccoli, et al. 2003. "New methods for digital modeling of historic sites." *IEEE Computer Graphics & Applications* 23: 32-41.
- Fan, Jingshi. 2000. "A Century of Research on Dunhuang Caves: A Review and Prospect." *Dunhuang Research* 186-187. doi:10.13584/j.cnki.issn1000-4106.2000.02.006
- Foni, Alessandro E., George Papagiannakis and Nedjma, et al. 2007. "Time-Dependent Illumination and Animation of Virtual Hagia-Sophia." *International Journal of Architectural Computing* 5: 283-301.
- Kim, J. S., S. J. Yoo and K. J. Li. 2014. "Integrating IndoorGML and CityGML for Indoor Space." In *Web and Wireless Geographical Information Systems: 13th International Symposium, W2GIS 2014, Seoul, South Korea, May 29-30, 2014. Proceedings*, ed. Dieter Pfoser and Ki-Joune Li. Lecture Notes in Computer Science 8470. Berlin: Springer, 184-96. doi:10.1007/978-3-642-55334-9\_12
- Ma, Yaxue, et al. 2019. "Theoretical Thinking of Urban Data Portrait from the Perspective of Digital Space." *Journal of the China Society for Scientific and Technical Information* 38: 62-71. doi:10.3772/j.issn.1000-0135.2019.01.007
- Mäkelä, Eetu, Eero Hyvönen and Tuukka Ruotsalo. 2012. "How to Deal with Massively Heterogeneous Cultural Heritage Data - Lessons Learned in Culturesampo." *Semantic Web* 3: 85-109.

- Moura, Aparecida, M., Monteiro and Drumond, S. 2014. "Knowledge Graph and 'Semantization' In Cyberspace: A Study of Contemporary Indexes." *Knowledge Organization* 41: 429-39.
- Ning, Q. 1986. "The Basic Structure of Dunhuang Grotto Art." *New Arts* 2: 56-58. doi:cnki:sun:xmsh.0.1986-02-015
- Noor, Salma, et al. 2018. "Modeling and representation of built cultural heritage data using semantic web technologies and building information model." *Computational & Mathematical Organization Theory* 25: 247-70.
- Pauwels, Pieter, Rens Bod, Danilo Di Mascio and Ronald De Meyer. 2013. "Integrating Building Information Modeling and Semantic Web Technologies for the Management of Built Heritage Information." In *Proceedings of the 2013 Digital Heritage International Congress*. Piscataway, NJ: IEEE. 1: 481-8. doi:10.1109/DigitalHeritage.2013.6743787
- Purday J. 2009. "Think Culture: Europeana.eu from Concept to Construction." *Bibliothek Forschung Und Praxis* 33: 170-80.
- Simeone, Davide, S Cursi and M Acierno. 2019. "BIM Semantic-Enrichment for Built Heritage Representation." *Automation in Construction* 97: 122-37.
- Singhal, Amit. 2012. "Knowledge Graph. Google Official Blog." <http://googleblog.blogspot.com.br/2012/05/introducing-knowledge-graph-things-not.html>.
- Wang, Xiaoguang, Xiaoxi Luo, Xu Tan, et al. 2020a. "Representation and Display of Digital Images of Cultural Heritage: A Semantic Enrichment Approach." *Knowledge Organization* forthcoming.
- Wang, Xiaoguang, Xilong Hou, Hanghang Cheng, et al. 2020b. "The Construction and Linked Data Publishing of Dunhuang Mural Thesaurus." *Journal of Library Science in China* 46: 69-84. doi:10.13530/j.cnki.jlis.2020032
- Whitfield, R and S Otsuka. *Dunhuang: Caves of the Singing Sands: Buddhist Art from the Silk Road*. Textile & Art Publications, 1995.
- Zeng, Marcia Lei. 2010. "Knowledge Organization Systems (KOS)." *Knowledge Organization* 35: 160-82. doi:10.5771/0943-7444-2008-2-3-160