Research on Knowledge Organization of Intangible Cultural Heritage Based on Metadata
Qing Fan, Guoxin Tan, Chuanming Sun, and Panfeng Chen

ABSTRACT
Metadata has been analyzed and summarized. Based on Dublin Core metadata, combined with the characteristics and forms of intangible cultural heritage, this article explores the metadata for intangible cultural heritage in knowledge organizations based on relevant resource description standards. The Wuhan woodcarving ship model is presented as an example of national intangible cultural heritage to control the application of metadata in intangible cultural heritage knowledge organizations. New ideas are provided for the digital development of intangible cultural heritage.

INTRODUCTION
Intangible cultural heritage includes traditions or living expressions inherited from our ancestors and passed on to our descendants. Digital storage and presentation of intangible cultural heritage resources is an inevitable requirement for the protection of China's long history and its culture in the information age. With the rapid development of artificial intelligence and big data, all kinds of massive data in the Internet age are expanding, necessitating the development of a database platform for the inheritance and protection of intangible cultural heritage. At the same time, organizations must consider how to deal with the intangible cultural heritage using complex data. Searching for data and visualizing the relationship with intangible cultural heritage is a current research hotspot. However, at this stage, there are still some problems in the construction of digital resources of intangible cultural heritage in China, such as the establishment of accurate and interoperable metadata. In this process, the diversity and uniqueness of intangible cultural heritage items needs to be fully considered, including the subsequent integration of digital resources and its existing digital resource system of intangible cultural heritage in China. Therefore, the construction of the intangible cultural heritage resource database is not only to simply organize and list the data, but more importantly, to reveal the relationships between the knowledge content and resources in the intangible cultural heritage field and to build a thorough and relevant knowledge system.

RESEARCH STATUS AT HOME AND ABROAD
Metadata is data that describes the attributes of a certain type of resource (or object). Metadata can be used to locate and manage the resource and display information about it. Metadata can also be structured data used to describe online information resources and strengthen the collection development, organization, and utilization of online information resources. From the perspective of knowledge organization, general metadata is used to describe the theme, content, and characteristics of information resources. The most common metadata format is Dublin Core.
(DC) metadata, which is structured and descriptive. The creation of metadata standards in the field of intangible cultural heritage must first combine the basic concepts and characteristics of cultural heritage to extract specific attributes and provide element definitions that describe the basic characteristics of intangible cultural heritage resources, that is, core metadata. This is not easy to achieve since intangible cultural heritage is traditional art, music, folklore, etc. Only by unifying intangible heritage resources of different expressions through metadata standards can a relatively standardized intangible cultural heritage resource library be formed.

The Visual Resources Association of America (VRA) created the VRA Core metadata standard to describe art, architecture, prehistoric artifacts, folk culture, and other artistic visual resources in the network environment.\(^3\) In terms of intangible cultural heritage material, Lan Xuliu et al. proposed the VRA Core as the foundation format and added elements from the Categories for the Description of Works of Art (CDWA) as the extended element metadata format of digital cultural resources.\(^4\) A sculpture of Abraham Lincoln was used as the basis for the metadata format. The example explains the specific use method of the proposed metadata format in practice. The solution does not extend the core elements and there is an overall lack of flexibility as users cannot customize the required elements. B. Murtha proposed a descriptive metadata architecture in the field of art and architecture, including the core category of ontology ID, and added a controlled vocabulary and classification system in the field of art and architecture to enrich the specific metadata model.\(^5\) It is mainly based on the theoretical discussion of metadata standards in this field, and there is no specific practice, but its method of formulating metadata from the perspective of user retrieval effects is worth learning. Yi Junkai et al. proposed the core metadata specification for digital museums as the basis for expansion, implemented the relevant methods in the metadata expansion rules, and finally formed a special metadata specification.\(^6\) This metadata specification system can guarantee the basic and personalized description of resources.

The metadata specification was developed and completed by the National Museum of China. To keep this specification consistent with the metadata description of other metadata specifications at home and abroad, the description method refers to the ISO-11179 standard.\(^7\) The National Museum metadata specification contains seven element sets, 60 elements, and 342 restricted elements. The seven metadata element sets are: collection resource entity, data resource entity, responsible entity, business entity, transaction entity, relationship entity, and save entity. Each metadata element of the museum’s digital resources defines several elements according to the concept of hierarchical structure; each element is defined and described by a group of attributes, such as name, version, logo, definition, type, and value range. There are 11 attributes of necessity, repeatability, lower-level elements, application scope, and annotations. The establishment of the metadata standard framework for museum digital resources is based on the digitization of museum collection resources. Collection resources are the core of museum work, and the content of museum collection resources is the core component of digital resources.\(^8\) The digital resources of these collections are related to communication, transmission, storage, or business activities.

Based on the characteristics of China’s existing intangible cultural heritage information resources, Li Bo proposed a compatible and interoperable metadata model. The description of intangible cultural heritage information resources was created on the basis of information structure and semantic component analysis.\(^9\) The ontological characteristics of each intangible cultural heritage information and related documents, characters, objects, spaces, and other entities are included in the construction of the intangible cultural heritage metadata model, which combines China’s non-material cultural heritage. The actual situation of the tangible cultural heritage database has a
certain degree of international generality. Ye Peng compared the DC metadata standard system with the needs of China’s intangible cultural heritage protection, proposed a metadata standard based on intangible cultural heritage resources, and gave the scope of application. This metadata standard contains multiple core metadata, corresponding to the relevant elements in DC. However, Ye Peng also pointed out that a major problem with this intangible cultural heritage metadata standard is that it is not compatible with China’s existing intangible cultural heritage database, such as information storage, digital mining, file retrieval, and multimedia distribution.

CONNOTATION AND DESIGN PRINCIPLES OF INTANGIBLE CULTURAL HERITAGE METADATA

**Connotation of intangible cultural heritage metadata**

The core metadata of this study will be designed based on the DC core metadata set, considering its versatility, scalability, easy conversion between metadata, interoperability between systems, and existing comparisons. Universal DC metadata is the most influential and widely used metadata standard in the field of information resource description under the network environment. Since the DC metadata standard is mainly aimed at the retrieval of network entity resources, it reveals common characteristics of digital entity resources but does not consider the cultural connotation and knowledge context of specific knowledge topics such as intangible cultural heritage. To reveal the originality of the object, the model proposed in this article will also combine the application and recording of China’s intangible cultural heritage items, reflecting the characteristics of specific intangible cultural heritage items, so as to facilitate compatibility and integration with existing information resources to form a unified interface standard with the existing intangible cultural heritage management system of the cultural sector, enabling the sharing of digital resources among cultural centers in different regions.

**Design principles of intangible cultural heritage metadata**

The design of the metadata model of intangible cultural heritage information resources should be fully compatible with popular metadata standards. Various metadata standards apply to different objects: DC is suitable for network resources, CDWA is suitable for artworks, and Federal Geographic Data Committee (FGDC) is suitable for geographic space.

When it comes to digital collections, the National Library of the Netherlands was one of the first institutions in the world to respond, starting in 1994 with the decision to collect digital publications and working with publishers and IT partners to make important contributions to digital collections research. The National Library of the Netherlands will develop a new global information network. The main approach of the system is to add DC data to all collected web pages. The new web page will require providers to add elements of the DC core set by themselves. Once submitted, the National Library of the Netherlands’ search engine will use these DC elements to assist in retrieval. In recent years, the art museum community has adopted several metadata standards such as CDWA and VRA Core to describe their collections of art works. Nam, Y. J. and Lee, S. M proposed a set of metadata elements customized to fit into the distinct context of small-scaled art museums in South Korea. A small art museum in Korea combines the existing CDWA, VRA Core, and DC standards and the proposed set of metadata elements is expected to support artistic resources. The metadata design of intangible cultural heritage resources should refer to the design cases of the Netherlands and South Korea. When applying the existing metadata standards, it is beneficial to fully reveal the characteristics of the described objects and decide whether to use the overall framework or the partial use, and must not be blindly used.
The design of the metadata should connect with existing intangible cultural heritage information sources. At present, China should refer to the relevant standards of world intangible cultural heritage digital resources and establish a management system for intangible cultural heritage resources in line with the national, provincial, municipal and county levels. The relevant cultural management functional departments have also established relevant information systems to form a unified set of authoritative and standardized data. Therefore, in terms of the elements and concepts used in the metadata model, special attention should be paid to the connection with these existing data models, so that as new resources are developed, these rich information sources can be shared through the mapping relationship between the elements.

The metadata model should have good scalability and strong descriptive ability containing more elements. Therefore, an element-rich metadata model has a strong influence on the organization and management of information resources and content disclosure. Data inspection should be flexible. Conversely, a metadata model with a lack of elements will be less flexible when technology is upgraded or user description requirements are expanded. A metadata model requires constant expansion and modification, and the practicality of the model will be greatly reduced. On the other hand, the design of the metadata model should have a mechanism that facilitates different types of users to expand elements according to different needs.

The design of metadata can show the relationship between intangible cultural heritage resource entities. With the development of information resource description technology at home and abroad, a batch of metadata standards for various types of information resources have been formed. The metadata standards for China’s intangible cultural heritage should aim for compatibility and integrate existing world standards based on current results. The metadata should further be expanded and developed in accordance with preserving intangible cultural heritage works. The metadata standards for intangible cultural heritage archives should describe resources while displaying the greatest degree of versatility, compatibility, and standardization. Therefore, combining the requirements of cultural heritage archiving and the characteristics of intangible cultural heritage, the DC metadata standard is used as the basic standard, and the advantages of other metadata standards are combined to determine the metadata standard of China’s intangible cultural heritage archives.

INTANGIBLE CULTURAL HERITAGE KNOWLEDGE ORGANIZATION

Definition of intangible cultural heritage metadata
Through semantic analysis, the core attributes and concepts involved in metadata can be obtained, and the specificity of attributes and concepts can be improved through metadata standards, which can make users’ cognition, retrieval, and evaluation of information more accurate and effective. At the same time, the normative concepts and common attributes in existing metadata schemes should be quoted as much as possible. According to the attribute characteristics of the object, close and similar conceptual entities can be selected from one or more common metadata schemes, so that the element definition has versatility and standardization. For intangible cultural heritage, according to the attributes and characteristics of the object, close and similar conceptual entities are selected from one or more general metadata schemes to make the element definition universal and normative. In the “Convention for the Safeguarding of Intangible Cultural Heritage,” UNESCO pointed out that the types of intangible cultural heritage include oral traditions, performing arts, social practices, festivals, traditional handicrafts. Based on the above-mentioned definitions of intangible cultural heritage types and the previous comparative research results on metadata
standards of various countries, combined with the research results of scholars, the DC standard metadata name and standard affix library incorporates a set of intangible cultural heritage archive metadata containing 23 elements and extended elements (table 1).

**Table 1.** Core metadata of intangible cultural heritage digital resources

<table>
<thead>
<tr>
<th>Category</th>
<th>Standard metadata name</th>
<th>Field name</th>
<th>Annotation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Content</td>
<td>Title</td>
<td>DC_Title</td>
<td>Name and content of intangible cultural heritage</td>
</tr>
<tr>
<td></td>
<td>category</td>
<td>DC_category</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Bintroduction</td>
<td>DC_Bintroduction</td>
<td></td>
</tr>
<tr>
<td>Creator</td>
<td>Mcreator</td>
<td>DC_Creator_Own</td>
<td>Creator identity information</td>
</tr>
<tr>
<td></td>
<td>Nation</td>
<td>DC_Creator_Nation</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sex</td>
<td>DC_Creator_Sex</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Age</td>
<td>DC_Creator_Age</td>
<td></td>
</tr>
<tr>
<td></td>
<td>area</td>
<td>DC_Creator_area</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Biography</td>
<td>DC_Creator_Biography</td>
<td></td>
</tr>
<tr>
<td>Category</td>
<td>Dance</td>
<td>DC_Category_Dance</td>
<td>Heritage list category</td>
</tr>
<tr>
<td></td>
<td>Song</td>
<td>DC_Category_Song</td>
<td></td>
</tr>
<tr>
<td></td>
<td>literature</td>
<td>DC_Category_literature</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Quyi</td>
<td>DC_Category_Quyi</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Art</td>
<td>DC_Category_Art</td>
<td></td>
</tr>
<tr>
<td>Resources</td>
<td>video</td>
<td>DC_Category_video</td>
<td>Resource type includes a description of resource content</td>
</tr>
<tr>
<td></td>
<td>Picture</td>
<td>DC_Resources_Picture</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Text</td>
<td>DC_Resources_Text</td>
<td></td>
</tr>
<tr>
<td></td>
<td>network</td>
<td>DC_Resources_network</td>
<td></td>
</tr>
<tr>
<td>Organization</td>
<td>area</td>
<td>DC_Organization_area</td>
<td>Organization Information</td>
</tr>
<tr>
<td></td>
<td>principal</td>
<td>DC_Organization_principal</td>
<td></td>
</tr>
<tr>
<td></td>
<td>OfficePhone</td>
<td>DC_Organization_OfficePhone</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Jobtitle</td>
<td>DC_Organization_Jobtitle</td>
<td></td>
</tr>
<tr>
<td></td>
<td>introduction</td>
<td>DC_Organization_Introduction</td>
<td></td>
</tr>
</tbody>
</table>

Intangible cultural heritage metadata standards unify the information format and mutual mapping relationship of intangible cultural heritage digital achievements. On the one hand, a single standard removes barriers to sharing metadata caused by having intangible cultural heritage information resources with different hardware, different platforms, and different formats. On the other, it enables the digital resources of intangible cultural heritage to be shared online. For example, the China Intangible Cultural Heritage Digital Museum ([https://www.ihchina.cn/](https://www.ihchina.cn/)) uses unified metadata to design this section, which solves the problem of integration and sharing of different resources.

The smooth conversion between new and old data is beneficial to the protection of intangible cultural heritage inventory data, avoiding duplication of work, and improving the efficiency and effectiveness of intangible cultural heritage storage. In addition, design of intangible cultural
heritage metadata standards must consider the versatility, compatibility, and individualization of the metadata system.

**Description of digital resources of intangible cultural heritage**

Through the analysis of the intangible cultural heritage project objects, we can provide content, management, resources, etc. These attributes can correspond to the elements of the metadata during the metadata design or serve as the semantic basis for the definition of the elements. The analysis and extraction of the core attributes and concepts of the object should first consider the full presentation of the object knowledge and resource content, and the concept should have a certain degree of specificity so that users can recognize, retrieve, and evaluate the information. Secondly, it is important to refer to the normative concepts and general attributes in the existing metadata schemes as much as possible. According to the attribute characteristics of the object, select close and similar conceptual entities from one or more general metadata schemes, so that the element definition is versatile and standardized. Therefore, the content description of intangible cultural heritage items should reflect unique cultural meanings and characteristics. At present, there are only general concepts such as name, category, subject, and region among several general metadata schemes. Figure 1 shows the metadata framework of intangible cultural heritage.

![Metadata framework of intangible cultural heritage](image)

**Figure 1.** Metadata framework of intangible cultural heritage.

In the content description, there are five elements which include names, types, subjects, regions, and protection levels as special attributes. In the metadata standard, the only elements that can be used in general are the name, subject, category, and region. The protection level means that the list of intangible cultural heritage is the object of national or provincial protection. The "National
Intangible Cultural Heritage Declaration Form” uses the five elements of content description, and then conducts resource description analysis based on the information organization structure of the intangible cultural heritage project object to construct intangible cultural heritage. The framework covers the main attributes and definitions involved in intangible cultural heritage objects, as well as their connections and hierarchical relationships. In the description framework, in addition to the attributes and definitions specified by the DC metadata, a set of custom elements is also set. Without changing the basic structure, users can customize elements according to standard needs to make the model extensible.

In the description of related resources, entities related to intangible cultural heritage are divided into four categories: inheritors, object categories, resources, and organizations. Among them, the inheritor-related attributes include six general attributes such as name, ethnicity, gender, region, age, and person profile. Object category attributes include dance, song, art, literature, video, network, etc. For intuitive objects, you can refer to the use of artistic works to describe the category or the core category of visual materials, and the documentation and materials can use the metadata defined by DC. This model does not specify the use of attributes and concepts in metadata. In a specific metadata solution, these attributes and concepts can correspond to metadata element names, or they can be modifiers, values, or metadata element definitions, such as the inheritor of shadow puppetry is Lin Shimin.

**Data association**

Linked data is a technical specification recommended by the World Wide Web Consortium (W3C). The relationship among linked data objects supports a greater degree of resource sharing and utilization, enabling users to efficiently and accurately locate needed resources on a larger scale. The release of linked data is to describe the metadata of cultural resources in the form of Resource Description Framework (RDF). After forming semantic associations, intelligent retrieval and data discovery services are provided on the intelligent application platform, so as to ensure the visual presentation and data sharing of intangible cultural heritage digital resources in knowledge organizations. Linked data publishing provides standardized data access specifications. The biggest advantage is that it can correlate data across platforms and establish links to different data, which is convenient for users to search for data in different repositories.

As far as the content of intangible cultural heritage is concerned, linked data presents unstructured, semi-structured, and structured data on the Internet in the form of RDF. RDF description refers to the transformation of metadata in resources into RDF triples through data and relationship mapping, and the formation of W3C-supported documents through semantic relationship construction. Visual presentation refers to the visual presentation of relevant content by users through network search with the support of the network architecture. In essence, the release of digital resource data is to realize the RDF description and sharing of metadata for intangible cultural heritage metadata by multiplexing the relationship. Its essence is the management application process of the database.

The linked data publishing process of intangible cultural heritage resources consists of three steps: (1) Converting the metadata of the repository into an RDF triple model and assigning a URI identifier to form an RDF document of linked data; (2) Establishing a semantic relationship and building relational links to form semantic associations; and (3) Mapping cultural resource data to the network through the URI access mechanism, and presenting data search results in a visual way.
through user SPARQ queries. Although there are differences in the structure of different data publishing tools, metadata-based linked data publishing follows these basic steps.

EXAMPLES OF METADATA APPLICATION OF INTANGIBLE CULTURAL HERITAGE KNOWLEDGE ORGANIZATIONS

Introduction to the Wuhan Wood Carving Ship Model intangible cultural heritage project

The Wuhan Woodcarving Ship Model is a unique art variety in Chinese woodcarving craftsmanship, with a history of more than 2,000 years. According to the Song Dynasty’s The History of Jin Shi·Zhang Zhongyan: “the craftsman did not know how to build the ship. When the boat was built, the craftsmen did not know how to build it. The boat model made by Zhang Zhongyan. It was only a few inches long and was very delicate. The front and rear of the boat could be spliced well without glue. The other craftsmen were all amazed.” As early as the 12th century, there were people in China who could carve small boats several inches long as models for making ships.

Hubei woodcarving boats are a national intangible cultural heritage project but the art and craft faces challenges. Like other intangible cultural heritage projects, development of the craftwork is weak. While younger generations in Hubei may recognize the form of wooden carving boat, few are willing to learn this art and more young people have not even heard of it. In order to better honor this long-standing tradition, this article focuses on the characteristics of intangible cultural heritage digital resources, combined with the relevant theories of knowledge organization, and adopts certain technical standards to organize the knowledge organization and construction of the metadata standards for Hubei woodcarving ships.

Knowledge organization construction based on metadata

To effectively use metadata in intangible cultural heritage, metadata specifications must be defined and described. RDF is metadata specification description language. It can semantically pay attention to the attributes of the ontology and the interrelationships between these attributes. By using RDF information, it can be easily exchanged between computers using different types of operating systems and application languages. RDF regulates the realization of semantics in a standardized and interoperable way. The web page can implement the invocation of RDF in a simple way, thereby facilitating the retrieval of network data and the discovery of related knowledge.

In this paper, the metadata system needs to use RDF to define the attributes, so that it can be better transformed into a language that the computer can understand. Intangible cultural heritage items have a certain relationship with inheritors, organizations, resource content, etc. In order to establish a complete intangible cultural heritage cultural resource database, these entities need to be described separately in RDF.

Wuhan Woodcarving Ship Model metadata definition

According to the RDF description, Wuhan Woodcarving Ship Model is used as a specific example to show the designed metadata scheme, that is, the relevant content of the example is filled into the defined resource description frame. For example, part of the RDF description of the Wuhan Woodcarving Ship Model intangible cultural heritage item can be found in the following code:
CONCLUSION

This article reviews the classification system of China’s intangible cultural heritage items and the integration of existing knowledge organizations and other types of resources for designing a set of more comprehensive and reasonable metadata standards with a certain degree of scalability and it is applied to the actual intangible cultural heritage knowledge organization. To effectively protect and use the digital resources of intangible cultural heritage, further research is needed for this study. Additional discussion on updating and promoting existing metadata specifications as well as multidimensional aggregation of existing resources to achieve knowledge discovery is needed. Through the integration of linked data and sharing existing digital resources, this article can encourage scholarship and conversation that leads to the preservation of China’s intangible cultural heritage.

FUNDING STATEMENT

This work was supported by the Hubei Key Laboratory of Big Data in Science and Technology. This work was also supported by the Palace Museum’s open project in 2021, Research on the Dissemination of Intangible Cultural Heritage of the Palace Museum from the Perspective of Artificial Intelligence. This subject has been funded by the Mercedes-Benz Star Wish Fund of China Youth Foundation.
ENDNOTES


