Research on the Semantic Architecture of Cultural Heritage Images Based on CIDOC CRM

Xiayu Zhou

Xiayu Zhou, School of Information Management, Wuhan University, China

Abstract. An image is a visual medium that describes information about objects. The amount of data in cultural heritage images is constantly amplified in the process of historical development and circulation. It has been significant challenges. The research is based on the international standard ontology reference model CIDOC CRM in the field of cultural heritage. It designs Semantic ontologies and specifications from the perspective of describing, correspondence paths and semantic architecture. The aim is to integrate and correlate the information of cultural heritage images with semantic information. And facilitate the integration and exchange of image information for traditional humanities and social science research.

1 Semantic Web and CIDOC CRM

Massive digital information resources meet the needs of the public but gradually show the deficiencies in information representation and utilization. Tim Berners-Lee proposed the Semantic Web (Semantic Web) in 2000. As an extension of the current network, the entire Internet has become an information exchange medium, through basic element semantic links. In actual research, even the same image symbol has different semantics in different time. Only the integration of interdisciplinary research information to achieve the semantics that computers can understand and accelerate the sharing of knowledge and information.

The ontology is an important technology to achieve semantic information retrieval. It describes the norms of concepts and the relationships between concepts. CIDOC CRM(CIDOC Conceptual Reference Model) is an object-oriented conceptual reference model developed by International Committee for Documentation. CIDOC CRM has become the international standard in the field of cultural heritage research in 2006, and it is the ontology of the domain of cultural heritage. It provides the semantic definitions and clarifications for the standard specification and description architecture of cultural heritage information, as an important information and reference model of the Semantic Web. Since 2006 it is official standard ISO 21127:2006. Current version is 6.26(June 2019, Published Version), including 98 entities, 168 Properties. It aims at providing the semantic definitions and clarifications needed to transform disparate, and localised information sources into a coherent global resource, be it within a larger institution, in intranets or on the Internet.[1]

CIDOC CRM is widely recognized and applied internationally. For example, the "Digital Storage of Folk Music and Dance Culture" project funded by the Slovenian State in 2007, integrated the conceptual design of the CIDOC CRM model and FRBR model to
provide structured data for its description. CIDOC CRM is widely recognized and applied internationally. Such as, the "digital storage of Slovene folk music and dance culture" project funded by the Slovenian State in 2007, integrated CIDOC CRM and FRBR to provide a flexible data model and implementation, for description of implicit and explicit concepts and relationships between concepts and events of production processes mentioned. [2] By 2010 CLAROS more than 20 million records had been made ‘interoperable’ using CIDOC-CRM, as Classical Art Research Online Research Services with Europe’s leading research centers for the art of ancient Greece and Rome.[3] In 2012, a SPARQL end-point to an rdf description of the National Gallery Collection is being developed. The National Gallery. It developed a mapping to the CIDOC CRM [4] Semantic Computing Research Group (SeCo) published large heterogeneous sets of data about the WW2 in Finland as Linked Open Data in 2016. [5]

2 Core element norms and frame construction

2.1 Core element specification of cultural heritage images

After being created, cultural heritage images have a long period of circulation, the history, geography, and characters involved are rich and complex. This study considers the event-based semantic model of CIDOC CRM knowledge ontology to form the core of cultural heritage images Specification of elements and attributes to build their semantic association.

Based on the reference of the mature semantic description metadata of CIDOC CRM, the specific semantic requirements of cultural heritage images, and the fact that the ontology vocabulary of CIDOC CRM are deleted and integrated, the core element specification of cultural heritage images proposed in this article defines the description elements of each module and corresponds to the CIDOC CRM standard from three aspects of the basic description, event elements and event description.

The details are shown in Table one. The basic description module has a total of thirteen core elements, namely name, type, document, condition state, inscription, language, dimension, material, symbol object, feature object, identifier, right and grade symbols, describing the intuitive basic external information of cultural heritage images. Since CIDOC CRM is based on information about historical facts, people, time, and places related to resource objects, and connects the information in the space-time dimension according to the attribute relationship, the basis for constructing the event description module lies in the event elements. Therefore, this article separately lists the key elements of time, space, and characters, and sets six core elements, namely time, time title, place, place title, actor, and actor title. The event description module is composed of nine core elements, design or procedure, beginning of existence, collection, part addition, part removal, comment, move, destruction and influence. It is used to reveal the conventional events experienced during the period from the creation of the cultural heritage image to the current collection.

| Module       | Element | Element description | CIDOC CRM entity mapping | CIDOC CRM Property mapping |
|--------------|---------|---------------------|--------------------------|---------------------------|
| Basic description | name    | the names assigned to works | E35 Title | P102 is title of |
|               | type    | vocabulary for description and classification | E55 Type | P2 is type of |
| **document** | information that make propositions about reality | E31 Document | P90 is documented in |
| **condition state** | the prevailing physical condition | E3 Condition State | P44 condition of |
| **inscription** | Recognizable short texts attached to instances | E34 Inscription | P3 has note |
| **language** | the natural languages in the sense of concepts | E56 Language | P72 has language |
| **dimension** | quantifiable properties that can be measured | E54 Dimension | P90 has value |
| **material** | the concepts of materials | E57 Material | P45 is incorporated in |
| **symbol object** | identifiable symbols and aggregation of symbols | E90 Symbolic Object | P106 is composed of |
| **physical feature** | identifiable features | E26 Physical Feature | P56 bears feature |
| **identifier** | strings or codes that make it uniquely and permanently | E42 Identifier | P48 is preferred identifier of |
| **right** | legal privileges, including reproduction and property rights | E30 Right | P105 right held by |
| **grade** | Protection level | E41 Grade | P141 graded |

**Event element**

| time | time range of the event | E2 Temporal Entity | - |
| time title | abstract temporal extents | E49 Time Appellation | - |
| place | the spatial extent of the event | E53 Place | - |
| Place title | specific common forms of place designation | E44 Place Appellation | - |
| actor | event performer | E39 Actor | - |
| actor title | names and symbols used for identification | E82 Actor Appellation | - |

**Event description**

| design or procedure | action plan for purpose | E29 Design or Procedure | P69 is associated with |
| beginning of existence | the object is created | E63 Beginning of Existence | P92 brought into existence |
| collection | collected by one or more character | E78 Collection | P109 current or former curator |
| part addition | be increased, enlarged or augmented by the addition of a part | E79 Part Addition | P111 was added by |
| part removal | be decreased by the removal of a | E80 Part Removal | P112 was diminished by |
2.2 Frame construction of cultural heritage images

As the base of the semantic framework, CIDOC CRM must realize the semantic integration of information, and must first correspond the existing explanatory information to the CIDOC CRM knowledge ontology. This article investigates the selection of representative museums and examines the structure of descriptive basic information on cultural heritage images. It is found that most of the descriptive basic information of museums are mostly presented in paragraph text descriptions, which still need to rely on manual methods. The conversion corresponds to CIDOC CRM. It requires high cognitive level and heavy workload. In his article, Smith B gave an example of the process in which CIDOC CRM knowledge ontology plays a role in the integration of descriptive basic information. Such as when we describe a person, a descriptive message "mother" appears. Semantically, this means that there is a mother-child relationship between the two. Corresponding to the CIDOC CRM, the path is two Actors, due to the Birth Event, and the attribute relationship is connected. This incident occurred at a specific time E2 Temporal Entity and location Place. [14]

We use this logic to convert the descriptive basic information into the corresponding path of CIDOC CRM, as shown in Table 3. According to the corresponding path in the table, this article further shows how the information is converted to the semantic architecture of CIDOC CRM. As shown in Figure 1, you can intuitively feel the relationship between descriptive basic information and events, and the descriptive basic information is expressed as an orderly semantic association of CIDOC CRM. It is worth noting that, as the core element specifications sorted out above show, the action of E39 Actor itself is in a certain time and space range, and the experienced events exist before and after themselves in their corresponding paths. The order relationship is its inherent nature, so no further cumbersome description will be given in this table.

Table 2. Descriptive basic information converted to CIDOC CRM corresponding path.

| name          | E39 Actor→P92 brought into existence→E63 Beginning of Existence→(P25 moved by→E9 Move→P109 current or former curator→E78 Collection→) P102 is title of→E35 Title |
|---------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| type          | E39 Actor→P69 is associated with→E29 Design or Procedure→(P15 was influenced by→E83 Type Creation→P67 is referred to by→E89 Propositional Object→) P2 is type of→E55 Type                                                                 |
| document      | E39 Actor→P69 is associated with→E29 Design or Procedure→P67 is referred to by→E89 Propositional Object→P90 is documented in→E31 Document                                                                 |
| condition state | E39 Actor→P92 brought into existence→E63 Beginning of Existence→(→P25 moved by→E9 Move→P109 current or former curator→E78 Collection→) P44 condition of→E3 Condition State |
inscription  E39 Actor→P92 brought into existence→E63 Beginning of Existence→(P111 was added by/ P112 was diminished by→E79 Part Addition/ E80 Part Removal→) P3 has note→E34 Inscription

language E39 Actor→P92 brought into existence→E63 Beginning of Existence→(P111 was added by/ P112 was diminished by→E79 Part Addition/ E80 Part Removal→) P3 has note→E34 Inscription→P72 has language→E56 Language

dimension E39 Actor→P92 brought into existence→E63 Beginning of Existence→(→P25 moved by→E9 Move→P109 current or former curator→E78 Collection→) P90 has value→E54 Dimension

material E39 Actor→P69 is associated with→E29 Design or Procedure→P45 is incorporated in→E57 Material

symbol object E39 Actor→P69 is associated with→E29 Design or Procedure→P106 is composed of→E90 Symbolic Object

physical feature E39 Actor→P69 is associated with→E29 Design or Procedure→P56 bears feature→E26 Physical Feature

identifier E39 Actor→P25 moved by→E9 Move→P109 current or former curator→E78 Collection→P48 is preferred identifier of→E42 Identifier

right E39 Actor→P25 moved by→E9 Move→P109 current or former curator→E78 Collection→P105 right held by→E30 Right

grade E39 Actor→P25 moved by→E9 Move→P109 current or former curator→E78 Collection→P141 graded→E41 Grade

Figure 1. Descriptive basic information converted to the semantic architecture of CIDOC CRM.

References

1. CIDOC CRM. Use Cases, 2016, [online] Available: http://www.cidoc-crm.org/Version/version-6.2.6

2. Gregor Strle, Matija Marolt, Slovenia. Conceptualizing the Ethnomuse: Application of CIDOC CRM and FRBR, 2007, [online] Available: http://www.edd.uio.no/cidoc/abstracts/Strle_Gregor.pdf.
3. CLAROS: The world of art on the semantic web, 2017, [online] Available: http://explore.clarosnet.org.

4. National Gallery API, 2012, [online] Available: http://research.ng-london.org.uk/wiki/index.php/National_Gallery_API

5. Eetu Mäkelä, Juha Törnroos, Thea Lindquist, Eero Hyvönen. WW1LOD: an application of CIDOC-CRM to World War 1 linked data [J]. International Journal on Digital Libraries, pp. 333-343, 2017.

6. Smith, B. The Blackwell Guide to the Philosophy of Computing and Information. [R], Malden, MA: Blackwell Publishing, pp. 155-166, 2003.