Cultural Mapping of Villages in India Using CIDOC-CRM

Toshant Sharma and Navjyoti Singh

International Institute of Information Technology Hyderabad,India
sharma.toshant@research.iiiit.ac.in,
navjyoti@iiit.ac.in

Abstract. The domain of cultural mapping deals with capturing and representing knowledge regarding diverse tangible and intangible aspects of cultural heritage, which involve hybrid data sets from incompatible sources with different terminologies and representational schemas. The complex schema integration problem in this domain can be addressed efficiently by use of formal ontologies which provide universal schema to represent data from heterogeneous sources. The paper describes an abstraction over the CIDOC CRM (ISO 21127:2014) ontology for the purpose of the cultural mapping of villages of India and discuss with case studies how cultural information in context of Indian villages can be captured, preserved and new relationships between cultural entities visualized using this ontology.

Keywords: Ontology, Cultural Mapping, RDF Triples, CIDOC-CRM

1 Introduction

In the domain of cultural mapping [14] the captured information, scattered over a large set of in-coherent sources (government publications, databases, web sites, surveys etc.) in different formats and can be accessed only in a fragmented manner as the interconnections between data sets are scanty. In the context of Indian Cultural Heritage, heterogeneity is one of the main characteristics and ontologies offer solutions to the semantic heterogeneity problem [1]. Using ontology in integration architecture as a global schema to which metadata from different sources can be mapped provides a new approach by which diverse cultural network semantics can be described independently of cultural data source characteristics and in turn provide a conceptual schema allowing cultural data exchange among heterogeneous information systems and users. With 68.84% people residing in villages having long standing oral-traditions and non-existence of external retention systems through ages, there is need to externalize both tangible and intangible aspects of cultural heritage of different communities. Therefore, in order to 'map' Indian culture it becomes imperative to capture the cultural dynamics operating at the granularity of constituent villages.

We propose an abstraction over CIDOC CRM [9,11,15] ontology for cultural mapping of the Indian villages. CIDOC CRM is a core generic ontology and ISO standard (ISO 21127) for the semantic integration of cultural information with library, archive and other information [9] but lacks domain specific specializations to better capture, represent different cultural contexts. Acting as the underlying framework, time and again both abstractions and extensions are made to the model [6,8] for knowledge representation in different contexts and sub-domains of cultural heritage. The paper makes the following contributions

- describing an abstraction over CIDOC CRM for specialized purpose of cultural mapping in the context of Indian villages
- providing examples of how information can be modeled with this model
- describing an implementation of the proposed ontology

2 Related Work

Doer [2] gives an overview of domain of cultural heritage, schema integration problem in cultural heritage, characteristics of cultural ontologies and reviews CIDOC-CRM, FRBRoo and other core ontologies in the domain. Stephan [15], Crofts et al.[9] describe the definition, structure and pertinence of CIDOC-CRM. Lieto et al.[12] propose to adopt and extend the CIDOC-CRM to describe some of the traits of the contemporary artworks by discussing three examples of an art installation in a museum, a live performance and, a physical modern artwork. The qualitative aspect of addition of 'score' as subclass of 'Design and procedure' is discussed and the class is proposed as an extension. Tan et al.[3] give representation of "Funeral Dance" based on CIDOC-CRM, wherein a domain ontology is created for the domain of funeral dance by modeling the domain in Protege using CIDOC-CRM as the base model. Hu et al.[4] apply the CIDOC Conceptual Reference Model to the construction of the entities and properties of the knowledge ontology of the Pang Wang Festival.

An extension of CIDOC-CRM, for the Invisibilia project, is given by Ng et al.[5] where the notion of performance is proposed for the integration into the ontology. Emphasizing that more specialized vocabularies are necessary, classes of performance activity, performance procedure, instrument, equipment and digital object are proposed for digital preservation purpose of interactive multimedia performances. Theodoridou et al.[6] propose an extension of CIDOC-CRM aimed at modeling the notion of reliability and provenance. By introducing four new specializations of material and immaterial items and six new specializations of events, the proposed extension is used to capture the modeling and the query requirements regarding the provenance of digital objects.

Hernandez et al.[13] describe the initiative for the construction of an ontology that compiles the knowledge around the cultural heritage related to region of Cantabria in Spain by converting unstructured high-quality information prepared by subject matter experts to semantic form employing RDF schema of existing ontologies.

3 Ontology Construction Methodology

The methodology involves ontology modularization of the domain and then iterative modeling through example representation. We first modularize the large domain of cultural mapping in context of Indian villages into small ontologies. To create modularized ontologies we use a normalization approach as proposed by Alan Rector [7]. In this approach each ontology is decomposed into disjoint skeleton taxonomies, which are restricted to simple graphs. These skeleton taxonomies are then integrated with CIDOC CRM, which acts as a common base to which different small branch ontologies are added. The changes to the model generated by mapping and adding skeleton ontologies to CIDOC are made using iterative example modeling in terms of the intermediate ontologies. Continuous ontology evolution form the basis for our iterative ontology modeling approach. Figure 1 depicts the steps carried out, ontology is available publicly at [19].

- Update skeleton Ontology. Based on the example representation in ontology update the skeleton ontology adding classes and properties minimally as necessary and possibly restructuring the class hierarchy to better reflect the knowledge.
- Integration with CIDOC-CRM. The next step is to integrate the smaller skeleton ontologies with the common base ontology i.e. CIDOC. The consistency and correctness of the model is to ensured i.e. disjoint classes can't have common individuals, the properties must exist between the entities which lie in their defined domains and ranges.
- Verification. The final step is verification that the applied changes actually implement what was desired. If the resulting ontology is not satisfactory or contains inconsistencies new changes are implemented into the skeleton ontologies and the process is repeated.

4 Extension Description

4.1 Performative Culture

The class performative culture is proposed as an abstraction to encapture the intangible aspects of culture, having spactio-temporal projections that can be expressed through medium of live performance for example music, dance, theatre. As an specialization (subclass) of the class propositional object (E89), this class is limited to cultural entities that have a changing extension in time or in other words have a performance stretched in time associated with them for example individuals belonging to music, theatre, rituals and dance. Figure 2 shows the schema for the class. In order to represent location associated with propositional object using CIDOC-CRM, first an event associated with an individual of that propositional object class is represented and that is linked to a location. Another way of representation is using P67 (refers to) but this does not capture the semantics fully for a statement of form "individual X of performative culture is local to locations l1 and l2". We capture this process through association of location with individual of class performance so as to prevent the loss of information.

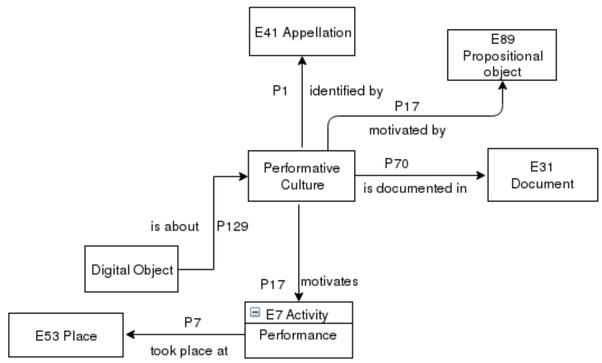

Fig. 1. Performative Culture

4.2 Performance

A relevant extension of cidoc-crm for capturing the performative aspects of the intangible cultural heritage is given by Ng et al.[5] where the notion of performance is proposed for the preservation of multimedia performances. As Lieto et al.[12] point out, a limitation of this contribution, however, is given by the fact that this extension is limited to the exclusive introduction of the class Performance in the ontology, without any specification about the relation that such type of class (and therefore the members of such class) entertain with the other ontological components (actors, roles, processes etc).

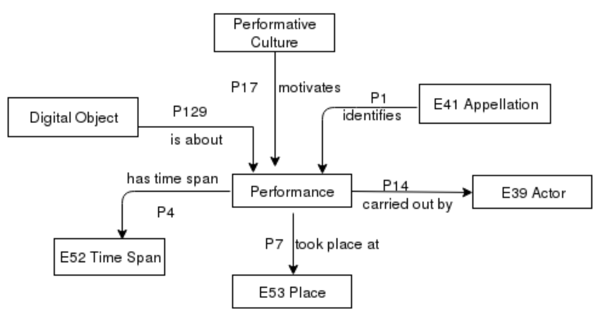

Fig. 2. Performance Schema

The performance class proposed by Ng et al. primarily aims at preservation rather than description of the interactive multimedia performance and lacks relationships to capture the cultural dynamics like motivation and context for performance, historical and literature references in performance active in the performance. We propose class Performance as a temporal construct and a specialization (subclass) of class E7 (activity), the schema is shown in figure 3. Here an object of performative culture motivates(P17) a performance which has associated actors, location and appellation. The performance can also be represented by digital object(P129).

4.3 Tangible Cultural Object

In order to capture the legacy of physical artifacts and aesthetic, historic, or social value associated with them a specialization of E22 (man made object) is proposed. This class comprises tangible cultural entities i.e. physical objects with real boundaries purposely created by human endeavor. The individuals of this class include cultural objects both moveable and stationary like paintings, cuisine, clothing, monuments etc. The individuals of class tangible cultural object(TCO) have associated individuals of E39 actors in the contexts of both producer of the cultural object and owner. P105 directly represents the actor holding the rights to the TCO. Property P108 connects TCO to the event production (subclass of E5 Event), P14 further describes the actor who carried out the production event of the TCO. The production event also has an associated time span with it.

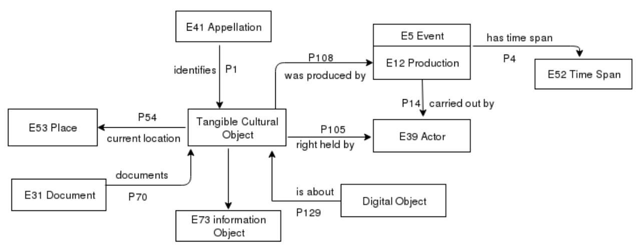

Fig. 3. Tangible Cultural Object

4.4 Survey

The class survey, encompasses not only the statistical studies of the size, structure, distribution of populations along with spatial or temporal changes in them but also represents the general methodologies to collect data about different tangible and intangible cultural aspects of an individual village. The class survey is described as a subclass of Design and Procedure. The property P106 links the individuals of class survey to its constituent documents(E31), design and procedures(E29) as individuals of class information objects(E73) which are sub-classes of symbolic object(E90).

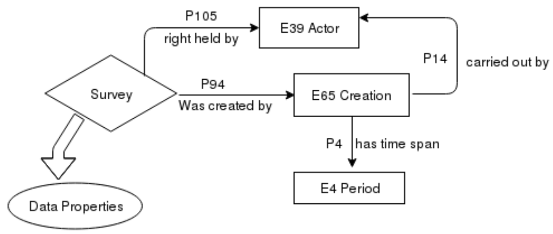

Fig. 4. Example of a Survey Object

4.5 Digital Object

We borrow the specialization 'digital object' of class E73 from Theodoridou et al [6].

4.6 Location Hierarchy

Using P89(contains) we define a hierarchy of subclasses of E53(Place) with E53 containing district containing sub-district containing villages . See[19] for details

5 Case Studies

5.1 Importing Census Data

Here we show a case study of importing and representing the demographic census data (Census of India 2001) using the proposed extension. The census of India 2001 contains 55 social, cultural and economic attributes in a relational table all having numerical values for around 628,000 villages of India. Figure 6 shows the schema used to represent the data using the model. To import the census data available in relational form into the ontology, we use Domain Semantics-driven Mapping Generation methodology [10] wherein the mappings are generated from RDB to RDF by incorporating domain semantics that are often implicit or not captured at all in the RDB schema. The census is defined as an individual of class survey, linked by P94 to a creation activity that has associated time span and individual. A set of sub properties of Data property PDN1 is defined to add census data to individuals of class village. Treating a specific village as an individual the attributes from census data are mapped to data properties of the entity village (sub properties of data property PDN1). Each attribute (total 55 of such attributes) defined in the census relational table having an associated numeric value is mapped correspondingly to a data property. The attribute values are stored as the literal values of the data property.

The schema is defined in Jena [16], to import and store the census data into the model we use RDF(resource description framework). The census data is first pre-processed from CSV (comma separated values) text to JSON format. Using this format the villages are first organized into a hierarchical tree like

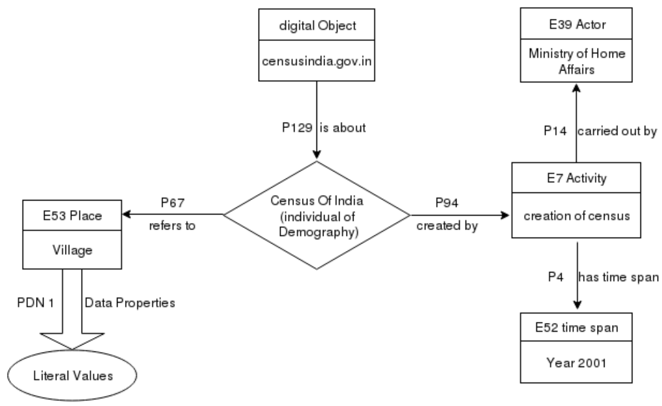

Fig. 5. Census Data Representation

structure with state, district, sub-district and village at different levels. While transforming into RDF form this hierarchy is maintained by use of the property P89(contains). Treating a village as subject of a resource the attributes from census data are mapped to properties in RDF format where each attribute defined in the table is mapped correspondingly to a data property. The attribute values are stored as the literal values of the property. The total number of triples added for the state of Punjab having 25,678 villages are 1,437,968. Refer [19] for detailed explanation.

5.2 Case Study of Ramlila Theatre

Ramlila is a theatrical performance based on the epic Ramayana. Ramlila consists of a series of scenes that include song, narration, recital and dialogue. A semantic graph is constructed for the folk drama of Ramlila wherein the ontological model captures the relationships between different entities and describe knowledge regarding Ramlila. Figure 7 describes concepts of the theatrical form of ramlila captured using the proposed model.

5.3 Cultural Semantic Graph of village Preet Nagar

Using the model we attempt to capture the cultural heritage of Preet Nagar village existing in the Indian state of Punjab. For that purpose, the cultural data is imported through means of extensive surveying. Figure 8 schema shows how a single example each from domains of surveys, tangible cultural object, performative culture are related to the village. The survey questionnaire consists of four tables to be filled up by takers about number and type of architectural locations, ethnographic constitution of village, number and type of artists existing in village and number and type of shops in village. The structure of the data collection tables had to be designed keeping in mind that the vast amount of variance in domain listed has to be captured. For each of the survey table an individual of class survey is declared and the attributes from the relational

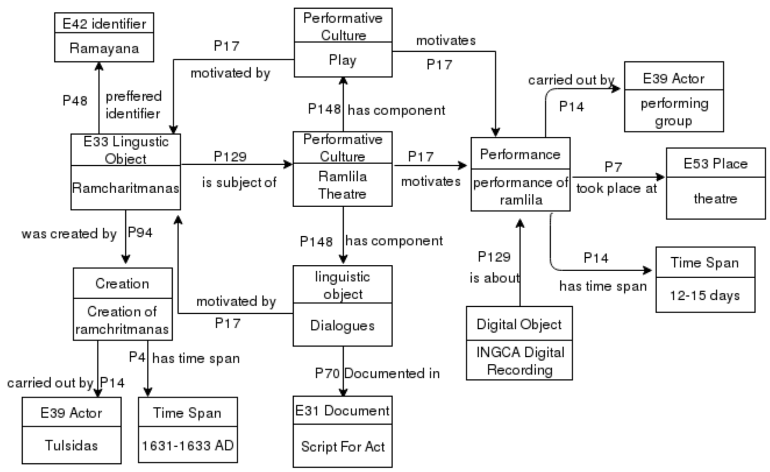

Fig. 6. Ramlila Theatre

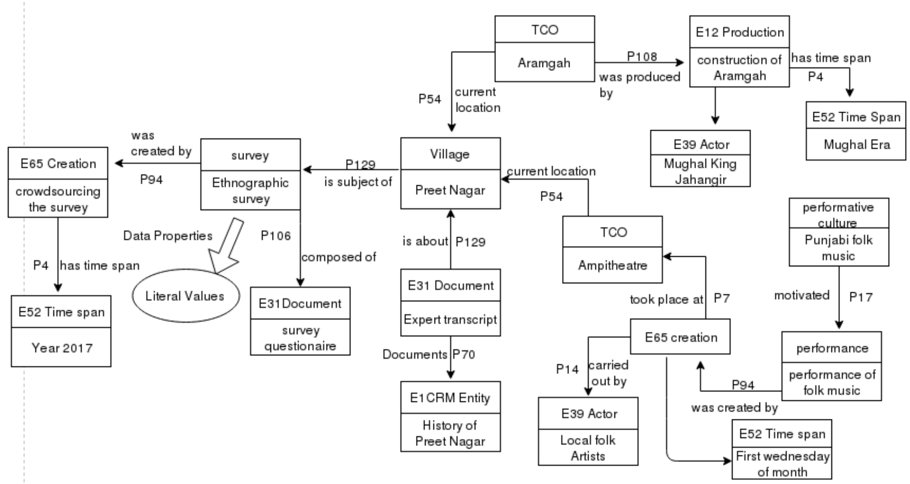

Fig. 7. Subsection of Cultural Graph

table are defined as the data type properties of the village. The relation of the village and the survey is represented through property P129 (is_subject_of). An example of this methodology would be, in ethnographic survey list all religions and castes (jatis) and then ask for numeric values of each existing in the village. However there are more than 3000 castes (jatis) in India [18], asking participants taking surveys to fill up huge tables is not practical, rather it is better to ask for the castes existing in the village only and construct a relational table for capturing that. The ethnographic survey table for a village had only three columns (religion, caste, number of households), the participants list the data accordingly.

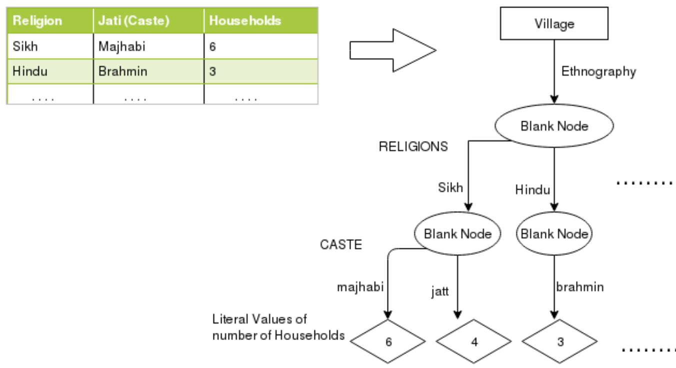

Fig. 8. Conversion Methodology

The tables are converted into RDF triples and imported into the ontology. To convert a table into equivalent RDF form we use blank nodes. A blank node (also called bnode) is a node in an RDF graph representing a resource for which a URI or literal is not given and can be used to describe multi-component resources [16]. The attributes stored in the table are mapped to predicates of blank nodes in RDF ending the chain at a literal value. The path from the village to the literal value describes what the literal value is representing. The process is straightforward and is illustrated in figure 9 which shows a section of the table of ethnographic decomposition and its corresponding model. The URI associated with village is appended by the predicates to generate URIs depicting different ethnographic groups in the village. Using the same methodology for conversion from survey tables to RDF form, we generate RDF triples for tables of type and number of artists, type and number of architectural sites, number and type of shops in the village Preet Nagar.

6 Conclusions And Future Work

One of the advantages of using RDF that has been stressed throughout is the way two separate graphs are automatically linked if they share as much as a single common resource node, which can provide insights into how different agents operating in a diverse and vivid cultures interact and manifest themselves. Relying upon expression of history as a series of events with associated time periods and above mentioned process of extracting relations from descriptive texts, a methodology and a use case for constructing a semantic graph of the history of villages are implemented as a prototype. This is yet to be developed further and not hence discussed here. Future works include improved cultural data collection and curation of collected content through the medium of a crowd sourcing web application focused on both surveys and textual descriptions. Further testing, specializations and example modeling, case studies using the model are open ended processes for the future.

References

1. Wache, Holger, et al *Ontology-based integration of information-a survey of existing approaches.* IJCAI-01 workshop: ontologies and information sharing. Vol. 2001.
2. Doerr, Martin. *Ontologies for cultural heritage.* Handbook on Ontologies. Springer Berlin Heidelberg, 2009. 463-486.
3. Tan, Guoxin, Chuanming Sun, and Zheng Zhong. *Knowledge representation of" funeral dance" based on CIDOC CRM* Knowledge Acquisition and Modeling, 2009. KAM'09. Second International Symposium on. Vol. 1. IEEE, 2009.
4. Hu, Jun, Yingchun Lv, and Mu Zhang. *The Ontology Design of Intangible Cultural Heritage Based on CIDOC CRM* International Journal of u-and e-Service, Science and Technology 7.1 (2014): 261-274.
5. Ng, Kia, et al. *Preservation of interactive multimedia performances.* International Journal of Metadata, Semantics and Ontologies 3.3 (2008): 183-196.
6. Theodoridou, Maria, et al. *Modeling and querying provenance by extending CIDOC CRM.* Distributed and Parallel Databases 27.2 (2010): 169-210.
7. Alan L. Rector *Modularisation of domain ontologies implemented in description logics and related formalisms including owl.* Proceedings of Knowledge Capture 2003, pages 121128. ACM, 2003.
8. ARIADNE Infrastructure. *Ariadne Reference Model* http://www.ariadne-infrastructure.eu/Resources/Ariadne-Reference-Model
9. Nick Crofts et al. *Definition of the CIDOC Conceptual Reference Model* http://cidoc.ics.forth.gr/official_release_cidoc.html
10. Sahoo, Satya S., et al. *A survey of current approaches for mapping of relational databases to RDF.* W3C RDB2RDF Incubator Group Report 1 (2009): 113-130.
11. CIDOC Conceptual Reference Model - A New Standard for Knowledge Sharing *26th International Conference on Conceptual Modeling (ER 2007)* Australian Computer Society, Inc. Vol. 83, 2007, pp.51-56.
12. Lieto, Antonio, Rossana Damiano, and Vanessa Michielon. *Conceptual Models for Intangible Art* https://mimesis.revues.org/690
13. Francisca Hernandez, Luis Rodrigo, Jesus Contreras, Francesco Carbone *Building a Cultural Heritage Ontlogy for Cantabria* 2008 Annual Conference of CIDOC
14. Sarah Court and Gamini Wijesuriya *People-Centred Approaches to the Conservation of Cultural Heritage: Living Heritage* Annexture, ICCROM, 2015 URL-http://personal.sirma.bg/vladimir/crm-tutorial/
15. Stephan Stead *CIDOC Conceptual Reference Model Special Group* ICS-FORTH, November 2008. URL-http://personal.sirma.bg/vladimir/crm-tutorial/
16. An Introduction to Jena RDF API *http://jena.sourceforge.net/.* URL http://jena.sourceforge.net/tutorial/RDF API/index.html
17. Tim Berners-Lee. *Relational Databases on the Semantic Web. Internet note.* http://www.w3.org/DesignIssues/RDB-RDF.html september 1998.
18. Jati-Caste System in India. URL http://asiasociety.org/education/jati-caste-system-india
19. https://github.com/toshan16/Semantic_Mapping