CURIOCITY: A Cultural Heritage Ontology for Urban Tourism

Alexander Pinto-De la Gala a,*, Yudith Cardinale a,b Irvin Dongo a,c and Regina Ticona-Herrera a,c

a Computer Science Department, Universidad Católica San Pablo, 04001 Arequipa, Perú
E-mails: alexander.pinto@ucsp.edu.pe, ifdongo@ucsp.edu.pe, rtherrera@ucsp.edu.pe
b Departamento de Computación y Tecnología de la Información, Universidad Simón Bolívar, 1080 Caracas, Venezuela
E-mail: ycardinale@usb.ve
c Univ. Bordeaux, ESTIA INSTITUTE OF TECHNOLOGY, 642010 Bidart, France

Editors: First Editor, University or Company name, Country; Second Editor, University or Company name, Country
Solicited reviews: First Solicited Reviewer, University or Company name, Country; Second Solicited Reviewer, University or Company name, Country
Open reviews: First Open Reviewer, University or Company name, Country; Second Open Reviewer, University or Company name, Country

Abstract. Urban tourism information available on Internet has been of enormous relevance to motivate the tourism in many countries. There exist many applications focused on promoting and preserving the cultural heritage, through urban tourism, which in turn demand a well-defined and standard model for representing the whole knowledge of this domain, thus ensuring interoperable and flexible applications. Current studies propose the use of ontologies to formally model such knowledge. Nonetheless, most of them only represent partial knowledge of cultural heritage or are restrictive to an indoor perspective (i.e., museum ontologies). In this context, we propose the ontology CURIOCITY (Cultural Heritage for Urban Tourism in Indoor/Outdoor environments of the CITY), to represent the cultural heritage knowledge based on UNESCO’s definitions. CURIOCITY ontology has a three-level architecture (Upper, Middle, and Lower ontologies) in accordance with a purpose of modularity and levels of specificity. In this paper, we describe in detail all modules of CURIOCITY ontology and perform a comparative evaluation with state-of-the-art ontologies. Additionally, to demonstrate the suitability of CURIOCITY ontology, we have developed a first version of a framework, which allows its automatic population, by transforming a museum data repository (in CSV format) into RDF triples of CURIOCITY ontology, to automatically populate the CURIOCITY repository. The framework also provides facilities to develop a set of tourism applications and services, following the UNESCO’s definitions.

Keywords: Ontology, Cultural Heritage, Urban Tourism, Automatic Population, Ontology Evaluation

1. Introduction

Urban tourism is one of the promising areas for the development of social and economic activities in urban environments [1–3]. Many cities in the world are historic city centers and constitute one of the most important elements of the cultural heritage. They are places that attract many visitors due to their relevance in terms of heritage. Actually, although cities are not necessarily historic centers, in general, part of a tourist trip itinerary includes activities related to different places of interest that can range from museums and parks to even medical centers [4]. Thus, urban tourism has become one of the core part of cultural heritage in many countries.

Cultural heritage includes representations of the value systems, beliefs, traditions, and lifestyles of communities; it expresses stories through the time and space, about a society and its culture. Urban tourism is a way to transmit and learn about cultural heritage
of countries [5]. Moreover, cultural heritage is supported by communication and information technologies for its preservation [6–9], being accessible to a wider public causing knowledge dissemination, covering more spaces, going beyond countries borders (e.g., web pages, wiki pages, virtual spaces, on-line information centres) [10, 11], and supporting tourist planning (e.g., e-tourism, recommendation systems) [12–17].

The huge amount of data that can be managed in such information tools and services, demands the use of more complex knowledge. Resources and objects that describe a specific heritage (e.g., a collection, a museum, a historical site) are related to their designer, creator, or owner, related to its convenience and function in a given time, and also can be related and extended to other knowledge organizations, inside and outside of the cultural heritage domain. All these relations generate an even more complex network of knowledge. In this sense, it is evident the necessity of a well-defined and standard model for representing the knowledge managed by these on-line services. Semantic Web seems to be a clear solution, from which we can take its organizational and relational capacity. It proposes concepts and tools such as ontologies, with the aim of creating a consensus of standard definitions and structures, in order to describe resources and define their relationships. Ontologies formalize complex knowledge networks aiming to facilitate the process of sharing and reusing information; they provide semantics of information sources that can be processed by computers and be communicated among different agents, both human and machines. Thus, an ontology is a formal way of capturing valid knowledge from a particular domain. Hence, this formal modeling of knowledge in a specific domain allows the development of interoperable services, which can be easily adapted to the particular requirements of different users.

In the context of cultural heritage, some studies have proposed ontologies to represent its partial knowledge. Thus, there exist ontologies to represent museums [18–23], improve exhibitions [24, 25], represent touristic points of interest [26–29], or represent curatorial narrative [30]. In particular, museums stand out as a knowledge source and are the main points of interest within urban tourist centers. However, the proposals addressed by authors to represent knowledge related to museums, vary according to characteristics of their research and particular interests. For example, they focus on aiming to deal with the current data and resource heterogeneity [19], allowing collaboration among a group of museums [22], guiding a visit according to a profile of interests [25], or providing the foundation for a virtual museum implementation [31]. This diversity of purposes means that ontology design, although it can start from some common or standard basis, must be adapted to optimally capture particular characteristics, which can be influenced by social aspects, by project’s technological requirements (e.g., distributed and distant data sources), by end users’ adaptability requirements (e.g., web page visitors or robot guide systems), among many other variables.

Moreover, most experiences about knowledge modeling of cultural heritage in museums, are usually found within an indoor perspective; however, cultural heritage concept is dynamic, thus it encompasses other concepts with not only cultural value, but also aesthetic, academic, economic, and recreational values, which are relevant to a society. In addition to this, urban tourism perspective, points out visitor’s interests are broad within city environments, which conform urban tourist centers with their own cultural heritage, with particular features and relationships, and therefore, they require a different knowledge organization.

To overcome these limitations, in a previous work we have proposed the ontology CURIOCITY (Cultural Heritage for Urban Tourism in Indoor/Outdoor environments of the CITY) [32], to represent the cultural heritage knowledge based on UNESCO’s definitions. In this work, we describe in detail its three-level architecture (Upper, Middle, and Lower ontologies) in accordance with a purpose of modularity and levels of specificity. CURIOCITY ontology is based on several standards of the cultural heritage domain, besides proposing the extension and integration with other domains in a way that allows the cultural heritage concept to be broadened within an urban tourist context. Based on a methodological process [33], we perform an evaluation taking into account our categorization of the cultural heritage knowledge [32], and compared it with state-of-the-art ontologies. Additionally, to demonstrate the utility and suitability of CURIOCITY ontology, we have developed a first version of CURIOCITY framework, which allows the ontology automatic population and provides facilities to develop a set of tourism applications and services, following the UNESCO’s definitions. CURIOCITY framework consists of the CURIOCITY repository (at the Repository layer) and a parser (at the Data Processing layer) to map and convert a museum repository (a CSV database) to RDF triples of CURIOCITY ontology. Hence, it is possible to automatically populate the
CURIOCITY repository, which constitutes the basis for the development of flexible, intelligent, and interoperable services and applications in the tourism domain.

The remainder of this paper is organized as follows. Section 2 presents a brief review about cultural heritage concepts. In Section 3, we discuss underlying standards and related work. Section 4 introduces CURIOCITY ontology. Section 5 presents and discusses the results of the evaluation of CURIOCITY ontology. Section 6 presents the design and first prototype of CURIOCITY framework applied in a case study. Section 7 discuss our conclusions and future work.

2. UNESCO’s definition of Cultural Heritage

UNESCO defines heritage as "our legacy from the past, what we live with today, and what we pass on to future generations". The concept of heritage is in continuous evolution; thus, richness and complexity of cultural heritage are evidenced by the semantic evolution of this concept. UNESCO classifies Cultural Heritage into two categories, Tangible and Intangible, and besides that, defines Natural Heritage and Armed Heritage categories, as follows [34]:

– Cultural Heritage:
  * Tangible Cultural Heritage:
    * Movable Cultural Heritage: paintings, sculptures, coins, manuscripts.
    * Immovable Cultural Heritage: monuments, archaeological sites.
    * Underwater Cultural Heritage: shipwrecks, underwater ruins and cities.
  * Intangible Cultural Heritage: oral traditions, performing arts, rituals.
– Natural Heritage: natural sites with cultural aspects.
– Armed Heritage: heritage in the event of armed conflict.

Loulanski [35] considers a previous classification, which includes other concepts like Handicrafts, Documentary, Digital and Cinematographic Heritage, Languages, Festive Events, Music and Songs, Traditional Medicine, Literature, Culinary Traditions, and Traditional Sports and Games. Loulanski also defines a spectrum of cultural heritage values in detail, such as:

– Cultural values, which consider that appreciation and conservation of heritage generate distinctiveness feelings at local, regional, and national levels.
– Educational and Academic values, that provide a way to understand the past of our own culture and with this knowledge to plan our future.
– Economic values, to assure that historical environments mean a contribution to economic development through tourism, and to represent how these vales create a better environment for community development.
– Resource values, that consider that long life buildings mean better use of resources and energy.
– Recreational values, which represent historical environments providing recreation and enjoyment.
– Aesthetic values, that reinforce the idea that historic buildings contribute to the aesthetic quality of urban and rural landscapes.

Thus, the value or significance of cultural heritage is recognized beyond the cultural area, even within economic, social, political, and scientific areas.

3. Related Work

The heterogeneity of concepts in cultural heritage domain has fostered the proliferation of different ontologies, particularly, in the context of tourism, to represent points of interest (POI) and museum knowledge. In this section, we survey the most recent and representative ontologies for both, POI and museums representations.

3.1. POI ontologies

Regarding ontologies for POI, several works have been proposed in the literature for the context of tourism.

The European Project “Harmonise” [26] proposes various technologies to solve the interoperability problem in the tourist domain. To do so, they propose an ontology, called IMHO (Interoperability Minimum Harmonisation Ontology), that considers basic concepts used for representing the content of information exchanges in tourism transactions [27]. Another ontology proposed under an EU funded project, is Qall-Me [28]. It is a domain-specific ontology for question answering in the domain of tourism. The tourism destinations, tourism sites, tourism events
and transportation are covered by this ontology. Qall-
Me is aligned with two upper ontologies, Word-
Net\(^1\) and SUMO\(^2\). In [29], an extension of the Qall-
Me ontology is proposed, by adding a new class
SiteCategory and three object properties for rela-
tionships, namely stronglyRelated, related, and
weaklyRelated. Using these properties, several
levels of relationships among sites can be ex-
pressed. For example, a museum can be strongly re-
lated to a tourist office, while it is weakly related to an
exhibition place.

The World Tourism Organization (UNWTO)\(^3\) was
created in 1975 for promoting the tourism, linked to
the United Nations a year later. As an effort of 20
years to standardize and utilize tourism terminolo-
gies, UNWTO proposed a multi-language thesaurus
(English, French, and Spanish) of the tourism domain
in 2001. Terms very specific to tourism were also ex-
tensively defined for a better interoperability. Based
on these concepts, Mondeca Tourism Ontology is pro-
posed. Tourism object profiling, tourism and cultural
objects, tourism packages, and tourism multimedia
content are described [37, 38]. HiTouch Ontology [37],
created under the IST/CRAFT European program, and
OnTour Ontology [28], developed by e-Tourism Work-
ning Group at Digital Enterprise Research Institute,
both also use the concepts of UNWTO. HiTouch rep-
resents additionally tourism products and customers’
tourism expectations, while OnTour adds descriptions
of leisure activities and geographic data.

DataTourisme\(^4\) ontology was created in 2017 by the
company PERFECT MEMORY in a french project.
The aim of this ontology is to centralize and publish as
Linked Open Data (LOD) travel information produced
by different tourist information systems in France. Ad-
ditionally, this ontology is connected to different exist-
ing ontologies as FOAF\(^5\), Schema\(^6\), GoodRelations\(^7\),
Dublin core\(^8\) to do not duplicate areas that are de-
scribed already and in this way, to facilitate links with
these open databases.

Local tourism ontologies for Australia [39], Thai-
lan [40, 41], Iran [42], and others, have also been pro-
mised to mainly develop applications such as recom-
modation systems [38, 42] and tourism planning [43].

3.2. Museum ontologies

Concerning museum knowledge, its variety and het-
erogeneity led to the establishment of various stan-
dards with the purpose of normalizing and creat-
ing bases for ontology development for particular purposes. Some popular standards in cultural her-
itage domain are, for instance, the thesaurus ICON-
CLASS [44], the paid service Resource Description and
Access (RDA) [45], the ISO Standard CIDOC CRM [46],
and the massive thesaurus and ontology service Finto [47].

ICONCLASS [44] is a classification system of ob-
ject definitions, people, events, and abstract ideas, ar-
ranged hierarchically, developed by the Netherlands
Institute for Art History, that can be used for indexing,
cataloging, and description of pictorial artworks, such
as paintings, reproductions, photographs. RDA [45] is
a set of elements, guidelines, and instructions for cre-
ation of metadata about library resources and cultural
heritage, according to international models focused on
linked data applications. RDA was created as a re-
placement for the Anglo American Cataloging Rules,
and has a most widespread application in the library
domain. RDA has a subscription cost.

The Committee for Documentation of the Interna-
tional Council of Museums (CIDOC) proposes the
Conceptual Reference Model (CIDOC CRM) [46],
which since December 2006, is recognized as an offi-
cial ISO 21127: 2014 standard. CIDOC CRM provides
definitions, structures, basic classes, and relationships
for describing cultural heritage documentation for the
querying and exploration of such data. It has exten-
sions that allow adapting it to particular uses, e.g., CR-
Ontologies, a tool that describes open databases.

Finto [47] is defined as a Finnish service for publish-
ing and using vocabularies, ontologies, and classifica-
tions. Finto is sponsored by various Finland govern-
ment entities and is the successor of FinnOnto [49], an
ambitious project that is the basis of metadata, onto-
logies, and LOD throughout Finland. The FinnOnto’s vi-
1WordNet is a lexical database for the English language - https://www.w3.org/2006/03/wn/wn20/
2SUMO: The Suggested Upper Merged Ontology, created for search, interconnection, and communication on the Semantic
Web [36]
3World Tourism Organization (UNWTO) - https://www.e-
unwto.org/doi/abs/10.18111/9789284404551
4https://info.datatourisme.gouv.fr/
5http://www.foaf-project.org/
6https://schema.org/
7http://www.heppnetz.de/projects/goodrelations/primer/
8https://dublincore.org/
to interconnect public and private organisms for intelligent exchange of content. Finto brings together ontologies from different domains, including Arts and Culture, that considers ontologies for Museum Domain and Applied Arts (MAO9/TAO10, terminology of Folklore, Cultural Anthropology and Ethnology (KULO)11, Music (MUSO)12, Musical Performance (SEKO)13, and Photography (VALO)14. These ontologies are based on YSO15, a general concept ontology. YSO provides an extensive number of concepts mainly arranged in a hierarchical structure, thus it has the capability to encompass a wide range of environments. However its massive thesaurus nature and parent-child structure could be overwhelming.

Usually, from these vocabularies and representation proposals, several authors have proposed ontologies in accordance with their research objectives, extending or integrating them.

MUSEUM FINLAND project [50], is a proposal for semantic integration of museums in Finland, based on seven domain ontologies: Artifacts, Materials, Actors, Situations, Locations, Times, and Collections. MUSEUM FINLAND uses the Finnish cultural content thesaurus (Museoalan asiasanasto - MASA)16 to create MAO ontology (now part of Finto service, as viewed before).

Europeana Data Model (EDM) [51], has the aim to standardize the representation of cultural heritage objects from different domains such as libraries, museums, and audiovisual archives. It is not built on a particular standard, but adopts a wide range of these, such as CIDOC LIDO17 for museums, EAD18 for archives, and METS19 for digital libraries; with the intention of being a Semantic Web framework between different domains.

ArCo (Arquitecture of Knowledge) [52] is an Italian project with the purpose of building a network of aligned ontologies to represent cultural heritage data and publish the General Catalogue proposed by the Italian Ministry of Culture. Arco ontology version 0.5 consists of seven modules: (i) arco, is the root of the network, it imports the other six modules and models top level cultural heritage concepts; (ii) core, represents orthogonal concepts imported by the other modules; (iii) catalogue, models catalogue records; (iv) location, represents spatial and geometry information; (v) denotative description, covers measurable characteristics and properties; (vi) context description, models the context covering information related to agents, activities, or situations; and (vii) cultural events, represents cultural events and exhibitions.

SCULPTEUR [18] project, under the support of the European Union, aims to develop a system for browsing and searching museum collections using textual metadata, in addition to content analysis and an ontological classification. The proposed architecture contains a semantic layer that consists of an ontology and information instances. SCULPTEUR is based on CIDOC CRM, and extends it to include concepts such as objects digital representations and their respective feature vectors, besides tools and algorithms used to produce and compare feature vectors, query construction, and digital media for displaying purposes.

CURATE [30] presents an approach for representing curatorial narratives, thus, an exhibition is enriched by stories, or even to conform an exhibition by themselves with support of physical media. The problem approached by authors of CURATE is that narrative meaning cannot be expressed or derived solely from the metadata of cultural heritage objects. Authors of CURATE base their research on the hypothesis that curatorial narrative has generic characteristics and properties that can be found in other narratives, such as novels or films, hence, concepts like Story, Plot, and Narrative can be adopted. CURATE is based on CIDOC CRM and DOLCE + DnS Ultralite (DUL) [53] ontologies.

MOM [19], is a top level ontology that also deals with the heterogeneous nature of cultural heritage and is based mainly on CIDOC CRM and EDM, in addition to ORE20, FOAF21, DC22, and SKOS23. MOM ontology takes from EDM, classes like Non-Information Resource, which includes Event, Time Span, Place; Information Resource, which includes Web Resource and Provided CHO (Cultural Heritage Object). From CRM, it takes Actor, 24

that includes Group and Person; Physical Thing, which considers Physical Man-Made Thing, Biological object, and Collection; Conceptual Object, that includes Appellation and Information Object, which in turn includes Procedure, Linguistic Object, Document, and Visual Item. The ontology is completed with own concepts, such as Role and Digital Information.

OntoMP [31, 54], is the foundation ontology of Museu da Pessoa (MP), a virtual museum which has the purpose of exhibiting stories about ordinary people. OntoMP is primarily based on CIDOC CRM, in addition to FOAF and DBPedia. OntoMP concepts are directly related to person nature (People, Ancestry, Offspring, etc.), life episodes (Childhood, Leisure, Marriage, Birth, etc.), abstract concepts (Dreams, Religion, Costumes, etc.); relationships (Receives, Visits, Performs, etc.). Some concepts are directly referred to CIDOC CRM, however some properties related to the person cannot be described naturally, thus FOAF concepts such as Gender, Person Names, and Person-images relations are included. From DBPedia, properties such as Religion, Profession, Education, Party, and Spouse are included.

Marchenkov et al. [20] propose an ontology aimed at developing a digital environment oriented to visitors and museum service staff. This environment offers personal recommendations based on user context and exhibition characteristics; in addition to the collaborative management of information contained in different museums. Authors propose a layer-based model in which the semantic layer is responsible for providing mainly three advanced services: (i) Visiting Service, consists of creating personalized exhibition of a set of museum objects, based on the available knowledge of the visitor; this service adapts itself dynamically during the museum tour; (ii) Exhibition Service to display descriptions and visual information on visitors’ personal screens and devices; thus, physical exhibition is enlarged by using digital media; and (iii) Enrichment Service to support the evolution of a semantic network, allowing to receive notes from visitors and staff in order to improve the database information. This ontological model is based on CIDOC CRM, but it is extended to be able to host a recommendation system, through a sub-ontology called Rank, which contains the Rank class that stores exhibits scores, in addition to Exhibit and Profile classes.

TOMS (Thailand Open Museum System) [22], is a project whose main objective is to enable collaboration and information exchange among Thailand national museums. It is based on LOD and CIDOC CRM. The project proposes a three layer architecture: Data Storage, Manipulation and Processing, and a System Interface Layer. Authors of TOMS detail how existing information is mapped to CIDOC CRM corresponding concepts. Finally, they make a qualitative evaluation based on the user’s experience and their satisfaction level. Some evaluation points are the improvement in work efficiency and the perceived system utility.

Lo Turco et al. [21], describe a research about relationships among cultural heritage, digital technologies, and visual models. Authors use CIDOC CRM classes and relationships for available data and the CR-Mig extension for the mapping of the documentation of the evaluative, analytical, deductive, interpretative, and creative decisions related to gathering data stage and then to computer-based visualization process.

3.3. Proposed Cultural Heritage Knowledge Categorization

In our previous work [32], we propose a categorization of the cultural heritage knowledge, based on the UNESCO’s definition and on relevant ontologies that represent some aspects of it. Our categorization considers the following aspects:

- Temporal Item:
  - Event: events, occasions, or situations.
  - Time-Span: historical period of time.

- Permanent Item:
  - Place: locations, physical areas.
  - Actor: people, roles, groups.
  - Physical Object: defines artifacts that can be human-made or from natural origin.
  - Material: defines the materials from which the artifacts are made.
  - Person Extended: refers to concepts linked to humans, such as their identity (names, nicknames, gender, etc.), to abstract elements (dreams, customs, profession, etc.), and human relationships such as politics or religion.

- Exhibition
  - Digital representations: defines the creation process and products of digitizing an exhibition (e.g., images, video, documents).
– Extended Cultural Heritage

* Performance: defines concepts linked to cultural activity carried out by people, such as speeches or dance.
* Site as Cultural Heritage: defines elements to extend the Place concept at Permanent Item to be able to include outdoor places with cultural interest as cities, landscapes, etc.
* Event as Cultural Heritage: defines elements to extend the Event concept at Temporal Item to be able to include social activities as festivals, rites, etc.
* Culinary Tradition: defines the process and products of food preparation with cultural interest.
* Music and Songs: defines concepts linked to music and its production as society cultural expression.

– Ranking: needed concepts to rank cultural heritage expressions (e.g., exhibitions, monuments, events, cultural sites) according to different criteria (e.g., visitors, reviews, comments).

Although this cultural heritage knowledge categorization covers our initial requirements (see Section 4), it can be further extended with other cultural heritage and urban tourism topics, such as Language and Traditional Medicine, as proposed in [35].

### 3.4. Comparison

Table 1 presents the comparison of the reviewed ontologies, in terms of the proposed knowledge categorization, extending the scope of our previous work [32]. The proposed classification does not attempt to compare the scope of each proposal within each concept. Table 1 displays only reviewed studies which have available information about resources, as concepts and properties that are part of them, leaving out works that do not present these details.

While standards are a good starting point, it could be they do not cover all the edges that can arise during the development of a project or their complexity may lead to adopt it partially. Main POI ontologies have been developed from European projects [26, 28, 37], during the first decade of the 2000s, and nowadays, some of the project webpages are not available, as well as their ontologies. DataTourisme is one of the most recent ontology and it is currently under support and constant updates. As shown in Table 1, for POI ontologies, concepts related to time (Temporal items) are partially covered, e.g., they do not cover activities such Production or Creation. Permanent Items are partially represented by such ontologies, due to they can model touristic places, but they do not have the interest on representing artworks that can be present in such places (i.e., physical objects, their materials). Actually, touristic places represented by these ontologies go beyond cultural heritage interest, they include, for example, hotels, restaurants, shopping centers. Since most of these ontologies support tourist recommendation systems, concepts related to ranking are mostly covered.

Concerning, museum ontologies, concepts related to Temporal items, Permanent items, and digital representations are the main targeted represented resources, while Curatorial Narrative and Ranking concepts are neglected. For both POI and museum ontologies, modeling Extended Cultural Heritage is not considered.

CURIOCITY ontology seeks to cover all aspects of our proposed cultural heritage categorization in a minimalist way, identifying concepts and properties, which serve as a nexus and points of integration and extension to other domains. It presents a modular conception with the intention of being flexible and adaptable according to the application characteristics. Since, it is based on standards, such as CIDOC CRM, and ICONCLASS, and on widely used ontologies such as Finto or DBPedia, the interoperability is guaranteed. For the design of CURIOCITY ontology, we gather the experience of previous proposals and take into consideration aspects already cataloged as necessary. It is possible to represent the concept of cultural heritage not only contained in museums but from a broader view according to the UNESCO categorization. This greater representation perspective allows other city elements to acquire a cultural, educational, economic, and recreational value; thus, its tourist attraction is enriched. CURIOCITY ontology is described in more detail in the following section.
To cover these requirements, we evaluated popular ontologies in this domain. We considered CIDOC CRM, FINITO, and ArCo as the closest to our objectives, because their degrees of knowledge coverage. Although we knew the complex nature of CIDOC CRM, as noted in [55], we decided to adopt it as the base ontology. It was not an easy decision. Event though FINITO and ArCo have a greater coverage of concepts, the adoption of CIDOC CRM was due to its status as a standard.

**CURIOCITY ontology** is mainly a subset of CIDOC CRM [46], which is focused on events; thus, it imposes a particular perspective of knowledge representation, that must be taken into account when integrating with other ontologies. According to RUTAS project requirements, we define five extensions to CIDOC CRM based on the UNESCO’s classification for a wider representation of the concept of cultural heritage: (1) Site as Cultural Heritage; (2) Event as Cultural Heritage; (3) Performing Arts; (4) Music; and (5) Culinary Tradition.

We considered CURATE [30], due to curatorial narrative has special interest to be applicable in the context of tourist guide robots, in order to provide them story narrative capabilities. 

**4. CURIOCITY Ontology: Our Proposal**

Our proposal is developed in the context of the project RUTAS (Robots for Urban Tourism Centers, Autonomous and Semantic based)\(^2\), aimed at developing tourist guide robots and services for the diffusion and preservation of cultural heritage and urban tourism. One of the RUTAS’s goal is to create a knowledge base of museums (indoor places) in Arequipa city in Peru, as well as the characterization of cultural and touristic elements present in its historical center (classified as cultural heritage by UNESCO), such as landscapes, monuments, buildings, which correspond to outdoor environments.

Urban tourism as cultural heritage must also be approached in the context of RUTAS project. Thus, artistic urban expressions, culinary art, urban cultural events, dance, music, etc., should be also represented.

Concepts related to handicrafts are also urban tourism expressions. However, they present similar characteristics to cultural heritage contained in museums; thus, no additional concepts or relations already modeled in the basis ontologies are required. Moreover, as in museum context, it is also needed concepts related to urban collections and exhibitions, such as ranking and curatorial activity.

---

\(^2\)https://github.com/JADA1979 under construction

\(^2\)https://giulianodelagala.github.io/CURIOCITY/

---

<table>
<thead>
<tr>
<th>Table 1: Comparison of Ontologies related to Cultural Heritage</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Name</strong></td>
</tr>
<tr>
<td>-----------------</td>
</tr>
<tr>
<td>CURIOCITY</td>
</tr>
<tr>
<td>Harmonise [26]</td>
</tr>
<tr>
<td>Qall-Me-Ext [29]</td>
</tr>
<tr>
<td>HITouch [37]</td>
</tr>
<tr>
<td>OnTour [28]</td>
</tr>
<tr>
<td>DataTourisme</td>
</tr>
<tr>
<td>Locals [39-42]</td>
</tr>
<tr>
<td>FINITO* [47]</td>
</tr>
<tr>
<td>MUSEUM FINLAND</td>
</tr>
<tr>
<td>EDM [51]</td>
</tr>
<tr>
<td>Acto [152]</td>
</tr>
<tr>
<td>SCULPTURE [18]</td>
</tr>
<tr>
<td>CURATE [36]</td>
</tr>
<tr>
<td>ERM [59]</td>
</tr>
<tr>
<td>OntoMP [54]</td>
</tr>
<tr>
<td>Manchenkov et al. [20]</td>
</tr>
<tr>
<td>TOMP [22]</td>
</tr>
<tr>
<td>Lo-Turco et al. [21]</td>
</tr>
<tr>
<td>Contains related concepts</td>
</tr>
</tbody>
</table>

\(^2\) Arts and Culture Category: MAA/TAO/KULO/MUSEO/VALO
cept and relationships. It is also included CRMdig extension [48] for digital representations of cultural elements, as well as other ontologies from domains of interest, such as music (e.g., MUSIC Ontology [57], DOREMUS [58], [59]) or food (FOODON [60]), in favor of identifying interconnection points, towards which they will be proposed the required extensions.

CURIOCITY ontology development was carried out using a top-down approach, which means identifying general terms and then going to specific ones [61]. We followed the simple but effective approach proposed by Methontology [62], which consists of seven phases: (i) Specification; (ii) Knowledge Acquisition; (iii) Conceptualization; (iv) Integration; (v) Implementation; (vi) Evaluation; and (vii) Documentation. We partially show the results of the Specification and Knowledge Acquisition phases in Section 3, that present our proposed knowledge categorization and the comparison with related studies. The result of the rest of phases are described in this section and the following one. We will keep iterating on these phases to reach a more extended CURIOCITY ontology version.

CURIOCITY ontology is defined in three levels of specificity: (i) Upper Ontology, that identify two main branches from which general concepts are derived (Persistent Item and Temporal Item modules); (ii) Middle Ontology, with classes and properties needed to extend the concept of cultural heritage (Extended Cultural Heritage module); and (iii) Low Ontology, providing a higher level of detail for the representation of artwork objects. These levels are not mutually exclusive, they are only intended to indicate an abstract division of specificity for the purposes of reasoning and concepts analyzed. CURIOCITY ontology is also enriched with axioms and inference rules, which are part of the Logic component. The whole proposed architecture is depicted in Figure 1. In the following we explain each level in detail.

4.1. Upper Ontology

The dichotomy of continuity and occurrence is taken as basis for entities hierarchy. Persistent Item represents things that have a persistent identity, which survive events. These can be people, objects, ideas, or concepts. While Temporary Entity, represents temporal concepts or phenomena whose nature is related to happening rather than being. From these two general concepts, it is defined the first general reasoning of CURIOCITY ontology, which conforms the Upper Ontology and is represented in Figure 2. We use the following prefixes to identify classes and relationships taken from the corresponding ontology: crm: for CIDOC CRM ontology, cur: for CURATE ontology, fin: for Fint ontologies, dbp: for concepts from DBPedia, mus: and foo: for concepts of music and food ontologies, respectively (e.g., MUSIC Ontology, DOREMUS, FOODON), and cit: for classes, properties, and relations added in CURIOCITY ontology.

crm:Event takes a central position, in accordance with CIDOC CRM proposal. crm:Event is a subclass of crm:Temporal Item and this in turn is a subclass of crm:Persistent Item, which is defined by crm:Time Lapse. A crm:Event occurs in a crm:Place (indirectly through crm:Period) and involves a crm:Persistent Item, which is superclass of crm:Actor, crm:Conceptual Thing, and crm:Physical Thing.

At this level of specificity, we identify general concepts from which our five required extensions must derive. We have extended crm:Event concept with subclasses that by themselves constitute a cultural heritage, such as festive events, traditions, rites, celebrations, and similar; thus, CURIOCITY ontology has cit:Event CH class (Event as Cultural Heritage), as a subclass of crm:Event to represent them.
In the same way, crm:Site is extended to make it possible to characterize places that constitute a heritage by themselves. CURIOCITY integrates Site CH (Site as Cultural Heritage), which may contain subclasses according to UNESCO’s classification such as: Historic Cities, Cultural Landscapes, Underwater Cultural Heritage, and Natural Sacred Sites.

Culinary traditions are considered as a crm:Physical Thing subclass, which is a non direct subclass of crm:Persistent Item. It is proposed cit:Food and cit:Food CH classes. In the case of music, songs, and performing arts, their extension is considered as subclasses of crm:Conceptual Thing. CURIOCITY ontology includes the concepts cit:Music and cit:Performing Arts.

We adopt the CRMDig extension (dig:Digital Object) to characterize digital representations of elements of digital exhibitions, such as virtual museum implementations. Also, cur:Curatorial Narrative concept, based on CURATE ontology, is integrated to cit:Event CH to be able to represent narrative as cultural heritage.

Additionally, the crm:Person class, which is rather limited on CIDOC CRM, is extended with properties from FOAF; thus a better representation of human characteristics is available. Furthermore, to improve expressiveness in the temporal domain, OWL-Time [63] and CRMgeo [64] concepts have been incorporated to CURIOCITY, and inference rules have been formulated from the temporal relations of Allen’s interval algebra [65].

Having identified these primary higher level elements, it can be specified the next level of reasoning that defines the specialized modules of CURIOCITY Middle Ontology.

4.2. Middle Ontology

This level presents classes and relationships that allow the extension and integration of CURIOCITY Upper Ontology with ontologies from other domains in order to enrich the representation of heritage knowledge. In this version of CURIOCITY we present five extensions according to our project requirements.

4.2.1. Site Middle Ontology Module.

Having identified place-related concepts, it is necessary to extend the knowledge, in such a way a site can also be a cultural heritage in its own right. This idea is reinforced by the special status UNESCO grants to certain cities to promote its conservation and protection. Cities, in turn, are home to places and points with cultural and touristic interest, therefore cit:Site CH (Site as Cultural Heritage) is a concept proposed in CURIOCITY ontology. Other place-related concepts have been identified as cit:Site CH subclasses, such as cit:Park, cit:Protected Area, and cit:Natural Landscape, which are adapted from DBPedia, but described with the additional characteristic of being items having cultural interest. Other related subclasses can be extended according to the needs of case of study (see Figure 3). It is also necessary to be able to describe these places of interest in a better way, using some additional concepts (e.g., extension, altitude, population, time zone) as instances of sub-classes (e.g., cit: Physical Measurements, cit:RGB Range) under the crm:Dimension class, represented by a crm:Measurement Unit, such as square kilometers, meters above sea level, or inhabitants. It is also necessary to describe non-exact characteristics as a cit:Quality of the place, such as cit:Climate. The whole reasoning for crm:Site can be seen in Figure 3.

Fig. 3. Reasoning about Extended Site as Cultural Heritage: Site Middle Ontology Module

4.2.2. Temporal Entity Middle Ontology Module.

crm:Event is defined, according to CIDOC CRM, as the coherent processes and delimited interactions of material nature in physical, social, or cultural systems. In this way, cit:Event CH (Event as Cultural Heritage) proposes an extension to characterize festive events and traditions as social activities, besides of rites and customs, that qualify as cultural heritage (e.g., the Inti Raymi Festival in Cusco-Perú or the Rio de Janeiro Carnival in Brazil). A Cultural Event such as a cit:Tradition is a subclass of cit:Event CH, which in turn has subclasses
such as cit:Rite, in which an crm:Actor participates, involves crm:Physical Things, has a classification or crm:Type (e.g., religious, sport, cultural), and is held in a crm:Place and in a crm:Time-Lapse. This reasoning is illustrated in Figure 4.

4.2.3. Music Middle Ontology Module.

Music and Songs are others cultural expressions considered by UNESCO. The concept of cit:Music is added to CURIOCITY ontology as a crm:Conceptual Thing subclass, which is already identified by CIDOC CRM. Music as cultural heritage requires other concepts which allow to extend beyond a music score contained in a museum and to be understood as a representative cultural expression of the people (e.g., peruvian Huayno). One effort for this integration is DOREMUS[58], an extensive project that among other contributions presents an ontology based in FRBRoo[59] and CIDOC CRM. DOREMUS aims to characterize music scores and recording data. Another proposal for the integration of music and cultural heritage is presented by Thalmann et al. [59], a model for physical and digital representation of music-related artifacts, as well as the paraphernalia of live music events, harmonizing CIDOC CRM, FRBR[59] and the Music Ontology. We include some minimal elements that permits the integration with more elaborated ontologies, such as the mentioned above. Music related activities like its crm:Creation by an crm:Actor, its cit:Performance by playing (performing) a mus:Musical Instrument; and its classification by a mus:Music Genre, which is a subclass of cit:Genre. Figure 5 illustrates this reasoning.

4.2.4. Performing Arts Middle Ontology Module.

In a similar way to cit:Music, cit:Performing Art class represents cultural activities that characterize people, including dances, theatrical performances, or similar. cit:Performing Art is a subclass of crm:Conceptual Thing and has also cit:Genre to classify these activities. Figure 6 illustrates this reasoning.

4.2.5. Food Middle Ontology Module.

The culinary tradition identifies and characterizes one society from another in a particular way. cit:Food CH as cultural heritage is proposed in a minimalist way that allows integration with ontologies of food domain, such as FoodOn [60]. It is also considered a cit:Food Product Type to classify ingredients, as a subclass of crm:Type, and cit:Preparation Process as an activity to describe the food preparation process, as a subclass of crm:Production. cit:Preparation Process is based on foo:Food Transformation Process, a more complex concept to represent different types of food processing. In a minimalist way, we can represent a typical dish such as Peruvian Ceviche as an instance of Food CH, product of the Process of Ceviche preparation (cit:Preparation Process), classified as marinated as cooking method (crm:Type), from ingredients such as green lemon, sea fish, red onion, etc. (instances of cit:Food). This reasoning is illustrated in Figure 7.
4.3. Low Ontology: Artworks and monuments

The reasoning about cultural heritage in its conventional form refers to elements contained in indoor environments (i.e., artworks in museums, historical churches, exhibitions, etc.) and outdoor environments (i.e., monuments and artworks in the city, in parks, etc.). This knowledge has been analyzed by standards, such as CIDOC CRM, which is continuously reviewed and improved by use and research experiences. Figure 8 depicts some of the concepts about artworks and monuments reasoning, which are considered in CURIOCITY ontology, e.g., an activity (subclass of crm:Event), such as the production of an utensil (crm:Persistent Item) like a basket case (Cesto), was carried out by a prehispanic culture such as the Nazca (crm:Group). Nowadays, this artifact is under the custody of a Peruvian Museum (crm:Group), exhibited in its Archeology and Ethnology department (crm:Place); located in Arequipa city, declared World Cultural Site (cit:Site CH) by UNESCO.

CURIOCITY ontology respects event-based CIDOC CRM approach, however it adds richness to the concepts that were defined in previous sections, i.e., an crm:Event is not only a link between crm:Actor, crm:Thing, and crm:Place, but crm:Event and crm:Place can represent by themselves an entity of cultural heritage. In this way, we have a reasoning not only towards events, but also from events. In the same way, a crm:Place not only delimit a space, but they are also cultural heritage that includes other cultural heritage elements.

5. CURIOCITY Ontology Evaluation

An evaluation of our proposal has been carried out in order to answer three main questions: (i) What percentage of elements do we keep in common with the standard?; (ii) Do the elements that constitute our proposal contribute to a better representation of the cultural heritage?.

To evaluate ontologies it is appropriate to follow a methodological process that provides metrics for qualitative and quantitative assessments. In this work, we follow the systematic approach proposed in [33], which in turn is based on well-known ontology evaluation strategies such as: golden standard [66], OQuaRE [67], OntoMetrics [68], andOops [69]. This methodology offers a comprehensive evaluation and even a comparative evaluation with similar available ontologies. It proposes a guideline to comparatively evaluate ontologies considering Correctness and Quality perspectives, based on three levels of comparison:

- Lexical: it includes linguistic, vocabulary, and syntactic aspects.
- Structural: it considers aspects related to taxonomy, hierarchy, relationships, architecture, and design.
- Domain Knowledge: it considers how effectively the knowledge has been covered and how the results of the application are aided by the use of the ontology.

Figure 9 illustrates the components of this evaluation framework. This comparative evaluation process assumes the existence of a reference, called golden standard, which can be represented by a knowledge categorization elaborated with the support of experts in the area, a base ontology, or a set of documents describing the domain knowledge. In our case, the golden standard is defined by the knowledge categorization based on UNESCO’s cultural heritage definition presented in Section 3.3 and the requirements from RUTAS project. We also conduct a structural comparative evaluation using OQuaRE methodology among CURIOCITY, CIDOC CRM Standard (available as ERLANGEN-CRM28), and ArCo Ontology [52].

5.1. Lexical Level

At this level, the evaluation is based on similarity metrics that allow analyzing the proximity of concepts and related vocabulary within the domain, from the ontologies evaluated.

---

28http://erlangen-crm.org/
Fig. 8. Reasoning about Middle and Low Ontology elements CURIOCITY - CIDOC CRM
A. Pinto-De la Gala et al. / CURIOCITY: A Cultural Heritage Ontology for Urban Tourism

To calculate these similarity metrics, we have developed a parser that allows the extraction of the ontology entities (i.e., classes, relationships, properties) from their RDF/XML language implementations. Thus, we have the lists of entities names of both ontologies.

To determine the percentage of reuse of ERLANGEN CRM elements adopted in CURIOCITY ontology, we utilize the Document Similarity using the Vector Space Model (VSM) to evaluate linguistic similarity between two ontologies [70]. In this sense, each ontology is represented as a document that consists of a bag of terms (conformed by the N terms that appear in any of the documents) extracted from the lists of entity’s names, labels, and comments in the ontologies. The term weighting function to calculate each component in the N-dimensional vector for each ontology is presented in Eqs. (1) to (3), where \( t \) is the number of times a term occurs in a document, \( T \) is the total terms in document, \( D \) is the total of documents to compare; and \( d \) denotes the number of documents where the term occurs at least once. Then, the Document Similarity between the two ontologies is calculated by taking the cosine dot product, as Eq. (4) shows, where \( VS_{O} \) are the term weighting vectors of the ontologies. To compute \( DocSim \), we use the class TfidfVectorizer of the scikit-learn library of Python\(^{29}\).

\[
\text{TermWeighting} = TF \times IDF \\
TF = \frac{t}{T} \\
IDF = \frac{1}{1 + \log \frac{D}{d}} \\
DocSim(O_i, O_j) = \frac{VS_{O_i} \cdot VS_{O_j}^T}{\|VS_{O_i}\| \|VS_{O_j}\|} \tag{4}
\]

According to \( DocSim(O_i, O_j) \) metric, CURIOCITY ontology has a 71.8% similar terms to ERLANGEN CRM; whereas the left percentage (28.2%) corresponds to the inclusion of other concepts and properties which conform the proposed extensions.

5.2. Structural Level

The structural level is mostly evaluated according to the relationships among entities, as ontologies are graph. We use OQuaRE methodology [67] to conduct a structural evaluation. OQuaRE metrics are calculated according to Eqs. (5) to (16). All OQuaRE’s characteristics (i.e., Structural, Functional Adequacy, Compatibility, Reliability, Transferability, Operability, and Maintainability) are scored according to the OQuaRE scale system (i.e., 1 means not acceptable, 3 is minimally acceptable, and 5 represents exceeds the requirements).

We implemented an application\(^{30}\) to perform automated metrics calculation, the assignment of the score according to OQuaRE charts, the results presentation, as well as graphics that allow a better comparative analysis. The application was developed in Python 3.8, using libraries such as Rdflib for the management of the ontology graph, as well as the generation of necessary queries; Numpy for numerical processing, Pandas for the generation of results tables, and Matplotlib for the creation of graphs.

\[
\begin{align*}
\text{LCOMOnto} &= \Sigma \text{PathLength}(C_{\text{thing}}, \text{Leaf}_{C_i}) / \Sigma \text{PathLength}(C_{\text{thing}}, \text{Leaf}_{C_j}) \tag{5} \\
\text{WMCOnto} &= \Sigma \text{PathLength}(C_{\text{thing}}, \text{Leaf}_{C_i}) / \Sigma \text{Leaf}_{C_j} \tag{6} \\
\text{DITOnto} &= \max(\text{PathLength}(C_{\text{thing}}, \text{Leaf}_{C_i})) \tag{7} \\
\text{NACOnto} &= \Sigma C_i \Sigma \text{Anc}_{C_j} / \Sigma \text{Leaf}_{C_j} \tag{8} \\
\text{NOCOnto} &= \Sigma C_i \Sigma \text{Sub}_{C_j} / (\Sigma C_i - \Sigma \text{Leaf}_{C_j}) \tag{9} \\
\text{CBOnto} &= \Sigma C_i \Sigma \text{Anc}_{C_j} / (\Sigma C_i - \Sigma \text{Leaf}_{C_j}) \tag{10} \\
\text{RFCOnto} &= (\Sigma C_i \Sigma \text{Pro}_{C_j} + \Sigma \text{Sub}_{C_j}) / \Sigma C_i \tag{11} \\
\text{NOMOnto} &= \Sigma C_i \Sigma \text{Pro}_{C_j} / \Sigma C_i \tag{12} \\
\text{RKOnto} &= \Sigma C_i \Sigma \text{Sub}_{C_j} / (\Sigma C_i \Sigma \text{Sub}_{C_j} + \Sigma C_i \Sigma \text{Pro}_{C_j}) \tag{13}
\end{align*}
\]


\(^{30}\)https://github.com/giulianodelagala/CURIOCITY/tree/master/Evaluation/OquaRE
where,

- \( C_i \): Ontology classes.
- \( R_{C_i} \): Relations of class \( C_i \).
- \( Pro_{C_i} \): Properties of class \( C_i \).
- \( Anc_{C_i} \): Direct ancestor of class \( C_i \).
- \( Sub_{C_i} \): Direct subconcept of class \( C_i \).
- \( C_{Thing} \): Ontology root.

Table 2 details the obtained metrics and their corresponding OQuaRE score to evaluate the structural level\(^{31}\), for CURIOCITY, ERLANGEN-CRM, and ArCo. Figure 10 depicts the comparison of the three ontologies according to each OQuaRE’s characteristic.

Figure 10(a) depicts the OQuaRE’s Structural characteristic, which evaluates ontology quality factors, such as Consistency, Formalization, and Entanglement. In this case, ontologies score similar for each sub-characteristic. The weakness of the ontologies is in Cohesion, whose LCOMonto metric shows that there is a strong dependency between components, mainly due to the complexity of the relationships between concepts. The other sub-characteristic with the lowest score is Formal Relationships, linked to the RRonto metric, which indicates that the ontologies present a lower number of sub-concepts versus the number of properties; it is not exactly a symptom of weakness of the ontologies, but an indicator of how they are structured.

Figure 10(b) represents the comparison of the two ontologies according to Functional Adequacy scores. CURIOCITY and ERLANGEN CRM get similar scores for each sub-characteristic. The weakness of both ontologies is in Clustering and Similarity sub-characteristics, because a wide range of properties of each concept makes clustering process difficult, whereas ArCo presents a better behavior. The other sub-characteristic with a low score is Results Representation, which indicates that all three ontologies are complex; therefore, they have a degree of analysis difficulty in the results they provide.

Figure 10(c) shows the comparison of the three ontologies according to the sub-characteristics corresponding to Compatibility (Replaceability), Reliability, and Maintainability characteristics. ArCo scores the best of the three ontologies. CURIOCITY scores higher than CIDOC CRM for each of these characteristics, which indicate that better performance is expected. This improvement in the overall scores is mainly due to a higher value of the WMonto metric; denoting that ArCo and CURIOCITY are less complex.

Figure 10(d) shows the Maintainability comparison. ArCo gets the highest score in all the sub-characteristics, followed by CURIOCITY. In the case of CURIOCITY, the weakest scores are in Analysability and Testability sub-characteristics, which can be understood as a degree of difficulty in diagnosing deficiencies and validation.

Table 3 and Figure 11 show the summary of the OQuaRE evaluation for the CURIOCITY, ERLANGEN CRM, and ArCo ontologies. ArCo scores higher overall mainly due to metrics such as WMonto (score 5), NOMonto (score 4), and DITonto (score 2), which show that ArCo has a less complex structure. CURIOCITY scores better than CIDOC CRM in Transferability, Reliability, Compatibility, Maintainability, and Operability characteristics, which implies being easier to adapt and maintain without losing interoperability with the standard ontology.

5.3. Domain Knowledge Level

At this level, the defined golden standard is used to assess coverage and correctness of the knowledge of the domain. Our golden standard is represented by the categorization based on UNESCO’s definition of the cultural heritage knowledge described in Section 3.3 and the requirements of RUTAS project. We subjected CURIOCITY ontology to experts evaluation, to confirm that golden standard is appropriate in this domain.

5.3.1. Experts’ opinions regarding the golden standard

The need of representing the concepts of cultural heritage that include not only artworks circumscribed in museums, but even those elements of the city that make up the attention of urban tourism, both of concrete and abstract nature, is the focus of this research. For this reason, experts in the area were consulted about the elements that are considered as heritage, and then contrasted with the golden standard, which is the base to identify those gaps that have been overlooked and should be part of CURIOCITY ontology.

To do so, two questionnaires were preliminarily developed through online forms. Opinions of 10 partici-

\(^{31}\)http://miuras.inf.um.es/evaluation/oquare/Metrics.html
Table 2
Structural comparison: CURIOCITY, ERLANGEN CRM, ArCo

<table>
<thead>
<tr>
<th>Metric</th>
<th>CURIOCITY</th>
<th>ERLANGEN-CRM</th>
<th>ArCo</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lack of Cohesion in Methods (LCOMOnto)</td>
<td>6.955</td>
<td>6.948</td>
<td>4.045</td>
</tr>
<tr>
<td>Weight method per class (WMCOnto)</td>
<td>10.097</td>
<td>15.980</td>
<td>4.359</td>
</tr>
<tr>
<td>Depth of subsumption hierarchy (DITOnto)</td>
<td>10</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>Number of Ancestor Concepts (NACOnto)</td>
<td>1.113</td>
<td>1.160</td>
<td>1.010</td>
</tr>
<tr>
<td>Coupling between Objects (CBOOnto)</td>
<td>1.135</td>
<td>1.184</td>
<td>1.005</td>
</tr>
<tr>
<td>Response for a concept (RFCOnto)</td>
<td>6.773</td>
<td>7.727</td>
<td>4.245</td>
</tr>
<tr>
<td>Number of properties (NOMOnto)</td>
<td>5.660</td>
<td>6.557</td>
<td>3.283</td>
</tr>
<tr>
<td>Relationship Richness (RROnto)</td>
<td>16.20%</td>
<td>15.02%</td>
<td>21.92%</td>
</tr>
<tr>
<td>Relationships per concept (INROnto)</td>
<td>109.43%</td>
<td>115.91%</td>
<td>92.17%</td>
</tr>
<tr>
<td>Tangledness (TMOnto)</td>
<td>1.113</td>
<td>1.170</td>
<td>0.962</td>
</tr>
</tbody>
</table>

Fig. 10. OQuaRe Characteristics Comparison
A first question presented to the experts was: When you think of Cultural Heritage, which concepts comes to your mind?; the expert is asked to give a level of relevance from “Unrelevant” to “Very relevant” for concepts related to: “Event, situations of interest”, “Artwork, handicraft”, “Performing art, theater, traditional dance”, “Music, Traditional songs”, “Festivities, traditions, customs”, “Typical food, culinary traditions”, “Monuments, buildings, squares”, “Landscapes, countryside, nature reserves”, “Sports, sporting events, children’s games”, and “Language, dialects, phrases”. We also included an open response alternative, to learn about other relevant concepts proposed by the interviewees, however we only received more specific concepts that can be related to the more general concepts above (e.g., photography could be related to Artwork).

As shown in Figure 12, concepts with the greatest relevance to the idea of cultural heritage are those related to works of art, monuments, and landscape, two of these three elements have a context related to outdoor environments (monuments and landscape), clearly identified as points of interest in an urban tourism context. The concepts related to performing art, music, festivities, typical food, and language (dialects) are rated as very relevant. Out of these, only language is not included in CURIOCITY. Finally, concepts related to events and sports are considered of medium-high relevance. Sports related concepts can be adapted from performing art concepts.

The second proposed question was: What information do you think is necessary to describe an element of Cultural Heritage?; the expert is asked to give a level of information necessity about: “Time: dates, periods, events”, “Place: location, spatial info”, “Person or Group: author, creator, culture”, “Exhibition, gallery, display”, “Curator, description, additional info”, and “Material, color, shape”; with the purpose of identifying the basic concepts on which a representation of cultural heritage elements should be based. Again, we also included an open response alternative, to learn about other relevant concepts proposed by the interviewees.
Fig. 12. Summary of answers to the question about concepts related to Cultural Heritage

about other descriptor elements proposed by the interviewees, however we only received more specific concepts related to the physical nature of an artifact (e.g., dimensions), these concepts are included in Material category.

Figure 13 summarizes the results of the survey to this question, from which it can be appreciated that the essential elements of the description are given by the space-time pair; in second place the information needed corresponds to the elements of the person, the exhibition and the material; finally, additional supporting data to the curator is required for representation of a cultural heritage element.

The questionnaire answers validate the necessity of an extension for cultural heritage representation from an urban tourism point of view, and validates our knowledge categorization as the golden standard, which in turn is defined from UNESCO’s cultural heritage classification (see Section 2) and on described knowledge from evaluated ontologies (see Section 3).

As shown in Table 1, CURIOCITY addresses some of the gaps in cultural heritage representation in the context of a city and urban tourism, according to this established golden standard and RUTAS project requirements.

5.4. Discussion

CURIOCITY ontology has been evaluated at three levels: Lexical, Structural, and Domain Knowledge. Lexical level evaluation results show that CURIOCITY ontology has around of 70% of similar concepts with CIDOC CRM base ontology (ERLANGEN CRM), indicating that a core has been preserved allowing interoperability with the standard, and a glossary of common terms, which facilitates the understanding and usefulness of the proposal.

Structural Level has OQuaRE metrics as guideline, as suggested by the evaluation methodology followed in this work. CURIOCITY, ERLANGEN CRM, and ArCo characteristics were compared.

CURIOCITY ontology shows similar conditions in comparison with CIDOC CRM in the Structural and Functional Adequacy characteristics, however CURIOCITY presents better conditions for the rest of characteristics, which could be understood that CURIOCITY has a better performance in maintenance and learning issues than the CIDOC standard. ArCo shows a less complex structure, whereas CURIOCITY and ERLANGEN CRM ontologies have complexity as a point to take into consideration for their use. On the other hand, all three ontologies present an optimal domain representation richness that translates into query consistency, good modularity, and adaptability.

Domain Knowledge level needs the intervention of domain experts; in this way, questionnaires were used to know their perception of the ontology’s adequacy. Results at this level, show that concepts needed for an adequate representation of cultural heritage and urban tourism domain have been considered. The perception of CURIOCITY ontology utility from a preliminary test also returns favorable responses. However, further tests on the quality of results and proposed inferences remain to be carried out.
6. Towards CURIOCITY Framework: An Application Case

RUTAS project aims to develop a system of robots as tour guides in urban centers, solving the problem of connectivity between robots and providing access to tourist information through a semantic repository available on the cloud. In the context of this project, data from indoor spaces (e.g., museums, historic churches, and libraries) and outdoor spaces (e.g., historical sites, squares, monuments) are being collected for art specialists in a repository called D-RUTAS32.

Fig. 14. CURIOCITY Framework Architecture

Towards the accomplishment of RUTAS project goals, we are developing CURIOCITY framework, which is roughly composed by three layers (see Figure 14): (i) Semantic Repository layer aimed at managing the semantic data, such as the CURIOCITY repository, with information about cultural heritage in urban tourism; other semantic repositories can also be included, such as an ontology to represent tourists (i.e., basic information, preferences, interests), an ontology to manage robots’ tasks (e.g., navigation, mapping, object detection); (ii) Data Processing layer in charge of processing sources of information (e.g., D-RUTAS repository, web pages, databases of museums, databases of government cultural institutions) to automatically populate the semantic repositories and generate a rich knowledge network; it also processes and executes queries from the Application layer; and (iii) Application layer, which provides interfaces for maintenance, updates, and navigation through the semantic repositories; services such as online artwork catalogs, virtual museums, tourism recommendation systems, web forms to collect data, are offered at this layer.

To demonstrate the suitability and flexibility of CURIOCITY ontology, we have developed a first version of this framework, whose code and documentation are available online33. Next sections describe the components developed of the first prototype of CURIOCITY framework and explain how CURIOCITY ontology supports its functionality.

6.1. Semantic Repository Layer

Currently, the Semantic Repository layer has a CURIOCITY ontology base version, mainly consisting of the Low Ontology, representing artworks and museums. This semantic repository is automatically instantiated from the Data Processing layer. In the current version, it has the information from four museums of Arequipa, Perú: Municipal Museum “Guillermo Zegarra Meneses”, Convent Museum “La Recoleta”, “Santa Catalina” Museum, and Convent Museum “Santa Teresa”, transformed into triples from D-RUTAS (see Section 6.2).

At this level, Apache Jena Fuseki is used as the SPARQL server, with TDB as the storage sublayer and OpenJena34 as the reasoner. Table 5 shows the number of records processed from D-RUTAS and the number of triples (before and after using the reasoner) that were generated. The largest number of instances obtained after applying the Pellet reasoner to the knowledge base supported by the CURIOCITY ontology (i.e., concepts, properties, inference rules), generates new knowledge from the initial data. The knowledge base obtained from the different steps of the instantiation are available at CURIOCITY’s repository35.

6.2. Data Processing Layer

This layer provides tools to automatically instantiate the Semantic Repository layer and generate SPARQL queries to support the Application layer. Nevertheless, for the current version of the framework, before using these tools, it is necessary to perform a manual mapping process of D-RUTAS concepts to CURIOCITY ontology classes and relations. This mapping configuration can be saved in a JSON file for later use. In the future, this mapping process will be also automated, by using, for example, string similarity, string matching, or natural language processing.

32https://github.com/dulcineo
33https://github.com/giulianodelagala/CURIOCITY
34https://github.com/giulianodelagala/CURIOCITY/blob/master/Instances
6.2.1. Mapping D-RUTAS to CURIOCITY Ontology

The mapping process consists of matching the D-RUTAS data contained in MS Excel spreadsheets to corresponding CURIOCITY ontology entities. Table 6 shows a summary of the mapping from museums description in D-RUTAS to classes and properties of CURIOCITY ontology. As instance of the extended cultural heritage concept, the class cit:Site CH is included to allow categorizing the museum and the city. Also, cit:Description is included in order to improve organization of additional notes for narrative purposes as well as virtual catalogs implementation.

Table 7 summarizes the mapping from object data contained in museums (i.e., artworks description in D-RUTAS) to CURIOCITY ontology. It takes as an example the artifact 'Cesto de la Cultura Nazca'.

6.2.2. Instantiation: CSV Parser

The implementation process of the instantiation begins with the spreadsheets containing the D-RUTAS museum data, which are exported to CSV format for manipulation. Parsing and generation of the RDF triplets is done through a parser developed with Python 3.8, Pandas 1.1.2, and RDFlib 5.0.0 libraries. D-RUTAS data are processed line by line, starting with general concepts with multiple references (e.g., RUTAS data are processed line by line, starting with

<table>
<thead>
<tr>
<th>Source</th>
<th>Records</th>
<th>Attributes</th>
<th>Instantiated triples</th>
<th>Instantiated triples (post-reasoner)</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;Guillermo Zegarra Meneses&quot; Municipal Museum</td>
<td>297</td>
<td>23</td>
<td>12633</td>
<td>29720</td>
</tr>
<tr>
<td>&quot;La Recoleta&quot; Convent Museum</td>
<td>166</td>
<td>22</td>
<td>6438</td>
<td>19781</td>
</tr>
<tr>
<td>&quot;Santa Catalina&quot; Convent Museum</td>
<td>25</td>
<td>22</td>
<td>1020</td>
<td>10307</td>
</tr>
<tr>
<td>&quot;Santa Teresa&quot; Convent</td>
<td>189</td>
<td>22</td>
<td>7368</td>
<td>17033</td>
</tr>
</tbody>
</table>

6.3. Application Layer

The current version of CURIOCITY framework offers three services for end users. The first one (Figure 16 (a)), is a general user oriented web page that allows users to browse the semantic repository, perform queries for searching under combined criteria, and obtain details of artifacts contained in museums. The second application (Figure 16 (b)) offers a 3D tour of virtual museums for end users, as well as a tour creator utility for developers to design and implement virtual visits to museums from the data and information kept in the semantic repositories. The third service (Figure 16 (c)) is a desktop application oriented to the administration of the semantic repository. The functionalities of this application allow administrators to configure the mapping of CSV tables (entities of the ontology and the generation of instances in an automated way from the data), to make queries to the semantic repository under combined criteria, to make particular

---

Table 5

Number of records processed from D-RUTAS and instantiated triples in CURIOCITY Ontology

<table>
<thead>
<tr>
<th>Source</th>
<th>Instantiated</th>
<th>Instantiated</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>triples</td>
<td>triples</td>
</tr>
<tr>
<td></td>
<td>(post-reasoner)</td>
<td>(post-reasoner)</td>
</tr>
<tr>
<td>&quot;Guillermo Zegarra Meneses&quot; Municipal Museum</td>
<td>297</td>
<td>12633</td>
</tr>
<tr>
<td>&quot;La Recoleta&quot; Convent Museum</td>
<td>166</td>
<td>6438</td>
</tr>
<tr>
<td>&quot;Santa Catalina&quot; Convent Museum</td>
<td>25</td>
<td>1020</td>
</tr>
<tr>
<td>&quot;Santa Teresa&quot; Convent</td>
<td>189</td>
<td>7368</td>
</tr>
</tbody>
</table>

---

36https://github.com/giulianodelagala/CURIOCITY/tree/master/Queries
queries in SPARQL format, and to update the semantic repository manually.

These applications allow final users to get information about museums of Arequipa, Perú, through graphical interfaces that automatically transform their requirements into queries and respective results, supported by the SPARQL QUERY Engine at the Data Processing layer, that in turn accesses the proper repository at the Semantic Repository layer. Similarly, the applications for developers (e.g., tour creator and desktop admin) are supported by graphical interfaces at the Application layer, that can access APIs, engines, scripts, etc. provided by the Data Processing layer, in order to query or manage the Semantic Repository.

The application experience of CURIOCITY ontology shows our progress in developing use cases and represents an indicator towards the successful realization of our requirements. We have shown that final
Table 7
Summary of mapping concepts and properties from D-RUTAS artwork data to CURIOCITY ontology

<table>
<thead>
<tr>
<th>Description</th>
<th>Example</th>
<th>Classes</th>
<th>Properties</th>
</tr>
</thead>
<tbody>
<tr>
<td>Identifier</td>
<td>m_01_s01_002</td>
<td>crm:E42 Identifier</td>
<td>crm:P48 has preferred identifier</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>crm:P139 has alternative form</td>
</tr>
<tr>
<td>Title</td>
<td>Cesto de la Cultura Nazca</td>
<td>crm:E22 Human Made Object crm:E35 Title</td>
<td>crm:P102 has title</td>
</tr>
<tr>
<td>Author</td>
<td>Cultura Nazca</td>
<td>crm:E74 Group crm:E39 Person crm:E12 Production</td>
<td>crm:P14 carried out by</td>
</tr>
<tr>
<td>Object utility</td>
<td>Contenedor de Objetos</td>
<td>crm:E55 Type</td>
<td>crm:P2 has type</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>crm:P127 has broader term</td>
</tr>
<tr>
<td>Owner</td>
<td>Museo Municipal “Guillermo Ze-</td>
<td>crm:E74 Group crm:E39 Person crm:E10 Transfer of Custody</td>
<td>crm:P52 has current owner</td>
</tr>
<tr>
<td></td>
<td>garra Meneses”</td>
<td></td>
<td>crm:P50 has current keeper</td>
</tr>
<tr>
<td>Dimensions</td>
<td>Height: 20 cm x Width 45 cm</td>
<td>crm:E55 Dimension crm:E58 Measurement Unit</td>
<td>crm:P43 has dimension</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>crm:P91 has unit</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>crm:P90 has value</td>
</tr>
<tr>
<td>Condition State</td>
<td>Bueno</td>
<td>crm:E3 Condition State</td>
<td>crm:P44 has condition</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>crm:P2 has type</td>
</tr>
<tr>
<td>Material</td>
<td>Vegetales</td>
<td>crm:E57 Material</td>
<td>crm:P45 consists of</td>
</tr>
<tr>
<td>Description</td>
<td>El cesto fue hecho por medio de ...</td>
<td></td>
<td>crm:P3 has note</td>
</tr>
<tr>
<td>Creation date</td>
<td>100-600 a.C.</td>
<td>crm:E12 Production crm:E52 Time-Span</td>
<td>crm:P4 has time-span</td>
</tr>
<tr>
<td></td>
<td></td>
<td>crm:E4 Period</td>
<td>crm:P82 at some time within</td>
</tr>
<tr>
<td>Period</td>
<td>Intermedio temprano</td>
<td>crm:E4 Period</td>
<td>crm:P4 has time-span</td>
</tr>
<tr>
<td>Acquisition</td>
<td>Donación</td>
<td>crm:E10 Transfer of Custody crm:E55 Type</td>
<td>crm:P29 custody received by</td>
</tr>
<tr>
<td></td>
<td></td>
<td>crm:E39 Person crm:E74 Group crm:E52 Time-Span</td>
<td>crm:P2 has type</td>
</tr>
<tr>
<td>Acquisition Actor</td>
<td>Edmundo Escomel</td>
<td>crm:E10 Transfer of Custody crm:E55 Type</td>
<td>crm:P28 custody surrendered by</td>
</tr>
<tr>
<td>Acquisition Date</td>
<td>1990/10/17</td>
<td>crm:E10 Transfer of Custody crm:E55 Type</td>
<td>crm:P4 has time-span</td>
</tr>
<tr>
<td>Museum Department</td>
<td>Arqueología y Etnología</td>
<td>crm:E53 Place</td>
<td>crm:P89 falls within</td>
</tr>
<tr>
<td>Physical Location</td>
<td>Sala 01 Cultura Prehispánicas</td>
<td></td>
<td>crm:P55 has current location</td>
</tr>
<tr>
<td>Link list</td>
<td><a href="http://www.objdetail.com">www.objdetail.com</a></td>
<td></td>
<td>crm:P3 has note</td>
</tr>
</tbody>
</table>

Fig. 16. Front-end Applications
users, as well as developers, can transparently access the Semantic Repository, through the utilities provided at the Processing Data layer (e.g., SPARQL engine, scripts, API, parsers) from comfortable and easy-to-use graphical interfaces at the Application layer. Thus, it is possible to represent the knowledge of cultural heritage and urban tourism domains of a city; besides being a semantic base for CURIOCITY framework services, in order to conform a level of abstraction of knowledge and use of semantic web technologies for final users.

7. Conclusions and Future Work

In this paper, we present a new formal representation of cultural heritage knowledge in the context of urban tourism. We describe the ontology CURIOCITY (Cultural Heritage for Urban Tourism in Indoor/Outdoor environments of the CITY) to represent the cultural heritage knowledge, following the UNESCO classification and mainly based on CIDOC CRM to keep the interoperability among applications. It is defined in three levels of specificity: (i) Upper Ontology, that identify two main branches from which general concepts are derived (Persistent Item and Temporal Item modules); (ii) Middle Ontology, with classes and properties needed to extend the concept of cultural heritage (Extended Cultural Heritage module); and (iii) Low Ontology, providing a higher level of detail for the representation of artwork objects and monuments. Following a methodological process, we evaluate and compare CURIOCITY ontology, showing that it is able of representing all aspects of cultural heritage, according to our proposed categorization of the domain knowledge and RUTAS project requirements. Moreover, we introduce and developed a very first version of CURIOCITY framework, to show the suitability of the ontology in a case study of four museums of Arequipa, Perú. The Data Processing layer allows the transformation of data from excel to RDF triples of CURIOCITY ontology (Semantic Repository layer), generating a richer repository. The Semantic Repository layer is the base of services and applications, such as on line catalog and virtual museum. In this scenario, we demonstrate that, by using CURIOCITY ontology, it is possible to represent the knowledge of cultural heritage and urban tourism domains of a city, as the basis for developing interoperable services and applications.

Even though the current version of CURIOCITY ontology covers the initial requirements related to RUTAS project, we plan to evaluate the inclusion of other urban tourism topics, such as Languages and Traditional Medicine, to give greater coverage to the concept of cultural heritage. It remains to evaluate CURIOCITY ontology through the applications proposed by the CURIOCITY framework, in order to know the level of satisfaction of the end user (expert and common user) in the activities supported by the ontology.

We also are working on the integration of CURIOCITY ontology with other ontologies developed in parallel by the RUTAS project (e.g., in the simultaneous location and mapping (SLAM) problem domain [71], in the user profile domain [72]), which will be part of the semantic repository of the CURIOCITY framework. We also continue working on the development of CURIOCITY framework to including more general ontology population techniques to serve different heterogeneous databases.

Acknowledgements

This research was supported by the FONDO NACIONAL DE DESARROLLO CIENTÍFICO, TECNOLÓGICO Y DE INNOVACIÓN TECNOLÓGICA - FONDECYT as executing entity of CONCYTEC under grant agreement no. 01-2019-FONDECYT-BM-INC.INV in the project RUTAS: Robots for Urban Tourism, Autonomous and Semantic web based.

References


