

OntoMP, an Ontology to build The Museum of the Person

Ricardo G. Martini¹, Cristiana Araújo¹, José João Almeida¹, and Pedro Rangel Henriques¹

Algoritmi Research Centre, Department of Informatics
University of Minho - Gualtar - 4710-057, Braga, Portugal
rgm@algoritmi.uminho.pt, decristianaaraujo@hotmail.com, {jj, prh}@di.uminho.pt

Abstract. This paper is concerned with the creation of a specific ontology for the knowledge repository of the **Museum of the Person** (Museu da Pessoa). The **Museum of the Person** assets are composed of several interviews (collected previously for a large cultural project) involving common people, to perpetuate their life stories.

The museum holds an heterogeneous collection of XML documents. In such format, the collection items are many times not recognizable and understandable by the visitors who wish to explore it. Therefore, we intend to use an ontology that allows a conceptual navigation over the available information, enabling the visitors to extract knowledge during the visit to these life stories.

So, this paper aims at presenting the ontology we have developed using CIDOC Conceptual Reference Model (CIDOC-CRM)[1] to enable visitors to lookup individual life stories, read them, and also intercross information among a cluster of life stories to build up the story of a company/institution or to study social behaviors and customs.

Keywords: Ontologies, Virtual Museums, Museum of the Person, CIDOC-CRM

1 Introduction

Virtual Museums are increasingly in vogue, because society is more and more concerned with preserving the cultural heritage and making it accessible to anyone with an interest in studying it.

According to Werner Schweibenz, a **Virtual Museum (VM)** is “a logically related collection of digital objects composed in a variety of media which, . . . lends itself to transcending traditional methods of communicating and interacting with visitors. . . ; it has no real place or space . . .” [2].

A **Virtual Museum**, analogously to a traditional museum, also acquires, conserves, and exhibits the heritage of humanity¹ creating a delightful environment

¹ In such cases, intangible objects, or immaterial things, according to: <http://www.unesco.org/culture/ich/index.php?lg=en&pg=00022#art2>. Accessed: 17-12-2015

for pleasure or enjoyment, as well as an appropriate place for learning, and research. This type of museum allows people to interrelate information among different stories, to study social phenomena, natural disasters, migrations, season's customs, among others.

The project discussed here is concerned with the creation of a VM that exhibits life stories of common citizens. More exactly, we are rebuilding the Portuguese version of the global Museum of the Person (depicted from here by MP) that connects individuals and groups through sharing their life stories². MP is located in Brazil, Portugal, USA, and Canada.

The museum's holdings are a set of XML documents and our aim is to build a virtual museum using CaVa³, a Learning Spaces (LS) generator that receives a specification about the LS (using the known vocabulary of the museum curators) as input and generates the final virtual museum web application. This vocabulary has the terms of the literature and the terms of the Cultural Heritage, based on the used ontology. So, a conceptual navigation is available to the visitor of the virtual exhibition rooms [3]. To generate the LS [4] with CaVa, the documents repository must be specified using an ontology [5, 6].

In this paper we will discuss the process of reverse engineering the MP structure in order to obtain a museum standard format view. This process can be split into the two steps described below, as depicted in Figure 1:

1. Reverse engineering of MP (from MP to MP ontology);
2. Mapping the museum standard format view (from MP ontology to CIDOC-CRM [7]).

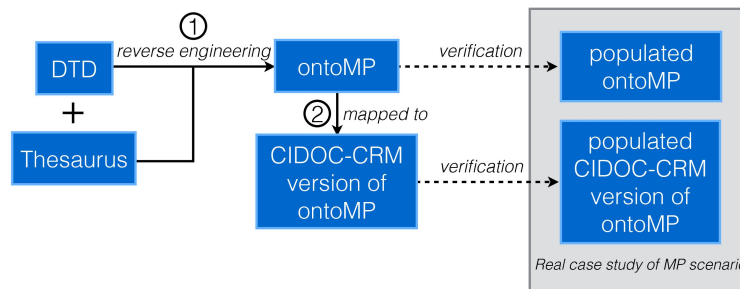


Fig. 1. Reverse Engineering and mapping of MP

Following the diagram of Figure 1, the paper is structured as follows. In Section 2 the Museum of the Person and its documents are described. Based on that description, on the *thesaurus*, and on the DTD's inherited, a first draft of

² <http://www.museumoftheperson.org/about/> Accessed: 17-12-2015

³ CaVa stands for the Portuguese term *Construção de Ambientes Virtuais de Aprendizagem*. In english: Learning Virtual Environments Generator

an ontology for the Museum of the Person is presented in Section 3. Then, in Section 4 that ontology is specified according to CIDOC-CRM standard. Finally, in Section 5, a synthesis of the study is presented and some suggestions for future work are discussed.

2 The Museum of the Person documents

MP was born in Brazil, São Paulo, in 1991, created by a group of historians who decided to build the country's history using testimonials of ordinary people [8–10]. This is still an alive project accessible at <http://www.museudapessoa.net>. MP aims at gathering testimonials from human being, famous or anonymous, to perpetuate his history [11, 9].

From the life stories of individuals, the objective is to write up the stories of families, communities, or institutions [9]. This museum deals with common people, human beings, not with physical objects usually composing the traditional museum assets. Its “*art collection*” is made up of intangible or immaterial things. In this case, the alive objects are used as informers, reporting the events and emotions they experienced [11]. Actually, the narrators, to report their life stories during a predefined structured interview⁴, remember events and particular situations they have participated in. These memories will act as a basic element for social research, because the set of life stories allows to reconstruct a social universe [11].

The workflow adopted by Museum of the Person technicians to acquire common people life stories is expressed below:

- a. The report of a participant is recorded (audio or video) by an interviewer. Although every interview is a unique thing, interviewers guide to some predefined topics in order to cover the entire life story;
- b. Interviews are transcribed;
- c. Transcriptions are annotated in XML, marking events, self contained stories, etc.;
- d. XML interviews may be used to produce several outputs.

Life stories are evidences in support of facts or statements attested by common people carrying a social and historical character, which must be preserved and processed to become an immeasurable human heritage [11].

The MP's collection consists of XML documents connected to each participant. Typically each interview is split into three parts [8]:

- mini-biography and personal data, such as name, date and place of birth, and job. This information is placed in a separate document, called BI;
- two versions of the interview: the text of the original interview, and the edited document;

The **interview** file refers to the raw interview; it contains all the questions asked and the narrator's answers.

⁴ Recorded in a tape or a film.

The **edited** file (an example can be seen in Listing 1.1) is a plain text, structured by themes that define small portions of a person’s life story. In this format, a life history may give rise to thematic stories (eg, dating, childhood, craft, among others).

Both **interview** and **edited** files contain metadata, tagging to define important testimony zones. Examples of this marks are given like institution names, jobs, places, etc.;

- photographs and their subtitles. The subtitle document contains a section for each photo or scanned document as file name and a caption. This caption includes a description of the image and the date, and wherever possible the name of the stakeholder.

Beside the interviews, there is also a *thesaurus* that includes key concepts mentioned in the stories. These concepts are linked by hierarchical relationships, namely: Broader term (BT) / Narrower term (NT) (superclass and subclass, respectively); Has (HAS) / Part-of (POF); Use instead (USE) / Used for (UF); Instance (INST) / Instance of (IOF); and Related term (RT).

As mentioned, the edited interview is in XML format, according to a Document Type Definition (DTD) defined specially for this purpose. It is composed of: identification of the deponent, episode, ancestry, descent, childhood, house, education, tradition, religion, quotidian, migration, place, dating, marriage, office, life’s philosophy, event, photograph, among others. Listing 1.1 displays a small excerpt from the MP DTD.

```
1 <?xml version="1.0"?>
2
3 <!ENTITY %histVidaElem "p | identificacao | episodio |
   educacao | ascendencia | descendencia | infancia | evento
   | ... ">
4
5 <!ELEMENT mp ((%histVidaElem;)+)>
```

Listing 1.1. Excerpt of the *Edited Interview* DTD

The root element (**mp**) of our Museum of the Person is a life story composed of life story elements. Each life history contains alternative elements, such as a paragraph (**p**), an identification (**identificacao**), an episode (**episodio**), an ancestry reference (**ascendencia**), an event (**evento**), etc. The event element (**evento**) is defined as follows (Listing 1.2):

```
1 <!ELEMENT evento (%texto;)>
2 <!ATTLIST evento
3   ano CDATA #REQUIRED
4   mes CDATA #IMPLIED
5   dia CDATA #IMPLIED
6   local CDATA #IMPLIED
7   eixo CDATA #IMPLIED
8   titulo CDATA #IMPLIED
9   descricao CDATA #IMPLIED
10  relevancia (Alta | Media | Baixa) "Media"
```

Listing 1.2. DTD excerpt to describe an Event (`Evento`)

An element (`evento`) is defined as (`%texto`), an XML Entity previously introduced also has a set of alternative elements: a paragraph (`p`), an episode, a theme, among others. As attributes of the element (`evento`) we have: year (`ano`), month (`mes`), day (`dia`), place (`local`), type (`eixo`), title (`titulo`), description (`descricao`), and relevance (`relevancia`), all of type CDATA (simple text).

Section 3 outlines how we have described MP's life stories as an ontology, aiming at creating a more abstract and formal level of description.

3 OntoMP, an ontology for Museum of the Person

From the DTD and *thesaurus* introduced and briefly explained in the previous section, it was possible to build OntoMP, an ontology for MP.

The first step was an exhaustive extraction of the concepts present in the life stories. After a long analysis phase we came up with the following list: people (`pessoa`), ancestry (`ascendência`), offspring (`descendência`), house (`casa`), job (`profissão`), education (`educação`), episode (`episódio`), dating (`namoro`), accident (`acidente`), migration (`migração`), festivity (`festividade`), political event (`evento político`), catastrophic event (`evento catastrófico`), marriage (`casamento`), birth (`nascimento`), dream (`sonho`), childhood (`infância`), uses (`costumes`), quotidian (`quotidiano`), leisure (`lazer`), religion (`religião`), life's philosophy (`filosofia de vida`).

In this phase relations were also identified: performs (`exerce`), depicted (`éRetratada`), visits (`visita`), lives (`vive`), receives (`recebe`), tells (`narra`), has (`tem`), enrolls (`participa`), has-type (`tipo`), occurs (`ocorre`), refers to (`dizRespeito`).

After the extraction phase we built an ontology for the MP using the concepts and relations above. Then we realized that we could still add some more concepts to make a more complete ontology. We add: marital status (`estadoCívil`), sex (`sexo`), literacy (`habilidades literárias`), political party (`partido político`), first communion (`primeira comunhão`), death (`morte`), baptism (`batismo`), and photos (`fotos`).

As illustrated in Figure 2, OntoMP has the ability to breakdown the raw story into logically related elements. In this way, the museum visitor can have a conceptual navigation over the collection.

4 CIDOC-CRM

CIDOC-CRM is a formal ontology planned to aid in the integration, mediation, and interchange of heterogeneous cultural heritage information [1].

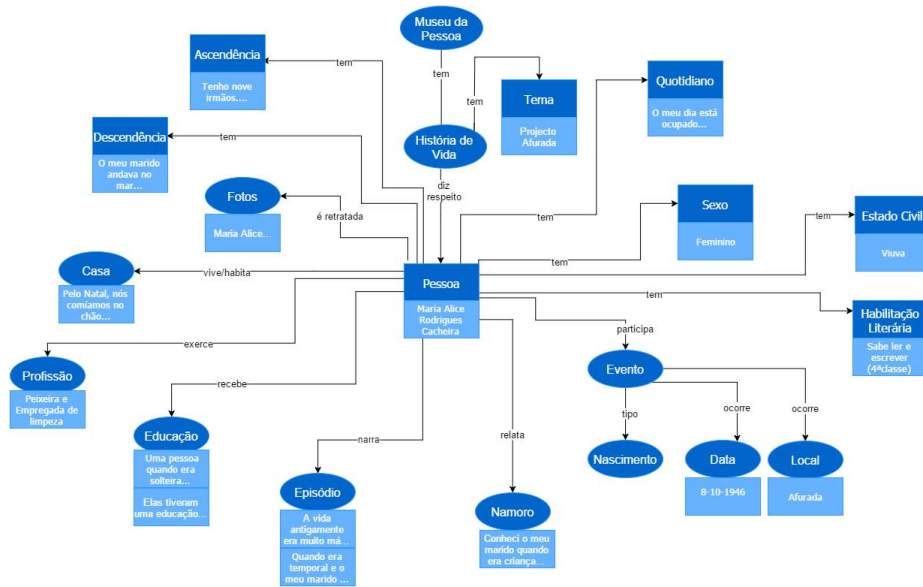


Fig. 2. An instance of OntoMP for Maria Cacheira life story (fragment)

More precisely, it specifies the semantics of museums documentation and document structures in cultural heritage in terms of a formal ontology. Summing up, CIDOC-CRM can be described as the curated knowledge of museums [1].

However, it is so broad and defined at a high abstract level that it can be adapted to any kind of museum. So this work uses CIDOC-CRM as a conceptual reference model to characterize Museum of the Person holdings, aiming to describe all dataset based on a standard ontology.

The core of CIDOC-CRM is based on seven concepts (*Temporal Entities*, *Events*, *Actors*, *Time-Spans*, *Conceptual Objects*, *Physical Things*, and *Places*). CIDOC-CRM is an event-based ontology, so its main concept is related with *Temporal Entities*. As an event-based ontology, it should contain *Time-Spans* and *Places* related with it. Besides, *Actors* and *Conceptual Objects* or/and *Physical Things* also should be related with an event.

CIDOC-CRM ontology can be synthesized⁵ as *Actors participate in Temporal Entities/Events, which affect or refer to Conceptual Objects/Physical Things within Time-Spans at Places*. All of these concepts can be related to *Appellations* and *Types*. *Appellations* denote a specific instance of some class, serving to identify it. *Types* are used to denote terms of controlled vocabularies that characterize and classify instances of the CIDOC-CRM ontology.

CIDOC-CRM ontology has name conventions that must be followed. The concepts and relations names start with the capital letters “E” (Entity) and “P”

⁵ Notice that CIDOC-CRM ontology is not composed only by these concepts. These are the core of the ontology.

(Property), respectively. Names are followed by identification numbers (e.g. E21 Person, E5 Event, P11 participated in, P152 has parents, etc.).

Section 4.1 explains how the Portuguese Museum of the Person collection was described in the standard CIDOC-CRM ontology.

4.1 A CIDOC-CRM version of OntoMP

After the construction of the OntoMP (see previous section), the next stage was to describe it in a standard ontology used for museums, CIDOC-CRM. The ontology was built up having in mind the need to keep it fully compatible with the original, avoiding to add or modify any class. This is demonstrated in Figure 3, which portrays the instantiation of CIDOC-CRM with a concrete life story (just a fragment is shown). For illustrative purposes, this fragment is the same that is depicted in Figure 2.

According to the abstract layer of the CIDOC-CRM ontology in Figure 3: *A life story (E31 Document) is about (P129) a person (E21 Person) and contains photos (E38 Image) concerning this person.*

Looking at the concrete layer of that ontology, we read that: *Maria Cacheira (E21 Person), participated in the (E5 Event) of her (E67 Birth). This event occurred at (E52 Time Span)—that is identified by (P78) 1946-10-08, an (E50 Date)—and at a (E53 Place)—that is identified by (P87) Afurada, an (E44 Place Appellation).*

In this fragment of *Maria Cacheira's* life story there are other concepts that were identified in Section 3. All these concepts, in CIDOC-CRM version, are described as (E55) Type because they characterize a (E21) Person and its instances. To exemplify that approach, consider the following case: *Maria Cacheira (E21 Person) has type (P2) Peixeira e Empregada de limpeza (E55 Type Job).*

To sum up, notice that Figure 3 shows a description of a specific life story (partial view) mapping the life story concepts and relations into standard CIDOC-CRM elements. In this way (following this standard representation) it will be possible to query systematically the knowledge repository to explore it.

The full diagrams for OntoMP (abstract ontology and instances) and their CIDOC-CRM version are available at www.di.uminho.pt/~gepl/OntoMP.

5 Conclusion

Along this paper we have proposed the development of an ontology to describe the collection of the Portuguese Museum of the Person, in order to enable a conceptual navigation over the museum's assets aiming at extracting knowledge from the different life stories (individually or intercrossed).

The work here reported is a part of a larger project aimed at the automatic generation of exhibition rooms (Learning Spaces) resorting to a framework called CaVa. The present case study was built upon a large repository of life stories collected in the past, in different places under different social research topics, and

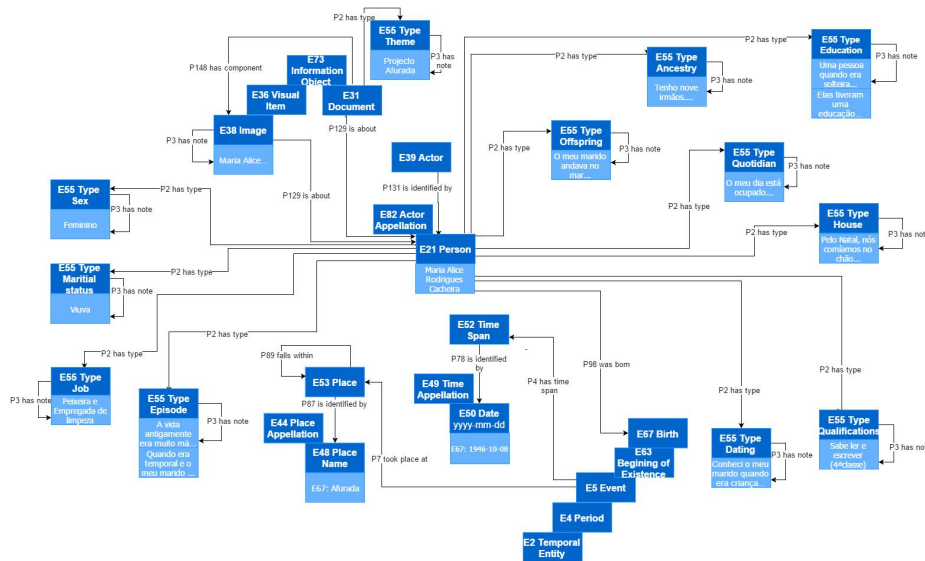


Fig. 3. An instance of CIDOC-CRM version of OntoMP for Maria Cacheira life story (fragment)

archived as structured eXtensible Markup Language documents. So, we did not start the ontology design from scratch; on the other way around, we performed a reverse engineering process on the referred legacy repository.

After building the abstract layer of the ontology that we call OntoMP, we instantiate it with various stories. This exercise was useful to tune OntoMP. Then we converted OntoMP to a CIDOC-CRM compatible form ontology with the objective of adopt a standard format used in the museums area. This way we preserve the interoperability of our repository for further exploration and knowledge extraction.

The next task, to realize the main project, is to build a virtual Learning Spaces (in this case, a Virtual Museum) to tell the world those picturesque and impressive life stories, and to extract knowledge about our society along the last decades connecting and relating the individual testimonials.

As future work, we intend to research how to extend CIDOC-CRM with FOAF⁶ (*Friend of a Friend*) network to better describe (i.e., to describe in a more natural way) some of the specific concepts intrinsic to the Museum of the Person. This is because FOAF is an ontology precisely oriented to talk about people; in that sense, FOAF has a descriptive vocabulary specific to individuals, their activities and their relations with other persons.

Acknowledgments: This work has been supported by FCT – Fundação para a Ciência e Tecnologia within the Project Scope: UID/CEC/00319/2013. The work of Ricardo G. Martini is supported by CNPq, grant 201772/2014-0.

⁶ Homepage: <http://www.foaf-project.org/>

References

1. ICOM/CIDOC: Definition of the cidoc conceptual reference model. Technical report, ICOM/CIDOC (May 2015)
2. Schweibenz, W.: The development of virtual museums. In: Virtual Museums. Volume 57(3). ICOM (2004)
3. Martini, R.: Formal Description and Automatic Generation of Learning Spaces based on Ontologies. Phd pre-thesis, Universidade do Minho (2015)
4. Goos, M.: Creating learning spaces. The Annual Clements/Foyster Lecture (2006)
5. Gruber, T.R.: Toward principles for the design of ontologies used for knowledge sharing. In: International Journal of Human-Computer Studies, Kluwer Academic Publishers (1993) 907–928
6. Studer, R., Benjamins, V.R., Fensel, D.: Knowledge engineering: Principles and methods. Data Knowl. Eng. (1998) 161–197
7. Oldman, D., Labs, C.: The CIDOC Conceptual Reference Model (CIDOC-CRM): PRIMER. International Council of Museums (ICOM) **1** (July 2014)
8. Simões, A., Almeida, J.J.: Histórias de Vida + Processamento Estrutural = Museu da Pessoa. In: XATA 2003 — XML: Aplicações e Tecnologias Associadas, Braga, Portugal, UM (2003) 16
9. Stafford, P.B.: Museum of person. Technical report (2015)
10. Worcman, K.: The museum of the person. In: Virtual Museums. Volume 57(3). ICOM (2004)
11. Almeida, J.J., Rocha, J.G., Henriques, P.R., Moreira, S., Simões, A.: Museu da Pessoa – arquitectura. In: Encontro Nacional da Associação de Bibliotecários, Arquivista e Documentalistas, ABAD’01, BAD (2001)