Architectural Approaches to build The Museum of the Person

Cristiana Araújo
Algoritmi Research Centre, Department of Informatics
University of Minho - Gualtar - 4710-057, Braga, Portugal
decristianaaraujo@hotmail.com

Ricardo G. Martini
Algoritmi Research Centre, Department of Informatics
University of Minho - Gualtar - 4710-057, Braga, Portugal
rgm@algoritmi.uminho.pt

Pedro Rangel Henriques
Algoritmi Research Centre, Department of Informatics
University of Minho - Gualtar - 4710-057, Braga, Portugal
prh@di.uminho.pt

José João Almeida
Algoritmi Research Centre, Department of Informatics
University of Minho - Gualtar - 4710-057, Braga, Portugal
jj@di.uminho.pt

Abstract — **The Museum of the Person (Museu da Pessoa, MP) is a virtual museum aimed at exhibiting life stories of common people. Its assets are composed of several interviews involving people whose stories we want to perpetuate. So the museum holds an heterogeneous collection of XML (eXtensible Markup Language) documents that constitute the working repository. The main idea is to extract automatically the information included in the repository in order to build the web pages that realize the museum's exhibition rooms. This project started by creating a specific ontology (OntoMP) for the knowledge repository of MP. That ontology is intended to allow a conceptual navigation over the available information. We will adopt the standard for museum ontologies CIDOC-CRM (CIDOC Conceptual Reference Model) refined with FOAF to represent OntoMP. The objective of this paper is to discuss different architectural approaches to build a system that will create the virtual rooms from the XML repository to enable visitors to lookup individual life stories and also inter-cross information among them. The first architecture is based on a TripleStore and uses SPARQL (SPARQL Protocol and RDF Query Language) technology to extract the information, while the second proposal is based on a Relational Database and uses CaVa Generator to query the repository and build the exhibition spaces.**

Keywords - Virtual Museums; Museum of the Person; Ontologies; RDF/SPARQL; Software Architectures.

I. INTRODUCTION

The society is more and more concerned with the preservation and the dissemination of Cultural Heritage, as works of art, ancient objects, and documents, among others.

Nowadays this can be achieved in a better way resorting to the information and communication technologies because they allow that the physical objects, on one hand, become accessible to anyone, and on the other hand, are not deteriorated rectos [1] [2] [3].

In this context of technological expansion, increasing the capability of extraction, storage and visualization of everyday life events, the museums have taken advantage to expand its field of action, as well as their own concept. They expand their geographical borders by providing information in their pages on the Internet and exhibiting their collections. On the other hand, completely virtual environments (called Virtual Museums, VM) appeared, without any references to physical spaces [1].

A Virtual Museum, such as a traditional museum, also acquires, conserves, and exhibits the heritage of humanity (in that case, intangible objects, or immaterial things, according to http://www.unesco.org/culture/ich/index.php?lg=en&pg=00022\#art2) creating a delightful environment for pleasure or enjoyment, as well as an appropriate place for teaching, and research.

On the other hand, the assets of the Museum of the Person (MP) contains several interviews that narrate the life stories of ordinary citizens. These citizens, to report their life stories, remember events and other particular situations they have participated in. MP resources are documental assets constituted by a collection of documents in XML (eXtensible Markup Language) format.

All the arguments presented above motivate the interest on building a virtual learning space (in that case, a Virtual Museum) to tell to the world those life stories and to extract knowledge about an epoch and a society connecting and relating them.

More precisely we aim at rebuilding npMP, the Portuguese branch of the Museum of the Person network (This network includes branches in Brazil, Portugal, USA, Canada, etc.) that connect individuals and groups through sharing their life stories (http://www.museumoftheperson.org/about/).

In this paper, and after a brief introduction to MP (Section I.A), we discuss the ontology built to hold the museum's knowledge repository (Section II), and then we present different technical approaches to implement the desired virtual museum (Section III).

Besides OntoMP, an ontology for the museum of the person that is new and a first contribution of this work, also the extension of the standard CIDOC-CRM for museums with FOAF concepts and properties is another contribution presented.

The last contribution of the paper is the detailed definition of a generic architecture for the implementation of a system that creates the museum exhibition rooms from the documents repository. Moreover we designed and propose two possible

implementations of that generic architecture, one more appropriate for situations where the repository is stored in a relational database, and the other to be used when the repository is archived in a triple-store. Our aim is to compare both approaches to understand the development effort involved in each one and to learn their benefits and drawbacks.

At the best of our knowledge there are not similar projects that use ontologies and tools to generate automatically virtual learning spaces from their specifications, neither in the scope of MP nor in the context of other virtual museums. So we will not include a section on related work. For the sake of space (necessary to introduce all the novelties of this paper) we decided not to include a state of the art section; the reader is referred to the authors pre-thesis [4] [5].

A. Museum of the Person, an overview

Museum of the Person aims at gathering testimonials from every human being, famous or anonymous, to perpetuate his history [2] [3].

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 (intangible or immaterial things). The interviewed are used as informers, reporting the events and emotions they experienced [2].

To report their life stories during a predefined structured interview, the narrators remember events and other particular situations they have participated in. These memories will act as a basic element for social research [2].

The Museum of the Person's collection consists of sets of XML documents, properly specified by a DTD (Document Type Definition created specially for that purpose and called MP-DTD) related with each participant.

Typically each interview is split into three parts [6]:

- **BI**: a brief biography and personal data, such as name, date and place of birth, and job;
- **interview**: two versions of the interview are built and saved -- the *interview* file refers to the raw interview and contains all the questions asked and the narrator's answers; the *edited* file is a plain text, structured by themes that define small portions of a person's life story.

  In this format, a life story may give rise to thematic stories (eg, dating, childhood, craft, among others). Both *interview* and *edited* files contain metadata tagging;
- **photographs** and their caption. This caption includes a description of the image, people depicted, place and the date.

Aside the interviews, there is also a *thesaurus* that includes key concepts mentioned in the stories.

Details about the elements that constitute each DTD will be mentioned in the next section that will discuss the development of MP's ontology (OntoMP). For more details on Museum of the Person please see [7].

II. THE CIDOC-CRM ONTOLOGY FOR MP, ONTOMP

A. OntoMP: Original Design

After an exhaustive analysis of all the documents (XML instances, respective DTD's, and the *thesaurus*) that belong to Museum of the Person, we could identify the concepts and relations involved in the life stories. This first step enabled us to design OntoMP, an ontology for the Museum of the Person. In this way, the museum visitor can have a conceptual navigation over the collection.

The main concepts extracted from the analysis phase are: 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*), leisure (*lazer*), quotidian (*quotidiano*), religion (*religião*), life's philosophy (*filosofia de vida*).

In a similar way we also identified the following relations: performs (*exerce*), depicted (*éRetratada*), visits (*visita*), lives (*vive*), receives (*recebe*), tells (*narra*), has (*tem*), enrolls (*participa*), has-type (*tipo*), occurs (*ocorre*), refers to (*dizRespeito*).

Then we realized that some more elements should be added to the ontology: marital status (estadoCivil), sex (sexo), literacy (habilitações literárias), political party (partido político), first communion (primeira comunhão), death (morte), baptism (batismo), and photos (fotos).

The ontology so far obtained is depicted in Figures 1, 2, and 3.
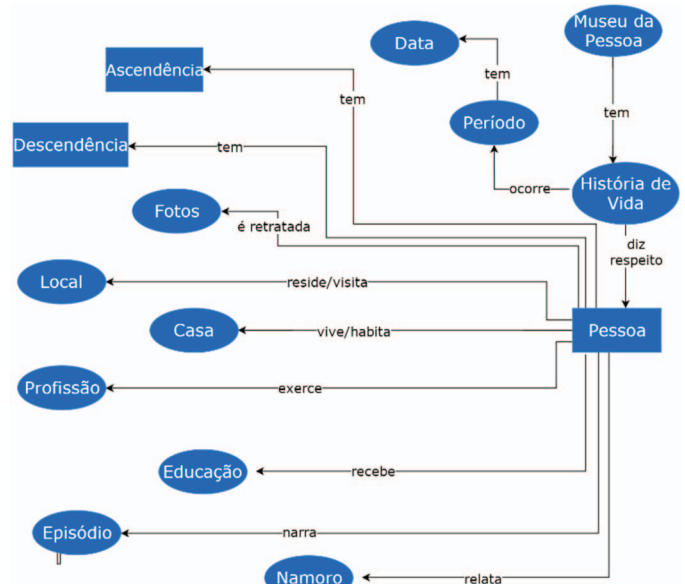

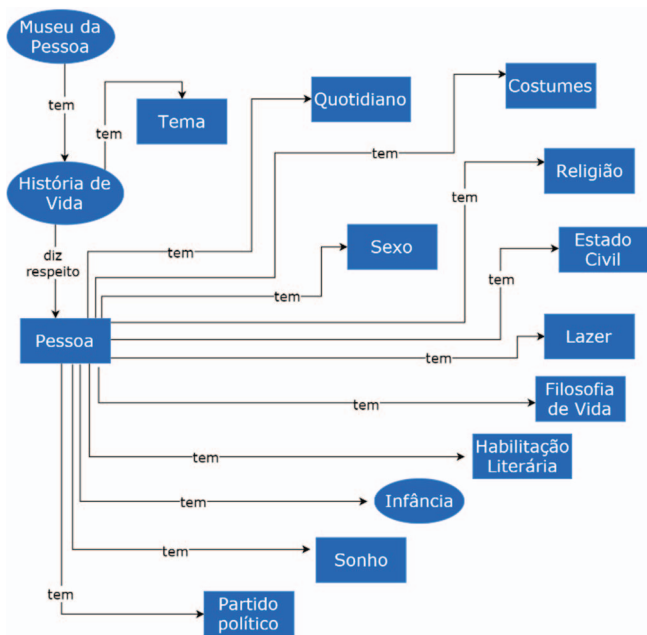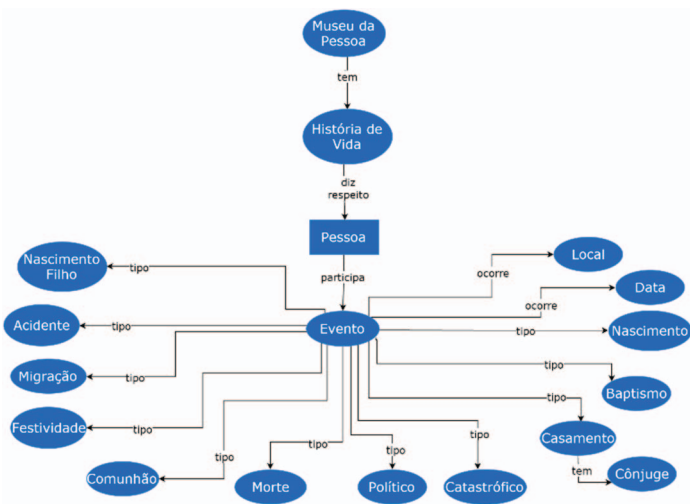Figure 1. A fragment of OntoMP

Figure 2. A second fragment of OntoMP

Figure 3. A last fragment of OntoMP

While Figures 1 and 2 show the general life concepts (ellipsis) related with 'Person' and his main data properties (rectangles), Figure 3 enhances 'Event' concept (a relevant component of OntoMP) and its different sorts (subclasses).

To validate the ontology designed, we created some instances using actual life stories picked-up from the MP collection, as can be consulted in the project's site at the URL http://www.di.uminho.pt/~gepl/npMP/. Notice that all those interviews were conducted in the past and we got written permissions to publish them.

B. OntoMP: CIDOC- CRM / FOAF Representation

After the validation and tuning of OntoMP, the next stage was to describe it in a standard ontology format used for museums, CIDOC-CRM (CIDOC Conceptual Reference Model) following the approach we adopted in the context of a project for another virtual museum [8].

CIDOC-CRM is a formal ontology planned to aid in the integration, mediation, and interchange of heterogeneous Cultural Heritage information [9]. It specifies the semantics of museums' documentation.

CIDOC-CRM is an event-based ontology. As an event-based ontology, it should contain *Time-Spans* and *Places* related with it. So, most of its concepts derive from the concept *Temporal Entity*. The core of CIDOC-CRM is based on seven concepts: *Temporal Entities*, *Events*, *Actors*, *Time-Spans*, *Conceptual Objects*, *Physical Things*, and *Places*. Besides, *Actors* and *Conceptual Objects* or *Physical Things* should also be related with *Event*.

The transformation of OntoMP, presented in previous subsection, to CIDOC-CRM standard was actually a straight-forward process because OntoMP is also based in events.

However we felt that some person properties could not be expressed in CIDOC-CRM in a natural way. So we decided to explore the combination with FOAF (Friend of a Friend) as it has a descriptive vocabulary specific to individuals, their activities and their relations with other people and objects [10] [11].

FOAF ontology describes two areas of digital identity information: biographical information and social network information [12].

After this investigation we decided to refine CIDOC-CRM adding FOAF definitions concerned with *gender* property, person names (*name, givenName, familyName* and *nick*) and person-image relations (*depicts* and *depiction*).

The final OntoMP built (also fully depicted in the project site referred above) is fully compatible with the original CIDOC-CRM and FOAF, that is, no classes were added or changed.

Once again we instantiated the general ontology with concrete life stories to validate it.

In Figure 4 we show an instance created from Maria Cacheira interview. Below we describe the CIDOC-CRM and FOAF fragment reproduced.

A life story *(E31 Document) is about (P129)* a person *(E21 Person)* and contains photos *(Image)* that *depicts* this person. *(E21 Person), gender* Female, *name* Maria Alice Rodrigues Cacheira (decomposed in *givenName* Maria Alice and *familyName* Rodrigues Cacheira), participated in *(E5 Event)* that is her birth *(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 other concepts can be identified. All these concepts, that characterize a *(E21 Person)*, are represented in CIDOC-CRM version, as *(E55*

Type). For example, *(E21 Person) has type (P2)* Peixeira e Empregada de limpeza *(E55 Type Job), has type (P2)* Viúva *(E55 Type Marital Status),* and *has type (P2)* "Sabe ler e escrever (4ª classe)" *(E55 Type Qualifications).*
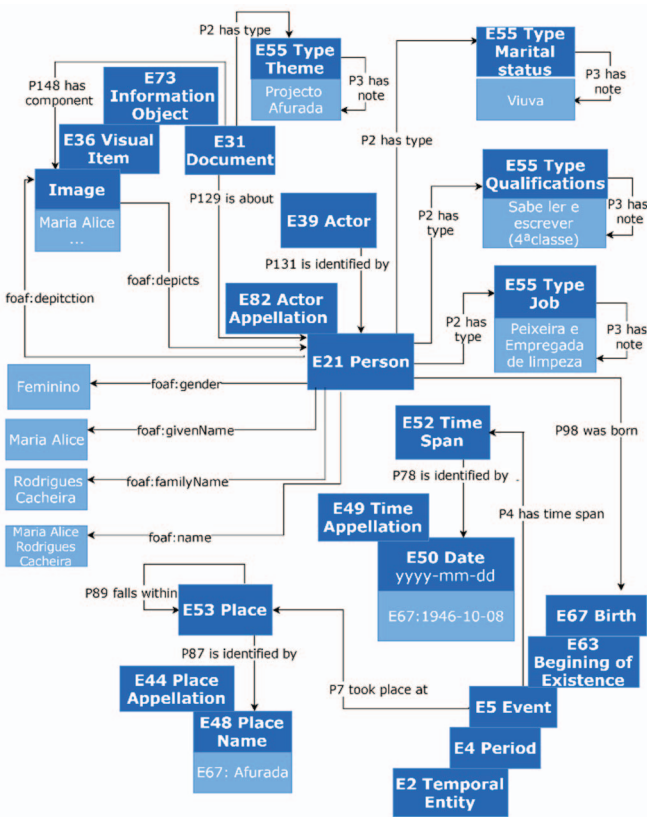

Figure 4. An instance of CIDOC-CRM / FOAF version of OntoMP for Maria Cacheira life story (fragment)

This CIDOC-CRM / FOAF ontology introduced is the heart of our project, as it describes MP's knowledge repository. Based on it, it is now possible to present the architecture of a system that builds the virtual museum from the documents collection.

III. PROPOSED ARCHITECTURES

This section presents a general approach for building the Museum of the Person.

This proposal is defined at an abstract level so that the main architectural blocks and their interactions can be clearly understood; the data flow and the main transformations will be emphasized without technological commitments.

After describing the general approach, two alternatives will be discussed to refine that architecture.

The general approach, illustrated in Figure 5, to build the Museum of the Person comprises: the repository; the Ingestion Function [M1] responsible for getting and processing the input data; a Data Storage [DS] that is the digital data archive; an

Ontology to map and link the concepts with the objects stored in [DS]; the Generator [M2] to extract data from [DS] and manage the information that will be displayed in Virtual Learning Spaces [VLS] (the final objective of this project).
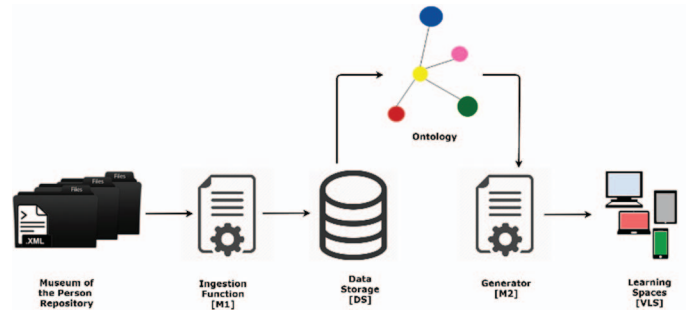

Figure 5. General Approach to build the Museum of the Person

This general approach will have two possible refinements, which are dependent on the [DS].

In approach 1 (Section A. Approach 1) the [DS] is a *TripleStore*, while in approach 2 (Section B. Approach 2) the [DS] is a *Relational Database.*

A. Approach 1

According to this decision, Ingestion Module and the Generator Module must be adapted; the first will transform the input XML documents into RDF triples, and the second will retrieve information from the RDF triples to create the museum web pages.

Figure 6 details the first module that is composed of three blocks:

- *Parser and Semantic Checker* that reads the repository documents and extracts the relevant data (annotated in XML), checking their semantic consistency;
- *Ontology Extractor* that identifies in the extracted data the concepts and relations that belong to the ontology creating in this way an instance of the abstract ontology (in another words, this component populates the ontology);
- *Triple Generator* that converts automatically the ontology triples (created in the preceding block) into triples in RDF notation appropriated to be stored in the [DS] chosen.
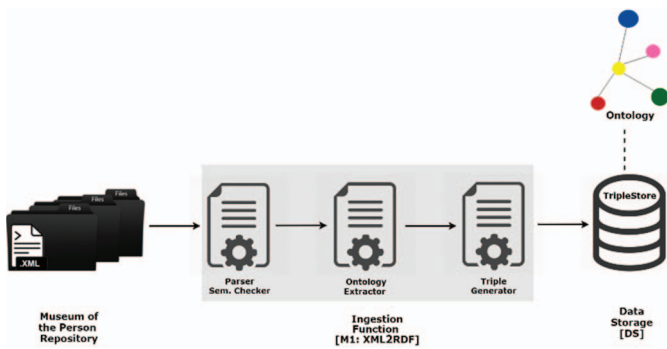
Figure 6. Module [M1] in Approach 1

The mapping between the domain ontology (previously defined) and the data extracted from the repository is automatically built by construction in second block, above.
This is, in this approach there is no need to create explicitly this mapping.
It means that the Generator [M2] can access directly the storage to obtain the conceptual information necessary to create the exhibition rooms.

To display in the Virtual Learning Spaces [VLS] the information stored in [DS] - TripleStore, the [VLS] Generator needs to send queries and process the returned data.

Figure 7 shows the second module [M2] (the Generator) that is composed of two blocks:
- *SPARQL Endpoint* that receives and interprets the SPARQL queries, accesses the TripleStore [DS] and returns the answers;
- *Query Processor* generates the SPARQL queries according to the exhibitions room requirements, sends them to the SPARQL Endpoint and after receiving the answer, combines the returned data to set up the Virtual Learning Spaces [VLS].
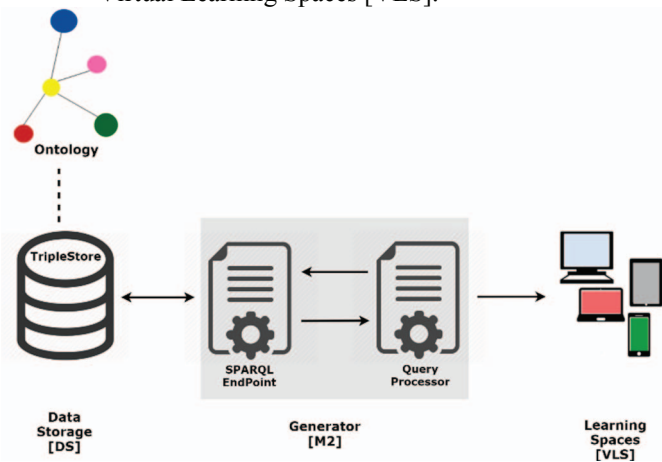

Figure 7. Module [M2] in Approach 1

In this approach, each Virtual Learning Space (a museum's exhibition room) is built fulfilling a web page template with the concrete data retrieved from the data store.

B. Approach 2

As in approach 1, the documents contained in the Repository of Museum of the Person must be stored in the Data Storage [DS]; however the upload process into a database is different from the previous one, and so the Ingestion Function [M1] must be adapted.
Now the input XML documents must be converted into SQL to populate the respective database.

Figure 8 describes the first Module [M1] that in this case has two blocks:
- *Parser and Semantic Checker* that reads the repository documents and extracts the relevant data (annotated in XML), checking their semantic consistency;
- *SQL Generator* that generates automatically the SQL statements that insert the retrieved data into the database tables.
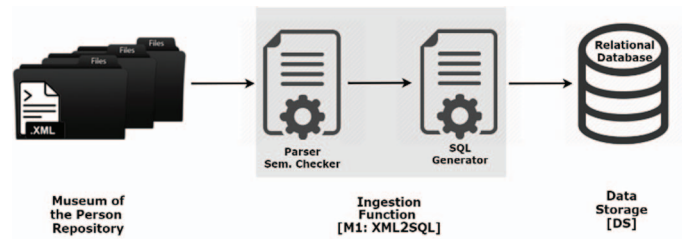

Figure 8. Module [M1] in Approach 2

After the two phases of Ingestion Function [M1] the documents data populated the Relational Database schema, due to the SQL statements generated.
As this schema is not directly related to the ontology, in this second approach an explicit mapping is necessary.
This is the first task that must be implemented by the second Module [M2].
After making this mapping available, it is possible to resort to CaVa (Criação de Ambientes Virtuais de Aprendizagem) system [4] to build automatically the Virtual Learning Spaces [VLS].
Notice that only the generator module of CaVa, CaVa^{Gen} will be used in this context.

Figure 9 sketches Module [M2] that in this case is composed of two parts:
- *DB2Onto Mapping* that associates concepts and relations belonging to the ontology with their respective instances stored in database (it allows to access database tables and fields to get the instances of the ontology concepts);

- CaVa^{Gen}, that generates automatically the Virtual Learning Spaces from their formal specification based on the ontology.
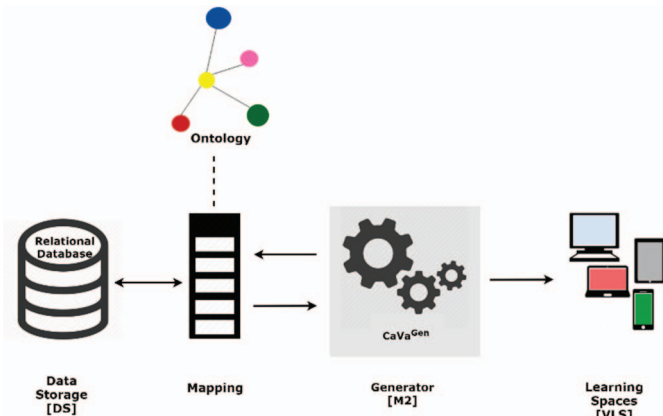

Figure 9. Module [M2] in Approach 2

In this second approach all the work concerned with the query generation according to the exhibition requirements and the answer processing to fulfill the rooms templates is left to CaVa^{Gen}.

Obviously this strategy saves a lot of development effort.

The only thing that is needed is the specification of the desired learning spaces in CaVa^{DSL}.

IV. CONCLUSION

Museum of the Person 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 [13]. This is a still alive project accessible at http://www.museudapessoa.net.

Our work concerns the Portuguese branch of such network of life stories museums, npMP. From the life stories of individuals, the objective is to write up the stories of families, communities, or institutions.

In this paper we propose different alternatives to build a software platform to create the museum's virtual exhibition rooms, as web pages, extracting information from the museum's repository, aiming at comparing them.

One approach uses a *TripleStore* to archive the ontology instances and resorts to SPARQL technology to query the repository and obtain the information that will be exhibited.

The other approach uses a *Relational Database* as archive and reuses CaVa framework to extract and display the information. CaVa is a novel proposal under development in the context of the PhD project of one of the authors, and our first objective was to use npMP as a second case study to test that framework.

After finishing the instantiation of the ontology we will start the implementation and testing of both architectures in order to realize positive and negative issues of each one, and to be able to provide different solutions to different repository implementations.

REFERENCES

[1] B. C. Rodrigues and G. Crippa, "Novas propostas e desafios das mediações culturais em museus virtuais," in El pensamiento museulógico contenporáneo. O pensamento museulógico contemporâneo, pages 599–608, ICOM, 2011, ISBN 978-92-9012-404-7.

[2] J. J. Almeida, J. G. Rocha, P. R. Henriques, S. Moreira, and A. Simões, "Museu da Pessoa – arquitetura," in Encontro nacional da associação de bibliotecários, arquivista e documentalistas, ABAD'01. BAD, 2001, URL http://repositorium.sdum.uminho.pt/handle/1822/585.

[3] P. B. Stafford, "The Museum of the Person: A global network for life stories," available at http://www.museumoftheperson.org, acessed in Feb. 2016.

[4] R. Martini, "Formal description and automatic generation of learning spaces based on ontologies," Phd pre-thesis, Universidade do Minho, 2015.

[5] C. Araújo, "An ontology for the Museum of the Person combining CIDOC-CRM with FOAF," MSc pre-thesis, Universidade do Minho, 2016.

[6] A. Simões and J. J. Almeida, "Histórias de vida + processamento estrutural = Museu da Pessoa," in XATA 2003 — XML: Aplicações e Tecnologias Associadas, page 16, Braga, Portugal, 2003, UM, URL http://repositorium.sdum.uminho.pt/handle/1822/629.

[7] R. Martini, C. Araújo, J. J. Almeida, and P. R. Henriques, "OntoMP, an ontology to build the Museum of the Person," in WorldCIST'2016 Proceedings, 2016.

[8] R. Martini, C. Araújo, G. Librelotto, and P. R. Henriques, "A reduced CRM-compatible form ontology for the virtual Emigration Museum," in WorldCIST'2016 Proceedings, 2016.

[9] ICOM/CIDOC, "Definition of the CIDOC conceptual reference model," Technical report, ICOM/CIDOC, May 2015.

[10] G. A. Grimnes, P. Edwards, and A. D. Preece, "Learning meta-descriptions of the foaf network," The semantic web - ISWC 2004: Third International Semantic Web Conference, Hiroshima, Japan, November 7-11, 2004, Proceedings, volume 3298, pages 152–165. Springer, 2004, ISBN 3-540-23798-4.

[11] D. Allemang and J. Hendler, "Semantic web for the working ontologist: effective modeling in RDFS and OWL," Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2 edition, 2011, ISBN 9780123859655, 9780123859662.

[12] M. M. Al-Mukhtar and A. T. Al-Assafy, "The implementation of foaf ontology for an academic social network," International journal of computer science engineering and technology (IJCSET), 4, 2014. ISSN 2231-0711.

[13] K. Worcman. "The museum of the person," in Virtual Museums, volume 57(3). ICOM, 2004.