

**Universidade do Minho**

Escola de Engenharia

Departamento de Informática

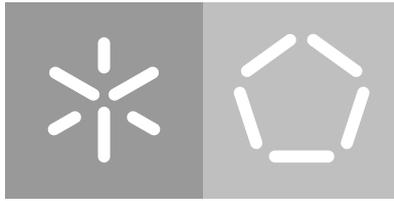
Cristiana Esteves Araújo

## **Building the Museum of the Person**

**Based on a combined  
CIDOC-CRM/ FOAF/ DBpedia Ontology**

December 2016





**Universidade do Minho**

Escola de Engenharia

Departamento de Informática

Cristiana Esteves Araújo

## **Building the Museum of the Person**

**Based on a combined  
CIDOC-CRM/ FOAF/ DBpedia Ontology**

Master dissertation  
Master Degree in Computer Science

Dissertation supervised by

**Supervisor:**

Pedro Manuel Rangel Santos Henriques

**Co-supervisor:**

Ricardo Giuliani Martini

---

## ACKNOWLEDGEMENTS

---

Throughout this Master's Dissertation I received collaboration and contribution, directly or indirectly, from various people to which I would like to address a few words of thanks and deep appreciation, in particular:

First of all, I would like to thank Professor Pedro Rangel Henriques by his unceasing presence, help, concern and encouragement along this journey. I appreciate all the availability ever revealed to me; for helping me to trace and focus on the most important goals. I want to also thank the coffee before our meetings, and for electing me "mummy" of our excellent working group. I will never forget all of this, and so a sincere and profound thanks for your friendship.

I want to thank also to Ricardo Martini for all the help, dedication and concern that has always shown to me. Thank you for helping me in doubts and difficulties that have emerged throughout this master's work.

I can not forget of course to thank Damien Vaz, his friendship, concern and company in the laboratory. Thanks Damien, Ricardo and Professor Pedro by the company in the coffee and the pleasant conversations over the same!

I would like to thank Professor José João Almeida for providing the assets of Museum of the Person. Without him this master's work was not possible. Thank you for all the help and support given, when asked to clarify doubts about the assets.

I want to give special thanks to My Father, My Mother and My Brother, a huge thank you for believing in me and in what I do, and all the teachings and values transmitted over the years. I will that this degree, I am achieving can reward you for all the love, support, dedication, concern and motivation to constantly give me.

Last but not least, I thank My Boyfriend for all the support, understanding and daily love, the transmission of confidence, encouragement and strength, at all times. Thank you for the joy, happy times that you provide me. Thank you for everything and for being always present!

---

## ABSTRACT

---

This document presents the work developed to fulfill the requirements for a Master Thesis in Software Engineering, in the areas of Virtual Museums, and Ontologies for knowledge representation and exploration.

The first objective of this thesis work was the creation of a specific ontology for the document repository of the Museum of the Person (Museu da Pessoa), using a standard for museums, CIDOC-CRM (Comité International pour la Documentation - Conceptual Reference Model), complemented with FOAF (Friend-of-a-Friend) and DBpedia that provides specific concepts and relations to deal with persons. This abstract ontology was then populated with life stories collected previously through of interviews of common people.

Two different approaches have been proposed to create the web pages for the Virtual Museum (VM), but only the approach 1 was implemented. A TripleStore was used as database to store all the information that constitutes the Museum assets; the VM was created consulting the datastore through SPARQL (SPARQL Protocol and RDF Query Language) queries.

In the dissertation will be discussed the design decisions, and provided the technical details; the project outcomes will be illustrated.

The npMP site created can be accessed at <http://npmp.epl.di.uminho.pt/> and complements this reading.

## KEYWORDS

Cultural Heritage; Virtual Museums; Museum of the Person; Ontologies; CIDOC-CRM; FOAF; DBpedia

---

## RESUMO

---

Este documento apresenta o resultado do trabalho relativo a uma Tese de Mestrado em Informática, nas áreas dos Museus Virtuais (VM, Virtual Museum), e Ontologias para representação do conhecimento e sua exploração.

O primeiro objectivo desta tese foi a criação de uma ontologia específica para o repositório de documentos do Museu da Pessoa (MP, Museum of the Person), utilizando um padrão para museus, CIDOC-CRM (Comité International pour la Documentation - Conceptual Reference Model), complementada com FOAF (Friend-of-a-Friend) e DBpedia as quais que fornecem conceitos e relações específicas para lidar com pessoas. Esta ontologia abstrata foi preenchida com histórias de vida recolhidas anteriormente através de entrevistas a pessoas comuns.

Dois abordagens diferentes foram concebidas para criar as páginas Web para o Museu Virtual, mas apenas a primeira abordagem foi implementada. Um TripleStore foi usado como base de dados para armazenar todas as informações que constituem o património do Museu; as salas de exposição do museu foram criadas através de consultas SPARQL (SPARQL Protocol and RDF Query Language) enviadas a esse TripleStore.

Na dissertação serão discutidas as decisões de design, e fornecidos os detalhes técnicos; os resultados do projeto serão também ilustrados. O site npMP criado, acessível em <http://npmp.epl.di.uminho.pt/>, complementa esta leitura.

## PALAVRAS-CHAVE

Herança Cultural; Museus Virtuais; Museu da Pessoa; Ontologias; CIDOC-CRM; FOAF; DBpedia

---

## CONTENTS

---

<b>1</b>	<b>INTRODUCTION</b>	<b>1</b>
1.1	Motivation	1
1.2	Objectives	2
1.3	Research Hypothesis	2
1.4	Thesis Organization	3
<b>2</b>	<b>MUSEUMS AND ONTOLOGIES</b>	<b>4</b>
2.1	Cultural Heritage	4
2.2	Ontologies	5
2.2.1	CIDOC-CRM Ontology	7
2.2.2	FOAF	9
2.2.3	DBpedia	10
2.3	Virtual Museums and Learning Spaces	13
2.4	Summary	13
<b>3</b>	<b>THE MUSEUM OF THE PERSON</b>	<b>15</b>
3.1	Maria Cacheira life story: An example	17
3.2	Summary	25
<b>4</b>	<b>PROPOSED APPROACHES TO BUILD THE MUSEUM OF THE PERSON</b>	<b>26</b>
4.1	System Architecture	26
4.2	Approach 1	27
4.3	Approach 2	29
4.4	Summary	31
<b>5</b>	<b>ONTOMP: ONTOLOGY FOR MUSEUM OF THE PERSON, EVOLUTION</b>	<b>32</b>
5.1	OntoMP: Original Design	34
5.2	OntoMP: CIDOC-CRM Representation	36
5.3	OntoMP: CIDOC-CRM / FOAF / DBpedia Representation	37
5.4	Summary	39
<b>6</b>	<b>DATA EXTRACTION AND STORAGE</b>	<b>41</b>
<b>7</b>	<b>QUERYING AND VISUALIZING</b>	<b>54</b>
<b>8</b>	<b>CONCLUSION</b>	<b>64</b>
8.1	Future Work	65
<b>A</b>	<b>ONTOMP: THE FULL ONTOLOGIES</b>	<b>71</b>
A.1	Original Design	71

## Contents

A.2	CIDOC-CRM Representation	73
A.3	CIDOC-CRM / FOAF / DBpedia Representation	75
B	VISITING MUSEUM OF THE PERSON	77

---

## LIST OF FIGURES

---

Figure 1	The core structure of the CIDOC-CRM ontology standard (Adapted from (Oldman and Labs, 2014))	8
Figure 2	FOAF Classes and Properties	10
Figure 3	Common DBpedia classes (Adapted from (Bizer et al., 2009))	12
Figure 4	Maria Alice Rodrigues Cacheira	20
Figure 5	General Approach to build the Museum of the Person	27
Figure 6	Module [M1] in Approach 1	28
Figure 7	Module [M2] in Approach 1	29
Figure 8	Module [M1] in Approach 2	30
Figure 9	Module [M2] in Approach 2	31
Figure 10	Reverse Engineering and mapping of Museum of the Person (Adapted from (Martini et al., 2016b))	33
Figure 11	An ontology <i>OntoMP: Original Design</i> for Museum of the Person (fragment)	35
Figure 12	An instance of <i>OntoMP: Original Design</i> for Maria Cacheira life story (fragment)	35
Figure 13	An instance of <i>OntoMP: CIDOC-CRM Representation</i> for Maria Cacheira life story (fragment)	37
Figure 14	An instance of <i>OntoMP: CIDOC-CRM / FOAF / DBpedia Representation</i> for Maria Cacheira life story (fragment)	39
Figure 15	Maria Cacheira life story triples in Turtle notation (fragment)	43
Figure 16	Architecture Ingestion Function [XML2RDF]	47
Figure 17	Response to the SPARQL query: <i>Interviewee and their photos</i>	60
Figure 18	Museum of the Person - Entrance Hall	62
Figure 19	Home Page Museum of the Person	63
Figure 20	An ontology <i>OntoMP: Original Design</i> for Museum of the Person	71
Figure 21	An instance of <i>OntoMP: Original Design</i> for Maria Cacheira life story	72
Figure 22	An ontology <i>OntoMP: CIDOC-CRM Representation</i> for Museum of the Person	73
Figure 23	An instance of <i>OntoMP: CIDOC-CRM Representation</i> for Maria Cacheira life story	74
Figure 24	An ontology <i>OntoMP: CIDOC-CRM / FOAF / DBpedia Representation</i> for Museum of the Person (fragment)	75

## List of Figures

Figure 25	An instance of <i>OntoMP: CIDOC-CRM / FOAF / DBpedia Representation</i> for Maria Cacheira life story	76
Figure 26	Workflow Visit to Museum of the Person	77
Figure 27	Main Page of the Museum of the Person	77
Figure 28	MP Visits: Entrance Hall	78
Figure 29	Room Projects and Life Stories	78
Figure 30	Room Projecto Afurada	79
Figure 31	Life Story of Interviewee: Maria Alice Rodrigues Cacheira	79
Figure 32	Types of Episodes of <i>Maria Alice Rodrigues Cacheira</i> life story	80
Figure 33	Room Interviewees by type of episode: General	81
Figure 34	Episode "A camisola castorina e os socos"	81

---

## LIST OF LISTINGS

---

3.1	Excerpt of the <i>Edited Interview</i> DTD . . . . .	16
3.2	DTD excerpt to describe an Event (Evento) . . . . .	17
3.3	BI of Maria Cacheira . . . . .	18
3.4	DTD of BI . . . . .	18
3.5	Identification of Maria Cacheira . . . . .	19
3.6	Photo caption . . . . .	20
3.7	DTD of Photo caption . . . . .	21
3.8	Episode - Christmas dinner on the floor (original interview) . . . . .	21
3.9	Episode - Christmas dinner on the floor (edited interview) . . . . .	22
3.10	Maria Cacheira's Wedding description (original interview) . . . . .	22
3.11	Maria Cacheira's Wedding description (edited interview) . . . . .	24
3.12	Event - River Douro's flood (original interview) . . . . .	24
6.1	Fragment of the RDF Triples for Maria Cacheira life story (Araújo et al., 2016a) . . . . .	41
6.2	An XML input document . . . . .	44
6.3	An RDF output document . . . . .	44
6.4	XML2RDF Lexer Grammar for AnTLR . . . . .	48
6.5	Lexer Grammar Photos main Mode . . . . .	48
6.6	Lexer Grammar Auxiliary Modes . . . . .	49
6.7	Lexer Grammar Print Mode . . . . .	51
7.1	Query SPARQL: 2. Interviewee and his photos . . . . .	54
7.2	Fragment of the Python script (first part) . . . . .	56
7.3	Fragment script Python (second part) . . . . .	57
7.4	Web form to interrogate Python script . . . . .	60

---

## INTRODUCTION

---

The society is more and more concerned with the preservation and the dissemination of Cultural Heritage, as works of art, ancient objects, documents, among others. 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 (Rodrigues and Crippa, 2011).

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 their 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 Museum) appeared, without any references to physical spaces (Rodrigues and Crippa, 2011).

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 . . .* (Schweibenz, 2004).

A Virtual Museum, such as a traditional museum, also acquires, conserves, and exhibits the heritage of humanity<sup>1</sup> creating a delightful environment for pleasure or enjoyment, as well as an appropriate place for teaching, and research. A virtual museum of people life stories allows visitors to interrelate information among different stories, to study social phenomena, natural disasters, migrations, season's customs, among others.

### 1.1 MOTIVATION

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 the life stories and to extract knowledge about an epoch and a society connecting and relating them.

---

<sup>1</sup> In that case, intangible objects, or immaterial things, according to <http://www.unesco.org/culture/ich/index.php?lg=en&pg=00022#art2>.

## 1.2. Objectives

More precisely we aim at rebuilding the Portuguese branch of the Museum of the Person network<sup>2</sup> that connect individuals and groups through sharing their life stories<sup>3</sup>.

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

### 1.2 OBJECTIVES

The masters' work here reported aims at the creation of the Portuguese MP (npMP) as a set of virtual Learning Spaces, using an ontology to describe the knowledge contained in the digital repository and to define the information that must be displayed. So the goals of this project are:

- To create a specific ontology for the document repository of the Museum of the Person, using a standard for museums, CIDOC-CRM (Comité International pour la Documentation - Conceptual Reference Model), enriched with concepts and relations defined in FOAF (Friend-of-a-Friend) and DBpedia;
- To populate the ontology with data collected from the interviews made in the context of some national projects;
- To store the ontology triples in a digital repository (a triple store or a relational database) the information collected;
- To build the Virtual Museum web pages to exhibit the collection of life stories, extracting the information from the stored ontology, and disseminating in this way the individual and collective knowledge held by the Museum of the Person.

### 1.3 RESEARCH HYPOTHESIS

This research project aims at demonstrating that it is possible to create automatically (in a systematic way) a Virtual Museum, the Museum of the Person, on the basis of an ontology that formally describes the documents repository (the museum's assets).

---

<sup>2</sup> This network includes branches in Brazil, Portugal, USA, Canada, etc.

<sup>3</sup> <http://www.museumoftheperson.org/about/>

## 1.4. Thesis Organization

### 1.4 THESIS ORGANIZATION

This document is organized in seven chapters. In Chapter 2 the concepts of Cultural Heritage, Ontologies, CIDOC-CRM, FOAF, DBpedia and Virtual Museum and Learning Spaces are introduced. The Museum of the Person and its documents, and an example of a life story are described in Chapter 3. In Chapter 4 a general architecture of the system is presented; two different instantiation approaches are then discussed. Chapter 5 proposes OntoMP, an ontology for the MP. This chapter presents the ontology original design, as well as its representation using the museums' standard CIDOC-CRM; finally an ontology representation in CIDOC-CRM complemented with FOAF and DBpedia elements is discussed. Appendix A will complement this chapter presenting the complete schema for all the three ontologies introduced, as well as the complete depiction of a full instantiation for each one. In Chapter 6 the extraction process and the storage strategy used for the Museum of the Person documents are described. The creation of the museum's web pages is discussed in Chapter 7; the querying process and data visualization are explained. Appendix B closes the dissertation offering a guided tour by the npMP exhibition rooms. Finally, in Chapter 8 a summary and directions for future research are presented.

---

## MUSEUMS AND ONTOLOGIES

---

In this chapter the main areas involved in this Master Thesis will be introduced and reviewed: Cultural Heritage, Ontologies, Virtual Museum and Learning Spaces.

The Cultural Heritage is an important term in this project because of the documentary collection of the Museum of the Person. The documents of Museum of the Person tell life stories of different common people, reflecting traditions, customs, memories . . . and so are Cultural Heritage of our country. All peoples make their contribution to the culture of the world. That is why it is important to respect and safeguard all Cultural Heritage. They represent history and identity; the link with the past, for the present and future.

Virtual Museums are referred because their construction is the target of this project. Learning Spaces are the building blocks for Virtual Museums in the perspective of the research group under which this project is being developed.

Ontologies are the formal description for knowledge domains crucial to the approaches that will be proposed.

### 2.1 CULTURAL HERITAGE

According the (UNESCO, 1989), Cultural Heritage *is the legacy of physical artifacts and intangible attributes of a group or society that are inherited from past generations, maintained in the present and bestowed for the benefit of future generations.*

*The Cultural Heritage is not just monuments and objects that have been preserved over the years. This also includes traditions, expressions of life, . . . , that people have inherited from their ancestors and vain passed from generation to generation (Waterton and Watson, 2013).*

*In other words, Cultural Heritage refers to cultural aspects, such as historic sites, monuments, folklore, activities and traditional practices, language, etc., which are considered vital to be preserved for future generations. Cultural Heritage offers people a connection to certain social values, beliefs, religions and customs. This allows them to identify with other mentality and similar backgrounds (Shi et al., 2008).*

## 2.2. Ontologies

*Cultural Heritage can provide an automatic sense of unity and belonging within a group and allows us to better understand previous generations and the story of where we came from (Shi et al., 2008).*

This way, Cultural Heritage can be characterized by physical artifacts, such as monuments, objects, among others, and intangible attributes, as expressions of life, traditions, among others. Thus, in this context, according to UNESCO, there are two types of Cultural Heritage (tangible<sup>1</sup> and intangible<sup>2</sup>):

- **Tangible** refers to significant places, worthy of preservation for the future, who defend the history and culture of the country. So material things, that is palpable. For example, monuments, mosques, shrines, monasteries, artifacts, among others.
- **Intangible** refers to aspects of a country that can not be touched or seen, so immaterial things. For example, traditional music, folklore, language, traditional craftsmanship, expressions, skills, performing arts, among others.

According to this definition, the "Intangible Cultural Heritage" concerns with the following five aspects: oral network and expressions, including languages as intangible cultural heritage medium; performing arts; social conventions, rituals, and festivals; knowledge and practices about nature and universe; traditional handicraft skills (Shi et al., 2008).

It should be noted that the Museum of the Person documents are part of the Intangible Cultural Heritage.

The next section will be introduced to the concept of Ontologies. In the field of Ontologies will be described three examples that will be used in this project: CIDOC-CRM, FOAF and DBpedia.

## 2.2 ONTOLOGIES

In the branch of philosophy, ontology refers to ... *the study; the area of metaphysics that relates to Being or essence of things, or Be in the abstract sense (Silva et al., 2009). The appropriation of the term "Ontology" of philosophy by the computing community was due to the fact the ontologies serve as means of organizing the things amenable to symbolic representation (formal representation), and from the formal representation, enable deductive reasoning through inferences rules (Silva et al., 2009).*

According to Gruber, *an ontology is an explicit specification of a conceptualization (Gruber, 1993). The term is borrowed from philosophy, where an Ontology is a systematic account of Existence. When the knowledge of a domain is represented in a declarative formalism, the set of objects that can be represented is called the universe of discourse. In such an ontology, definitions associate the*

<sup>1</sup> Available at: <http://www.unesco.org/new/en/cairo/culture/tangible-cultural-heritage/>

<sup>2</sup> Available at: <http://www.unesco.org/culture/ich/en/convention#art2>

## 2.2. Ontologies

*names of entities in the universe of discourse (e.g., classes, relations, functions, or other objects) with human-readable text describing what the names mean, and formal axioms that constrain the interpretation and well-formed use of these terms.*

Borst also features a very accepted, by the community, definition of ontology: *a formal of a shared conceptualization*(Borst, 1997). Guarino describes an ontology *as an artifact of engineering, consisting of intentional vocabulary related to a certain reality, together with explicit assumptions in the form of first-order logic, representing concepts and relationships between concepts*(Guarino, 1997).

According McGinness and Harmelen, there are five items that describe briefly the reasons to develop and create an ontology (Noy and McGinness, 2001):

- To share common understanding of the structure of information among software agents or people;
- To enable reuse of domain knowledge;
- To make domain conjecture explicit;
- To separate domain knowledge from the operational knowledge;
- To analyze domain knowledge.

The ontologies can be set for various languages. (McGinness and Harmelen, 2004) in agreement with an ontology can be represented in:

- XML (eXtensible Markup Language), which provides a surface syntax for structured documents, but imposes no semantic constraints on the meaning of these documents.
- XML-Schema, is a language that restricts the structure of XML documents and also extends XML with datatypes.
- RDF (Resource Description Framework) is a standard model for data interchange on the Web. This model has objects and relations between them and provides a simple semantic.
- RDF Schema is a vocabulary that provides a data-modelling vocabulary for RDF data, with a semantics for generalization-hierarchies of such properties and classes.
- OWL (Web Ontology Language) adds more vocabulary for describing properties and classes, for example, relations between classes cardinality, equality, richer typing of properties, characteristics of properties and enumerated classes;
- SKOS (Simple Knowledge Organization System) is a standard way to represent knowledge organization systems using the RDF. Encoding this information in RDF allows it to be passed between computer applications in an interoperable way.

## 2.2. Ontologies

The following Sections (2.2.1, 2.2.2 and 2.2.3) describe three examples of the ontologies used in this project, the CIDOC-CRM, the FOAF, and the DBpedia, respectively.

### 2.2.1 CIDOC-CRM Ontology

CIDOC-CRM came from the CIDOC Documentation Standards Group in the International Committee for documentation of the International Council of Museums, and was accepted as ISO 21127, in 2006. CIDOC-CRM (Comité International pour la Documentation - Conceptual Reference Model) is an ontology in the domain of Cultural Heritage, created as a reference ontology for integration of information (Oldman and Labs, 2014; ICOM/CIDOC, 2015).

*The CRM model aims to provide a common language for heterogeneous information systems, and allows their integration, despite possible semantic and structural incompatibilities. In this way, Cultural Heritage information can be exchanged and retrieved, and Cultural Heritage institutions can make their interoperable information systems, without compromising their specific needs or the current level of accuracy of your data (Oldman and Labs, 2014; ICOM/CIDOC, 2015).*

*CIDOC-CRM fosters a shared understanding of Cultural Heritage information by providing a common and extensible semantic framework, in which any Cultural Heritage information can be mapped. It is intended to be a universal language for domain experts and implementers, aiming to formulate requirements for information systems and serves as a guide to good practice of modeling concepts. Thus, CIDOC-CRM can provide a "bridge" required to do the mediation between different sources of information of Cultural Heritage, such as those published by the museums, libraries and archives (Oldman and Labs, 2014; ICOM/CIDOC, 2015).*

*CIDOC-CRM is a top-level ontology that allows the integration of Cultural Heritage information and its correlation with information of museums, libraries, and archives which, therefore, can be easily converted in other formats readable by machine, such as RDF/XML. Probably the most ambitious application of CIDOC-CRM is to develop integrated query tools, mediation systems and data storage (Oldman and Labs, 2014).*

There exist a huge amount of information stored in library catalogs, inventories, and collections of museums, that is difficult to consult and that is isolated from the related collections. *Different sources of information typically need to be consulted individually, and the links between the systems are scarce. The ability to reconcile and integrate information from various sources has the potential to add significant value to existing data facilitating the research and improving the quality of the user experience (Crofts et al., 2002).*

Figure 1 illustrates the main entities of the CIDOC-CRM.

CIDOC-CRM is event-based. At the core of this event model are *Temporal Entities*, that, the things that happen in a given period of time (past, present and future):

- **Temporal Entity** is a class used to represent anything (particularly events) related to the events duration, to allow the definition of time slots. Actor, Conceptual Object

## 2.2. Ontologies

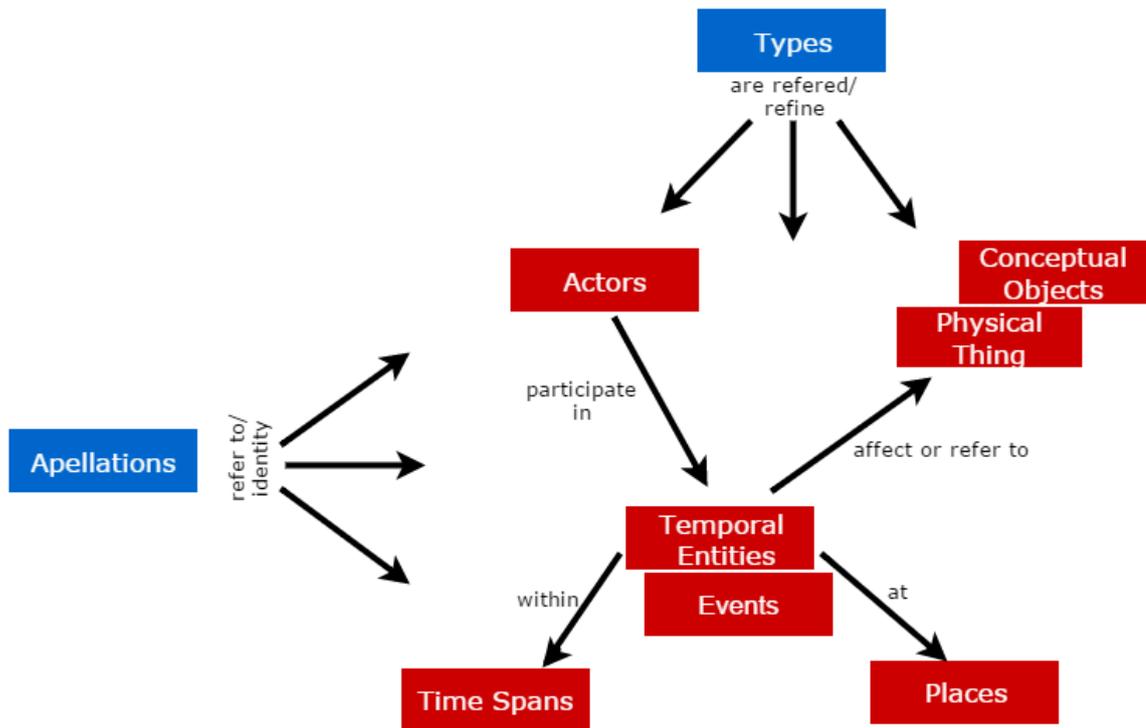


Figure 1.: The core structure of the CIDOC-CRM ontology standard (Adapted from (Oldman and Labs, 2014))

and Physical Thing may not be directly related to time. Thus, they must be associated with an event, this is, a temporal entity;

- **Place** is the location's description, such as a geographical location, an address, a country, among others;
- **Actor** can be an individual or a group. An individual can be a person and a group can be a company, for example. Actors can interact with things (Conceptual Object and Physical Things) through events;
- **Physical Thing** is something that can be physically destroyed or transformed into something else;
- **Conceptual Object** it is something that cannot be destroyed (immaterial);
- **Appellation** is used for naming things in CIDOC-CRM, for example, names of people, places, among others;
- **Type** is used to sort things through properties associated with these things, for example, floods, marriage, birth of a person, etc.

## 2.2. Ontologies

As CIDOC-CRM does not describe some characteristics of people, has the intention of re-searching how to extend CIDOC-CRM using FOAF<sup>3</sup> (Friend-of-a-Friend) complemented with DBpedia<sup>4</sup>, since both contain a vocabulary specific to describe individuals, their activities and their relations with other people and objects.

In the next section will be described ontology, the FOAF, which will be used to build Museum of the Person.

### 2.2.2 FOAF

FOAF<sup>5</sup> (Friend-of-a-Friend), *is a project devoted to linking people and information using the Web; like the Web itself, is a linked information system. FOAF defines a dictionary of people-related terms that can be used in structured data environments; actually it is a dictionary of named properties and classes using W3C's RDF technology. It is built using decentralized Semantic Web technology, and is designed to allow integration of data through a wide variety of applications, such as websites and software services and systems, among others (Brickley and Miller, 2016).*

FOAF *is a machine-readable ontology describing persons, their activities and their relations to other people and objects (Grimnes et al., 2004; Allemang and Hendler, 2011). The FOAF ontology describes two areas of digital identity information: biographical information and social network information (Al-Mukhtar and Al-Assafy, 2014).*

*It is important to stress that the FOAF vocabulary is not a standard in the sense of ISO Standardization, or that associated with W3C Process. FOAF depends on W3C's standards work, specifically on XML, XML Namespaces, RDF and OWL. FOAF is a descriptive vocabulary expressed in RDF (Resource Description Framework) and OWL (Web Ontology Language) (Grimnes et al., 2004).*

*The FOAF vocabulary can be described by the terms (RDF classes and properties) that constitute it, so that Semantic Web systems can use these terms in a variety of RDF-compatible document formats and applications. FOAF vocabulary is defined by the namespace <http://xmlns.com/foaf/0.1/> (Grimnes et al., 2004; Allemang and Hendler, 2011).*

FOAF collects a variety of terms that describe the characteristics of individuals and social groups, and documents that are independent of time and technology. These terms may be used to describe the basic information about the people in our day, history, Cultural Heritage and digital library contexts (Brickley and Miller, 2016).

Figure 2 shows in alphabetic order the classes and properties belonging to the FOAF collection.

In the next section, will be described the ontology DBpedia, which will be used in conjunction with FOAF to describe people of Museum of the Person.

<sup>3</sup> Homepage: <http://www.foaf-project.org/>

<sup>4</sup> Homepage: <http://http://wiki.dbpedia.org/>

<sup>5</sup> Homepage: <http://www.foaf-project.org/>

## 2.2. Ontologies

Classes	Agent	Document	Group	Image
	LabelProperty	OnlineAccount	OnlineChatAccount	OnlineEcommerceAccount
	OnlineGamingAccount	Organization	Person	PersonalProfileDocument
	Project			
Properties	account	accountName	accountServiceHomepage	age
	aimChatID	based_near	birthday	currentProject
	depiction	depicts	dnaChecksum	familyName
	family_name	firstName	focus	fundedBy
	geekcode	gender	givenName	givenname
	holdsAccount	homepage	icqChatID	img
	interest	isPrimaryTopicOf	jabberID	knows
	lastName	logo	made	maker
	mbox	mbox_sha1sum	member	membershipClass
	msnChatID	myersBriggs	name	nick
	openid	page	pastProject	phone
	plan	primaryTopic	publications	schoolHomepage
	sha1	skypeID	status	surname
	theme	thumbnail	Tipjar	title
	topic	topic_interest	Weblog	workInfoHomepage
	workplaceHomepage	yahooChatID		

Figure 2.: FOAF Classes and Properties

### 2.2.3 DBpedia

The DBpedia project is a community effort to extract structured information from Wikipedia and to make this information accessible on the Web (DBpedia, 2016; Lehmann et al., 2015).

The DBpedia knowledge base currently consists of around 274 million RDF Triples, which have been extracted from the English, German, French, Spanish, Italian, Portuguese, ..., versions of Wikipedia. It features labels and short abstracts in 30 different languages; 609,000 links to images; 3,150,000 links to external web pages; 415,000 Wikipedia categories, and 286,000 YAGO categories (DBpedia, 2016; Bizer et al., 2009; Lehmann et al., 2015).

The resulting knowledge base currently describes more than 2.6 million entities, covering 198,000 persons, 328,000 places, 101,000 musical works, 34,000 films and 20,000 companies. The knowledge base contains 3.1 million links to external web pages; and 4.9 million RDF links in other Web

## 2.2. Ontologies

*data sources. For each of these entities DBpedia defines a unique identifier that can be dereferenced (according to the principles Linked Data) through the web in a rich description of the entity RDF (DBpedia, 2016; Bizer et al., 2009).*

Figure 3 shows an overview of the most common DBpedia classes, and displays the number of instances, as well as some example properties for each class.

The DBpedia project contributes to the development Web linked data, in the following aspects (Bizer et al., 2009):

- develops an information extraction framework that converts Wikipedia content into a rich multi-domain knowledge base. By accessing the Wikipedia live article update feed, the DBpedia knowledge base timely reflects the actual state of Wikipedia. A mapping from Wikipedia infobox templates to an ontology increases the data quality.
- defines a Web-dereferenceable identifier for each DBpedia entity. This helps to overcome the problem of missing entity identifiers that has hindered the development of the Web of Data so far and lays the foundation for interlinking data sources on the Web.
- publishes RDF links pointing from DBpedia into other Web data sources and support data publishers in setting links from their data sources to DBpedia. The result was the emergence of a Web of Data around DBpedia.

*The base DBpedia knowledge has several advantages over the other knowledge bases: it covers many domains, it represents real community agreement, it automatically evolves as Wikipedia changes, it is truly multilingual, and it is accessible on the Web (DBpedia, 2016; Bizer et al., 2009).*

Next section will cover key concepts, which are the ultimate objective of this thesis, Learning Spaces and Virtual Museum.

## 2.2. Ontologies

Ontology Class	Instances	Example Properties
Person	198,056	name, birthdate, birthplace, employer, spouse
Artist	54,262	activeyears, awards, occupation, genre
Actor	26,009	academyaward, goldenglobeaward, activeyears
MusicalArtist	19,535	genre, instrument, label, voiceType
Athlete	74,832	currentTeam, currentPosition, currentNumber
Politician	12,874	predecessor, successor, party
Place	247,507	lat, long
Building	23,304	architect, location, openingdate, style
Airport	7,971	location, owner, IATA, lat, long
Bridge	1,420	crosses, mainspan, openingdate, length
Skyscraper	2,028	developer, engineer, height, architect, cost
PopulatedPlace	181,847	foundingdate, language, area, population
River	10,797	sourceMountain, length, mouth, maxDepth
Organisation	91,275	location, foundationdate, keyperson
Band	14,952	currentMembers, foundation, homeTown, label
Company	20,173	industry, products, netincome, revenue
Educ.Institution	21,052	dean, director, graduates, staff, students
Work	189,620	author, genre, language
Book	15,677	isbn, publisher, pages, author, mediatype
Film	34,680	director, producer, starring, budget, released
MusicalWork	101,985	runtime, artist, label, producer
Album	74,055	artist, label, genre, runtime, producer, cover
Single	24,597	album, format, releaseDate, band, runtime
Software	5,652	developer, language, platform, license
TelevisionShow	10,169	network, producer, episodenummer, theme

Figure 3.: Common DBpedia classes (Adapted from (Bizer et al., 2009))

### 2.3. Virtual Museums and Learning Spaces

#### 2.3 VIRTUAL MUSEUMS AND LEARNING SPACES

According to ICOM<sup>6</sup>, a *Traditional Museum is a non-profit institution, which has a permanent service to society and its development, open to the general public. The Traditional Museum acquires, conserves, researches, communicates and exhibits the tangible and intangible heritage of humanity and its environment for the purposes of education and study.*

A recent concept of museum, other than the Museum of "brick and mortar", is becoming in vogue as it requires no physical space: it is known as virtual museum (Schweibenz, 2004).

Virtual Museum can be a collection of digital objects composed of a variety of media which, because of its ability to provide connectivity and multiple points of access, transcends the traditional methods of communication and interaction with visitors. This type of museum, without a physical location its objects and related information can be disseminated around the world easily and in a broader way (Schweibenz, 2004).

The Virtual Museum is no competitor to the museum of "brick and mortar" because, being digital its nature, it cannot provide real objects to the visitors, as the Traditional Museum. But on the other way around, it can extend the ideas and the concepts of collections to the digital space, and thus reveal the essential nature of the museum. Moreover, the Virtual Museum will reach the virtual visitors who cannot visit a particular museum personally (Schweibenz, 2004).

A concept that can be related to the Virtual Museum is the concept of Virtual Learning Spaces (LS). *Virtual Learning Spaces are virtual spaces that offer different points of access to their virtual visitors. The information is presented in a manner geared to the context rather than being oriented to objects. In addition, the virtual space providing diversified linked information in a attractive interface captivates easily the attention of the visitor and can be seen as a teacher motivating him to learn a specific topic—in this sense, that space can be thought as a Virtual Learning Space (Schweibenz, 2004).*

#### 2.4 SUMMARY

Throughout this chapter some crucial areas for this masters thesis were introduced. The Cultural Heritage is an important concept because the documents of Museum of the Person are about life stories of different common people, reflecting memories, traditions, among others. Virtual Museums are important because their construction is the objective of this project, and the building blocks for Virtual Museums are the Learning Spaces. Rua and Ferreira, in their recent dissertations in the context of the research work for the Digital Museum of U. Porto, define in more detail the concepts that were discussed in this chapter

---

<sup>6</sup> Available at: <http://icom.museum/the-vision/museum-definition/>

#### 2.4. Summary

(Cultural Heritage, Traditional Museum, and Virtual Museum, as well as the related concept of Digital Museum) (Rua, 2016; Ferreira, 2016). Ontologies are the formal description for knowledge domains crucial to the approaches that will be proposed. Among all ontologies some of them deserved a deeper study because they were chosen to be used in this project: CIDOC-CRM universal standard for museums; FOAF and DBpedia to describe people.

After the presentation and discussion of the main concepts involved in this masters' work—aiming at enhancing the importance of ontologies to build web learning spaces to create virtual museums—next chapter will be devoted to the introduction of Museum of the Person and the analysis of its documents.

---

## THE MUSEUM OF THE PERSON

---

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 common people (Simões and Almeida, 2003; Worcman, 2004; Stafford, 2015). This is a still alive project accessible at <http://www.museudapessoa.net>. Museum of the Person aims at gathering testimonials from every human being, famous or anonymous, to perpetuate his history (Almeida et al., 2001; Stafford, 2015).

From the life stories of individuals, the objective is to write up the story of families, communities, or institutions (Stafford, 2015). This museum deals with common people, human beings, not with the 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 (Almeida et al., 2001). Actually, the narrators, to report their life stories during a predefined structured interview<sup>1</sup>, remember events and other 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 (Almeida et al., 2001).

The workflow adopted by MP technicians to acquire common people life stories is expressed below:

1. the report of a participant is recorded (audio or video) by an interviewer. Although every interview is a unique thing, interviewers conduct it following a sequence of predefined topics in order to cover the entire life story;
2. interviews are transcribed;
3. transcriptions are annotated in XML (eXtensible Markup Language), marking events, self contained stories, etc.;
4. XML interviews may be used to produce several outputs.

---

<sup>1</sup> Recorded in a tape or a film.

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 (Almeida et al., 2001).

The Museum of the Person's collection consists of a set of XML documents connected to each participant. Typically each interview is split into three parts (Simões and Almeida, 2003):

- 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<sup>2</sup>;
- 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.  
The edited file (an example can be seen in Listing 3.1) 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 marking are *given or names, jobs, places*, etc.;
- photographs and their legend. The legend document contains a section for each photo or scanned document as file name and a caption. This legend includes a description of the image and the date, and wherever possible the name of the stakeholder.

Aside 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) designed specially for that purpose called MPDTD. 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 3.1 displays a small excerpt from the MPDTD.

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

Listing 3.1: Excerpt of the *Edited Interview* DTD

<sup>2</sup> Bilhete de Identidade, in Portuguese.

### 3.1. Maria Cacheira life story: An example

The root element (`mp`) of the Museum of the Person DTD is a life story composed of one or more life story elements. Each life history element (`histVidaElem`) 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 3.2):

```
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"
11 >
```

Listing 3.2: DTD excerpt to describe an Event (Evento)

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

Chapter 5 outlines how the Museum of the Person's life stories were described as an ontology, aiming at creating a more abstract and formal level of description.

Before going into that topic, this chapter is closed with the description of a concrete life story that will be used as a running example.

#### 3.1 MARIA CACHEIRA LIFE STORY: AN EXAMPLE

Projecto Afurada was one of the sub-projects carried out by Museum of the Person- Portuguese branch (npMP) in the context of Porto 2001, the European Capital of Cultural event. The main objective of this project was the reviving of some ancient boroughs and traditional folk culture. In this project, Maria Alice Rodrigues Cacheira was one of the persons interviewed.

Listing 3.3 displays the bi that contains a mini-biography of Maria Cacheira, containing personal data, date and place of birth, job, among others.

### 3.1. Maria Cacheira life story: An example

```
1 <?xml version="1.0" encoding="ISO-8859-1"?>
2 <bi>
3   <projecto>Projecto da Afurada</projecto>
4   <depoente>Maria Alice Rodrigues Cacheira</depoente>
5   <entrevistador>Ana Pereira; Guilherme Soares e Maria Jose da
6     Cunha</entrevistador>
7
8   <biografia>Maria Alice Rodrigues Cacheira nasceu na Afurada, a
9     8 de Outubro de 1946. Andou na escola ate a <habilitacoes
10      quem="Maria Cacheira" nivel="EB"><escolaridade>quarta
11      classe</escolaridade> </habilitacoes>.
12
13     Perdeu a mae aos 11 anos e abandonou os estudos. Como era a filha
14     mais velha, passou a infancia a cuidar dos tres irmaos e a
15     ajudar o pai, que era pescador, a vender o peixe. Passados 18
16     meses o pai casou com uma tia e dessa relacao nasceram sete
17     filhos.
18
19     Casou aos 21 anos com um ajudante de motorista de barco. Apos
20     seis anos de casamento o marido morreu com leucemia deixando
21     tres filhas ainda pequenas. Comecou a trabalhar num matadouro
22     e, posteriormente, a vender peixe. Tem seis netos e trabalha
23     na <local de="Local de trabalho" quem="Maria Cacheira" >junta
24     de freguesia de Sao Pedro da Afurada</local>.
25
26     <local de="Local de trabalho" quem="Maria Cacheira" >Ao sabado
27     ainda vende peixe no Arrabida.</local></biografia>
28
29     <data ano="2000" mes="12" dia="12"/>
30     <profissao>peixeira; empregada de limpeza</profissao>
31     <nascimento onde="Afurada" ano="1946" mes="10" dia="08"/>
32 </bi>
```

Listing 3.3: BI of Maria Cacheira

This BI was written according to the DTD shown in Listing 3.4. The root element is BI and can be composed of the names of the project and of the deponent, the interviewer, date, biography, address, job, birth, photo, etc. The photo element identifies a photograph that can be chosen to be shown along with the mini-biography.

```
1 <?xml version="1.0" encoding="ISO-8859-1"?>
2 <!ELEMENT bi (projecto, ((depoente, entrevistador?, transcritor?,
3   editor?) | autor), foto?, titulo?, biografia?, data?,
4   profissao?, nascimento?, morada?, notas?,rel*)>
5 <!ELEMENT foto (#PCDATA)>
```

### 3.1. Maria Cacheira life story: An example

```
4 <!ATTLIST foto
5   ficheiro CDATA #REQUIRED
6 >
7 <!ELEMENT projecto (#PCDATA)>
8 <!ELEMENT titulo (#PCDATA)>
9 <!ELEMENT notas (#PCDATA)>
10 <!ELEMENT depoente (#PCDATA)>
11 <!ELEMENT entrevistador (#PCDATA)>
12 <!ELEMENT editor (#PCDATA)>
13 <!ELEMENT transcritor (#PCDATA)>
14 <!-- NOTA: biografia deve ler lida como 'resumo' no caso de
15     historias -->
16 <!ELEMENT biografia (#PCDATA)>
17 <!ELEMENT autor (#PCDATA)>
18 <!ELEMENT data EMPTY>
19 <!ATTLIST data
20   dia CDATA #IMPLIED
21   mes CDATA #REQUIRED
22   ano CDATA #REQUIRED
23 >
24 <!ELEMENT profissao (#PCDATA)>
25 <!ELEMENT nascimento EMPTY>
26 <!ATTLIST nascimento
27   dia CDATA #IMPLIED
28   mes CDATA #IMPLIED
29   ano CDATA #REQUIRED
30   onde CDATA #IMPLIED
31 >
32 <!ELEMENT morada (#PCDATA)>
33 <!ELEMENT escolaridade (#PCDATA)>
34 <!ATTLIST depoente
35   estadocivil (solteiro | casado | viuvo | divorciado | junto)
36   #IMPLIED
37 >
38 <!ELEMENT rel (#PCDATA)>
39 <!ATTLIST rel tipo CDATA #REQUIRED>
```

Listing 3.4: DTD of BI

In Listing 3.5, the short identification of Maria Cacheira is composed of a brief self-introduction and the link to her photo with a dated caption.

```
1 <identificacao><p>Chamo-me Maria Alice Rodrigues Cacheira. Nasci
2   e vivi sempre na Afurada. Nunca conheci outra terra.</p>
```

### 3.1. Maria Cacheira life story: An example

```
2 <foto ficheiro="090-F-02.b.jpg">Maria Alice Rodrigues Cacheira a  
12 de Julho de 2001 na Afurada</foto> </identificacao>
```

Listing 3.5: Identification of Maria Cacheira

Figure 4 is the photo of Maria Cacheira referred above in Junta de Freguesia da Afurada.

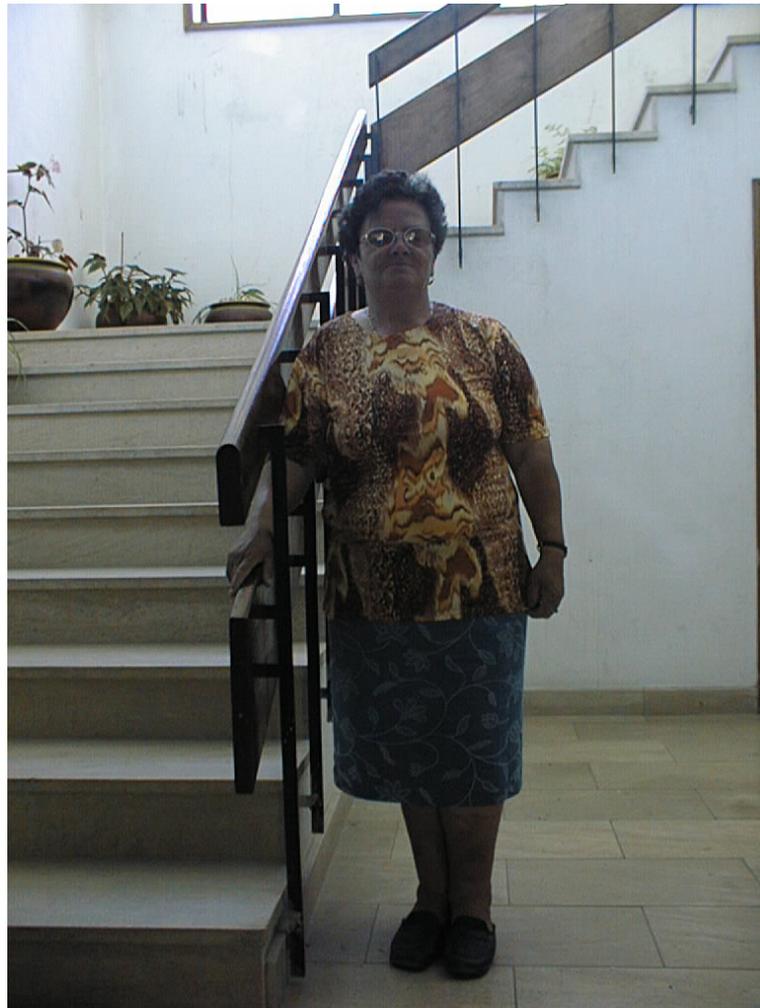


Figure 4.: Maria Alice Rodrigues Cacheira

The caption of this photo (Figure 4) contains the name of the person in the photo (Maria Alice Rodrigues Cacheira), the place (Junta de Freguesia da Afurada) and the date on which it was taken the picture (2001-07-12), as can be seen in Listing 3.6.

```
1 <?xml version="1.0" encoding="ISO-8859-1"?>  
2 <fotos>
```

### 3.1. Maria Cacheira life story: An example

```
3 <foto ficheiro="090-F-01.jpg"><quem>Maria Alice Rodrigues  
Cacheira</quem><onde>Junta de Freguesia da  
Afurada</onde><quando data="2001-07-12"></quando></foto>  
4 </fotos>
```

Listing 3.6: Photo caption

The DTD used to describe a photo (as illustrated above) is displayed in Listing 3.7. The root element is `photos` that can contain more than just one photo and more than one author. The photo element consists of the people depicted in the photo, the location and date the photo was taken, the author, etc.

```
1 <!ELEMENT fotos (autor*,foto*)>  
2 <!ELEMENT foto (quem, onde?, quando?, facto?, legenda?,autor?)>  
3 <!-- ATTLIST foto ficheiro CDATA #REQUIRED -->  
4 <!ELEMENT quem (#PCDATA)>  
5 <!ELEMENT quando (#PCDATA)>  
6 <!-- ATTLIST quando data CDATA #IMPLIED -->  
7 <!ELEMENT onde (#PCDATA)>  
8 <!ELEMENT facto (#PCDATA)>  
9 <!ELEMENT legenda (#PCDATA)>  
10 <!ELEMENT autor (#PCDATA)>  
11 <!-- ATTLIST autor tipo CDATA #IMPLIED -->
```

Listing 3.7: DTD of Photo caption

It is possible to read in Listing 3.8 an episode narrated by Maria Cacheira. This excerpt (markup in XML) was taken from the original interview.

As mentioned in the beginning of Chapter 3 the interview contains some key concepts that are part of the *Thesaurus*, for example `<rel tipo="festa tradicional">Natal</rel>`. There are also metadata markup, in this case we have, for example, a date `quando="24 de Dezembro"` and a term definition `<expressao significado="tecido grosseiro feito de materiais vegetais entrelacados">esteiras</expressao>`.

```
1 <infancia titulo="Comer no chao" quando="24 de Dezembro">  
2 <pergunta>E o seu pai, nao olhava pelos filhos tambem?</pergunta>  
3 <resposta><p> O meu pai tambem tinha uma vida muito apurada. O  
meu pai ia para o mar, podia ir a meia noite, podia ir a uma  
hora, nao tinha hora. Podia ir de dia, todo o dia, tinha de  
pescar a noite. Depois mesmo quando nao pescasse nao podia  
estar a olhar pelos filhos. Mas gostava muito dos filhos.  
Quando comia, deitava-se um bocadinho no chao a descansar e os  
filhos e os netos iam para cima dele, deitado no chao. E era  
assim! Pelo Natal, nos comamos no chao. Ainda mantenho a  
tradicao na minha casa.</p></resposta>
```

### 3.1. Maria Cacheira life story: An example

```
4
5 <pergunta>Comer no chao no Natal?</pergunta>
6 <resposta><p>Comer no chao no dia de <rel tipo="festa
   tradicional">Natal</rel> porque se usava dantes. Andava para
   ai um senhor a vender umas <expressao significado="tecido
   grosseiro feito de materiais vegetais
   entrelacados">esteiras</expressao>, e nos juntavamos as
   esteiras e comiamos no chao. Ainda hoje mantenho a tradicao.
7 Uma tradicao diferente!</p>
8
9 <p>Eu tenho duas mesas, uma mesa na sala de jantar e uma mesa de
   cozinha. Mas eu arrumo a casa e boto uns cobertores no chao,
   uns cobertores mais escuros e quero comer no chao. Quero
   manter a tradicao. E nao quero outra comida senao bacalhau com
   batatas e couves.
10 Essa e tradicao!</p></resposta></infancia>
```

Listing 3.8: Episode - Christmas dinner on the floor (original interview)

Listing 3.9 shows the same excerpt of Listing 3.8 but after being edited. Between the original and the edited interview there are a few differences, the most notorious is the fact that edited no longer contains the question-answer structure.

```
1 <casa titulo="Comer no chao">
2 <p>Pelo Natal, nos comiamos no chao. Ainda mantenho a tradicao
   na minha casa.
3 Comer no chao no dia de Natal porque se usava dantes. Andava para
   ai um senhor a vender umas esteiras, e nos juntavamos as
   esteiras e comiamos no chao. Ainda hoje mantenho a tradicao.
4 Uma tradicao diferente!</p>
5
6 <p>Eu tenho duas mesas, uma mesa na sala de jantar e uma mesa de
   cozinha.
7 Mas eu arrumo a casa e boto uns cobertores mais escuros no chao e
   quero comer no chao. Quero manter a tradicao. E nao quero
   outra comida senao bacalhau com batatas e couves.
8 Essa e tradicao!</p></casa>
```

Listing 3.9: Episode - Christmas dinner on the floor (edited interview)

Listing 3.10 is a description of Maria Cacheira marriage. Here she tells how old she was at marriage, the birth of children, the disease and death of her husband, among other episodes.

### 3.1. Maria Cacheira life story: An example

```
1 <pergunta>E casada?</pergunta>
2 <resposta><p>Sou viuva ha 26 anos.</p></resposta>
3
4 <pergunta>E o seu marido era pescador?</pergunta>
5 <resposta><p>Era ajudante de motorista numa traineira da
6 sardinha.</p></resposta>
7
8 <pergunta>Quer contar-nos um bocadinho como e que conheceu o seu
9 marido, como e que casou?</pergunta>
10 <resposta><p>Foi em crianca, foi na escola que comecei a namorar.
11 A minha mae quando morreu, eu tinha 11 anos menos vinte e tres
12 dias e ela tinha saudades do genro. Nos ja namoravamos. Depois
13 os pais dele nao queriam, mas depois nos casamos. Estive
14 casada seis anos e cinco meses.</p></resposta>
15
16 <pergunta>Com quantos anos e que casou?</pergunta>
17 <resposta><p>Casei aos 21.</p></resposta>
18
19 <pergunta>E o seu marido morreu como?</pergunta>
20 <resposta><p>Uma leucemia no sangue.</p></resposta>
21
22 <pergunta>E a senhora tem filhos?</pergunta>
23 <resposta><p>Tenho tres. Fiquei com uma de dois meses, uma de 2
24 anos e uma de 6 anos. Hoje ja tem uma 33, uma 28 e uma
25 26.</p></resposta>
26
27 <pergunta>Olhe, e depois de casar, como e que fez? Saiu de casa
28 de seus pais?</pergunta>
29 <resposta><p>Sai de casa ja antes, dois ou tres meses. Ja tinha
30 uma casa alugada e fui morar com o meu marido. Depois passado
31 dois meses de casar nasceu a minha primeira filha, que eu ja
32 ia de bebe. Foi servico adiantado. Agora nao se usa mais levar
33 filhos antes de casar porque elas tomam comprimidos ou outras
34 coisas, senao usava-se. Mas eu adiantei o servico.</p>
35 <p>Fui feliz no tempo em que vivi com o meu marido. Mas a minha
36 sogra era muito intrometida na minha vida. Ja morreu essa
37 macaca. Vou todas as semanas ao cemiterio e a beira da campa
38 dela nunca vou.</p></resposta>
39
40 <pergunta>Mas foi feliz com o seu marido?</pergunta>
41 <resposta><p>Pouco tempo, mas fui. Mas se eu soubesse que era
42 para ficar viuva tao nova nao tinha casado.</p></resposta>
```

Listing 3.10: Maria Cacheira's Wedding description (original interview)

### 3.1. Maria Cacheira life story: An example

In the edited version of the excerpt from the marriage of Maria Cacheira (Listing 3.11) can be seen several examples of metadata markup, as the date of marriage `<data tipo="casamento" ano="1967" mes="00" dia="00">Casei aos 21 anos</data>`, the birth of the first daughter `<nascimento quem="primeira filha">Depois passado dois meses de casada nasceu a minha primeira filha, que eu ja ia de bebe.</nascimento>`, the date of death of the husband `<data deque="morte marido" ano="1973" mes="00" dia="00">seis anos e cinco meses</data>` and husband's job `<profissao quem="marido" classe="marinheiro">ajudante de motorista numa traineira da sardinha</profissao>`.

```
1 <casamento><p>Os pais dele nao queriam, mas depois nos casamos.  
   <data tipo="casamento" ano="1967" mes="00" dia="00">Casei aos  
   21 anos</data>.  
2 Sai de casa dois ou tres meses antes do casamento, para ir morar  
   com o meu marido.  
3 <nascimento quem="primeira filha">Depois passado dois meses de  
   casada nasceu a minha primeira filha, que eu ja ia de  
   bebe.</nascimento> Foi servico adiantado. Agora nao se usa  
   mais levar filhos antes do casamento porque elas tomam  
   comprimidos ou outras coisas, senao usava-se. Mas eu adiantei  
   o servico.</p>  
4  
5 <p>Fui feliz no tempo em que vivi com o meu marido. Mas a minha  
   sogra era muito intrometida na minha vida. Ja morreu essa  
   macaca. Vou todas as semanas ao cemiterio e a beira da campa  
   dela nunca vou.</p>  
6  
7 <p> Estive casada <data deque="morte marido" ano="1973" mes="00"  
   dia="00">seis anos e cinco meses</data>. O meu marido morreu  
   com uma leucemia no sangue. Sou viuva ha vinte e seis anos.  
8 O meu marido era <profissao quem="marido"  
   classe="marinheiro">ajudante de motorista numa traineira da  
   sardinha</profissao>. Quando ele faleceu fiquei com tres  
   filhas: uma de dois meses, uma de 2 anos e uma de 6.</p>  
9  
10 <p>Fui feliz pouco tempo, mas fui. Mas se eu soubesse que era  
   para ficar viuva tao nova nao tinha casado!</p></casamento>
```

Listing 3.11: Maria Cacheira's Wedding description (edited interview)

In the next excerpt (Listing 3.12), Maria Cacheira narrates an event: *a flood in Douro River that invaded her house*.

### 3.2. Summary

```
1 <episodio termo="cheias" titulo="Sem casa" onde="rio Douro">
2 <pergunta>Sempre viveu aqui?</pergunta>
3 <resposta><p>Sempre nasci e vivi aqui, nunca conheci outra terra.
   </p>
4 <p>Ha nove anos, veio uma cheia e eu morava na <ref
   tipo="local">Rua 27 de Fevereiro, Afurada </ref>e a propria
   agua invadiu-me a casa. Eu esperei vinte meses por uma casa
   camararia. Agora moro para la da ponte, num bairro camarario
   chamado <ref tipo="bairro">Bairro do Cavaco</ref>. Ha o Bairro
   1 e o Bairro 2. Eu moro no Bairro 2.</p></resposta>
5 </episodio>
```

Listing 3.12: Event - River Douro's flood (original interview)

Once again, it should be noticed, the key concepts that are part of the *Thesaurus*, as for example: the street where she lived `<ref tipo="local">Rua 27 de Fevereiro, Afurada </ref>`, the neighborhood where she was staying after flood `<ref tipo="bairro">Bairro do Cavaco</ref>` and the references to the flood itself `<episodio termo="cheias" titulo="Sem casa" onde="rio Douro">`.

### 3.2 SUMMARY

This chapter was devoted to introduce and describe the concrete museum that will be our case study, the Museum of the Person; namely the Portuguese branch of such network of virtual museums—o *núcleo português do Museum of the Person* (npMP)—was analyzed. The repository of this museum is composed of documents annotated in XML and photos, that describe life stories of different common people.

Maria Cacheiras life story set of documents is an example of the raw material that constitutes npMP's assets. Its description also illustrates the outcomes of some main steps of the npMP workflow, described in the beginning of the chapter. Their deep study was crucial to understand the museum objects in order to create the ontology and and to enable their processing in order to build the virtual museum exhibition rooms.

After the study and analysis of important concepts for the development of this thesis (Chapter 2), and after introducing the key notions related to the Museum of the Person (Chapter 3), Chapter 4 will discuss the proposal to build the Museum of the Person.

---

## PROPOSED APPROACHES TO BUILD THE MUSEUM OF THE PERSON

---

In this chapter the technical and architectural approaches proposed to build the Portuguese Museum of the Person from the annotated document repository are discussed.

### 4.1 SYSTEM ARCHITECTURE

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 that are interviews collection annotated in XML; the Ingestion Function [M1] responsible for getting and processing the input data; a Data Storage [DS] that is the data digital 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) (Araújo et al., 2016b).

This general approach will have two possible refinements, which are dependent on the [DS]. In approach 1 (Figures 6 and 7) the [DS] is a *TripleStore*, while in approach 2 (Figures 8 and 9) the [DS] is a *Relational Database*. According to the kind of storage chosen, the ingestion function and the learning spaces generator will require different designs (Araújo et al., 2016b).

Sections 4.2 and 4.3 described in more detail the approach 1 and the approach 2, respectively.

#### 4.2. Approach 1

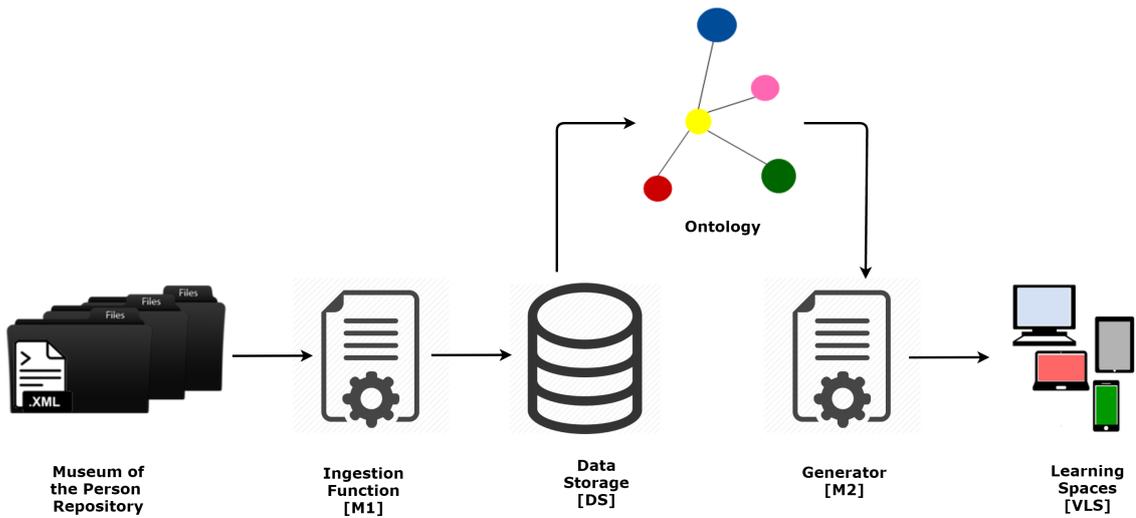


Figure 5.: General Approach to build the Museum of the Person

#### 4.2 APPROACH 1

As said above, Approach 1 is based on the decision of using a TripleStore as data storage [DS]. According to this decision, Ingestion Module and the Generator Module must be adapted; the first will transform the input XML (eXtensible Markup Language) documents into RDF (Resource Description Framework) triples, and the second will retrieve information from the RDF triples to create the museum web pages.

Figure 6 details the first module [M1] that is composed of three blocks (Araújo et al., 2016b):

- *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.

The mapping between the domain ontology (previously defined) and the data extracted from the repository is automatically built by construction in the 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 (Araújo et al., 2016b).

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

#### 4.2. Approach 1

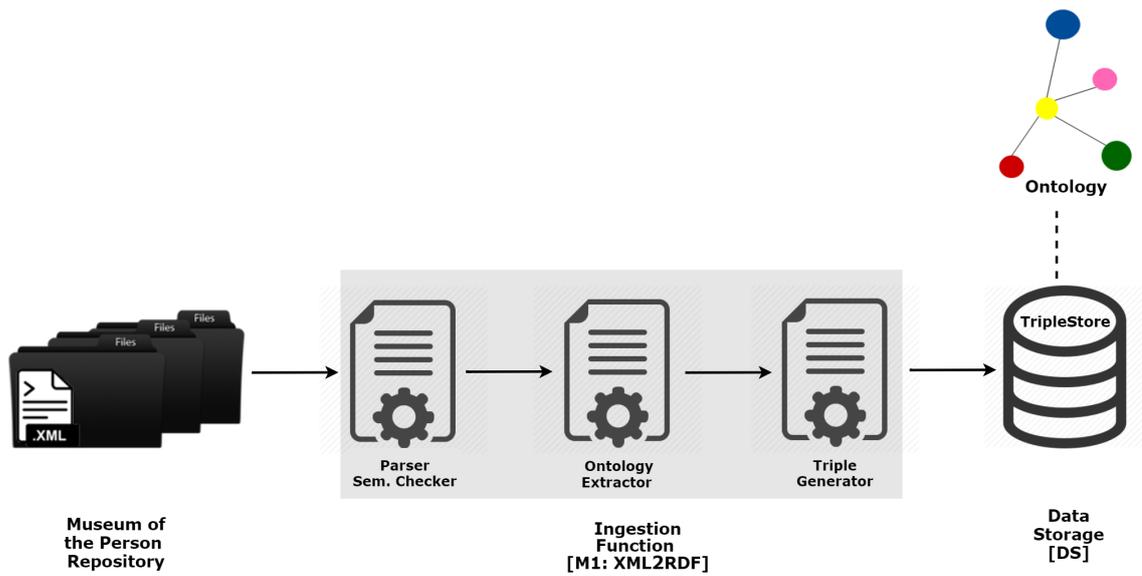


Figure 6.: Module [M1] in Approach 1

the second module [M2] (the Generator) that is composed of two blocks (Araújo et al., 2016b):

- *SPARQL Endpoint* that receives and interprets the SPARQL queries, accesses the Triple-Store and returns the answers;
- *Query Processor* generates the SPARQL queries according to the exhibition 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].

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.

### 4.3. Approach 2

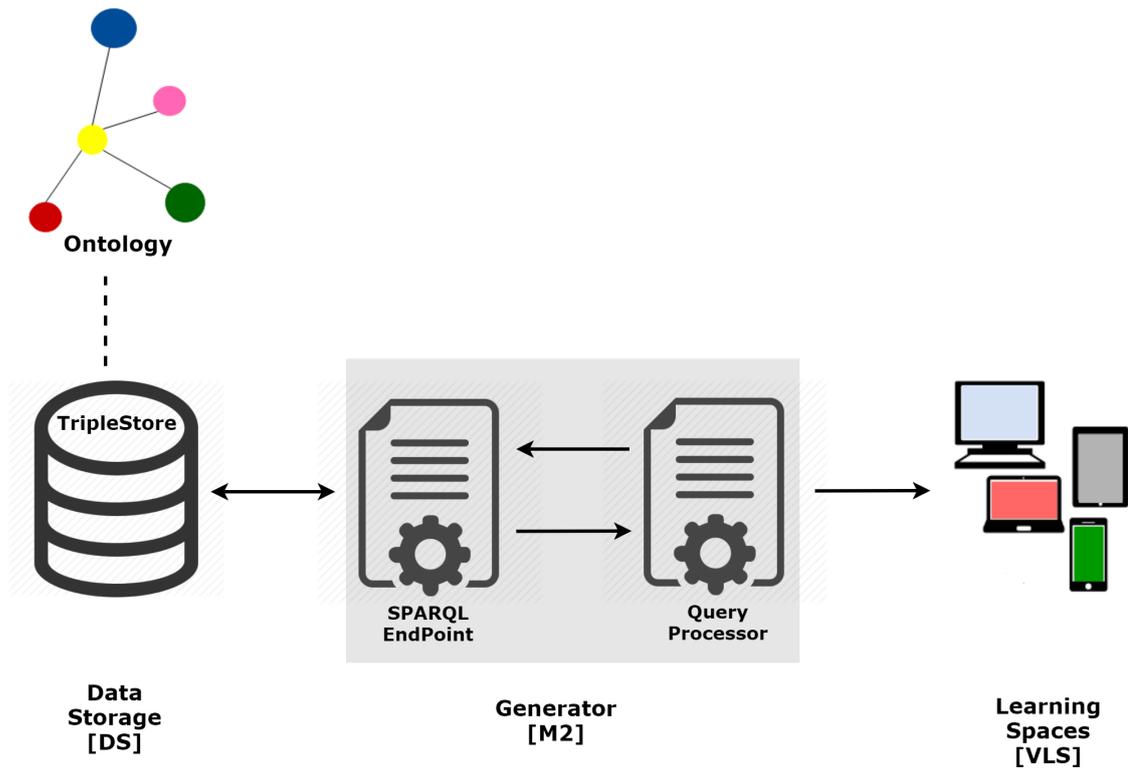


Figure 7.: Module [M2] in Approach 1

### 4.3 APPROACH 2

As in approach 1, the documents contained in the Repository of MP must be stored in a Data Storage [DS] but this time the choice is to resort to a traditional relational database; in this case 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 (Araújo et al., 2016b).

Figure 8 describes the first Module [M1] that in this case has two blocks (Araújo et al., 2016b):

- *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.

After the two phases of Ingestion Function [M1], the documents data populate 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

#### 4.3. Approach 2

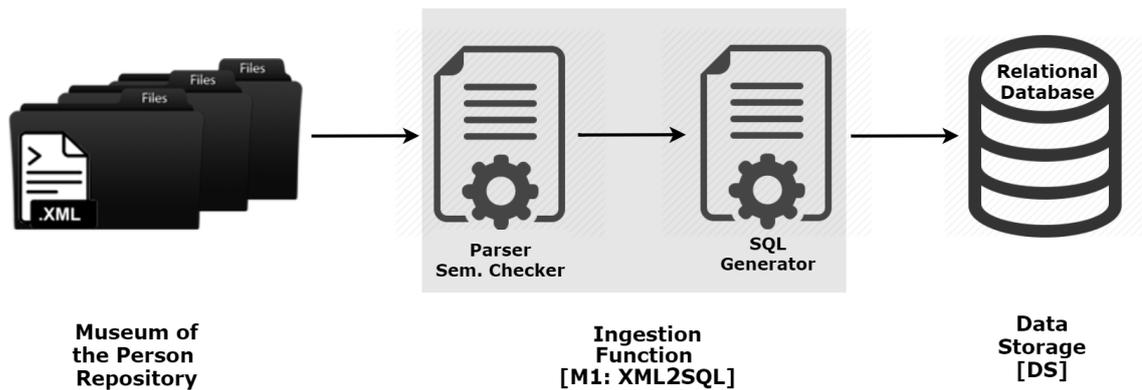


Figure 8.: Module [M1] in Approach 2

this mapping available, it is possible to resort to CaVa (Criação de Ambientes Virtuais de Aprendizagem) system (Martini, 2015) to build automatically the Virtual Learning Spaces [VLS]. Notice that only the generator module of CaVa,  $CaVa^{Gen}$ , will be used in this context (Araújo et al., 2016b).

Figure 9 sketches Module [M2] that in this case is composed of two parts (Araújo et al., 2016b):

- *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.

In this second approach all the work concerned with the query generation according to the exhibition requirements and the answer processing to fulfill the room 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 Virtual Learning Spaces in  $CaVa^{DSL}$  (Araújo et al., 2016b).

#### 4.4. Summary

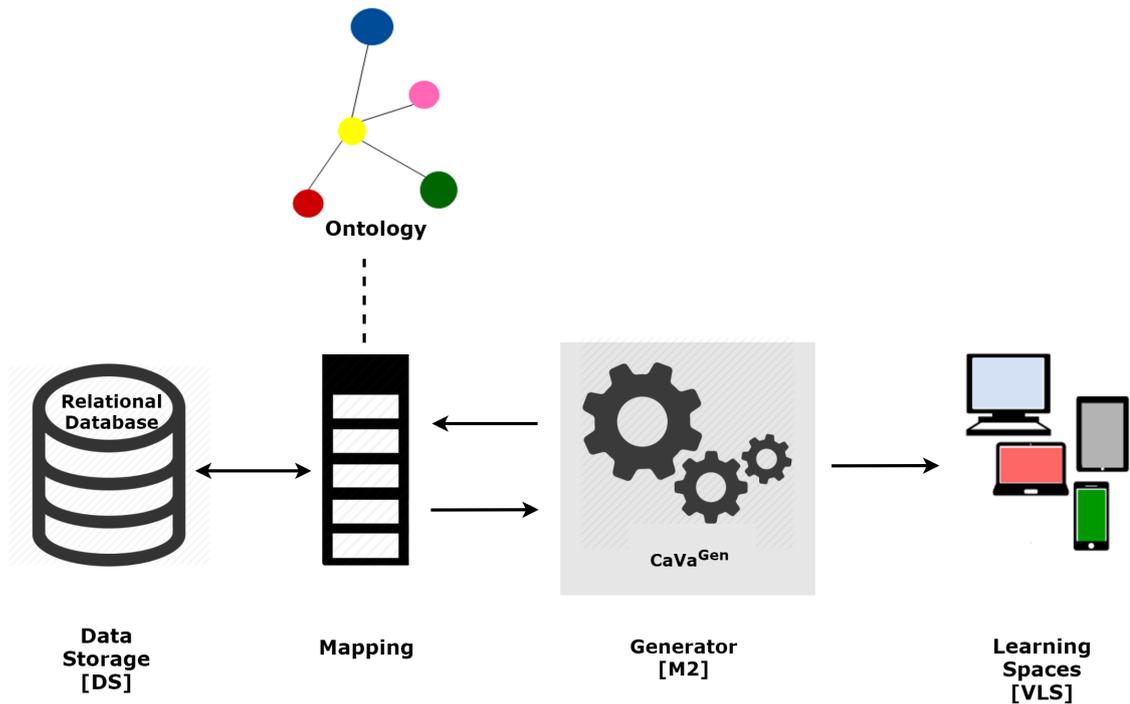


Figure 9.: Module [M2] in Approach 2

#### 4.4 SUMMARY

In this chapter, a general architecture to build systematic and automatically the Museum of the Person was discussed; two specific approaches able to realize that general architecture were also proposed. The proposed general architecture comprises: the repository, the Ingestion Function [M1], a Data Storage [DS], an Ontology, the Generator [M2], and Virtual Learning Spaces [VLS].

Depending on the Data Storage [DS], the general architecture will have two possible refinements: in approach 1 the [DS] is a *TripleStore*, while in approach 2 the [DS] is *Relational Database*. The choice to implement the [DS] determines the technical solutions to deal with the other modules, as explained above. In this way, these two approaches cover the two most usual types of data storage, the *TripleStore* and the traditional *Relational Database*.

Chapter 5 will present the ontology built for Museum of the Person (OntoMP), and illustrate a fragment of OntoMP instantiated with the life story of Maria Cacheira.

---

## ONTOMP: ONTOLOGY FOR MUSEUM OF THE PERSON, EVOLUTION

---

In this chapter are discussed the three stages to build the ontology for the Museum of the Person (OntoMP).

### DEVELOPEMENT METHODOLOGY

The ontologies created were developed using the methodology set out by Noy and McGuinness. This method comprises the following steps (Noy and McGuinness, 2001):

- The first step is to determine the domain and scope of the ontology, which involves answering questions such as: What is the domain that the ontology will understand?, What is the use foreseen for the ontology?, what kind of questions the ontology should provide an answer? and who will use and maintain the ontology?.
- The second step involves considering the use of existing ontologies in order to prevent repeated work already done.
- The third step is to identify the important terms in the ontology, for which a list of all terms that can make statements about the problem domain shall be done.
- The fourth and fifth steps are related and are done in parallel.  
In the fourth step, classes and their hierarchy shall be defined using strategies top-down, bottom-up or a combination thereof ((Uschold and Gruninger, 1996)); whereas in the fifth step, the attributes of the classes are defined, which can be intrinsic, extrinsic.
- In the sixth step the constraints of the attributes are defined in terms of domain and range, their cardinality and type (string, number, logical enumeration or instance).
- Finally, are created instances of the hierarchy by setting each individual composing the ontology.

To validate the ontology created, a service from W3C, called RDF Validation Service <sup>1</sup>, can and shall be used. So the overall process followed to create OntoMP can be split into the three steps described below, as depicted in Figure 10 (Martini et al., 2016b):

1. Reverse engineering of Museum of the Person (from Museum of the Person to Museum of the Person ontology);
2. Mapping the museum standard format view (from Museum of the Person ontology to CIDOC-CRM).
3. Extending the CIDOC-CRM version (from CIDOC-CRM to FOAF and DBpedia).

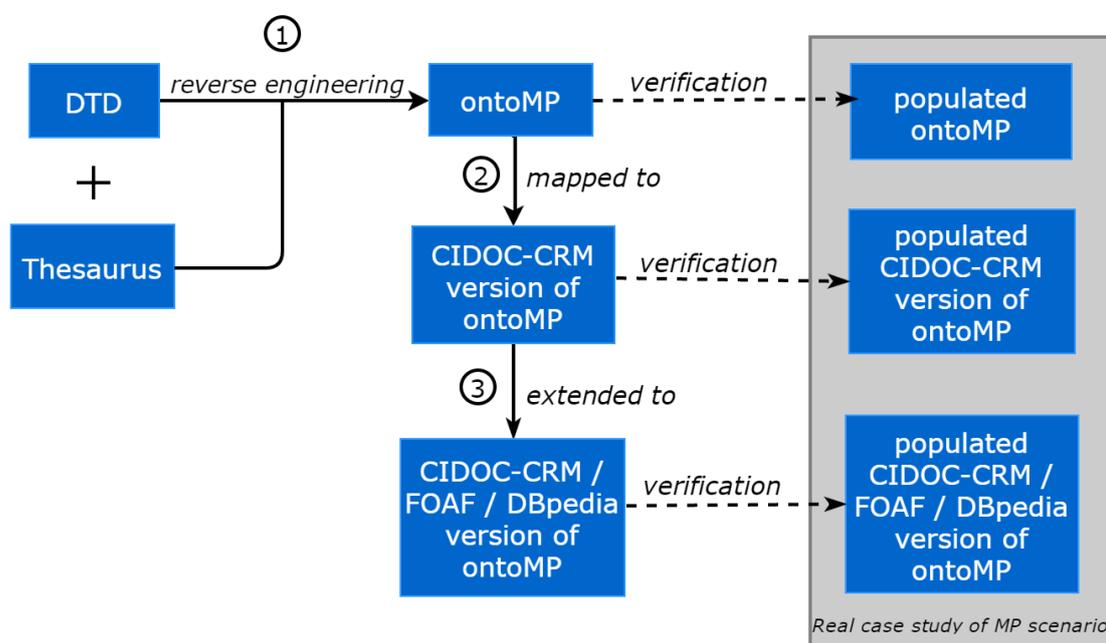


Figure 10.: Reverse Engineering and mapping of Museum of the Person (Adapted from (Martini et al., 2016b))

Following the diagram of Figure 10, the present chapter is structured as follows. Based on the museum description presented in Chapter 3, on the thesaurus, and on the DTD's inherited, a first draft of an ontology for the Museum of the Person is presented in Section 5.1. Then, in Section 5.2 that ontology is specified according to CIDOC-CRM standard. Finally in Section 5.3, an extended version of the original CIDOC-CRM ontology refined with FOAF and DBpedia concepts and properties will be presented.

<sup>1</sup> <https://www.w3.org/RDF/Validator/>

## 5.1. OntoMP: Original Design

### 5.1 ONTOMP: ORIGINAL DESIGN

From the DTD and *Thesaurus* introduced and briefly explained in the Chapter 3, it was possible to build OntoMP, the ontology for Museum of the Person.

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), job (profissão), house episode (episódio casa), education episode (episódio educação), dating episode (episódio namoro), general episode (episódio geral), childhood episode (episódio infância), leisure episode (episódio lazer), religious episode (episódio religioso), accident (evento acidente), migration (migração), life's philosophy (filosofia de vida), festivity (festividade), political event (evento político), catastrophic event (evento catastrófico), marriage (casamento), birth (nascimento), dream (sonho), uses (costumes), religion (religião) (Martini et al., 2016b; Araújo et al., 2016b).

In that phase, relations were also identified: performs (exerce), depicted (éRetratada), visits (visita), lives (vive), receives (recebe), tells (narra), has (tem), has-type (tipo), enrolls (participa), occurs (ocorre), refers to (dizRespeito) (Martini et al., 2016b; Araújo et al., 2016b).

After, building an ontology for the Museum of the Person using the concepts and relations above, it was realized that some more concepts could be added to make a more complete ontology. The concepts added were: marital status (estadoCívil), spouse (cônjuge), widowhood (viuvez), sex (sexo), literacy (habilitações literárias), political party (partido político), first communion (primeira comunhão), death (morte), baptism (batismo), child's birth (nascimento do filho), photos (fotos), description (descrição) and file (ficheiro) (Martini et al., 2016b; Araújo et al., 2016b).

After all these phases, an ontology for the Museum of the Person, called *OntoMP: Original Design* was built up. A fragment of *OntoMP: Original Design* defined is presented in Figure 11.

In order to verify that *OntoMP: Original Design* is appropriate to describe Museum of the Person repository, the ontology was instantiated with various interviews. In Figure 12 a fragment of *OntoMP: Original Design* instantiated with Maria Cacheira life story is shown.

As illustrated in Figure 12, *OntoMP: Original Design* 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.

### 5.1. OntoMP: Original Design

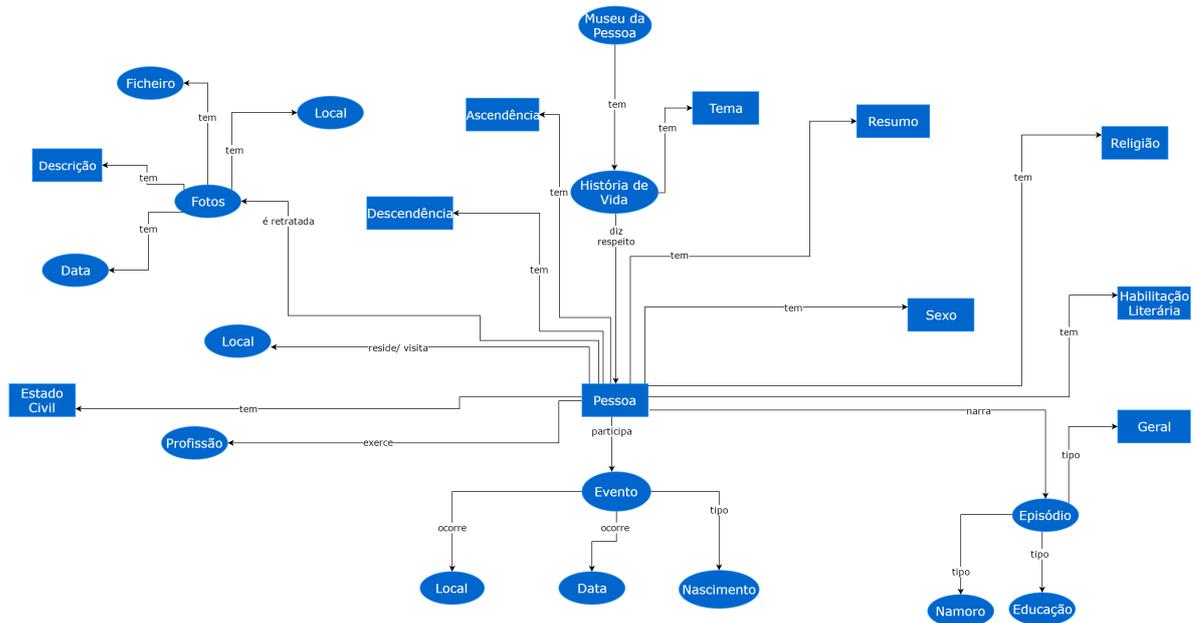


Figure 11.: An ontology *OntoMP: Original Design* for Museum of the Person (fragment)

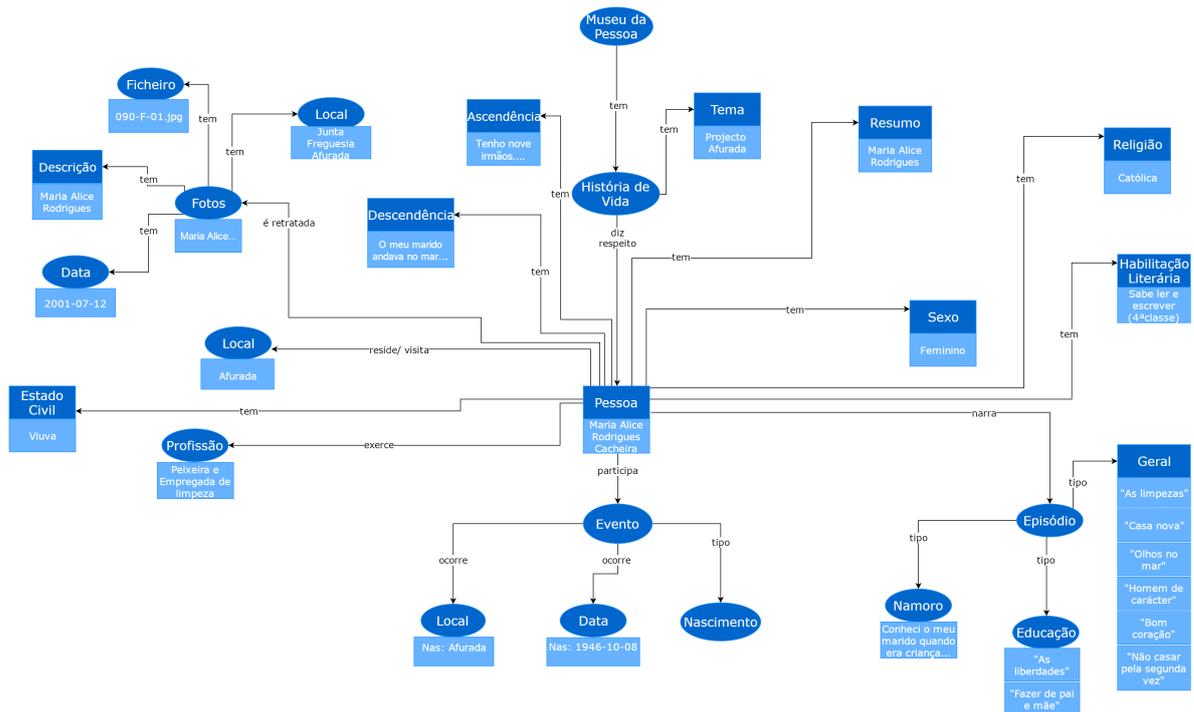


Figure 12.: An instance of *OntoMP: Original Design* for Maria Cacheira life story (fragment)

In the next section the mapping of *OntoMP: Original Design* ontology to a CIDOC-CRM compliant format is discussed.

## 5.2. OntoMP: CIDOC-CRM Representation

### 5.2 ONTOMP: CIDOC-CRM REPRESENTATION

After the construction of the *OntoMP: Original Design* (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 13, 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 12.

Looking at the concrete layer of that CIDOC-CRM ontology, it can be read, in Figure 13, that: *A life story (E31 Document) is about (P129) a person (E21 Person) and contains photos (E38 Image) concerning this person.*

*The photo is identified by (P1) 090-F-01.jpg (E41 Apellation), has note (P3) Maria Alice Rodrigues Cacheira, refers to (P67) Maria Alice Rodrigues Cacheira (E55 Type Description), and was taken in a (E52 Time Span) — that is identified by (P78) 2001-12-07, an (E50 Date) — and at a (E53 Place) — that is identified by (P87) Junta de Freguesia da Afurada, an (E44 Place Appellation).*

*Maria Cacheira (E21 Person), participated in (P11) 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 5.1.

All these concepts, in CIDOC-CRM version, are described as (E55) Type because they characterize a (E21) Person and his instances. To exemplify that approach, consider the following case: *Maria Cacheira (E21 Person) has type (P2) Peixeira e Empregada de limpeza (E55 Type Job), has type (P2) Viúva (E55 Type Marital Status), has type (P2) Sabe ler e escrever (4ª classe) (E55 Type Qualifications), has type (P2) Católica (E55 Type Religion).*

To sum up, notice that Figure 13 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.



### 5.3. OntoMP: CIDOC-CRM / FOAF / DBpedia Representation

After the refinement of CIDOC-CRM ontology with FOAF and DBpedia elements, we got a simpler notation (descriptions became less verbose); moreover the original was enriched conceptually, this is more details about people's stories can be included in the knowledge base (Araújo et al., 2016a).

The final OntoMP was once again instantiated with concrete data extracted from the real life stories. It was possible to validate it once more.

In Figure 14 is shown an instance of the ontology created with data extracted from Maria Cacheira interview. Notice that, the fragment of the instantiated ontology shown is the same of the one presented in Figure 13, now modified (incremented) with the concepts and relations imported from FOAF and DBpedia. Below is described the CIDOC-CRM, FOAF and DBpedia fragment reproduced (Araújo et al., 2016a).

*A 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 a (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) (Araújo et al., 2016a).*

*This person (E21 Person) is depicted in the photo (E38 Image). This photo is identified by (P1) 090-F-01.jpg (E41 Apellation), has note (P3) Maria Alice Rodrigues Cacheira, refers to (P67) Maria Alice Rodrigues Cacheira (E55 Type Description), and was taken in a (E52 Time Span) — that is identified by (P78) 2001-12-07, an (E50 Date) — and at a (E53 Place) — that is identified by (P87) Junta de Freguesia da Afurada, an (E44 Place Appellation) (Araújo et al., 2016a).*

*A person (E21 Person) has education Sabe ler e escrever (4ª classe), professes the religion Católica and has profession Peixeira e Empregada de limpeza (Araújo et al., 2016a).*

Besides the concepts described above, there are still others which are represented in CIDOC-CRM version, as (E55 Type). For example, *(E21 Person) has type (P2) Viúva (E55 Type Marital Status) (Araújo et al., 2016a).*

## 5.4. Summary

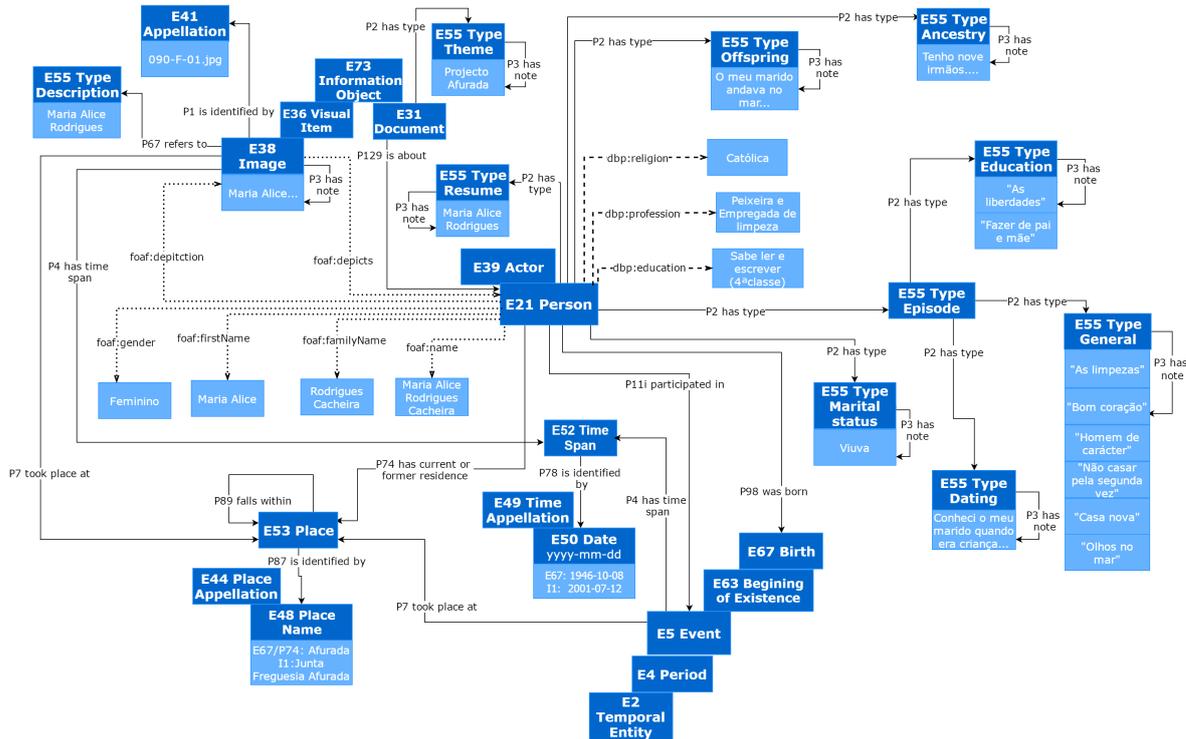


Figure 14.: An instance of *OntoMP: CIDOC-CRM / FOAF / DBpedia Representation* for Maria Cacheira life story (fragment)

The person's properties imported from FOAF (above identified) are emphasized in Figure 14 using dotted line. Similarly, DBpedia properties used are enhanced as dashed line.

This CIDOC-CRM ontology enriched with FOAF and DBpedia elements can describe appropriately the knowledge repository of the Museum of the Person.

## 5.4 SUMMARY

After a careful analysis of npMP documents, an ontology for this museum, *OntoMP*, was designed; the reverse engineering process followed was discussed. *OntoMP* was validated with various examples of life stories.

The transformation of *OntoMP* to the standard ontology used for museums, CIDOC-CRM was discussed. This ontology is fully compatible with the original. Then it was shown that some properties related with person could not be expressed in CIDOC-CRM in a simple and natural form. So it was decided to explore the combination with FOAF (Friend-of-a-Friend) and DBpedia, since both contain a vocabulary specific to describe individuals, their activities and their relationships with other people and objects.

After refining the first version of the CIDOC-CRM ontology with FOAF and DBpedia elements, a simpler notation was reached, as shown above; moreover the original was enriched

#### 5.4. Summary

conceptually, this is more details about people's stories can be included in the knowledge base.

The next chapter presents, in detail, the implementation of approach 1 (actually its first part).

---

## DATA EXTRACTION AND STORAGE

---

After the construction and validation of the ontology, the next phase is to implement the Ingestion Module [M<sub>1</sub>] following the approach 1 described in Section 4.2 of Chapter 4. The decision to follow that approach comes directly from the fact that in our case study, the virtual museum npMP, the digital repository is a collection of XML documents and in this approach the Ingestion Module [M<sub>1</sub>], part 1 of our general architecture, will transform the input XML documents into RDF/XML Triples—so the conceptual design fits directly in our reality. Remember that approach 2 is more appropriate to deal with cases where the digital repository is a relational database, and that is not the case.

At an early stage, to realize the kind of information contained in the documents and how it would be represented, it was decided to conduct the analysis and extraction of information from documents manually. In other words, the three phases of Ingestion Module were accomplished by hand.

As said above, to implement the Museum of the Person it was decided to use a TripleStore to archive the ontology triples, because it is the more natural way to store the data extracted from the XML documents. Although there are other notations for ontology description, it was chosen RDF because we use CIDOC-CRM, FOAF and DBpedia that are described, in its original form, in RDF; moreover, RDF is completely suitable to accomplish all the project needs. Listing 6.1 shows an excerpt of the RDF Triples built (Araújo et al., 2016a).

```
1 <!-- Description Interviewee 1 -->
2 <rdf:Description rdf:about="%ecrm;Interviewee_1">
3     <rdf:type rdf:resource="%ecrm;E21_Person"/>
4     <rdf:type rdf:resource=
5     "http://dbpedia.org/ontology/Person"/>
6     <rdf:type rdf:resource="%foaf;Person"/>
7     <foaf:firstName rdf:datatype="xsd:string">Maria
Alice</foaf:firstName>
8     <foaf:name rdf:datatype="xsd:string">Maria Alice
Rodrigues Cacheira</foaf:name>
```

```

9      <foaf:familyName rdf:datatype="&xsd:string"> Rodrigues
Cacheira</foaf:familyName>
10     <P98i_was_born rdf:resource="&ecrm;B1"/>
11
12     <foaf:gender rdf:datatype="&xsd:string">
Feminino</foaf:gender>
13     <foaf:depiction rdf:resource= "&ecrm;I1_Interviewee_1"/>
14
15     <dbp:profession rdf:datatype="&xsd:string"> Peixeira e
empregada de limpeza</dbp:profession>
16     <dbp:religion rdf:datatype="&xsd:string">
Catolica</dbp:religion>
17     <dbp:education rdf:datatype="&xsd:string"> Sabe ler e
escrever (quarta classe)</dbp:education>
</rdf:Description>
18
19 <!-- Event Birth Interviewee 1 (B1) -->
20 <rdf:Description rdf:about="&ecrm;B1">
21     <rdf:type rdf:resource="&ecrm;E67_Birth"/>
22     <P98_brought_into_life rdf:resource= "&ecrm;Interviewee
_1"/>
23     <P4_has_time-span rdf:resource="&ecrm;TS1"/>
24     <P7_took_place_at rdf:resource="&ecrm;PL1"/>
25 </rdf:Description>
26
27 <!-- Description Photo Interviewee 1 (I1) -->
28 <rdf:Description rdf:about="&ecrm;I1_Interviewee_1">
29     <rdf:type rdf:resource="&ecrm;E38_Image"/>
30     <rdf:type rdf:resource="&foaf;Image"/>
31     <foaf:depicts rdf:resource="&ecrm;Interviewee_1"/>
32     <P67_refers_to rdf:resource=
"&ecrm;I1_Description_Interviewee_1"/>
33     <P3_has_note rdf:datatype="&xsd:string"> Maria Alice
Rodrigues Cacheira</P3_has_note>
34     <P1_is_identified_by rdf:resource= "&ecrm;090-F-01.jpg"/>
35     <P4_has_time-span rdf:resource="&ecrm;TS7"/>
36     <P7_took_place_at rdf:resource="&ecrm;PL8"/>
37 </rdf:Description>

```

Listing 6.1: Fragment of the RDF Triples for Maria Cacheira life story (Araújo et al., 2016a)

The triple fragment shown in Listing 6.1 contains information about life story of Maria Cacheira. The first section (line 1-17) displays the biographic information about Maria Cacheira, as name (first, last name and full name), birth, sex, photo, profession, religion,

and education. The second section (line 19-25) describes the birth event of Maria Cacheira, date and place of it. The last section (line 27-37) contains specific information about the photo of the interviewee, such as description, caption, file, date and place (Araújo et al., 2016a).

To clarify the organizational form of the triples created, Figure 15 shows then using Turtle notation. This ontology representation is simpler and so it becomes easier to read the ontology instance and understand the number of triples and how they are formed (Araújo et al., 2016a).

```

:Interviewed_1 a          rdf:Description , dbp:Person ,
                        :E21_Person , foaf:Person ;

    foaf:name            "Maria Alice Rodrigues Cacheira"^^xsd:string ;
    foaf:firstName       "Maria Alice"^^xsd:string ;
    foaf:familyName     "Rodrigues Cacheira"^^xsd:string ;
    :P98i_was_born      :B1 ;

    foaf:gender          "Feminino"^^xsd:string ;
    foaf:depiction       :I1_Interviewed_1 ;

    dbp:profession       "Peixeira e empregada de limpeza"^^xsd:string ;
    dbp:religion         "Católica"^^xsd:string ;
    dbp:education        "Sabe ler e escrever (4ª Classe)"^^xsd:string .

```

Figure 15.: Maria Cacheira life story triples in Turtle notation (fragment)

In Figure 15 various triples can be seen, each one containing three parts: subject, predicate and object. A triple example is: *Interviewee\_1* (subject) *foaf: name* (predicate) *"Maria Alice Rodrigues Cacheira"^^xsd:string* (object). Notice that the remaining triples in this fragment follow the same logic and in all of them the subject is *Interviewee\_1* (Araújo et al., 2016a).

After building the RDF Triples manually, it was necessary to validate them to ensure that the very long textual description produced contained no errors. For this, it was used the W3C online tool RDF Validator <sup>1</sup>, which checks the consistency of the RDF Triples and displays them in a table with three columns 'subject', 'predicate' and 'object' (Araújo et al., 2016a).

The feedback obtained, after loading RDF file, was *"VALIDATION RESULTS: Your RDF document validated successfully"* that is just we want to get from that tool. The huge triples table was also useful for a quick visual reference (Araújo et al., 2016a).

After validated successfully by RDF Validator, the next step was to store them in a data set. For this it was used Apache Jena TDB, which is an RDF database (Araújo et al., 2016a).

*TDB is a component of Jena (free and open source Java framework for building Semantic Web and Linked Data applications) for RDF storage and query, and can be used as a high performance RDF store on a single machine. A TDB store can be accessed and managed with the provided command line*

<sup>1</sup> <https://www.w3.org/RDF/Validator/>

scripts and via the Jena API. When accessed using transactions a TDB dataset is protected against corruption, unexpected process terminations and system crashes. Apache Jena Fuseki component provides a SPARQL server to be used with TDB (JENA, 2016b).

After performing the three phases of the Ingestion Function [M1], it was understood how to make the extraction and analysis of semantic concepts and how to convert the triple ontology into RDF Triples. However, we came to the conclusion that it was a very time-consuming work to be done manually for all documents in the repository, so it was decided to develop a text filter able to scan all files that compose an interview (BI, Edited Interview and Photograph Captions), extract relevant information and convert into a single RDF Triples file. The text filter was developed by a group of students of LCC bachelor in the context of their final project (Ferreira and Dias, 2016).

To develop the text filter, it was used the compilers generator system ANTLR (Another Tool for Language Recognition) (ANTLR, 2016) integrated in ANTLRWorks tool, version 2.1, a plugin for NetBeans IDE. ANTLR is a powerful parser generator for reading, and processing structured text files; so it is extensively used to build language-based tools, and frameworks<sup>2</sup>.

In our case, from a set of Regular Expressions (RE), ANTLR generates a Lexical Analyzer that realizes the desired text filter for data extraction.

That text filter, or extractor, will accept an input document, like the one exemplified in Listing 6.2, and, after analyzing and processing it, will output a RDF description, like the one shown in Listing 6.3.

```
1 <?xml version="1.0" encoding="ISO-8859-1"?>
2 <fotos>
3   <foto ficheiro="090-F-01.jpg">
4     <quem>Maria Alice Rodrigues Cacheira</quem>
5     <onde>Junta de Freguesia da Afurada</onde>
6     <quando data="2001-07-12"/>
7   </foto>
8   <foto ficheiro="090-F-02.jpg">
9     <quem>Maria Alice Rodrigues Cacheira</quem>
10    <onde>Junta de Freguesia da Afurada</onde><
11    <quando data="2001-07-12"/>
12  </foto>
13 </fotos>
```

Listing 6.2: An XML input document

```
1 <rdf:Description rdf:about="&ecrm;090-F-01.jpg">
2   <rdf:type rdf:resource="&ecrm;E41_Appellation"/>
```

<sup>2</sup> <http://www.antlr.org/index.html>

```

3 </rdf:Description>
4
5 <rdf:Description rdf:about="&ecrm;I0_Interviewee_1"/>
6   <rdf:type rdf:resource="&ecrm;E38_Image"/>
7   <rdf:type rdf:resource="&foaf;Image"/>
8   <foaf:depicts rdf:resource="&ecrm;Interviewee_1"/>
9   <P67_refers_to rdf:resource="&ecrm;I0_Description_Interview_1"/>
10  <P1_is_identified_by rdf:resource="&ecrm;090-F-01.jpg"/>
11  <P4_has_time-span rdf:resource="&ecrm;TS1"/>
12  <P7_took_place_at rdf:resource="&ecrm;PL1"/>
13 </rdf:Description>
14
15 <rdf:Description rdf:about="&ecrm;2001-07-12">
16   <rdf:type rdf:resource="&ecrm;E49_Time_Appellation"/>
17 </rdf:Description>
18
19 <rdf:Description rdf:about="&ecrm;TS1">
20   <rdf:type rdf:resource="&ecrm;E52_Time-Span"/>
21   <P78_is_identified_by rdf:resource="&ecrm;2001-07-12"/>
22 </rdf:Description>
23
24 <rdf:Description rdf:about="&ecrm;PL1">
25   <rdf:type rdf:resource="&ecrm;E53_Place"/>
26   <P87_is_identified_by rdf:resource="&ecrm;Place1"/>
27 </rdf:Description>
28
29 <rdf:Description rdf:about="&ecrm;Place1">
30   <rdf:type rdf:resource="&ecrm;E48_Place_Name"/>
31   <P3_has_note rdf:datatype="&xsd;string">Junta de Freguesia da
32     Afurada</P3_has_note>
33 </rdf:Description>
34
35 <rdf:Description rdf:about="&ecrm;I0_Description_Interview_1">
36   <rdf:type rdf:resource="&ecrm;E55_Type"/>
37   <P2_has_type rdf:resource="&ecrm;Description"/>
38   <P3_has_note rdf:datatype="&xsd;string">Maria Alice Rodrigues
39     Cacheira</P3_has_note>
40 </rdf:Description>
41
42 <rdf:Description rdf:about="&ecrm;090-F-02.jpg">
43   <rdf:type rdf:resource="&ecrm;E41_Appellation"/>
44 </rdf:Description>

```

```
44 <rdf:Description rdf:about="&ecrm;I1_Interviewed_1"/>
45   <rdf:type rdf:resource="&ecrm;E38_Image"/>
46   <rdf:type rdf:resource="&foaf;Image"/>
47   <foaf:depicts rdf:resource="&ecrm;Interviewed_1"/>
48   <P67_refers_to rdf:resource="&ecrm;I1_Description_Interview_1"/>
49   <P1_is_identified_by rdf:resource="&ecrm;090-F-02.jpg"/>
50   <P4_has_time-span rdf:resource="&ecrm;TS1"/>
51   <P7_took_place_at rdf:resource="&ecrm;PL1"/>
52 </rdf:Description>
```

Listing 6.3: An RDF output document

To implement the text filter, three files were created:

- XML2RDF.g4,
- Person.java,
- MainLexerXML2RDF.java.

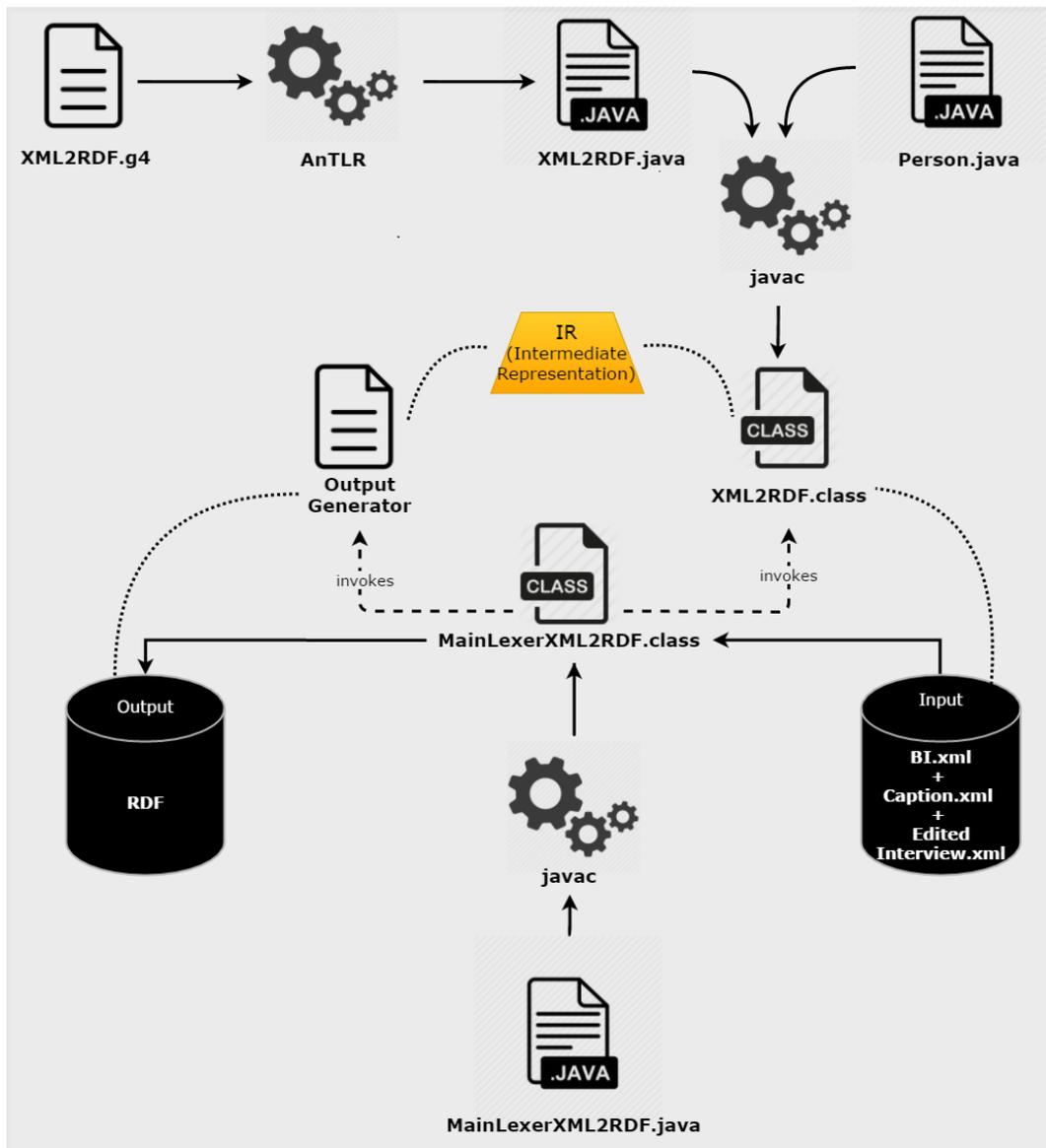


Figure 16.: Architecture Ingestion Function [XML2RDF]

Figure 16 depicts the architecture of this module [M1], the data extractor and RDF Generator, based on those three files. `XML2RDF.g4` is sent to `AnTLR` that generates `XML2RDF.java`. This file will be compiled including `Person.java` to create the desired processor `XML2RDF.class`. `MainLexerXML2RDF.java` after compilation gives rise to the main module, that reads the XML input documents, invokes the processor and writes the RDF output.

That automatic transformation is obtained using a specification (an *AnTLR Lexer grammar*) illustrated in Listing 6.4. The fragment shown in Listing 6.4 is a sequence of transformation rules that corresponds to the beginning of the global specification (the specification part not included will be discussed below). Each rule has a name and a pair composed of a Regular

Expression (RE) and a Semantic Action (SE) written in Java. A rule is interpreted from left to right: if the Regular Expression is found in the input, then the corresponding Semantic Action is triggered. The RE defines the text pattern that shall be found in the input, and the SE specifies how the concrete text found shall be transformed.

Moreover, AnTLR lets the programmer to set up modes that group the specific rules to address each sub block in the input file.

In Listing 6.4 it can be seen the three rules (namely, *Cabec*, *Fotos e MP*) corresponding to the three input files (*BI*, *Photography Captions*, and *Edited Interview*), respectively. When the extractor reads a XML tag defining the beginning of one of these three documents, it enters a special AnTLR mode to process that document's content.

```

1 lexer grammar XML2RDF;
2
3 Cabec : '<[Bb][Ii]>' -> mode(sBI)
4       ;
5 Fotos : '<[Ff][Oo][Tt][Oo][Ss]>' -> mode(sFOTOS)
6       ;
7 MP    : '<[Mm][Pp]>' -> mode(sMP)
8       ;
9 Default: . { ; }
10      ;
11
12      ...
13 .....Modes specification.....
14      ...

```

Listing 6.4: XML2RDF Lexer Grammar for AnTLR

Listing 6.5 shows the main mode to process the *Photography Caption* XML documents. The listing illustrates the general approach adopted: when a block opening tag is found, the appropriate mode is entered to consume the block contents; when the block closing tag is found, the processor exits the mode and returns to the initial mode.

The four auxiliary modes, called from the main one (see lines 13-19), contains the specific rules used to extract information from the four main blocks of the *Photography Captions* input document.

```

1 mode sFOTOS;
2 GetSFOTOS : '<foto' -> mode(sFOTO)
3           ;
4 OutFOTOSSAVE : '</fotos>' -> mode(DEFAULT_MODE)
5           ;
6 DefaultsFOTOS : . { ; }
7           ;

```

```

8
9
10 mode sFOTO;
11 GetFOTO      :    [ ]+'ficheiro="'    -> mode (sFICHEIRO)
12              ;
13 GetQUEM      :    '<quem>'            -> mode (sQUEM)
14              ;
15 GetQUANDO    :    '<quando>'          -> mode (sQUANDO)
16              ;
17 GetFACTO     :    '<facto>'           -> mode (sFACTO)
18              ;
19 GetONDE      :    '<onde>'            -> mode (sONDE)
20              ;
21 OutFOTOS     :    '</>'               -> mode (sPRINTTUDO)
22              ;
23
24 DefaultsFOTO :    .    { ; }
25              ;

```

Listing 6.5: Lexer Grammar Photos main Mode

Listing 6.6 contains the rules executed (the modes activated at line 11 - 20 of Listing 6.5) to analyze and extract information from XML documents repository. This mode record the name of the person in variable *quem*, then checks if it exists in the HashMap (*people*) previously created. If the *quem* variable does not exist in the *people* class, it is added. The variable *newCountKeyQuem* counts the number of occurrences of *quem* and this mapping saves that value. When all variables are assigned, they will be printed in the RDF file by *mode sPRINTTUDO* (Listing 6.7).

The modes *mode sFICHEIRO*, *mode sQUANDO*, *mode sFACTO* and *mode sONDE* follow the same logic of *mode sQUEM*.

```

1 mode sFICHEIRO;
2 GetFoNa      :    ~('"')+    {ficheiro=quem=data=facto=onde="";
3                               ficheiro=getText();
4                               if(!(fotos.containsKey(ficheiro))){
5                                   fotos.put(ficheiro, countKeyFicheiro);
6                                   newCountKeyFicheiro=countKeyFicheiro;
7                                   countKeyFicheiro++; }
8                               else{
9                                   newCountKeyFicheiro=fotos.get(ficheiro);}}
10             ;
11 OutFoNa     :    '">'      -> mode (sFOTO)

```

```

12         ;
13
14 mode sQUEM;
15 GetSQUEM : ~('<')+ {quem=getText();
16             if(!(pessoas.containsKey(quem))) {
17                 pessoas.put(quem, countKeyQuem);
18                 newCountKeyQuem=countKeyQuem;
19                 countKeyQuem++;
20                 whoAdded=true; }
21             else{
22                 newCountKeyQuem=pessoas.get(quem);
23                 whoAdded=false; }}
24         ;
25 OutSQUEM : '</quem>' -> mode(sFOTO)
26         ;
27
28 mode sQUANDO;
29 GetQU : [ ]+'data="' {;}
30         ;
31 GetQUADA : ~('"')+ {data=getText();
32             if(!(datas.containsKey(data))){
33                 dateAdded= true;
34                 datas.put(data, countKeyData);
35                 newCountKeyData=countKeyData;
36                 countKeyData++; }
37             else{
38                 newCountKeyData=datas.get(data);
39                 dateAdded=false; }}
40         ;
41 OutQU : '">' -> mode (sFOTO)
42         ;
43
44 mode sONDE;
45 GetSONDE : ~('<')+ {onde=getText();
46             if(!(lugares.containsKey(onde))){
47                 lugares.put(onde, countKeyLugar);
48                 newCountKeyLugar=countKeyLugar;
49                 countKeyLugar++;
50                 whereAdded=true; }
51             else{
52                 newCountKeyLugar=lugares.get(onde);
53                 whereAdded=false; }}
54         ;

```

```

55 OutSONDE      :   '</onde>'      ->  mode (sFOTO)
56              ;
57
58 mode sFACTO;
59 GetSFACTO    :   ~('<')+      {facto=getText(); }
60              ;
61 OutSFACTO    :   '</facto>'     ->  mode(sFOTO)
62              ;

```

Listing 6.6: Lexer Grammar Auxiliary Modes

Listing 6.7 contains the rules (just a fragment is shown) executed at the end of the processing (mode activated at line 21 of Listing 6.5) to print out the RDF Triples built in the internal representation. This grammar fragment is actually responsible for the generation of the RDF output file.

```

1  mode sPRINTTUDO;
2  GetsPRINTTUDO : 'foto'      {
3
4      pessoa.AddImage("I"+newCountKeyFicheiro+"_Interviewee_"+
5      countinterview);
6
7      System.out.print("<rdf:Description rdf:about=\"&ecrm;");
8      System.out.println(ficheiro+"\">");
9      System.out.println("\t<rdf:type
10     rdf:resource=\"&ecrm;E41_Appellation\"/>");
11     System.out.println("</rdf:Description>\n\n");
12     System.out.println("<rdf:Description
13     rdf:about=\"&ecrm;I"+newCountKeyFicheiro+"_Interviewee_"+
14     countinterview+"\"/>");
15     System.out.println("\t<rdf:type
16     rdf:resource=\"&ecrm;E38_Image\"/>");
17     System.out.println("\t<rdf:type
18     rdf:resource=\"&foaf;Image\"/>");
19     System.out.println("\t<foaf:depicts
20     rdf:resource=\"&ecrm;Interviewee_"+countinterview+"\"/>");
21
22     if(!quem.equals("")){
23         System.out.println("\t<P67_refers_to
24         rdf:resource=\"&ecrm;I"+newCountKeyFicheiro+
25         "_Description_Interview_"+ countinterview+"\"/>");}
26
27     if(!facto.equals("")){

```

```

18         System.out.println("\t<P3_has_note
rdf:datatype=\"&xsd:string\">"+facto+"</P3_has_note>"); }
19
20     System.out.println("\t<P1_is_identified_by
rdf:resource=\"&ecrm;"+ficheiro+"\"/>");
21
22     ...
23
24     if(!quem.equals("")){
25         if(whoAdded){
26             System.out.println("<rdf:Description
rdf:about=\"&ecrm;I"+newCountKeyQuem+"_Description_Interview_"+
countinterview+"\">");
27             System.out.println("\t<rdf:type
rdf:resource=\"&ecrm;E55_Type\"/>");
28             System.out.println("\t<P2_has_type
rdf:resource=\"&ecrm;Description\"/>");
29             System.out.println("\t<P3_has_note
rdf:datatype=\"&xsd:string\">"+quem+"</P3_has_note>");
30             System.out.println("</rdf:Description>\n\n"); }
31     }
32 }
33
34
35 OutsPRINTTUDO : '>'      -> mode(sfOTOS)
36

```

Listing 6.7: Lexer Grammar Print Mode

#### SUMMARY

In approach 1, the Ingestion Module [M1] transforms the input XML documents into RDF Triples. Initially, to understand the type of information contained in the repository documents and how to do the XML transformation to RDF, it was decided to perform this process manually.

To store the ontology triples in an TripleStore, RDF notation was used, because the standard ontologies chosen, are described in its original form in RDF.

After building the RDF Triples manually, the online tool RDF Validator was used to check it, this is, to verify if the ontology built contain errors.

The next step, after validated successfully by RDF Validator, was to store the RDF triples in a RDF database. The RDF database used was Apache Jena TDB.

However, after handwritten the Ingestion Module [M1], it was realized that it was a lengthy and time-consuming process to be performed manually. So it was decided, to develop the text filter to scan all files that compose an interview, extract important information and convert into a single RDF Triples file, as discussed along this chapter..

In the next chapter, the second part of approach 1 will be described , and explained how the RDF triples, stored in Apache Jena TDB, are queried to produce the virtual exhibition rooms.

---

## QUERYING AND VISUALIZING

---

As previously said, in Chapter 4, the modules that compose the second part of the system [M2] retrieve information from the RDF TripleStore to create the museum web pages. This task is performed by the Virtual Learning Spaces Generator [M2].

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

To run SPARQL queries, sent by the Query Processor, it is necessary to resort to a SPARQL Endpoint. The SPARQL Endpoint used was Apache Jena Fuseki (version 2.0) (Araújo et al., 2016a).

*Apache Jena Fuseki is a SPARQL server, that can run as an operating system service, as a Java web application (WAR file), and as a standalone server. It provides security (using Apache Shiro) and has a user interface for server monitoring and administration. Fuseki is tightly integrated with TDB to provide a robust, transactional persistent storage layer, and incorporates Jena text query and Jena spatial query. It can be used to provide the protocol engine for other RDF query and storage systems (JENA, 2016a).*

Some queries were built to test whether the answer is the expected or not. Examples of the queries issued are: 1. Interviewee and respective Project; 2. Interviewee and his photos; 3. Interviewee, his profession, marital status and spouse's name; 4. Interviewee and his events; 5. Interviewee by sex and residence; 6. Interviewee by type of episode; 7. Total Interviewee by Project; 8. Life story of Interviewee. All these queries were executed successfully in Fuseki web interface (Araújo et al., 2016a).

Listing 7.1 displays the SPARQL query: 2. *Interviewee and his photos* (Araújo et al., 2016a).

```
1 PREFIX : <http://erlangen-crm.org/150929/>
2 PREFIX foaf: <http://xmlns.com/foaf/0.1/>
3 PREFIX dbp: <http://dbpedia.org/ontology/>
4 PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
5 PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
6
7 SELECT DISTINCT ?nome ?ficheiro ?personDesc ?legenda ?dataPhoto
   ?placePHOTO
```

```

8
9 WHERE {
10
11   ?pessoa a :E21_Person;
12     :P129_is_subject_of ?doc;
13     foaf:name ?nome;
14     foaf:depiction ?fotos.
15       ?fotos :P1_is_identified_by ?ficheiro.
16   OPTIONAL{
17     ?fotos :P3_has_note ?legenda.}
18   OPTIONAL{
19     ?fotos :P67_refers_to ?personDesc.
20     ?personDesc :P3_has_note ?personDescricao.}
21   OPTIONAL{
22     ?fotos :P4_has_time-span ?data.
23     ?data :P78_is_identified_by ?dataPhoto.}
24   OPTIONAL{
25     ?fotos :P7_took_place_at ?placePhoto.
26     ?placePhoto :P87_is_identified_by ?placePhoto1.
27     ?placePhoto1 :P3_has_note ?placePHOTO.}
28
29
30 }ORDER BY ?nome ?ficheiro

```

Listing 7.1: Query SPARQL: 2. Interviewee and his photos

The code block between lines 10 and 14 of Listing 7.1 is intended to search for all respondents (*E21\_Person*) and photos in which each one is represented (*foaf:depiction*). The full name of each interviewee is described by the property *foaf:name* (Araújo et al., 2016a).

Actually the photos does not contain all the possible properties that would be desirable to retrieve. Therefore, in the query the argument *OPTIONAL* is used to cope properly with that variability in the legend (lines 15-16), in the description (lines 17-19), in the date (lines 20-22) and in the photo place (lines 23-26) (Araújo et al., 2016a).

For more information on the results and executed queries, please see: <http://npmp.ep1.di.uminho.pt>. Reusing the SPARQL queries manually created, that script selects those needed for a concrete request and sends the queries to the SPARQL endpoint. It was created a Python script that reuses the SPARQL queries manually created, that script selects those needed for a concrete request and sends them to the Fuseki (SPARQL Endpoint) and after receiving the answer, combines the data returned to configure the Virtual Learning Spaces [VLS]. This Python script also includes HTML (Hyper Text Markup Language) and CSS (Cascading Style Sheets) to create and format the web pages (Araújo et al., 2016a).

Listing 7.2 shows a part of the Python script.

```
1 #!/usr/bin/env python3.4
2 #*encoding: utf-8*
3 import sys
4 from SPARQLWrapper import SPARQLWrapper, JSON
5 import cgi
6 import cgitb
7
8 cgitb.enable()
9 form = cgi.FieldStorage()
10
11 sparql = SPARQLWrapper("http://localhost:40000/ds/sparql")
12 queryString = ""
13 PREFIX : <http://erlangen-crm.org/150929/>
14 PREFIX foaf: <http://xmlns.com/foaf/0.1/>
15 PREFIX dbp: <http://dbpedia.org/ontology/>
16 PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
17 PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
18
19 SELECT DISTINCT ?nome ?ficheiro ?personDesc ?legenda ?dataPhoto
20 ?placePHOTO
21 WHERE {
22     ?pessoa a :E21_Person;
23     :P129_is_subject_of ?doc;
24     foaf:name ?nome;
25     foaf:depiction ?fotos.
26     ?fotos :P1_is_identified_by ?ficheiro.
27     OPTIONAL{
28         ?fotos :P3_has_note ?legenda.}
29     OPTIONAL{
30         ?fotos :P67_refers_to ?personDesc.
31         ?personDesc :P3_has_note ?personDescricao.}
32     OPTIONAL{
33         ?fotos :P4_has_time-span ?data.
34         ?data :P78_is_identified_by ?dataPhoto.}
35     OPTIONAL{
36         ?fotos :P7_took_place_at ?placePhoto.
37         ?placePhoto :P87_is_identified_by ?placePhoto1.
38         ?placePhoto1 :P3_has_note ?placePHOTO.}
39
```

```

40 }ORDER BY ?nome ?ficheiro
41 """
42 sparql.setQuery(queryString)
43 sparql.setReturnFormat(JSON)
44
45 try:
46     results = sparql.query().convert()
47     requestGood = True
48 except Exception as e:
49     results = str(e)
50     requestGood = False

```

Listing 7.2: Fragment of the Python script (first part)

This Python script uses the *SPARQLWrapper* library to query the SPARQL endpoint about all *Interviewee and his photos* (query shown in Listing 7.1). It requests the results in a JSON (JavaScript Object Notation) format so that it can easily iterate through the returned data. After preparing the SPARQL query to send off, the script sends the query within a *try/except* block so that it can check for communication problems before attempting to render the results.

As is a CGI (Common Gateway Interface) script, it creates HTML and sends it to a browser, a *Content-type* header for sending the actual web page. Listing 7.3 shows HTML and CSS code to create and format the web page. The listing illustrates (using code fragments) the three main parts: creation of the header, body and footer of the web page. The script also includes Python code to return the results of the query.

```

1 sys.stdout.buffer.write(bytes("Content-type: text/html
2 <!DOCTYPE html>
3 <html>
4 <!-- HEADER -->
5 <head>
6     <meta charset="utf-8">
7     <meta http-equiv="X-UA-Compatible" content="IE=edge">
8     <meta name="viewport" content="width=device-width,
9     initial-scale=1">
10    <title>Museum of the Person |</title>
11    ...
12 <!-- BODY -->
13 <body>
14     ...
15
16    <div class="col-md-12 col-sm-12 col-xs-12">

```

```

17     <div class="x_panel">
18         <div class="x_title">
19             <h2>Interviewee and their photos</h2><br/><br/>
20         </div>
21         <div class="x_content"><br> <br>
22             <table class="table table-striped responsive-utilities
jambo_table bulk_action">
23                 <thead>
24                     <tr class="headings">
25                         <th class="column-title" width="18%">Interviewee</th>
26                         <th class="column-title" width="15%">File</th>
27                         <th class="column-title" width="25%">Caption</th>
28                         <th class="column-title" width="22%">Description</th>
29                         <th class="column-title" width="8%">Date</th>
30                         <th class="column-title" width="20%">Place</th>
31                     </tr>
32                 </thead>"; "utf-8"))
33 if requestGood == False:
34     sys.stdout.buffer.write(bytes("<h1>Problem communicating
with the server</h1>", "utf-8"))
35     sys.stdout.buffer.write("<p>" + results + "</p>", "utf-8"))
36 elif (len(results["results"]["bindings"]) == 0):
37     sys.stdout.buffer.write("<p>No results found</p>", "utf-8"))
38
39 else:
40     entArray = []
41     for result in results["results"]["bindings"]:
42         entrevistado= result["nome"]["value"]
43         entrevistado =
entrevistado.replace("http://erlangen-crm.org/150929/", "")
44         sys.stdout.buffer.write("<tr><td", "utf-8"))
45         if entrevistado not in entArray:
46             sys.stdout.buffer.write(entrevistado)
47             entArray = entrevistado
48             sys.stdout.buffer.write("</td>", "utf-8"))
49
50         ficheiros= result["ficheiro"]["value"]
51         ficheiros =
ficheiros.replace("http://erlangen-crm.org/150929/", "")
52         sys.stdout.buffer.write(bytes("<td>" + ""<div
class="highslide-gallery"><a
href="images/galleryInterviews/" + ficheiros + "" "
class="highslide" onclick="return hs.expand(this)"></a></div> "" "</td>", "utf-8"))
53
    try:
54         legendas= result["legenda"]["value"]
55         legendas =
56         legendas.replace("http://erlangen-crm.org/150929/", "")
57         sys.stdout.buffer.write(bytes("<td>" + legendas +
"</td>", "utf-8"))
58     except KeyError:
59         sys.stdout.buffer.write(bytes("<td>" + " " + "</td>",
"utf-8"))
60
61         ...
62
63 sys.stdout.buffer.write(bytes(" "
64     </table>
65     </div>
66     </div>
67     </div>
68     </div>
69     <div class="clearfix"></div>
70 <!-- FOOTER -->
71     <footer>
72     <div class="">
73     <p class="pull-left">Copyright &copy; npMP development team |
2016</p>
74     <p class="pull-right">Universidade do Minho</p>
75     </div>
76     <div class="clearfix"></div>
77     </footer>
78     </div>
79     </div>
80 </div>
81 </div></body></html>"" , "utf-8"))

```

Listing 7.3: Fragment script Python (second part)

Figure 17 displays the web page created by Python script (Listings 7.2 and 7.3). In other words, it shows the response to the query: (2.) *Interviewee and his photos*.

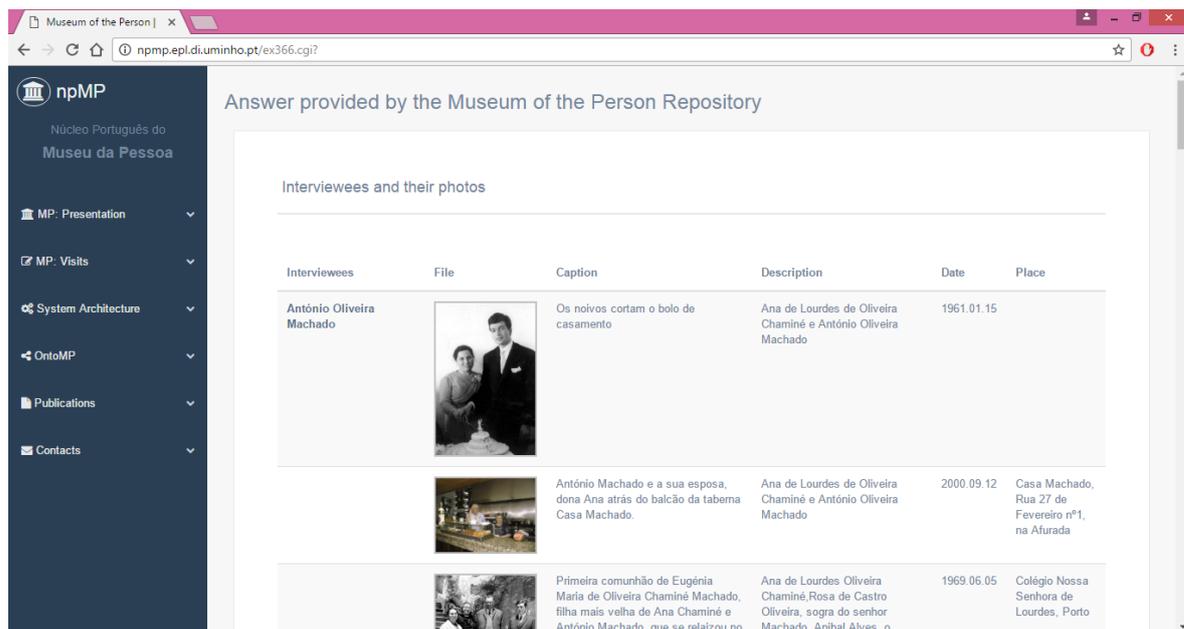


Figure 17.: Response to the SPARQL query: *Interviewee and their photos*

The next step was to create a web form to execute the query presented in Python script (Listings 7.2 and 7.3). Listing 7.4 shows the web form, in HTML, to list the *Interviewee and his photos*. This listing also illustrates (using code fragments) the three main parts: creation of the header, body and footer of the web page.

```

1 <!DOCTYPE html>
2 <html lang="en">
3 <!-- HEADER -->
4 <head>
5     <meta http-equiv="Content-Type" content="text/html;
6     charset=UTF-8">
7     <meta charset="utf-8">
8     <meta http-equiv="X-UA-Compatible" content="IE=edge">
9     <meta name="viewport" content="width=device-width,
10    initial-scale=1">
11    <title>Museum of the Person | </title>
12    ...
13 <!-- BODY -->
14 <body>
15     ...
16     <div class="right_col" role="main">
17         <div class="">
18             <div class="page-title">

```

```

19         <div class="title_left">
20             <h3>Consulting the Museum of the Person
Repository</h3>
21         </div>
22     </div>
23     <div class="clearfix"></div>
24     <div class="row">
25
26         ...
27
28         <div class="col-md-12 col-sm-12 col-xs-12">
29             <div class="x_panel">
30                 <div class="x_title">
31                     <h2>Interviewee and their photos?</h2>
32                     <div class="clearfix"></div>
33                 </div>
34                 <div class="x_content"><br />
35                     <form class="form-horizontal form-label-left"
role="form" action="ex366.cgi" method="get">
36                         <div class="col-md-12 col-sm-12 col-xs-12">
37                             <div class="form-group"> <br />
38                                 <button type="submit" class="btn
btn-primary">Submit</button>
39                             </div>
40                         </div>
41                     </form>
42                 </div>
43             </div>
44         </div>
45
46         ...
47
48     </div>
49 </div>
50 <!-- FOOTER -->
51     <footer>
52         <div class="">
53             <p class="pull-left">Copyright &copy; npMP
development team | 2016 </p>
54             <p class="pull-right">Universidade do Minho</p>
55         </div>
56         <div class="clearfix"></div>
57     </footer>

```

```

58     </div>
59     </div>
60 </div>
61 </body></html>

```

Listing 7.4: Web form to interrogate Python script

Figure 18 displays the web page created in HTML (Listing 7.4).

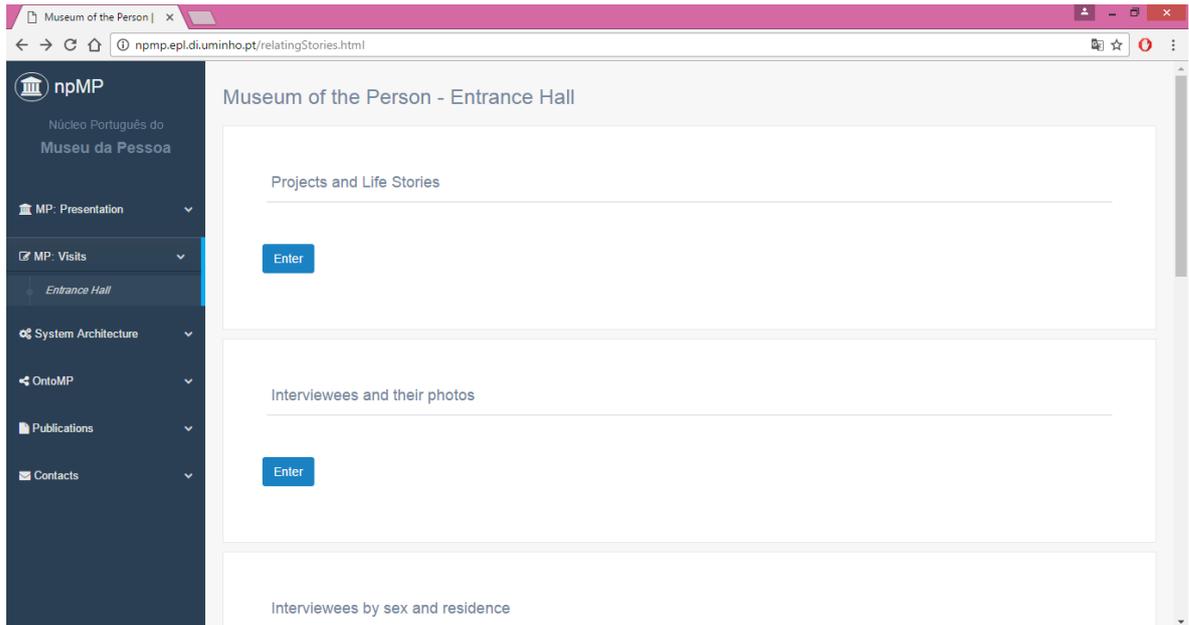


Figure 18.: Museum of the Person - Entrance Hall

Finally, Figure 19 displays the Museum of the Person entry page.

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.



Figure 19.: Home Page Museum of the Person

## SUMMARY

In this chapter the Generator [M2], that sends the queries and processes the returned data to display in the Virtual Learning Spaces [VLS], was explained.

The SPARQL Endpoint chosen to run SPARQL queries sent by the Query Processor was Apache Jena Fuseki. Some queries were built and placed in SPARQL Endpoint, to test whether the result was successfully obtained.

Finally, in this chapter it was also discussed the Python script built to select the concrete queries, send them to the Endpoint and process the returned data to create the Learning Spaces.

---

## CONCLUSION

---

The ultimate goal of this project is the systematic creation of a Virtual Museum for Museum of the Person. Web Learning Spaces, that implement the museum's exhibition rooms, will be generated through the population of HTML page templates with data extracted from a digital data storage that contains the museum assets. This process is guided by an ontological description of that repository.

Initially, concepts such as Cultural Heritage, Virtual Museum, Learning Spaces, and Ontologies were studied. In particular, CIDOC-CRM, FOAF and DBpedia ontologies were described.

The document collection of Museum of the Person (interviews (original and edited), *Thesaurus*, DTD's and photographs) was presented and analyzed in this dissertation.

Thereafter, a general system architecture, has been proposed, to build a software platform to create the museum's virtual exhibition rooms, as web pages, extracting information from the museum's digital repository. This architecture has two approaches that depend on the Data Storage (TripleStore or Relational Database). The general architecture and its refinement approaches are a novelty, considered as a contribution of this research work. However, to refine this architecture, it was chosen approach 1, that uses a TripleStore to archive the ontology instances. This implementation decision was taken considering that the actual npMP digital repository is a collection of XML documents (containing the information about each interviewee and his life story) and that in this concrete case the most natural way to populate the ontology is to create directly the instance triples with the data extracted from those original documents (the use of a relational database, as foreseen in approach 2, seems not a natural solution for this museum).

After analyzing the documents that make up the repository, an ontology, called OntoMP, for the Museum of the Person was designed. This ontology was created based on MP holdings, already referred: interviews (edited), *Thesaurus*, and the photographs.

The next stage after the validation and tuning of OntoMP (the first contribution of this masters' thesis) was to describe it in a standard ontology format used for museums, CIDOC-CRM (Comité International pour la Documentation - Conceptual Reference Model). However, after perform the mapping to CIDOC-CRM and validation with various interviews, it was concluded that some properties related with person could not be expressed in CIDOC-CRM

## 8.1. Future Work

in a simple and natural form. So, it was decided to complement and enrich CIDOC-CRM with some pertinent FOAF and DBpedia concepts and properties. The extension of CIDOC-CRM with FOAF and DBpedia elements is one more contribution of this masters' work.

As stated above, it was chosen approach 1 to implement the general architecture proposed. Initially, the Ingestion Function [M1] (first module of Approach 1) was implemented by hand, but it was concluded, that it was extremely slow. So, it was decided to implement a text filter to perform the three parts of [M1] automatically. This can be pointed out as another contribution of this masters' work.

After implementing the module [M1] (the Ingestion Function), the next phase was to implement the second module, Generator [M2], to configure the Virtual Learning Spaces [VLS] according to the exhibition requirements. To achieve this a Python script, complemented with HTML and CSS code, was used. Reusing the SPARQL queries manually created, that script selects those needed for a concrete request and sends the queries to the SPARQL endpoint. The returned data is processed and then displayed in the [VLS]. In this way, the information stored in the TripleStore [DS] will be automatically exhibited to the museum visitor. Module [M2] is novel and a final contribution of the research project here discussed.

In Chapter 7 and in Appendix B are presented some illustrative pictures of the museum rooms created as a proof of concept, some that presented the individual life stories of the interviewees and others that allow to interrelate life stories to understand social phenomena. The attainment of such an outcome prove the initial research hypothesis.

Along this year the work developed in the context of the Museum of the Person was published in different conferences and journals: (Martini et al., 2016b); (Araújo et al., 2016b); and Araújo et al. (2016a). An extension of one of the conference papers was submitted to another journal. A previous work involving the creation and instantiation of a CIDOC-CRM ontology for the Virtual Emigration Museum has also been presented at an international conference (Martini et al., 2016a) and (Martini et al., 2016c).

Additionally it was created a project web site <http://npmp.ep1.di.uminho.pt/> that provides access to the museum's exhibitions rooms but also makes available general information about Museum of the Person and on all the technical details of the project.

## 8.1 FUTURE WORK

All the objectives, previously defined at the beginning of this master's work, have been fulfilled. However, there are points of improvement that raise issues which can be taken as future lines of further investigation. As future work is proposed:

### 8.1. Future Work

- improve the visual interface of the Museum of the Person increasing the interactivity and exploiting geographical information applications following what is being done in the newest Museu Marítimo de Sesimbra<sup>1</sup>;
- rethinking the architectures and technologies used to improve the system performance;
- implement Approach 2;
- combine the first module [M1: XML2RDF] (Ingestion Function) of Approach 1, with the second module [M2: CaVa<sup>Gen</sup>] (Generator) of Approach 2;
- investigate new standards and other notations for ontology description;
- integrate and develop new tools to automatize the life cycle of the Museum of the Person. As a first step, integrate in our platform a mobile application that was built to record the interviews (Gylytskyy et al., 2016); then, create an application to transcribe audio interviews to text, and to markup the important concepts. Thus, it was possible to enrich easily the Museum of the Person with new interviews;
- compare the approach developed by us with other approaches (either at the level of the museum specification, or at the level of the output produced), such as Oda (Objetos de Aprendizaje), a software system created at Universidad Complutense de Madrid (UCM) (Crespo et al., 2014) to build Virtual Museums, Repositories of Learning Objects, etc.;
- apply all the knowledge acquired and implemented on Approach 1 to another case study, such as the Museu do Brinquedo Português<sup>2</sup> or the Centro de Interpretação do Território de Ponte de Lima<sup>3</sup>.

---

<sup>1</sup> <http://www.cm-sesimbra.pt/pages/1544>

<sup>2</sup> <http://www.cm-pontedelima.pt/ver.php?cod=0M1B>

<sup>3</sup> <http://www.cm-pontedelima.pt/noticia.php?id=1729>

---

## BIBLIOGRAPHY

---

- Mumtaz M Ali Al-Mukhtar and Ahmed T Abbas Al-Assafy. The implementation of foaf ontology for an academic social network. *International Journal of Science, Engineering and Computer Technology*, 4(1):10, 2014. ISSN 2231-0711.
- Dean Allemang and James Hendler. *Semantic Web for the Working Ontologist: Effective Modeling in RDFS and OWL*. Elsevier Science, 2011. ISBN 9780123859662.
- José João Almeida, Jorge Gustavo Rocha, Pedro Rangel Henriques, Sónia Moreira, and Alberto Simões. Museu da Pessoa – arquitectura. In *Encontro Nacional da Associação de Bibliotecários, Arquivista e Documentalistas, ABAD'01*. BAD, 2001. URL <http://repositorium.sdum.uminho.pt/handle/1822/585>.
- ANTLR. ANTLR. <http://www.antlr.org/>, 2016. Accessed: 2016-09-14.
- Cristiana Araújo, Ricardo Martini, Pedro Rangel Henriques, and José João Almeida. Building the Museum of the Person from RDF Triples and SPARQL. *Communications and Innovations Gazette (ComInG)*, 1(2):1–14, 2016a. ISSN 2448-1904. doi: <http://dx.doi.org/10.5902/2448190422984>.
- Cristiana Araújo, Ricardo G. Martini, Pedro Rangel Henriques, and José João Almeida. Architectural Approaches to build The Museum of the Person. In Álvaro Rocha, Luís Paulo Reis, Manuel Pérez Cota, Octavio Santana Suárez, and Ramiro Gonçalves, editors, *Sistemas y Tecnologías de Información — Atas da 11<sup>a</sup> Conferência Ibérica de Sistemas e Tecnologias de Informação*, volume Vol. I — Artículos de la Conferencia, pages 383–388. AISTI–Associação Ibérica de Sistemas e Tecnologias de Informação, June 2016b. ISBN 978-989-98434-6-2. doi: 10.1109/CISTI.2016.7521367.
- Christian Bizer, Jens Lehmann, Georgi Kobilarov, Sören Auer, Christian Becker, Richard Cyganiak, and Sebastian Hellmann. Dbpedia - a crystallization point for the web of data. *Web Semant.*, 7(3):154–165, September 2009. ISSN 1570-8268. doi: 10.1016/j.websem.2009.07.002. URL <http://dx.doi.org/10.1016/j.websem.2009.07.002>.
- Willem Nico Borst. *Construction of Engineering Ontologies for Knowledge Sharing and Reuse*. PhD thesis, 1997. URL <http://www.ub.utwente.nl/webdocs/inf/1/t0000004.pdf>.
- Dan Brickley and Libby Miller. FOAF. <http://www.foaf-project.org/>, 2016. Accessed: 2016-04-10.

## Bibliography

- Alfredo Fernández-Valmayor Crespo, Ana Fernández-Pampillón Cesteros, and Varadero Software Factory. *OdA v2.5, Guía de Gestión*. Technical report, Universidad Complutense de Madrid, 2014.
- Nicholas Crofts, Patrick Le Buf, and Odile Artur. Elag presentation ontologies. Semantic web and libraries. In , 2002. URL <http://www.ifnet.it/elag2002/papers/pap9.html>.
- Dbpedia. Ontology. <http://wiki.dbpedia.org/>, 2016. Accessed: 2016-06-15.
- Diogo Ferreira and Pedro Dias. Tradutor de documentos XML para ontologia em triplos RDF. Development report, Universidade do Minho, 2016.
- Filipe Daniel Silva Ferreira. *Uma Plataforma Comum para os Museus da U. Porto: Gestão, Processos e Tecnologia*. M.Sc. Dissertation, Universidade do Porto, July 2016.
- Gunnar Aastrand Grimnes, Peter Edwards, and Alun D. Preece. Learning meta-descriptions of the foaf network. In Sheila A. McIlraith, Dimitris Plexousakis, and Frank van Harmelen, editors, *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.
- Thomas R. Gruber. Toward principles for the design of ontologies used for knowledge sharing. In *International Journal of Human-Computer Studies*, pages 907–928. Kluwer Academic Publishers, 1993.
- N. Guarino. Understanding, building and using ontologies: A commentary to using explicit ontologies in kbs development. *International Journal of Human and Computer Studies*, pages 293–310, 1997.
- Alexey Gylytskyy, Frederico Ribeiro, and Luís Vieira. *Museu da Pessoa - Assistente de Entrevistas*. Development report, Universidade do Minho, 2016.
- ICOM/CIDOC. Definition of the CIDOC Conceptual Reference Model. Technical report, ICOM/CIDOC, May 2015. URL [http://www.cidoc-crm.org/docs/cidoc\\_crm\\_version\\_6.2.pdf](http://www.cidoc-crm.org/docs/cidoc_crm_version_6.2.pdf).
- APACHE JENA. Apache Jena Fuseki. <https://jena.apache.org/documentation/fuseki2/index.html>, 2016a. Accessed: 2016-06-01.
- APACHE JENA. TDB. <https://jena.apache.org/documentation/tdb/index.html>, 2016b. Accessed: 2016-06-01.
- Jens Lehmann, Robert Isele, Max Jakob, Anja Jentzsch, Dimitris Kontokostas, Pablo N Mendes, Sebastian Hellmann, Mohamed Morsey, Patrick van Kleef, Sören Auer, et al.

## Bibliography

- Dbpedia—a large-scale, multilingual knowledge base extracted from wikipedia. *Semantic Web*, 6(2):167–195, 2015.
- Ricardo Martini. Formal Description and Automatic Generation of Learning Spaces based on Ontologies. Phd pre-thesis, Universidade do Minho, 2015.
- Ricardo Martini, Cristiana Araújo, and Pedro Rangel Henriques. Uma Ontologia CIDOC-CRM para o Museu Virtual da Emigração. *Communications and Innovations Gazette (Com-InG)*, 1(1):79–90, 2016a.
- Ricardo G. Martini, Cristiana Araújo, José João Almeida, and Pedro Rangel Henriques. *New Advances in Information Systems and Technologies: Volume 2*, chapter OntoMP, An Ontology to Build the Museum of the Person, pages 653–661. Springer International Publishing, Cham, 2016b. ISBN 978-3-319-31307-8. doi: 10.1007/978-3-319-31307-8.67. URL [http://dx.doi.org/10.1007/978-3-319-31307-8\\_67](http://dx.doi.org/10.1007/978-3-319-31307-8_67).
- Ricardo G. Martini, Cristiana Araújo, Giovani R. Librelotto, and Pedro R. Henriques. *New Advances in Information Systems and Technologies*, chapter A Reduced CRM-Compatible Form Ontology for the Virtual Emigration Museum, pages 401–410. Springer International Publishing, Cham, 2016c. ISBN 978-3-319-31232-3. doi: 10.1007/978-3-319-31232-3-38. URL [http://dx.doi.org/10.1007/978-3-319-31232-3\\_38](http://dx.doi.org/10.1007/978-3-319-31232-3_38).
- D. L. McGuinness and F. Van. Harmelen. Owl web ontology language overview, 2004.
- Natalya F. Noy and Deborah L. McGuinness. Ontology development 101: A guide to creating your first ontology. Technical report, 2001.
- Dominic Oldman and CRM Labs. The CIDOC Conceptual Reference Model (CIDOC-CRM): PRIMER. *International Council of Museums (ICOM)*, 1, July 2014.
- Bruno Cesar Rodrigues and Giulia Crippa. Novas Propostas e Desafios Das Mediações Culturais em Museus Virtuais. In *El Pensamiento Museológico Contemporáneo neo. O Pensamento Museológico Contemporâneo*, pages 599–608. ICOM, 2011. ISBN 978-92-9012-404-7.
- João Pedro Almeida Rua. Digitalização, Preservação Digital e Acesso em Instituições de Memória: Contributos para o Projecto Museu Digital da U. Porto. M.Sc. Dissertation, Universidade do Porto, July 2016.
- Werner Schweibenz. The development of virtual museums. In *Virtual Museums*, volume 57(3). ICOM, 2004.
- Yuanwu Shi, Jianfeng Hao, and Shouqian Sun. The digital protection of intangible cultural heritage: The construction of a digital museum. In *Computer-Aided Industrial Design and Conceptual Design, 2008. CAID/CD 2008. 9th International Conference on*, pages 1196–1199, Nov 2008. doi: 10.1109/CAIDCD.2008.4730777.

## Bibliography

- Daniela Lucas Silva, Eliana Antonia Demarques, Renato Rocha Souza, and Gercina A. B. de Oliveira Lima. Ontologias e unified modeling language: uma abordagem para representação de domínios de conhecimento. *DataGramaZero - Revista de Ciência da Informação*, 10(5), 2009. URL [http://www.dgz.org.br/out09/Art\\_05.htm#R1](http://www.dgz.org.br/out09/Art_05.htm#R1).
- Alberto Simões and José João 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>.
- Philip B. Stafford. Museum of person. Technical report, 2015.
- UNESCO. *Draft Medium Term Plan 1990 1995*. United Nations Educational, Scientific and Cultural Organization, 1989. URL <http://unesdoc.unesco.org/images/0008/000825/082539eb.pdf>.
- Mike Uschold and Michael Gruninger. Ontologies: Principles, methods and applications. *Knowledge Engineering Review*, pages 93–136, 1996.
- E. Waterton and S. Watson. *HERITAGE AND COMMUNITY ENGAGEMENT: Collaboration or Contestation?* Taylor & Francis, 2013. ISBN 9781317986577. URL <http://books.google.pt/books?id=PQ7dAAAAQBAJ>.
- Karen Worcman. The museum of the person. In *Virtual Museums*, volume 57(3). ICOM, 2004.



### A.1. Original Design

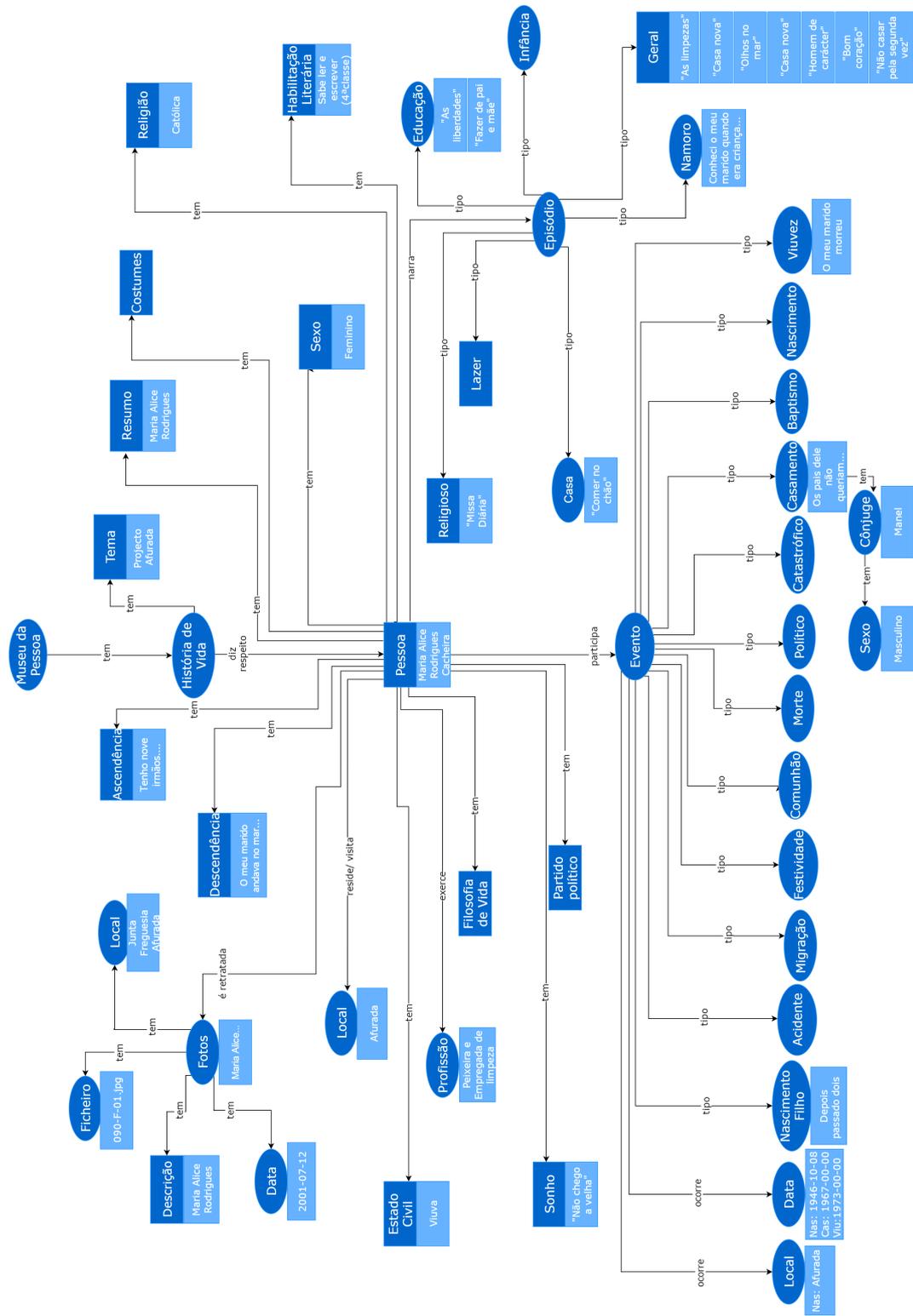


Figure 21.: An instance of *OntoMP: Original Design* for Maria Cacheira life story



## A.2. CIDOC-CRM Representation

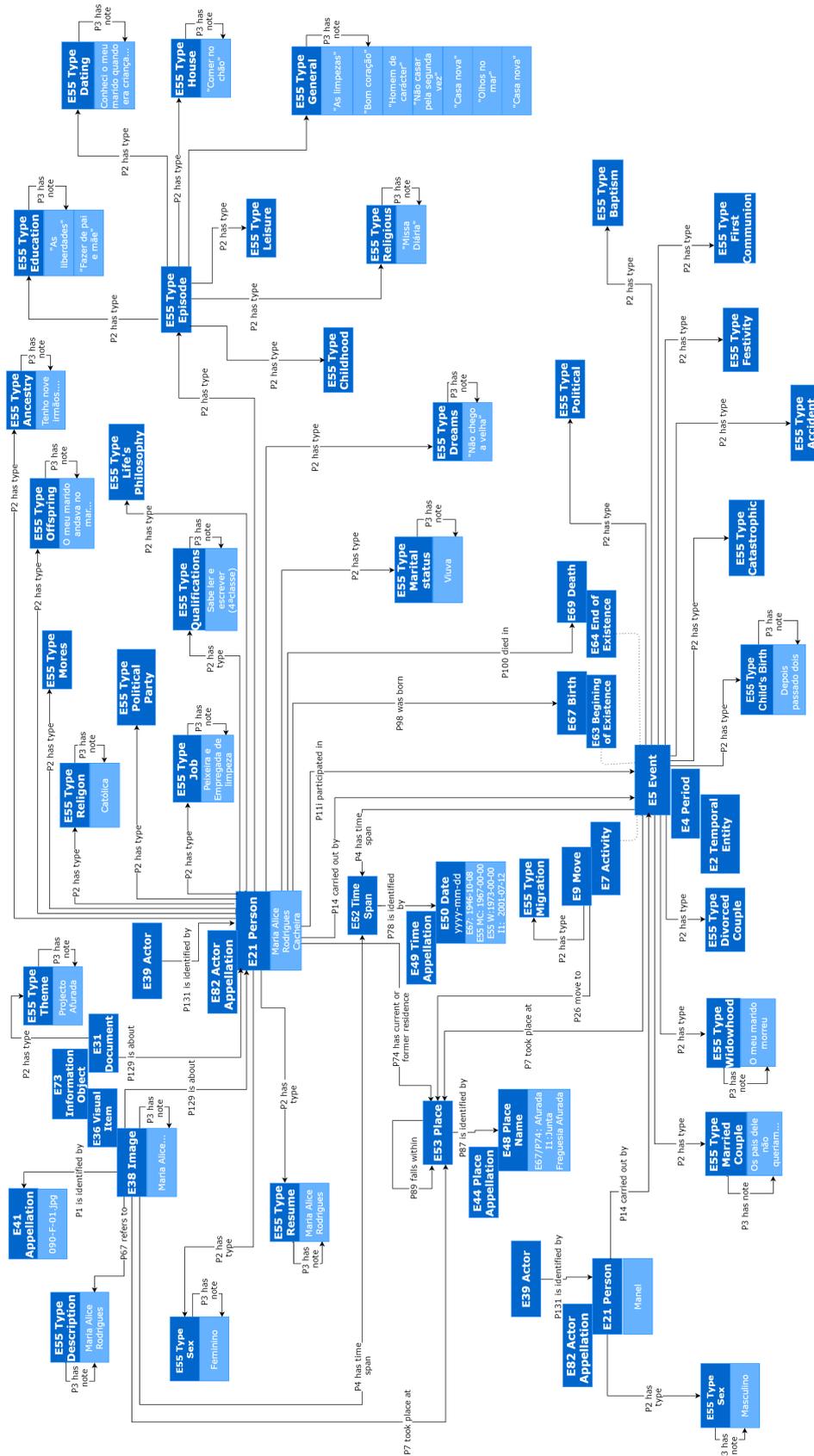


Figure 23.: An instance of *OntoMP*: CIDOC-CRM Representation for Maria Cacheira life story





---

## VISITING MUSEUM OF THE PERSON

---

To better illustrate the main outcome of this masters project, in this appendix the visitor will be guided through the Museum of the Person.

This visit comprises the following steps:

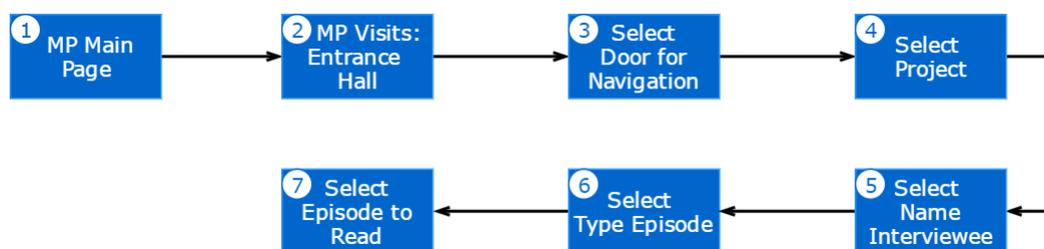


Figure 26.: Workflow Visit to Museum of the Person

The visit starts at: <http://nmp.epl.di.uminho.pt/index.html>. At this address the main page of the Museum of the Person (Figure 27) will be seen – step (1).



Figure 27.: Main Page of the Museum of the Person

At this point, selecting the MP: Visits option in the menu on the left hand side and then the option Entrance Hall, we go to the museum's Entrance Hall (Figure 28) – step (2).

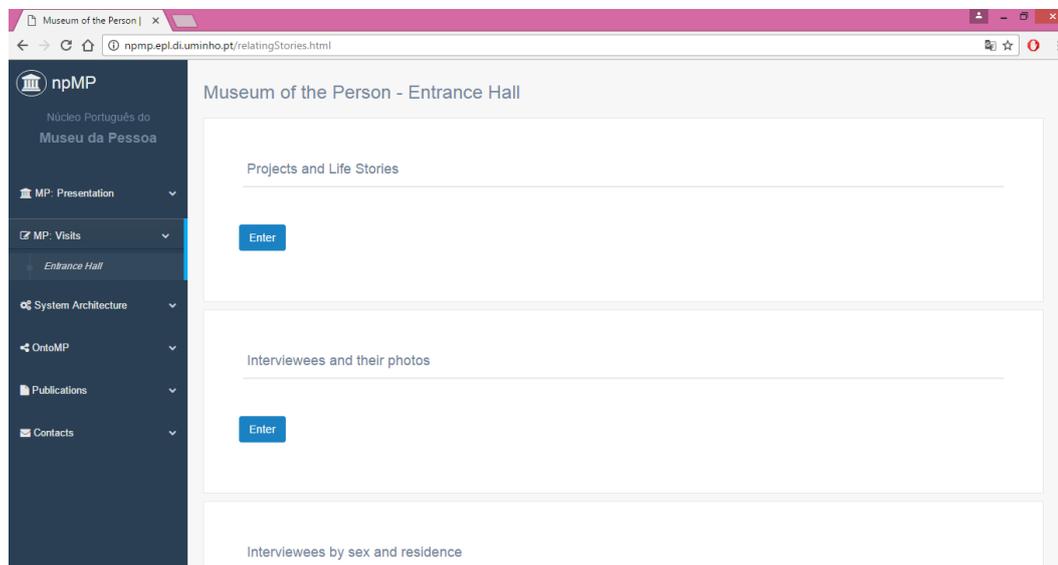


Figure 28.: MP Visits: Entrance Hall

As previously stated, this page contains multiple entrance doors to start the navigation. In step (3) we will visit the room Projects and Life Stories (Figure 28).

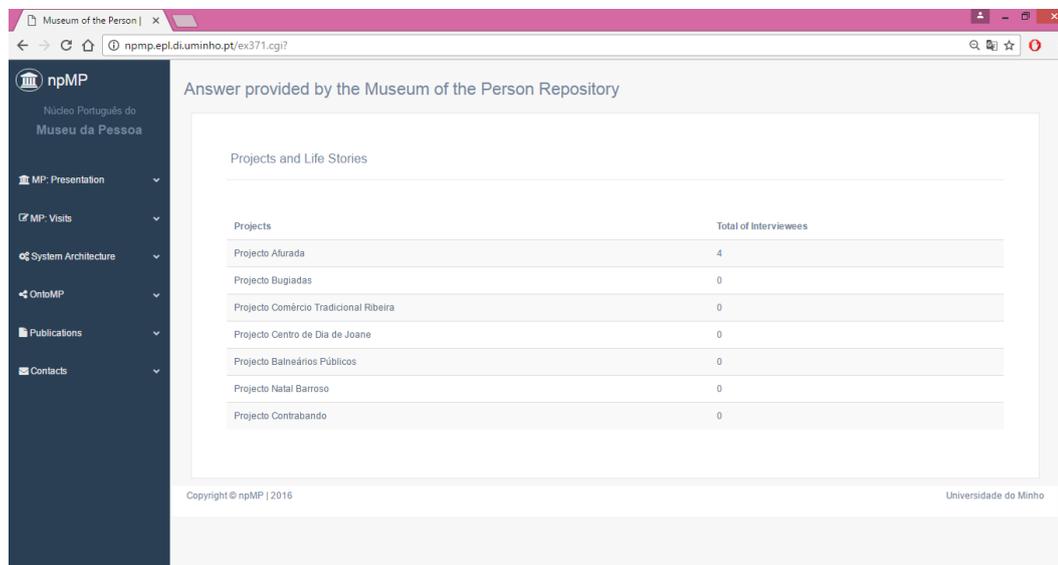


Figure 29.: Room Projects and Life Stories

The room Projects and Life Stories displays the list of all projects in the repository and the number of life stories per project (Figure 29). The step (4) of this visit is to select

the project from that list. In this case, we will visit Projecto Afurada. Clicking on that item, Projecto Afurada, we enter the room of this project (Figure 30).

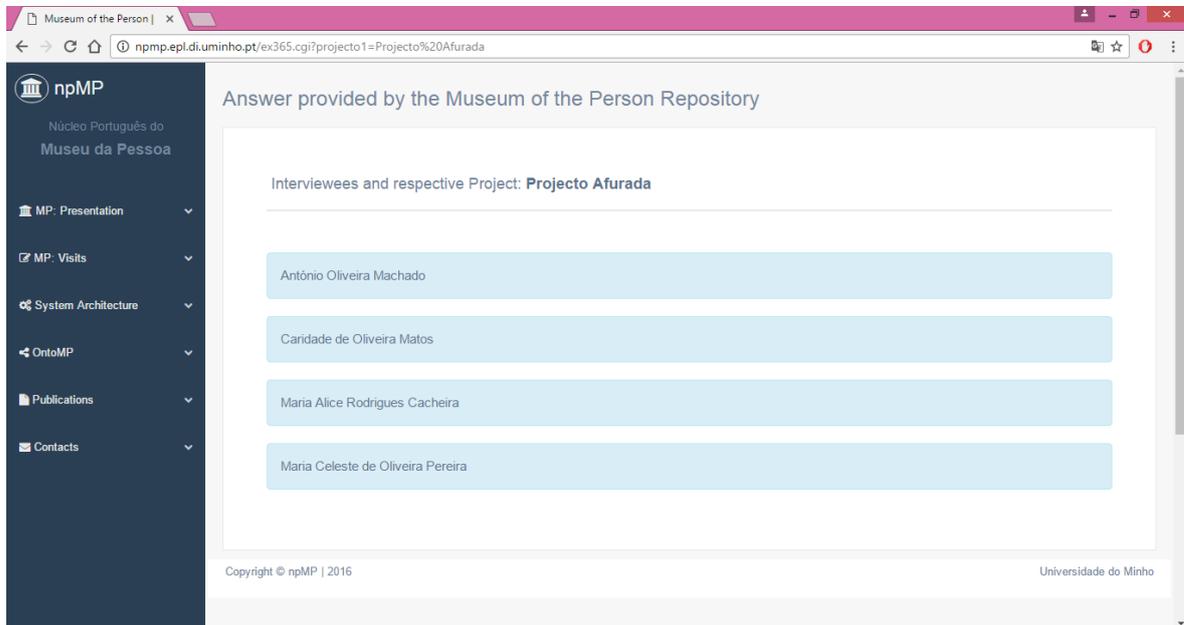


Figure 30.: Room Projecto Afurada

The room Projecto Afurada lists the name of all the people interviewed in the context of that project. To read the life story of someone we just click on the person's name – step (5). The person chosen in this visit is Maria Alice Rodrigues Cacheira (Figure 31).

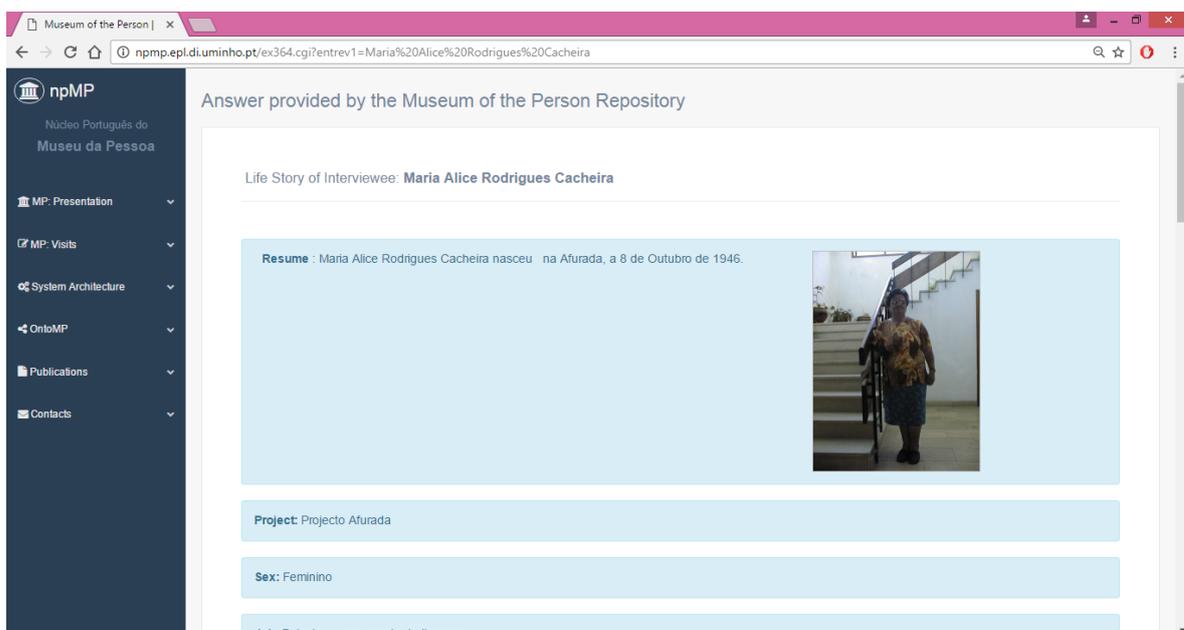


Figure 31.: Life Story of Interviewee: Maria Alice Rodrigues Cacheira

In this room the life story of Maria Alice Rodrigues Cacheira is told. Details, such as date of birth and place of birth, profession, qualifications, ancestry, photos, events in which she participated and episodes narrated, can be found and read. There are several types of episodes, such as educational, religious, general, etc. (Figure 32)

The screenshot shows a web browser window with the URL `npmp.epl.di.uminho.pt/ex364.cgi?entrev1=Maria%20Alice%20Rodrigues%20Cacheira`. The page displays a list of episodes categorized by type. The categories and their corresponding descriptions are as follows:

Episode	Description
Dating	Conheci o meu marido quando era criança ...
Educational	"As liberdades"
	"Fazer de pai e mãe"
General	"As limpezas"
	"Bom coração"
	"Casa nova"
	"Homem de carácter"
	"Não casar pela segunda vez"
	"Olhos no mar"
	"Peixeira"
House	Pelo Natal nós comiamos no chão ...
Religious	Missa Diária

Below this list, there is another table with columns for Event, Description, Date, Place, and Spouse. The first row in this table is:

Event	Description	Date	Place	Spouse
Child's Birth	Depois nascado dois			

Figure 32.: Types of Episodes of *Maria Alice Rodrigues Cacheira* life story

In that room (Figure 32), where the life story is displayed, besides being possible to read all the episodes narrated by Maria Alice Rodrigues Cacheira, it is also possible to see their types. Choosing a type (by clicking over the type name) the visitor will learn more about the episodes of that type narrated by all the interviewees. In concrete, we will choose the type `General` and will navigate to a new room that displays all the episodes of that kind (`General`) for all the people interviewed (Figure 33) – step (6). This room allows to related people and stories regarding the episodes they lived.

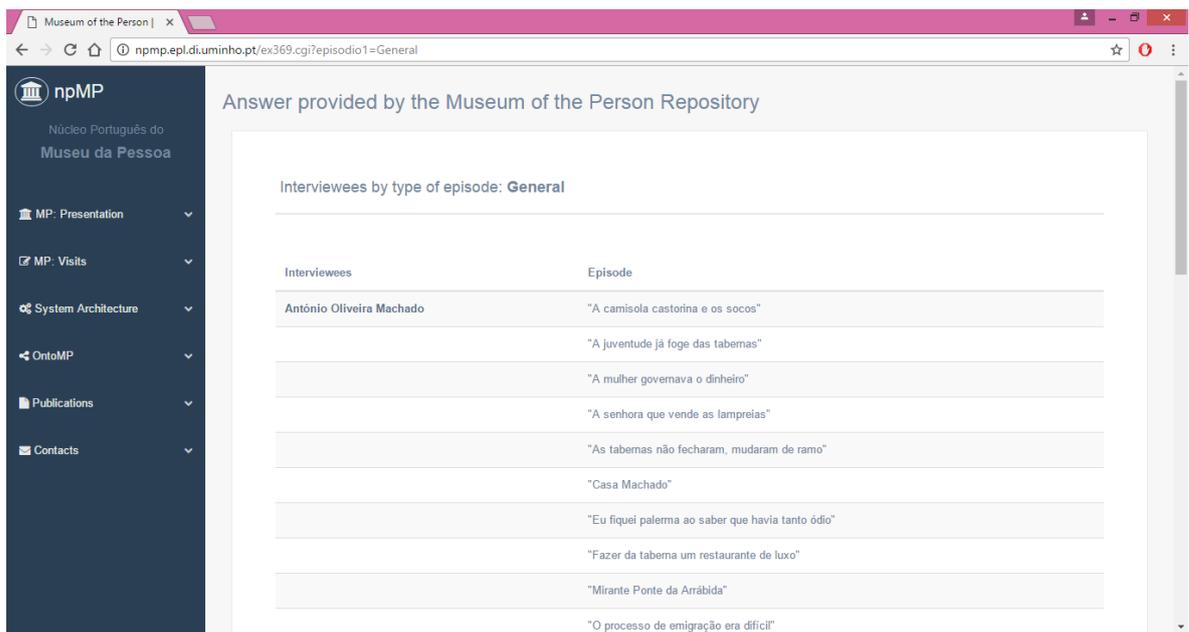


Figure 33.: Room Interviewees by type of episode: General

If the visitor is interested in reading completely an episode of an interviewee, it is enough to click over the episode's name/title (Figure 34) – step (7). To finish the visit we will read, for instance, the episode "A camisola castorina e os socos" experienced by António Oliveira Machado.

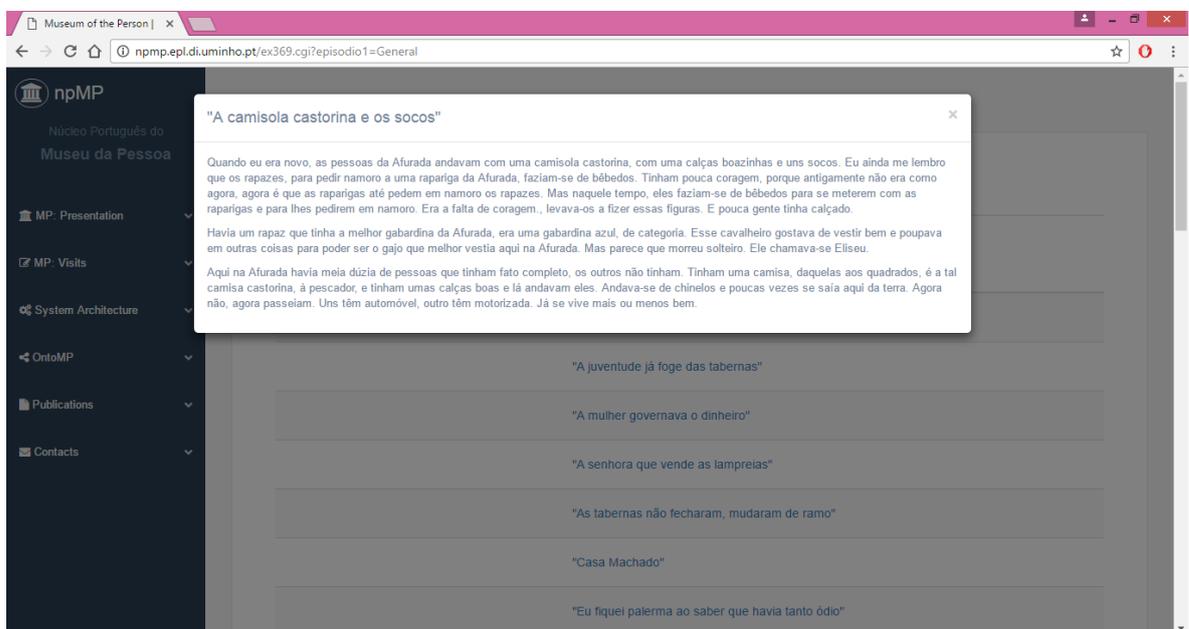


Figure 34.: Episode "A camisola castorina e os socos"