



Universidade do Minho  
Escola de Engenharia

# Building the Museum of the Person

Based on a combined

CIDOC-CRM/ FOAF/ DBpedia Ontology

---

Cristiana Esteves Araújo PG27769

MEI 15/16

## Context

---

- The society is more and more concerned with the *preservation* and the *dissemination* of Cultural Heritage;
- This can be achieved resorting to the *information and communication technologies* that
  - allow physical objects to become *accessible* to anyone and *not deteriorated*.
- Increasing the capability of *information extraction, storage and visualization* Museums have taken advantage to expand its field of action, as well as their own concept:
  - Expanding their geographical borders by providing information on the web to exhibit their collections;
  - Giving birth to completely virtual environments, called **Virtual Museums**.

# Motivation

- Our main research challenge is concerned with the creation of **Virtual Museums**
  - using directly the digital repositories (the documents or databases) that constitute the Museums' assets.
- Apply this research to a concrete case study:
  - the **Museum of the Person (MP)** that contains citizen's interviews that narrate their life stories.



# Goals

- Create a specific **ontology** for the document repository of the Museum of the Person, based on:
  - **CIDOC-CRM** (Comité International pour la Documentation - Conceptual Reference Model);
  - enriched with concepts and relations defined in **FOAF** (Friend-of-a-Friend) and **Dbpedia**;
  - populate the ontology with the interviews, to verify if the ontology is appropriate for the document collection.
- Store in a digital repository (a **TripleStore** and a **Relational database**) the information that constitutes the Museum assets
  - **automatizing** this process (data extraction from XML documents and ontology population);
- Create automatically the **Virtual Museum** Web pages to exhibit the collection of life stories, extracting the information from the stored ontology, using SPARQL queries.

# Museum of the Person, an overview

---

- Museum of the Person was born in Brazil, São Paulo, in 1991
  - created by a group of historians who decided to build the country's history using testimonials of ordinary people.
- Museum of the Person aims at gathering testimonials from every human being, famous or anonymous, to perpetuate his history.
- Persons are used as informers, reporting the events and emotions they experienced and other particular situations they have participated in;
- Its "art collection" is made up of intangible or immaterial things;
- These memories will act as a basic element for social research, because the set of life stories allows to rebuild a social universe.

## Museum of the Person Documents

---

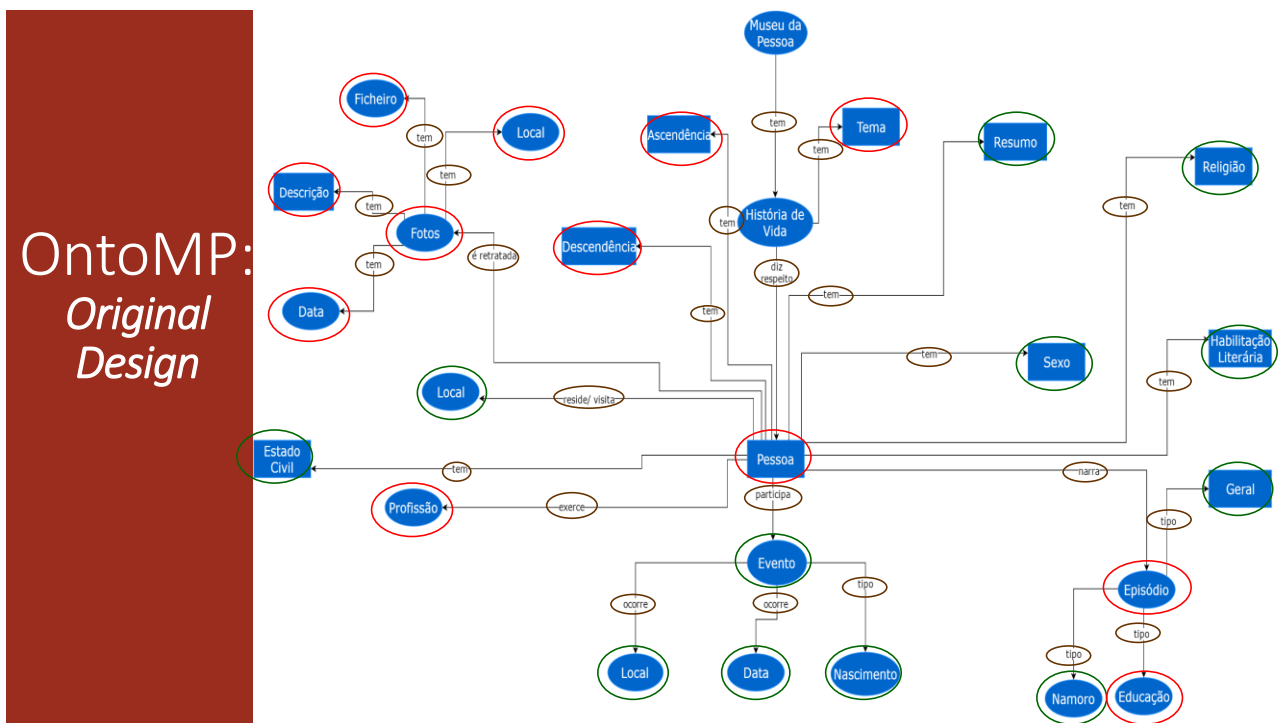
The Museum of the Person's collection consists of XML documents connected to each participant:

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

Additionally a **Thesaurus** that includes key concepts mentioned in the stories.

# OntoMP: an ontology for MP

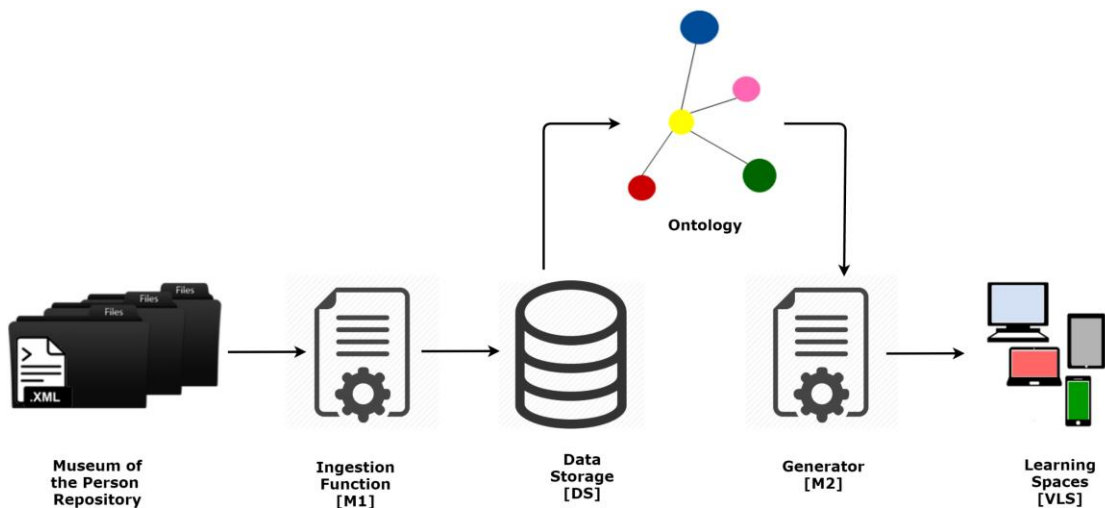
- The first step was the design on a specific **ontology** for the document repository of the Museum of the Person, and its specification based on
  - **CIDOC-CRM**
  - enriched with concepts and relations defined in **FOAF** and **DBpedia**.
- Then the ontology was populated with documents content from the collection.





## MP architecture

- We need an Information System
  - to extract the data from the stored ontology,
  - and build the Virtual Museum Web Pages to exhibit the collection of life stories.
- We propose a **generic architecture** and **two possible instances**



Proposal – General Approach

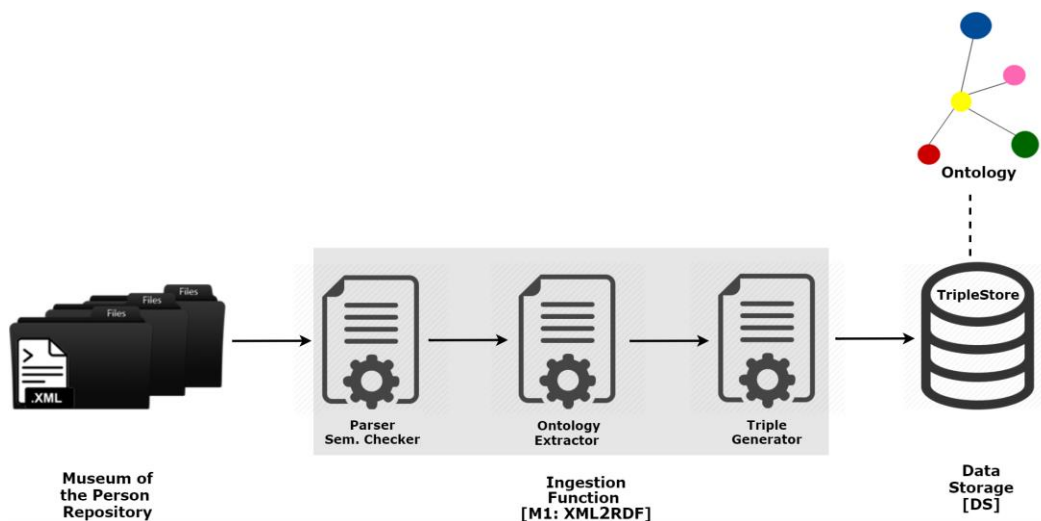
# General Approach

## Approach 1

- Supported on annotated documents and a direct representation of the Ontology (**TripleStore**)

## Approach 2

- Supported on a **Relational Database** and a *mapping to the Ontology*



## Approach 1 - Module 1

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<fotos>
  <foto ficheiro="090-F-01.jpg">
    <quem>Maria Alice Rodrigues Cacheira</quem>
    <onde>Junta de Freguesia da Afurada</onde>
    <quando data="2001-07-12"/>
  </foto>
</fotos>
```

## Data Extraction and Storage

### XML Photo

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

<rdf:Description rdf:about="&ecrm;I0_Interviewee_1"/>
  <rdf:type rdf:resource="&ecrm;E38_Image"/>
  <rdf:type rdf:resource="&foaf;Image"/>
  <foaf:depicts rdf:resource="&ecrm;Interviewee_1"/>
  <P67_refers_to rdf:resource="&ecrm;I0_Description_Interview_1"/>
  <P1_is_identified_by rdf:resource="&ecrm;090-F-01.jpg"/>
  <P4_has_time-span rdf:resource="&ecrm;TS1"/>
  <P7_took_place_at rdf:resource="&ecrm;PL1"/>
</rdf:Description>

<rdf:Description rdf:about="&ecrm;I0_Description_Interview_1">
  <rdf:type rdf:resource="&ecrm;E55_Type"/>
  <P2_has_type rdf:resource="&ecrm;Description"/>
  <P3_has_note rdf:datatype="&xsd:string">Maria Alice Rodrigues
  Cacheira</P3_has_note>
</rdf:Description>
```

```
<rdf:Description rdf:about="&ecrm;2001-07-12">
  <rdf:type rdf:resource="&ecrm;E49_Time_Appellation"/>
</rdf:Description>

<rdf:Description rdf:about="&ecrm;TS1">
  <rdf:type rdf:resource="&ecrm;E52_Time-Span"/>
  <P78_is_identified_by rdf:resource="&ecrm;2001-07-12"/>
</rdf:Description>

<rdf:Description rdf:about="&ecrm;PL1">
  <rdf:type rdf:resource="&ecrm;E53_Place"/>
  <P87_is_identified_by rdf:resource="&ecrm;Place1"/>
</rdf:Description>

<rdf:Description rdf:about="&ecrm;Place1">
  <rdf:type rdf:resource="&ecrm;E48_Place_Name"/>
  <P3_has_note rdf:datatype="&xsd:string">Junta de Freguesia da
  Afurada</P3_has_note>
</rdf:Description>
```

## Data Extraction and Storage

### RDF Photo



```
lexer grammar XML2RDF;

Cabec  : '<[Bb][Ii]>'          -> mode(sBI)
;
Fotos  : '<[Ff][Oo][Tt][Oo][Ss]>' -> mode (sFOTOS)
;
MP      : '<[Mm][Pp]>'          -> mode (sMP)
;
Default: . { ; }
;
...
.....Modes specification.....
...
```

```
mode sFOTOS;
GetSFOTOS : '<foto>' -> mode(sFOTO)
;
OutFOTOSSAVE : '</fotos>' -> mode(DEFAULT_MODE)
;
DefaultsFOTOS : . { ; }
;

mode sFOTO;
GetFOTO : [ ]+'ficheiro=' -> mode (sFICHEIRO)
;
GetQUEM : '<quem>' -> mode (sQUEM)
;
GetQUANDO : '<quando>' -> mode (sQUANDO)
;
GetFACTO : '<facto>' -> mode (sFACTO)
;
GetONDE : '<onde>' -> mode (sONDE)
;
OutFOTOS : '</>' -> mode (sPRINTTUDO)
;
```

## Translation System

### XML -> RDF

```
mode sPRINTTUDO;
GetsPRINTTUDO : 'foto' {

    pessoa.AddImage("I"+newCountKeyFicheiro+"_Interviewee_"+
countinterview);

    System.out.print("<rdf:Description rdf:about=\"&ecrm;\"");
    System.out.println(ficheiro+">");
    System.out.println("\t<rdf:type
rdf:resource=\"&ecrm;E41_Appellation\"/>");
    System.out.println("</rdf:Description>\n\n");
    System.out.println("<rdf:Description
rdf:about=\"&ecrm;I"+newCountKeyFicheiro+"_Interviewee_"+
countinterview+"\"/>");
    System.out.println("\t<rdf:type
rdf:resource=\"&ecrm;E38_Image\"/>");
    System.out.println("\t<rdf:type
rdf:resource=\"&foaf;Image\"/>");
    System.out.println("\t<foaf:depicts
rdf:resource=\"&ecrm;Interviewee_"+countinterview+"\"/>");

    if(!quem.equals("")){
        System.out.println("\t<P67_refers_to
rdf:resource=\"&ecrm;I"+newCountKeyFicheiro+
"_Description_Interview_"+ countinterview+"\"/>");
        System.out.println("\t<P3_has_note
rdf:datatype=\"&xsd:string\">"+facto+"</P3_has_note>"); }
}
```

```
System.out.println("\t<P1_is_identified_by
rdf:resource=\"&ecrm;"+ficheiro+"\"/>");

...

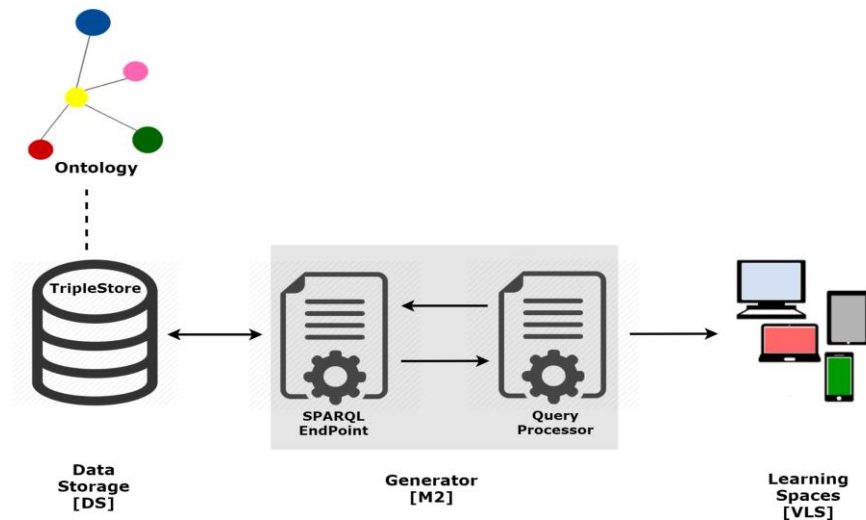
if(!quem.equals("")){
    if(whoAdded){
        System.out.println("<rdf:Description
rdf:about=\"&ecrm;I"+newCountKeyQuem+"_Description_Interview_"+
countinterview+"\"/>");
        System.out.println("\t<rdf:type
rdf:resource=\"&ecrm;E55_Type\"/>");
        System.out.println("\t<P2_has_type
rdf:resource=\"&ecrm;Description\"/>");
        System.out.println("\t<P3_has_note
rdf:datatype=\"&xsd:string\">"+quem+"</P3_has_note>");
        System.out.println("</rdf:Description>\n\n"); }
    }

    .

OutsPRINTTUDO : '>' -> mode(sFOTOS)
;
```

## Translation System

### Print RDF



## Approach 1 - Module 2

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

<rdf:Description rdf:about="%ecrm;I0_Interviewee_1"/>
  <rdf:type rdf:resource="%ecrm;E38_Image"/>
  <rdf:type rdf:resource="%foaf;Image"/>
  <foaf:depicts rdf:resource="%ecrm;Interviewee_1"/>
  <P67_refers_to rdf:resource="%ecrm;I0_Description_Interview_1"/>
  <P1_is_identified_by rdf:resource="%ecrm;090-F-01.jpg"/>
  <P4_has_time-span rdf:resource="%ecrm;TS1"/>
  <P7_took_place_at rdf:resource="%ecrm;PL1"/>
</rdf:Description>

<rdf:Description rdf:about="%ecrm;I0_Description_Interview_1">
  <rdf:type rdf:resource="%ecrm;E55_Type"/>
  <P2_has_type rdf:resource="%ecrm;Description"/>
  <P3_has_note rdf:datatype="%xsd:string">Maria Alice Rodrigues
    Cacheira</P3_has_note>
</rdf:Description>
```

```
sys.stdout.buffer.write(bytes("Content-type: text/html\n\n"))
<!DOCTYPE html>
<html><head>
  <meta charset="utf-8">
  <meta http-equiv="X-UA-Compatible" content="IE=edge">
  <meta name="viewport" content="width=device-width, initial-scale=1">
  <title>Museum of the Person |</title>
  ...
<body>
  ...
  <div class="x_title">
    <h2>Interviewee and their photos</h2><br/><br/>
  </div>
  <div class="x_content"><br/><br/>
    <table class="table table-striped responsive-utilities jambo-table bulk_action">
      <thead><tr class="headings">
        <th class="column-title" width="18%">Interviewee</th>
        <th class="column-title" width="15%">File</th>
        <th class="column-title" width="25%">Caption</th>
        <th class="column-title" width="22%">Description</th>
        <th class="column-title" width="8%">Date</th>
        <th class="column-title" width="20%">Place</th>
      </tr></thead>
      ...
    </table>
  ...
  sys.stdout.buffer.write(bytes("\n\n"))
  <div class="pull-left">Copyright &copy; npMP development team | 2016</div>
  <div class="pull-right">Universidade do Minho</div>
</div></body></html>
```

## Generating the Learning Spaces (Exhibition Rooms)

RDF -> HTML

```

PREFIX : <http://erlangen-crm.org/150929/>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX dbp: <http://dbpedia.org/ontology/>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>

SELECT DISTINCT ?nome ?ficheiro ?personDesc ?legenda ?dataPhoto
?placePHOTO

WHERE {
  ?pessoa a :E21_Person;
    :P129_is_subject_of ?doc;
    foaf:name ?nome;
    foaf:depiction ?fotos.
    ?fotos :P1_is_identified_by ?ficheiro.
  OPTIONAL{
    ?fotos :P3_has_note ?legenda.}
  OPTIONAL{
    ?fotos :P67_refers_to ?personDesc.
    ?personDesc :P3_has_note ?personDescricao.}
  OPTIONAL{
    ?fotos :P4_time-span ?data.
    ?data :P78_is_identified_by ?dataPhoto.}
  OPTIONAL{
    ?fotos :P7_took_place_at ?placePhoto.
    ?placePhoto :P87_is_identified_by ?placePhoto1.
    ?placePhoto1 :P3_has_note ?placePHOTO.}
}ORDER BY ?nome ?ficheiro

```

```

entArray = []
for result in results["results"]["bindings"]:
    entrevistado= result["nome"]["value"]
    entrevistado =
    entrevistado.replace("http://erlangen-crm.org/150929/", "")
    sys.stdout.buffer.write("<tr><td> ", "utf-8")
    if entrevistado not in entArray:
        sys.stdout.buffer.write(entrevistado)
        entArray = entrevistado
        sys.stdout.buffer.write("</td> ", "utf-8")
        ficheiros= result["ficheiro"]["value"]
        ficheiros =
        ficheiros.replace("http://erlangen-crm.org/150929/", "")
        sys.stdout.buffer.write(bytes("<td> " + "<div
class='highslide-gallery'><a
href='images/galleryInterviews/'"+ ficheiros + " "
class='highslide' onclick='return hs.expand(this)'><img
class='img-responsive' src='images/galleryInterviews/'"+
ficheiros + " " "></a></div> " " "</td> ", "utf-8"))
    try:
        legendas= result["legenda"]["value"]
        legendas =
        legendas.replace("http://erlangen-crm.org/150929/", "")
        sys.stdout.buffer.write(bytes("<td> " + legendas +
"</td> ", "utf-8"))
    except KeyError:
        sys.stdout.buffer.write(bytes("<td> " + " " + "</td> ",
"utf-8"))
...

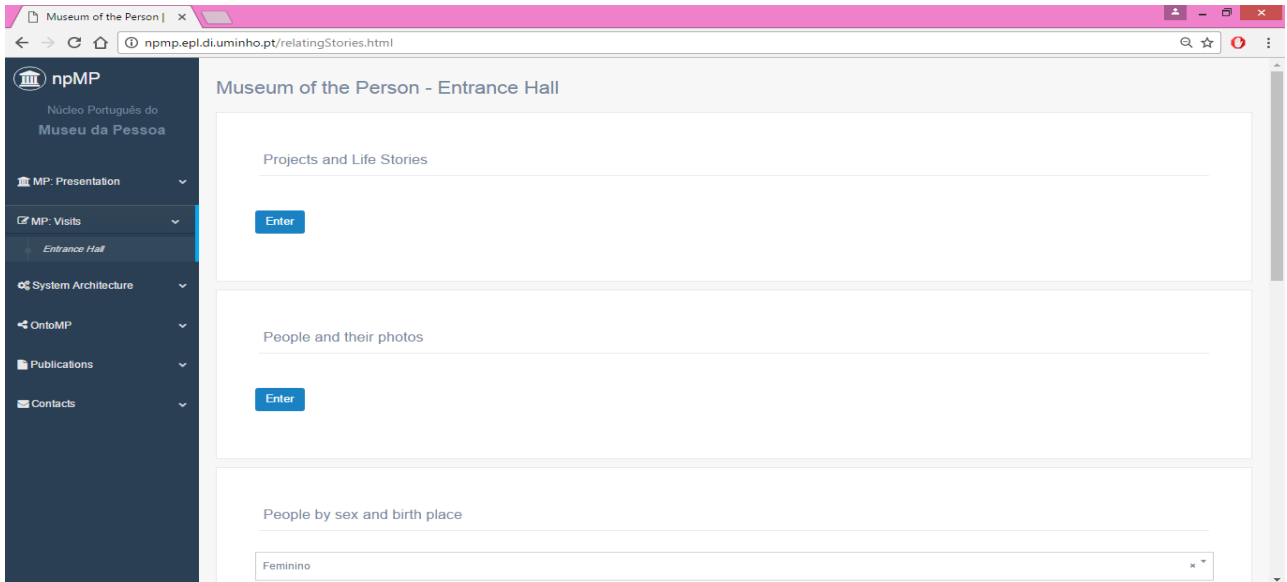
```

## Querying and Displaying

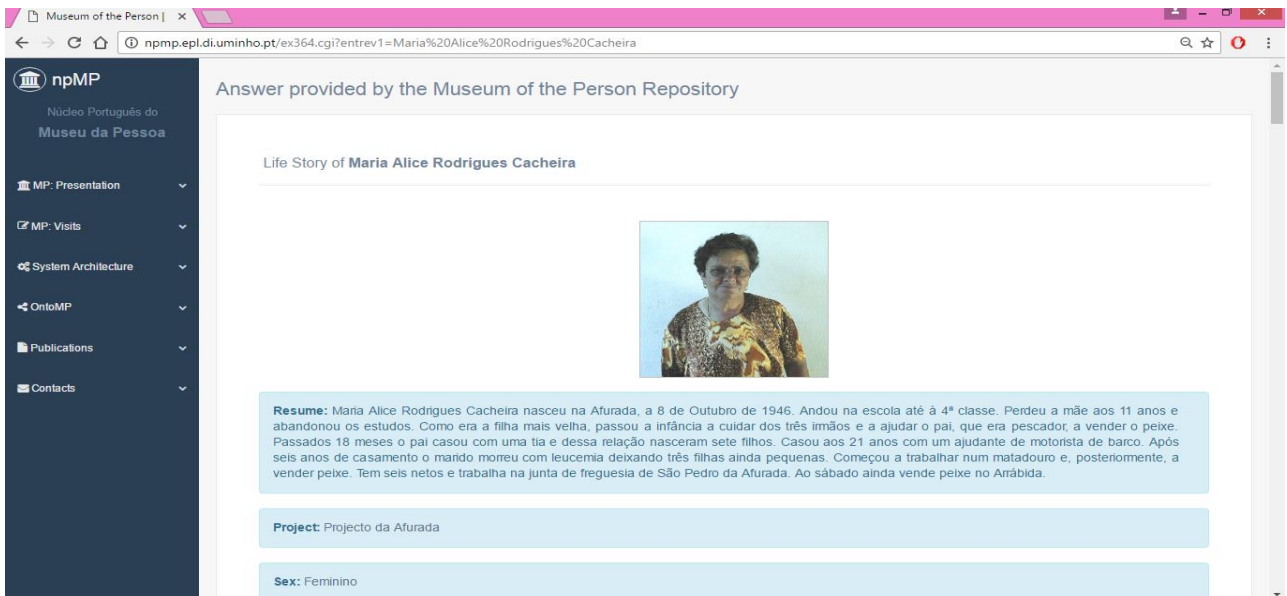


The screenshot shows a web browser window with the URL `npmp.epl.uminho.pt/index.html`. The page title is "Núcleo Português do Museu da Pessoa". On the left, there is a dark sidebar with the npMP logo and a menu with items: "MP: Presentation", "MP: Visits", "System Architecture", "OntoMP", "Publications", and "Contacts". The main content area features a large black and white photograph of a group of people standing in front of a building. At the bottom of the page, there is a copyright notice "Copyright © npMP | 2016" and the text "Universidade do Minho".

npMP Project site  
<http://npmp.epl.uminho.pt>



*npMP Entrance Hall*  
<http://npmp.epl.uminho.pt>



*npMP Life Stories Room (1 person)*  
<http://npmp.epl.uminho.pt>

# Conclusion

---

The main contributions of this masters' work are:

- The Documents collection of npMP (interviews, Thesaurus, DTD's and photographs) was analyzed.
- An Ontology for the Museum of the Person , called OntoMP, was designed
  - OntoMP was described in a standard ontology format used for museums, CIDOC-CRM enriched with pertinent FOAF and DBpedia concepts and properties.
- A General Architecture was proposed to create the museum's virtual exhibition rooms, as web pages, extracting information from the museum's digital repository.
  - Two instantiation Approaches were designed, depending on the Data Storage (TripleStore or Relational Database).
  - Approach 1 was fully implemented

# Conclusion

---

- The Ingestion Function [M1], to automatically extract data from the XML documents and generate the RDF triples, was implemented by a specially tailored text filter.
- The Generator [M2], to configure the Virtual Learning Spaces according to the exhibition requirements, was implemented:
  - using a Python script, complemented with HTML and CSS code.
  - reusing the SPARQL queries manually created, those needed for a concrete request are selected and sent to the SPARQL endpoint.
  - processing the returned data to be displayed in the [VLS].

# Future Work

---

Some directions for future work are:

- Improve the visual interface of the npMP increasing the interactivity and exploiting geographical information applications following what is being done in Museu Marítimo de Sesimbra;
- Improve the system performance, resorting to other technologies or adapting the system's architecture;
- Implement Approach 2;
- Combine the Ingestion Function [M1: XML2RDF] of Approach 1 with the Generator [M2: CaVaGen] of Approach 2;

# Future Work

---

Some directions for future work are (cont.):

- Investigate new standards and other notations for ontology description;
- Integrate and develop new tools to automatize the complete life cycle of npMP:
  - integrate in our platform a mobile application that was built to record the interviews;
  - create an application to transcribe audio interviews to text, and to markup the important concepts.
- 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) 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 or the Centro de Interpretação do Território de Ponte de Lima.