

# Romaria De N<sup>a</sup> Sr<sup>a</sup> D'Agonia: Building a Digital Repository and a Virtual Museum

Sara Cristina Freitas Queirós ✉

ALGORITMI Research Centre / LASI, DI-University of Minho, Braga, Portugal

Cristiana Araújo ✉ 🏠 

ALGORITMI Research Centre / LASI, DI-University of Minho, Braga, Portugal

Pedro Rangel Henriques ✉ 🏠 

ALGORITMI Research Centre / LASI, DI-University of Minho, Braga, Portugal

---

## Abstract

*Romarias* are Christian pilgrimages that occur in order to celebrate a specific saint. *Romaria de N<sup>a</sup> Sr<sup>a</sup> d'Agonia* (RNS Agonia, for short) occurs aimed at celebrating *Nossa Senhora da Agonia*, patron of all Fishermen, at Viana do Castelo, Portugal. RNS Agonia is a very old event that surely belongs to the Minho's Cultural Heritage. There are many written documents, of various types, that describe the event; so, their digital preservation is mandatory. However, digital preservation is not restricted to a database of digital images obtained by scanning the documents. In this paper we are concerned with digital repositories of XML-based annotated documents from which we can extract automatically data to build a virtual museum that helps on disseminating information about RNS Agonia. Such a Web resource is crucial to support people wishing to know more about that pilgrimage, and also as a booster for tourism. The paper describes the different stages of this project, including the documents annotation process, data extraction mechanisms, the creation of a triple storage to archive the knowledge built from the sources analyzed, and the virtual museum implementation. The methodological approach devised for the project under discussion is based on the creation of an ontology that describes the RNS Agonia domain completely. The idea is to define the XML dialect, to be used in the annotation, from the ontology. Moreover the ontology will also lead the definition of the triple store used to set up the knowledge base that feeds the museum.

**2012 ACM Subject Classification** Applied computing → Document preparation; Information systems → Ontologies

**Keywords and phrases** Ontology, XML, Romaria, Pilgrimage, Digital Knowledge Repository, Triple Storage Database, Virtual Museum

**Digital Object Identifier** 10.4230/OASICS.SLATE.2023.16

**Funding** This work has been supported by FCT – Fundação para a Ciência e Tecnologia within the R&D Units Project Scope: UIDB/00319/2020.

## 1 Introduction

With the increasing transition from the analog to the digital world, paper documents have become obsolete and, in consequence, all the past knowledge about the evolution ends up in oblivion. Portuguese *Romarias* are one of the main folk events that have been affected by this change. Alongside, with time, customs and traditions have been changing and adapting to this new reality.

**Romaria de N<sup>a</sup> Sr<sup>a</sup> d'Agonia** occurs every year in Viana do Castelo, being known as the biggest *romaria* in the country. With faith and tradition, between the devotion and the feast of a people, this is how, year after year, for more than two centuries, the RNS Agonia has been made. Like all the others, it has a strong historical past, built up from a set of events that have been expanding, particularly in the last decade.



© Sara Cristina Freitas Queirós, Cristiana Araújo, and Pedro Rangel Henriques; licensed under Creative Commons License CC-BY 4.0

12th Symposium on Languages, Applications and Technologies (SLATE 2023).

Editors: Alberto Simões, Mario Marcelo Berón, and Filipe Portela; Article No. 16; pp. 16:1–16:16

OpenAccess Series in Informatics



OASICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

All these events have been documented in articles, books, news, and even transmitted orally through generations, so that all the information is spread over different sources. Therefore, this project aims to collect these diverse materials such as newspaper news, posters, flyers, postcards, photographs and build a documentary repository. From this repository, it will be possible to gather the historical and cultural information, which will allow the recreation of those events, since the beginning of this pilgrimage up to the present day.

To guide the collection and data extraction tasks, we built an ontology – OntoRomaria – to describe the domain of festivities of RNS Agonia. This ontology is composed of the concepts that describe the domain of Pilgrimage and the relationships that link the concepts, building the triples. OntoRomaria aims to structure for archival purposes, in a triple-storage format, all the information collected about the Pilgrimage. In order to publicly display the collected information, we built a Virtual Museum. In this Virtual Museum the users can conceptually browse the knowledge repository about RNS Agonia, a pilgrimage that belongs to Minho's cultural heritage illustrating their strong roots and traditions.

This paper is organized into 9 sections. In Section 2, the RNS Agonia and the main concepts that characterize it are presented, as well as the events that take place throughout the pilgrimage. Section 3 defines two crucial concepts in this project: Virtual Museums and Ontologies. Section 4 presents the ontology built to describe the domain of RNS Agonia. Section 5 shows the construction of the knowledge repository. In Section 6 the architecture of the Virtual Museum is discussed. Section 7 describes the development of the Virtual Museum. Section 8 shows the interface of the Virtual Museum of *Romaria de N<sup>a</sup> Sr<sup>a</sup> d'Agonia*. Finally, Section 9 presents conclusions and future work.

## 2 Romaria de Nossa Senhora da Agonia

*Romaria de N<sup>a</sup> Sr<sup>a</sup> d'Agonia* occurs every year in Viana do Castelo, being known as the biggest *romaria* in the country. Pride in essence, faith, tradition and devotion are the main reasons that drive people, for more than two centuries, to plan, prepare and celebrate this christian pilgrimage [11]. This pilgrimage is devoted to *Nossa Senhora da Agonia*. *Agonia*, which translates to “agony”, was chosen as a surname for Our Lady (in Portuguese, “Nossa Senhora”) due to the suffering that the constant sea hazards, such as shipwrecks, caused to fishermen's families. The sea in the region was very rough, causing frequently shipwrecks, and the families of the brave fishermen used to wait for them on docks, distressed by the struggle that these men were fighting for their lives. Kneeling, they prayed fervently to Our Lady of Agony. As a way to relieve their pain, they would kneel and pray vigorously to *Nossa Senhora*, hoping to put an end to the agony of the uncertainty about their loved ones [9]. Nowadays, northern portuguese fishermen and their families, carry on the tradition of visiting *Nossa Senhora da Agonia* sanctuary to pray for protection for the times to come and thank for all the givens of that year.

Nevertheless, this *romaria* is not only about praying, but also about proudly exhibit old costumes and other traditions. All festivities lasts four to five days full of events. As mentioned before, over the last decade this event has become increasingly more publicized, attracting more than 1 million visitors per year, improving the region's tourism.

## 2.1 Main Concepts

To be able to understand the concept of this Christian Pilgrimage, it is necessary to explore a little bit all the concepts that are associated with it:

1. **Mordoma** is a core term when RNS Agonia is mentioned. It describes usually young and single **women** from Viana do Castelo that wear the traditional costumes (“trajes”) as a way to pay tribute to the time that gave origin to the beginning of the celebration of these festivities. They are responsible for being the “messengers” by introducing and inviting everyone to gather and celebrate.
2. **Gold** is the most relevant ornament used by *Mordomas* in their *Trajes*, without it, it’s incomplete. At the beginning of this traditions, gold jewelry was passed down from generation to generation in the form of inheritance. It had its greatest highlight in christian festivities, trips to the city of Viana do Castelo, where young woman would proudly exhibit all that gold on their chest, representing their families social status.
3. **Chieira** it describes pride and vanity (on the positive side of this adjectives) for wearing the costumes and gold.
4. **Traje** is a portuguese word used to reference old, traditional and historical costumes. Within the context of this pilgrimage, the costumes in question were worn by women from the old times, in the various occasions of their lives such as work, mourning, celebration, engagement, etc.

## 2.2 Main Events

During the RNS Agonia, five relevant events take place:

- **Parade of *Mordomas***: parade through the city where women from Viana do Castelo proudly exhibit their costumes.
- **Historical/Ethnographic parade** is an highly theatrical moment where traditional celebrations and old habits are represented by each parish of Viana do Castelo.
- **Procession to the sea**: The image of *Nossa Senhora da Agonia* is carried through the streets decorated with salt tapestries to the sea, where it is loaded in one of the flowery boats and sails part of the river Lima.
- **Folk festival**: all the vast variety and richness of *trajes* for all moments, traditional music and folk related traditions are exhibited.
- **Fireworks**: On the end of the festivities of each pilgrimage day fireworks are set off. The last day is marked by a special set of fireworks, that are let off on Eiffel’s bridge and boats on the river, leading to a spectacular show of lights that also reflects on the river. This is usually called as “*Serenata*” (Serenade) because it’s accompanied by soulful songs.

## 3 Virtual Museums and Ontologies

To develop the solution it was necessary to understand two concepts: Virtual Museum and Ontology.

On the one hand, **virtual museums** are a digital extension on the Internet, basically, a museum without walls, capable of extending ideas and concepts into the digital space, trying to reveal the essential nature of the museum. At the same time it allows people, who might never be available of visiting a certain museum in person, to reach it out [7].

On the other hand, **ontology** is an artifact used for representation, where these are intended to designate some combination of universal concepts, defined classes, and certain relationships between them [2].

Virtual museums developed using ontologies have gained increasing popularity in recent years as a way to preserve and present cultural heritage in a digital format, since an ontology is a formal representation of a set of concepts within a domain and the relationships between those concepts. One example of a virtual museum developed using ontologies is the Virtual Museum of Canada (VMC) (it was decommissioned in order to evolve out to Digital Museums Canada in 2020<sup>1</sup>). The VMC used the Cultural Heritage Ontology (CHO) to represent the information about the museum's collection. At a national level is possible to mention *Museu da Emigração e Comunidades* (Immigration and Communities Museum)<sup>2</sup>. However, to the best of our knowledge, all of these examples developed the ontology based on annotations made on the documents. The approach presented in this paper aims to reverse the direction of this workflow and systematize the annotation of documents from an ontology already formulated for this purpose.

One of the key advantages of using ontologies in virtual museums is that they provide a standardized way of representing and organizing the information, which makes it easier to search, browse, and access the museum's content. Ontologies also enable interoperability between different virtual museums, allowing information to be shared and reused across different platforms around the world [8].

## 4 OntoRomaria

The first step towards designing the virtual museum was to define an ontology to describe the domain under work. In order to do it formally, it was used OntoDL+, a Domain Specific Language that was created to allow the rigorous description of ontologies by non-programmer users. OntoDL+ is not a markup language like other ontology description languages (for instance the well known standards OWL or RDF-Schema). It was designed to be easy to write and clear to read, aiming at being used with simplicity by anyone from Science to Humanities. OntoDL+ Compiler not only checks the specification to assure well-formedness and its semantic correctness, but also can convert it to other ontology representation languages [4]. That language just requires the identification of the relevant *concepts* present in the domain, the *attributes* that characterize them, the *relations* that will link concepts to define the domain meaning, and the *triples* (two concepts, the Subject and the Object, and a relation, the Predicate, that connects them) [1]. Optionally a fourth block can be added to declare the *individuals* that will instantiate the concepts, allowing for the ontology population. Listing 1 illustrates that DSL exhibiting a fragment of a OntoDL+ template.

In order to build the ontology, the first step was to read and analyze multiple documents such as newspaper news, posters, flyers, postcards, photographs, among others, highlight the relevant data present in them, and understand which concepts are more important and what are the relevant relations among them. Complementary to that bibliography research process, and in order to enrich the acquisition of knowledge concerning RNS Agonia, many conversations were held with experts to recover the oral tradition.

After that long manual process of information retrieval, the next step was to design and specify in OntoDL+ language the envisaged OntoRomaria to describe RNS Agonia. Listing 2 shows a short extract of the ontology developed.

<sup>1</sup> Available at <https://www.digitalmuseums.ca>

<sup>2</sup> Available at <https://epl.di.uminho.pt/~ritafaria/MEC/index.php>.

■ **Listing 1** A template exhibiting OntoDL+ Syntax.

```

Ontology ontologyName
attributes { %attrs declaration block (optional)
  attribA: string,
  attribB: enum( va, vb, vc )
}
concepts { %concepts declaration block
  ConcA[attribA, attribC:int],
  ConcB[attribB],
  ConcC, ConcD, ConcE
}
individuals { %individuals declaration block
  indA, indB
}
relations { %relations declaration block
  % is-a, iof, pof are predefined
  relA[domain: ConcA, codomain: ConcB],
  rel1, rel2
}
triples { %triples declaration block
  ConcA =relA=> ConcB;
  ConcB =[ rel1 => ConcC, ConcD;
          rel2 => ConcE ];
  indA =iof=> ConcA[attribA="val1", attribC=3];
  indB =iof=> ConcB[attribb=vb]
}

```

## 5 Virtual Museum, building the Repository

As previously said, one of the objectives of this project is to create a Virtual Museum extracting the information from a digital repository, which should be populated automatically from documents using a data extraction process. However to enable that process, the documents to be used as information sources have to be annotated in XML.

XML is a markup language used worldwide due to the fact that it is a software and hardware independent tool, easy to be read by humans and computers that allows to emphasize information in a normal text by just wrapping data between tags [10] (annotation elements that may be complemented by attributes). The XML language has no predefined tags. This meta-language just establishes a small set of rules defining what can be a tag, the way annotations open and close, and the way elements can contain other elements. So, when a new XML dialect is desired, it is up to the user to define the set of tags he needs and the element structure he desires.

Listing 3 is an example of a text (that is concerned with RNS Agonia, our specific working domain) annotated in our XML dialect.

As explained above, to start the process of document annotation, it was necessary to define the appropriate tags. The recurrent problem that information engineers face in that stage, to perform that creative task, is how to deal with that subjective concept: *appropriateness*; how to choose XML elements (tags, attributes and relations among elements) that are the most adequate for a domain. Our contribution to overcome this difficulty is the proposal of an *ontological approach*: XML elements shall be defined taking OntoRomaria (conceptual level, without individuals) into account.

■ **Listing 2** OntoRomaria ontology in OntoDL+ (fragment).

```

Ontology OntoRomaria

concepts {
  Romaria[dataInicio: string, dataFim: string, local: string,
          data: string, sinon: string],
  Evento[local: string, data:string,
         fotografia: string, descricao: string, sinon: string],
  Desfile_Mordomia[nr_mordomas: int],
  Mordoma[descricao: string, explicacao: string, sinon: string],
  Traje[material: string, adereco: string, cor: string,
        significado: string, tipo: string, sinon: string],
}
individuals {
  Romaria_S_Agonia,
  Traje_Mordoma, Traje_Trabalho, Traje_Feirar
}
relations {
  usa, enverga, ocorre, desfila
}
triples {
  Romaria_S_Agonia =iof=> Romaria[dataInicio=...];
  Romaria_S_Agonia = ocorre => Evento;

  Desfile_Mordomia =is-a=> Evento;
  Desfile_Mordomia = desfila => Mordoma;
  Mordoma = usa => Traje;

  Traje_Mordoma =iof=> Traje[material=...],
  Traje_Trabalho =iof=> Traje[material=...];
  Traje_Feirar =iof=> Traje[material=...]
}.

```

Therefore, based on the ontology defined in the previous phase, a group of XML elements was created to markup documents concerned with the RNS Agonia.

For each concept in the ontology a XML element is created; the tag is the concept name. If the concept has attributes, the respective element will have a similar list of attributes, not required to make the annotation more flexible. For each relation, an empty element is created; once again the tag is the relation name.

Using the DTD created according to the principles above, the annotation process followed the guidelines below:

- A sentence that describes a specific 'Concept' is annotated inside the tags: "`<Concept></Concept>`".
- When 'Concept' is characterized by specific property 'att', it is annotated inside tags with the attribute 'att' associated to its contextual value 'txt-value': "`<Concept att='txt-value'></Concept>`".
- When two concepts are annotated according to the rules above, and the former is the subject of a relation to the later, which is the object (target) of such a relation, then a tag "`<verb/>`" is inserted between the concepts; notice that 'verb' denotes the referred relation. This strategy enables an easy extraction of the ontology triples.
- When a concept has properties whose values are long texts, such as a *description* or

*explanation*, these attributes are treated as an element '`<Description>`' instead of an attribute. Care must be taken to enclose that tag within the concept tag. Here follows an example:

```
<Procissao_ao_Mar>
  <Descricao>vénia da Senhora a todos os que a veneram
</Descricao>
</Procissao_ao_Mar>
```

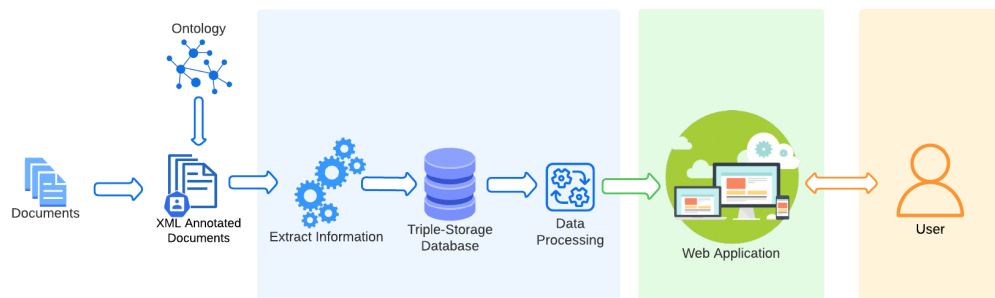
■ **Listing 3** Example of an annotated document with XML tags based on OntoRomaria.

```
<Romaria_Sra_Agonia>
  <ocorre/>
  Desfile da Mordomia
  <Desfile_Mordomia nr_mordomas='mais_de_meio_milhar'> Preparam-se
    durante dias, vestem-se rigorosamente e
    saem a rua para mostrar a historia da <desfila/><Mordoma
    sinon='mulher_vianense'> mulher vianense</Mordoma>.
    Mais de meio milhar de mulheres
    desfilam orgulhosamente, espalhando a tradicional chieira
    pela cidade no Desfile da Mordomia. Envergando os mais
    belos <usa/><Traje><Descricao>trajes das freguesias de
    Viana do Castelo</Descricao></Traje> e carregando
    <adornado_com/><Ouro quantidade='quilos'><Descricao>
    quilos de ouro ao peito</Descricao></Ouro>, com varias
    <usa/><Adereco qual='pecas_seculares_de_familia'>pecas
    seculares de familia</Adereco>, estas mordomas
    representam, de forma espontanea, a maior montra de ouro
    ao ar livre do mundo.
    <desfila/><Mordoma> Trajam-se a rigor e com minucia
    na arte de bem vestir e ourar, cumprindo todos os detalhes
    herdados de geracoes anteriores, mantendo a tradicao para
    desfilar com orgulho.</Mordoma>
  </Desfile_Mordomia>
  <Desfile_Mordomia>O Desfile da Mordomia e a historia viva
    dos trajes, desde as lavradeiras as varinas, ate aos mais
    requintados trajes das mordomas, usados pelas raparigas e
    senhoras de Viana do Castelo. Muitas das pecas de ouro sao
    unicas e so sao mostradas em publico neste dia, redobrando a
    chieira das mulheres que as usam. Uma tradicao cada vez mais
    enraizada entre as mulheres de Viana do Castelo, que junta
    varias geracoes, no mesmo quadro, para mostrarem o esplendor
    e a riqueza dos trajes e do ouro de Viana do Castelo.
  </Desfile_Mordomia>
</Romaria_Sra_Agonia>
```

Once XML elements are defined in an appropriated dialect created for RNS Agonia, it is possible to start the document annotation process using the referred dialect (as illustrated in Listing 3). This process enables the Information Extractor to perform its work based on the predefined set of tags.

## 6 Virtual Museum, Architecture

The architecture describes the fundamental organization of a system and its components. In our case, these components are the (human or software) modules responsible for all processing, from documents treatment to the presentation of the information extracted in a web interface.



■ **Figure 1** System Architecture.

Figure 1 allows for an easier understanding of the different components and how they interact in order to present the final product to the user. To make the image clearer, a color scheme was used to distinguish them.

The leftmost, non-colored, part refers to an external process that has to be done prior to the automatic processing. The source documents are simple articles, newspapers news, flyers, photos, etc, that were gathered in order to extract valuable information. As already explained, the ontology is used for the annotation process, specifically in tags creation, since that process involves identifying instances of important concepts and their respective relations. At this first stage the documents are manually annotated with the mentioned XML tags.

After this process, the information processing represented by the blue component begins. In this module, the Data Extractor will traverse the XML tree (the internal representation of the annotated documents) and get sets of triples: subject, predicate and object that describe in an ontological style the information obtained. In our case, these components refer respectively to the concepts of the pilgrimage and relationships with other concepts and their attributes. After extracting them they will be stored in a triple-storage database. That database will be processed to answer the requests that will have origin on the web application.

As Figure 1 shows, the green component refers to the Web application, responsible for interacting with the user. It receives the requests and converts them into appropriate queries that are send to the triple database and collects the responses to display to the users. The final user, in the orange component, has the chance to conceptually navigate over the information displayed.

## 7 Virtual Museum, Development

The development is divided in four phases as described in Figure 2. The first three will be described in this section and the Data presentation is discussed in Section 8.





■ **Figure 2** Stages of data processing evolution.

As mentioned before, from the designed ontology, a set of XML tags were created and adapted to manually annotate documents. Once those rules for standardizing annotations were in place, the next step was data extraction, taking them into account.

## 7.1 Data Extraction

To extract information from the annotated documents, using Python and the *xml.etree* library, the following data structures were created:

1. Two classes, one for each entity: concepts and relations. Within each of that class, were defined generic attributes to store the relevant information about that entity. For “concept” class those attributes were id, properties (attributes), description, explanation and relations.

Once that classes were defined, an array of objects for each class was created to store the data. Each object in the array represents a specific instance of the entity type, and contains the values of its corresponding attributes.

Storing the data in this way, it is easily accessible and manageable using Python object-oriented programming capabilities. For example, it is possible to search for all instances of a particular entity type, or filter the data based on certain attributes and easily add information while documents are processed.

Overall, using this approach allows for efficiently and effectively store the extracted data from the documents, while also providing flexibility and ease of use for further analysis and processing.

2. One stack to keep track of the hierarchical relation and to inheritance among concepts. Once the closing tag of the concept in question is found, this one is popped out of that stack and the “parent” of that concept is maintained to keep analysing and storing further relations.

Given as input an annotated document like the example shown in Listing 3, the data structure produced by the extractor is:

```
Romaria_Sra_Agonia
  Relacoes: [('ocorre', 'Desfile_Mordomia')]
  Atributos: {}

Desfile_Mordomia
  Relações: [('desfila', 'Mordoma'), ('destaca', 'Mordoma'), ('enverga', 'Traje')]
  Atributos: {'descricao': [Desfile da Mordomia e a historia ...
                           riqueza dos trajes e do ouro de Viana do Castelo.]}

Mordoma
  Relacoes: [('usa', 'Traje'), ('usa', 'Adereco')]
  Atributos: {'sinonimo': ['mulher vianense'], nr_mordomas: ['mais de
                           meio milhar'], 'descricao': [Trajam-se a rigor e com minucia na
                           arte de bem vestir e ourar ... desfilar com orgulho.]}
```

```

Traje
  Relacoes: [('adornado_com', 'Ouro')]
  Atributos: {'descricao': ['trajes das freguesias de Viana do Castelo']}
Ouro
  Relacoes: []
  Atributos: {'quantidade': ['quilos'], 'descricao': ['quilos de ouro ao
peito']}

```

## 7.2 Data Serialization

The main goal in terms of data extraction and storage was to use an appropriate triple-storage approach. In the context of triple-storage, RDF stands for *Resource Description Framework*. It is a standard model for data interchange on the web, using a graph-based data model to represent information about resources in the form of subject-predicate-object triples. Each triple in RDF represents a statement about a resource, where the subject is the resource being described, the predicate is the property or attribute of the resource being described, and the object is the value of the property or attribute. The triples can be stored in a variety of ways, such as in a triple-store database or as a text file in one of the RDF serialization formats like Turtle or RDF/XML, providing a flexible and inter-operable way to represent and exchange data on the web. To this end, among the many formats that are available, Turtle was chosen to serialize the data for its representation through RDF graphs [3]. In order to perform this serialization, two main steps must be performed:

1. Convert the ontology OntoRomaria into Turtle format.
2. Convert also the internal representation of the extracted data (described in Subsection 7.1) into Turtle.

### Convert OntoRomaria to Turtle

Since the first step is to convert OntoRomaria to Turtle, that process was made by using a tool accessible online, OntoCnE-Nave<sup>3</sup> that provides an environment to visualize and explore ontologies [6]. The referred navigation tool has an extra, but actually useful, functionality that converts ontologies written in OntoDL to Turtle format (required, as input format, by that ontology processing environment). OntoRomaria was written in OntoDL+ and that could be a hindrance to the use of the tool. However, as OntoDL+ is a superset of OntoDL, that conversion was straightforward. Given OntoRomaria as input to OntoCnE-Nave, the translator produced a Turtle version proper for its visualization as a graph.

### Convert Extracted Data to Turtle

After extracting data from the sources and storing it in internal structures, it was necessary to clean and prepare the extracted data and convert it to Turtle. For that purpose, it was necessary to execute the following tasks:

1. A prefix for the ontology has to be established to make IRIs shorter, allow the reuse, clarity and compatibility between ontologies, tools and software based on RDF. In this particular situation, the predefined prefix is <http://example.org/ontoromaria#>. It is important to note that the mentioned domain<sup>4</sup> is public and can be used in this context.

<sup>3</sup> Available at <https://computationalthinking4all.ep1.di.uminho.pt/OntoCnENave>.

<sup>4</sup> Available at <http://example.org/>.

2. Along side with the previous prefix, it is important to establish other three that will help to specify types and labels:

```
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
```

3. Every concept has an IRI appropriate to the scope of the pilgrimage and two associated relationships: one defining its label according to the RDFS prefix and the other its type, according to the RDF prefix. If it has additional relations with other concepts they will be presented linking directly to those concepts IRIs, as it's possible to see:

```
<http://example.org/ontoromaria#Romaria_Sra_Agonia> rdf:type
ontoromaria:Romaria_Sra_Agonia;
  rdfs:label "Romaria Sra Agonia";
  ontoromaria:ocorre <http://example.org/ontoromaria#Procissao_ao_Mar>;
  ontoromaria:ocorre <http://example.org/ontoromaria#Desfile_Mordomia>;
  ontoromaria:descricao "Ao meio-dia,o som de ... Revista de Gigantones
e Cabeçudos";
```

4. Specific cases such as dates need to be treated properly so, when the converter detects that the object of a triple is a date, and follows the usual format (yyyy-mm-dd), adds that type information type to the triple, using the XSD prefix.

Following all the guidelines above, the data stored in the internal data structures can be converted to Turtle. Therefore, it is ready to be moved to the storage phase, as described in Subsection 7.3.

Listing 4 exhibits the Turtle triples generated for the input document shown in Listing 3 and converted to the internal representation illustrated in Subsection 7.1.

■ **Listing 4** Example of an annotated document converted to Turtle.

```
@prefix ontoromaria: <http://example.org/ontoromaria#> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .

<http://example.org/ontoromaria#Romaria_Sra_Agonia> rdf:type
ontoromaria:Romaria_Sra_Agonia;
  rdfs:label "Romaria_Sra_Agonia";
  ontoromaria:descricao "Ao_meio-dia...e_Cabecudos";
  ontoromaria:ocorre
  <http://example.org/ontoromaria#Desfile_Mordomia>;

<http://example.org/ontoromaria#Desfile_Mordomia> rdf:type
ontoromaria:Evento;
  rdfs:label "Desfile_Mordomia";
  ontoromaria:desfila
  <http://example.org/ontoromaria#Mordoma>;
  ontoromaria:usa
  <http://example.org/ontoromaria#Traje>.

<http://example.org/ontoromaria#Mordoma> rdf:type
ontoromaria:Mordoma;
  rdfs:label "Mordoma";
  ontoromaria:sinonimo "mulher_vianense";
  ontoromaria:nr_mordomas "mais_de_meio_milhar";
```

```

ontoromaria:adereco "pecas_seculares_de_familia";
ontoromaria:descricao "Trajam-se...orgulho.".

<http://example.org/ontoromaria#Traje> rdf:type
ontoromaria:Traje;
  rdfs:label "Traje";
  ontoromaria:descricao "trajes...Viana_do_Castelo".

```

### 7.3 Data Storage

After extracting the data available in the annotated sources and serializing it in Turtle, it was necessary to study the graph databases (which allow triple storage) available systems. For this, three options were considered:

- GraphDB<sup>5</sup>
- Neo4j<sup>6</sup>
- StarDog<sup>7</sup>

GraphDB, Neo4j and Stardog are all popular graph databases with similar functionality but also have some differences. Some key factors to consider when comparing these three graph databases[5] are listed below:

1. **Performance and Scalability:** GraphDB and Neo4j are known for their high performance and scalability, while Stardog has focused more on providing a rich set of semantic reasoning features.
2. **Query Language Support:** GraphDB supports the SPARQL query language, while Neo4j uses the Cypher query language, and Stardog supports both SPARQL and a subset of SQL. This factor is exclusively dependent on the programmer's choice depending on his familiarity with the languages.
3. **Reasoning and Inference:** Stardog is known for its strong support of semantic reasoning and inference, while GraphDB and Neo4j have more limited support for these features. If reasoning and inference are a critical requirement, Stardog may be the better choice.
4. **Integration with Other Tools and Frameworks:** Integration with Other Tools and Frameworks: All three graph databases have strong integration with various tools and frameworks, including popular programming languages, data integration tools, and machine learning frameworks.

Given this general insight, summarized as a table in Figure 3, **Stardog was considered the best option** since it focus on a rich set of semantic reasoning features and supports SPARQL as query language. This is a plus since it allows to infer relations that are not explicitly declared in the database.

## 8 Virtual Museum of Romaria de N<sup>a</sup> Sr<sup>a</sup> d'Agonia

After gathering and processing all the information, the goal is to provide a conceptual navigation regarding the Romaria domain. Thus, Figures 4, 5, 6 and 7 illustrate a navigation flow through the information exposed allowing to explore concepts and their relationships.

<sup>5</sup> For details, see <https://www.ontotext.com/products/graphdb/>

<sup>6</sup> For details, see <https://neo4j.com/>

<sup>7</sup> For details, see <https://www.stardog.com/>

	Neo4j	GraphDB	Stardog
Performance and Scalability	High Performance	High Performance	Rich set of semantic reasoning features
Query Language Support	Cypher	SPARQL	SPARQL + Subset of SQL
Reasoning and Inference	Limited support	Limited support	Strong support of semantic reasoning and inference
Integration	Strong integration	Strong integration	Strong integration

■ **Figure 3** Summary of DB Features.

As an example, consider the case where the user wants to explore *Desfile da Mordomia*. After landing in the museum's opening page (Figure 4), the user can choose to start the navigation in order to explore the knowledge about RNS Agonia. Clicking on the main button, he will access the museum's Homepage (Figure 5) that displays some general information and the buttons that allow for the navigation to the subpages where the main concepts are detailed. Choosing the button Mordomia, the user is redirected to a page (Figure 6) that allows to explore that concept. In its explanation there are several other concepts that can also be visited. By clicking on one of these concepts, *Mordomas*, for example, another page is displayed with the information necessary for the user to understand the domain (Figure 7).

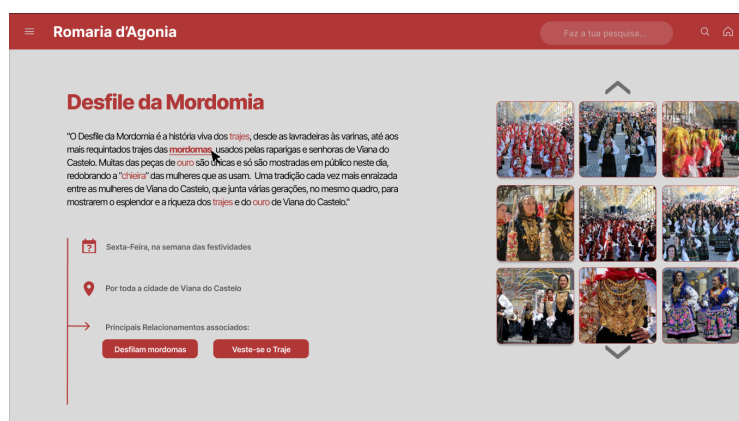


■ **Figure 4** Landing Page on Virtual Museum for RNS Agonia.

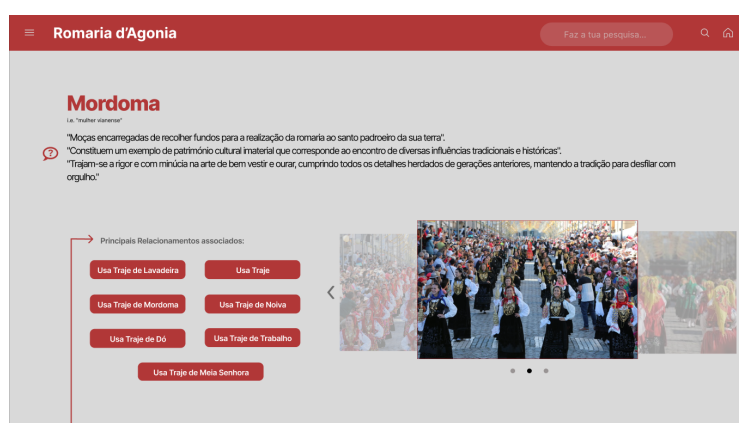
## 16:14 Romaria De N<sup>a</sup> Sr<sup>a</sup> D'Agonia: Building a Digital Repository and a Virtual Museum



■ **Figure 5** Home Page on Virtual Museum for RNS Agonia, showing the main concepts.



■ **Figure 6** *Desfile da Mordomia* Page on Virtual Museum for RNS Agonia.



■ **Figure 7** *Mordomas* Page on Virtual Museum for RNS Agonia.

## 9 Conclusion

The main goal of the project discussed in the paper is the creation of a documents repository and the development of a Virtual Museum for the ancestral Portuguese pilgrimage *Romaria de N<sup>a</sup> Sr<sup>a</sup> d'Agonia* (denoted as RNS Agonia, for short).

To achieve the established objectives it was necessary to deepen the knowledge related to the *Romaria* domain (as summarized in Section 2), and study Virtual Museum and Ontology concepts (discussed in Section 3).

RNS Agonia was described in Section 2 presenting some elements of its history, namely its origin and evolution, its festive events, main symbols and remarkable characteristics that distinguish it from all others.

After defining the problem domain using an ontology designed specifically (OntoRomaria), it was discussed how to create a XML dialect to markup documents describing the *Romaria*. To cope with the annotated texts, a data extraction process was implemented to build up a digital repository supported by a triple data store.

It was also explained how the storage, structured according to OntoRomaria, is queried to create the envisage virtual museum.

The hypothesis that we intend to prove with that project is that the web space built from the data related according to the ontology will provide a conceptual navigation over the pilgrimage information that offers to the virtual museum Visitor a chance to easily create knowledge about the *Romaria*.

The engine built to automatize the information extraction and archival, and the Web site implementation, the ontology created and the markup process are the contributions of the work here exposed.

To the best of our knowledge, there is no other system built according to the ontological approach here proposed. We are aware that the process still includes some manual tasks, but in the future we plan to automatize all the phases.

The first prototype of the referred engine and a small repository to be used as a test scenario have been developed. More tests and optimizations must be done in the near future, as well as more documents shall be annotated to increase the knowledge repository.

Of course experiments have to be designed and conducted to assess the users satisfaction and the platform effectiveness.

After that, we intend to find partners and start cooperation with local public or private institutions that can support the project in order to maintain that virtual museum to promote *Romaria de N<sup>a</sup> Sr<sup>a</sup> d'Agonia*.

---

## References

- 1 Dean Allemang and Jim Hendler. *Semantic Web for the Working Ontologist*. Morgan Kaufmann, 2011.
- 2 Robert Arp, Barry Smith, and Andrew D. Spear. *Building Ontologies with Basic Formal Ontology*. Massachusetts Institute of Technology, 2015.
- 3 Richard Cyganiak, David Wood, and Markus Lanthaler. Rdf 1.1 concepts and abstract syntax. <https://www.w3.org/TR/rdf-concepts/>. accessed April 10, 2022.
- 4 Alexandre Dias. Ontodl+. [https://ep1.di.uminho.pt/~gepl/GEPL\\_DS/OntoDL/index.html](https://ep1.di.uminho.pt/~gepl/GEPL_DS/OntoDL/index.html), 2020. accessed November 8, 2022.
- 5 Bob DuCharme. *Learning SPARQL: Querying and Updating with SPARQL 1.1*. O'Reilly Media, 2013.
- 6 Daniela Fonte, Alda Lopes Gançarski, Daniela da Cruz, and Pedro Rangel Henriques. Conversor de OntoDL para Turtle. Project report, Dep. of Informatics, University of Minho, June 2022.
- 7 Werner Schweibenz. Virtual museums. *The Development of Virtual Museums, ICOM News Magazine*, 3(3), 2004.
- 8 Werner Schweibenz. The virtual museum: an overview of its origins, concepts, and terminology. *The Museum Review*, 4(1):1–29, 2019.

## 16:16 Romaria De N<sup>a</sup> Sr<sup>a</sup> D'Agonia: Building a Digital Repository and a Virtual Museum

- 9 VisitPortugal. Santuário da senhora da agonia. <https://www.visitportugal.com/pt-pt/content/santuario-da-senhora-da-agonia>, 2020. accessed September, 2022.
- 10 W3Schools. Introduction to xml. [https://www.w3schools.com/xml/xml\\_what\\_is.asp](https://www.w3schools.com/xml/xml_what_is.asp). accessed November 9, 2022.
- 11 Álvaro Campelo. *A Falar de Viana (Volume I, Série 2)*, chapter As festas da Sra. D' Agonia "Virgem do Mar, Mordoma da Terra: um contraste de rostos, uma festa de emoções!". Biblioteca Municipal de Viana do Castelo, 2012.