

Implementing an RDF/OWL Ontology on Henry the III Fine Rolls

Jose Miguel Vieira, Arianna Ciula

Centre for Computing in the Humanities, King's College London, Kay House, 7 Arundel
Street, WC2R 3DX London, England, United Kingdom
{jose.m.vieira, arianna.ciula}@kcl.ac.uk <http://www.kcl.ac.uk/cch/>

Abstract. This paper will describe the creation of an ontology to model information contained in historical documents. It will describe the structure of the RDF/OWL ontology and how the model facilitated the expression and analysis of implicit and hidden information/associations existing in the sources. Some aspects of data processing and delivery based on the ontology, like the generation of indices and the implementation of search mechanisms, will also be described.

Key words: Ontologies, RDF/OWL, authority lists, indices, search, medieval documents, family history, career study, Henry the III.

Introduction

The Henry III Fine Rolls is a three-year collaborative project between King's College London and the National Archives (UK) that aims to represent the complexity of historical documents known as the Fine Rolls. [1]

The fine rolls record offers of money to the king Henry III of England for a wide variety of concessions and favours to individuals and corporate bodies, both municipal and religious. They originate from the period of intensive innovation in record keeping around the turn of the thirteenth century. For the historian of the reign of Henry the III the fine rolls can be of prime importance in the study of political, social, and economic history and of government and administration at a local and national level. [2]

In total there are 64 rolls compiled in Latin by a handful of scribes and containing 730 membranes of parchment, there exists one roll for almost all of the fifty-six years of Henry III's reign from 1216-72; the project covers the first thirty-two years of the reign, down to 1248.

The purpose of the Henry the III Fine Rolls project is to represent and model the information contained in the fine rolls for historical research purposes, in a way that can be widely and easily accessible by everyone with a computer and a web browser, and to create a print edition too.

The digital edition of the fine rolls is being created using XML, according to the TEI¹ guidelines, and includes digitised images of the rolls themselves [3][4]. The outcome will be an electronic version of the rolls, modeled according to the format of the traditional printed calendar – an English summary of records, plus a set of indices. To respond to these requirements the need for the creation of an authority list, that would store the information about persons, places and subjects, arose.

Each roll is marked-up as an XML document as to include information about:

- The physical structure of the roll – e.g. membranes, *marginalia*;
- The structure of the calendar and its content – date and place of the record, body of each entry, witness list;
- The semantic content of the roll – names of individuals, places and subjects, together with their identification.

While this information is enough for the creation and presentation of the basic edition it was necessary to carry out deeper analysis in order to:

- Associate an occurrence of a person, a subject, or a place in the text to its correspondent instance of a logical authority of a Person, Subject or Place;
- Express complex relationships between two or more authorities.

In order to meet these requirements an OWL ontology was developed.

RDF/OWL

At the moment there are several standards available that can be used for the creation of authority lists. After conducting a comparison between MADS², Topic Maps³ and RDF⁴/OWL⁵, we decided to choose RDF/OWL for the following reasons:

- It is a W3C standard for the Semantic Web, it is widely recognized and supported;
- The number of existing tools is greater for RDF/OWL than for the other technologies we analyzed;
- It can be expressed as XML, facilitating the process of data delivery: this makes it easy to create indices of people, places and subjects using XSLT.
- It allows for the expression of relationships among instances (of a person, a place or a subject, for instance) mentioned in the fine rolls source materials.

Henry the III Fine Rolls Ontology

The FRH3 ontology⁶ was created with the purpose of serving as an authority list for people, places and subjects described in the fine rolls. The need to express complex

¹ <http://www.tei-c.org/>

² Metadata Authority Description Schema – <http://www.loc.gov/standards/mads/>

³ <http://www.topicmaps.org/>

⁴ <http://www.w3.org/RDF/>

⁵ <http://www.w3.org/2004/OWL/>

⁶ Initially developed in collaboration with Gautier Poupeau (gpoupeau@enc.sorbonne.fr) from École Nationale des Chartes (<http://www.enc.sorbonne.fr/>)

associations to study things like the careers of people, genealogical information, associations of people to places, and other connections could be fulfilled thanks to the characteristics of RDF/OWL.

Throughout the definition of our ontology, we created inverse, symmetric and transitive properties; created intersection classes; added restriction to the allowed values for certain classes. All of this was carried out with the objective of maximizing the quantity and richness of information that could be extracted from the fine rolls.

The use of transitive properties was especially useful to represent the geopolitical organization of places. For instance by using the property `cidoc:P88B.forms_part_of`, we can infer that if Village A forms part of Hundred B and Hundred B forms part of County C then Village A forms part of County C. Still within the structure of places, various intersection classes were created to define the semantics of places that are common to two classes.

The following sections describe the most relevant classes in our ontology.

Structure

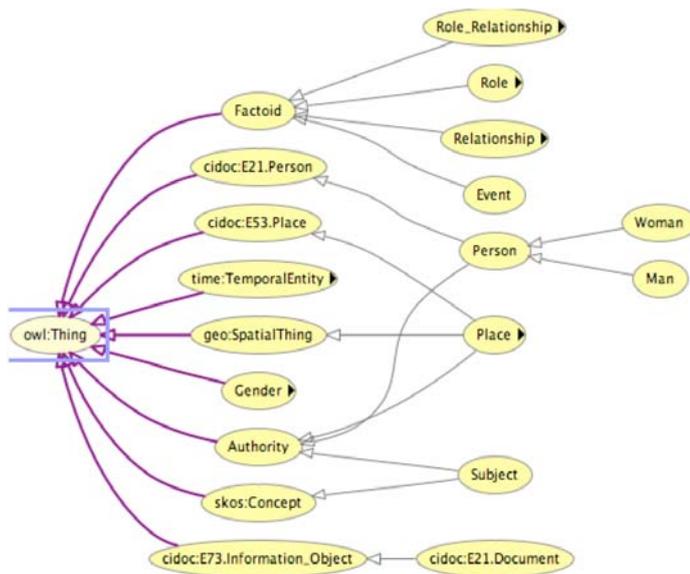


Fig. 1. Top-level structure of the ontology

The specific and most relevant classes and subclasses for the FRH3 project are:

- Authority – defines a relevant authority in the fine rolls.
- Factoid⁷ – defines something about an authority as described in the fine rolls.
- Gender – defines the gender of a person.

⁷ The Factoid concept was borrowed from the PASE Project (<http://www.pase.ac.uk/>)

Namespace

The namespace for the FRH3 ontology is <http://www.cch.kcl.ac.uk/xmlns/frh3#>, for both the ontology definition and the individuals instances. The namespace prefix used is **frh3**.

Other Ontologies

OWL enables the reuse and extension of existing ontologies and whenever possible we tried to use some of the predicates already existing in other ontology vocabularies.

The FRH3 ontology uses predicates from the following existing ontologies:

- CIDOC-CRM⁸
- Dublin Core⁹ – provides an element set for describing networked resources;
- Geo¹⁰ – provides a standard way to represent information about spatially located things. Two predicates were used, `geo:lat` and `geo:long`, to define the latitude and longitude respectively, of a place;
- Simple Knowledge Organization System (SKOS)¹¹.
- Time¹² – The W3C time ontology was used to define time instants or time intervals of events that are asserted in the fine rolls.

CIDOC-CRM

Provides definition and a formal structure for describing concepts and relationships used in cultural heritage documentation.

We decided to use some of the classes from CIDOC-CRM because of its ability to cover contextual information, in our case historical information, making it useful for the description of the events in the fine rolls.

We used the following classe from CIDOC-CRM:

- `cidoc:E53.Place` – Is used to define extents in space, in our case places that are present in the fine rolls. We use some object properties that help define that a Place is part of another Place, or to define that a Place was a stage for a certain event.
- `cidoc:E21.Person` – It is used to include real persons that are stated in the fine rolls. The object properties that we used from this class allow us to state that a Person participated in a certain event, or to define that a Person is identified by a Role or a Relationship.
- `cidoc:E73.Information_Object`
 - `cidoc:E21.Document` – In our ontology every authority, Person, Place, Subject is referred to by a document. These two classes are used to serve this purpose. We also use some object properties that help define that a document is composed of several other documents, or to say that a document refers to a

⁸ <http://cidoc.ics.forth.gr/>

⁹ <http://dublincore.org/>

¹⁰ <http://www.w3.org/2003/01/geo/>

¹¹ <http://www.w3.org/2004/02/skos/>

¹² <http://www.w3.org/TR/owl-time/>

specific authority. We also use some predicates from Dublin Core that enable us to set a title and an identifier for a document.

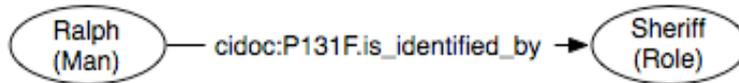


Fig. 2. Example of usage of cidoc:P131F.is_identified_by property.

SKOS

Provides a standard way to represent knowledge organisation systems. This was especially useful for the Subjects hierarchy. Some of the object properties we used allow us to define that a Subject has a more general meaning than another Subject; or that a Subject has a more specific meaning than another Subject; or that a Subject is included in another Subject.

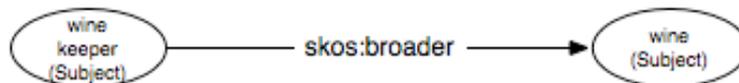


Fig. 3. Example of usage of skos:broader property.

FRH3 Classes

The following sections will describe the main FRH3 classes.

Authority Class

The authority class is used to define any useful and relevant authority in the fine rolls. It has three subclasses, Person, Place and Subject that are used to define the individual in the fine rolls.

It defines several objects properties that allow the creation of relationships between two Authority individuals, or between an Authority and a cidoc:E73.Information_Object.

Factoid Class

The Factoid class defines something about an Authority (Person, Place or Subjects), as asserted in the fine rolls. It has three subclasses, Role, Relationship and Role_Relationship.

This class only has object properties to help relate a Factoid to another Factoid or Authority; to relate a Factoid to a cidoc:E73.Information_Object.

Role Class

The Role class is a subclass of the Factoid class. Besides inheriting the properties from its superclass, the Role class defines new properties and adds restrictions to some others.

It represents an occupation/status a Person had as defined in the fine rolls. The Role class contains an extensive list of subclasses, each one representing a role (occupation/status) that is described in the fine rolls (e.g. Abbot, King, Sheriff).

- Some of the inherited properties from the Factoid class are redefined to include additional restrictions; for instance the property `cidoc:P11F.had_participant` for the class Nun only accepts individuals of the class Woman.

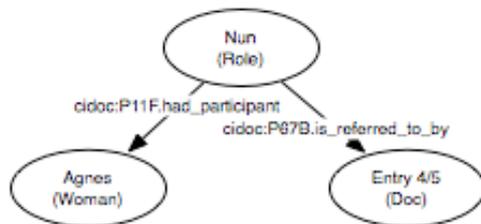


Fig. 4. Example of a Role

Relationship Class

We use a class to represent relationships instead of properties, because we want to have the possibility to define in which entry in the fine rolls the relationship is mentioned.

The Relationship class is a subclass of the Factoid class. It represents any sort of relationship between two persons in the fine rolls. A set of all the possible combinations of kinship, and legal relationships, was created for each relationship. Each relationship has been defined as a subclass of the Relationship class; some of the relationships created are: Husband_Wife, Mother_Daughter, Benefactor_Heir, to name but a few.

The Relationship classes inherit all the properties from the Factoid class and also add some specific properties according to the relationship being defined. For instance, in a Mother_Son class we defined the properties `frh3:has_mother` and `frh3:has_son`. The `frh3:has_mother` only accepts individuals of the class Woman, while the `frh3:has_son` only accepts individuals of the class Man.

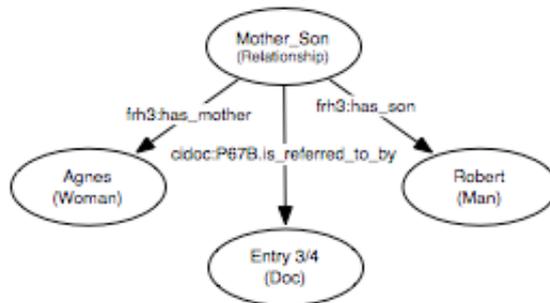


Fig. 5. Example of a Mother_Son relationship.

Role_Relationship class

The Role_Relationship class is a subclass of the Factoid class. It relates a Role or a Person to a Role; a subclass was created for each possible combination in the relationships of a Role or a Person to a Role.

The Role_Relationship classes inherit all the properties from the Factoid class and also add some specific properties according to the role relationship being defined. For instance, in a King_Servant class we defined the properties frh3:has_master and frh3:has_servant. Both these properties accept individuals of classes Role or Person.

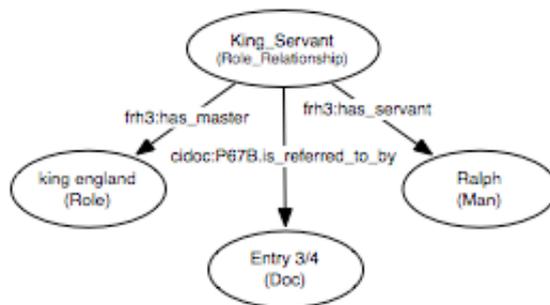


Fig. 6. Example of a King_Servant role relationship

Person class

This class is an intersection between the Authority class and the cidoc:E21.Person class. Besides inheriting the properties from its superclasses it also defines new properties that are specific to the FRH3 domain.

Properties like frh3:toponym_name allow the association of the Person surname to a Place. Other properties allow the association of a Person surname with a

Relationship, for cases where a Person is only known by her patronymic name or by the association to a kin.

The Person class has two subclasses, Man and Woman. Some properties are restricted so that they either belong to Man or Woman. For instance, the property `frh3:is_father` only has as range the class Man.

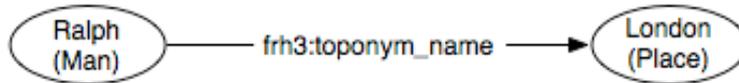


Fig. 7. Example of a Person with a toponymic name

Place class

The Place class is the intersection between the classes Authority, `cidoc:E53.Place` and `geo:SpatialThing`. This class has a great set of subclasses, one for each type of place described in the fine rolls (e.g. County, Country, Abbey).

New properties have been defined in the Place class to permit the definition of uncertainty. For example a place might be identifiable, but not identified with certainty. In this case we can use properties like `frh3:might_form_part_of` and `frh3:might_consist_of`.

Subject class

The Subject class is an intersection between the Authority class and the `skos:Concept` class. It doesn't define any new properties, it only inherits the ones from its superclasses.

The class has been defined in a way that allows the representation of all the subject in the fine rolls both as a flat and as a hierarchical structure through the use of the `skos` properties described above.

Outcomes

One of the major benefits, of using an ontology to define the authority list, is the ability to create a heavily structured and rich index (for persons, places and subjects) with relatively ease. This happens for two main reasons.

First because the ontology can be expressed as XML, we used XSLT to rapidly produce and customize the indices.

Second, because of the characteristics of RDF/OWL (inverse properties, symmetric properties, transitive properties, the expressiveness for relationships, the ability for knowledge discover, etc.) it allows us to infer/use information that initially was only implied in the fine rolls and can now be used to enrich the indices.

The ontology can also help in the creation of the search mechanism, because it makes it possible to process the data logically and infer information that is not explicit in the markup of the fine rolls, allowing users to search for information that initially is not available.

For example, if a Man M is described as the Husband of a Woman W in the fine rolls, without the use of the ontology, we could only search for occurrences of husband. Thanks to the ontology, and the use of inverse properties, we can also say that if M is Husband of W, then W is Wife of M. This will enable the creation of a much richer search mechanism.

Conclusion

This paper described the use of a RDF/OWL ontology to model complex associations and how this was used to produce richer indices and search mechanisms, to help researchers of the period of King Henry III in the interpretation of the source documents.

The approach of extracting data from the XML files and plotting them into an ontological model has shown several advantages, but had at least two limitations.

First of all, the construction of an ontology that is based on XML encoded documents presents a problem of data synchronization between the source documents and the ontology itself. While the integration of information from the XML to the ontology is straight forward and can be done automatically, the other way around may pose some problems in the future. Indeed, while editing the ontology the researchers have a different overview on the material than the one they have while editing the linear text in XML. Therefore they can make new interpretations (e.g. merge individuals or split them up) that can be expressed easily in the ontology, but that will have to be consistently imported into the XML source documents.

Secondly, although we have tried to exploit existent vocabularies and we hope our work could be re-used in other projects, the Fine Rolls has a specific structure developed for a limited context

This ontology has been developed with the intention of being extended both in terms of data and application. Indeed, it will be integrated with new data for the whole life of the project (including the Fine Rolls until 1248) and it will possibly be reused in conjunction with other projects that deal with data from the same period.

References

1. Boot, Peter. "Decoding Emblem Semantics". *Literary and Linguistic Computing* 2006 21:15-27.
2. Ciula, Arianna. "Searching the Fine Rolls: A Demonstration of the Electronic Version" (paper presented to the International Medieval Congress 2006, University of Leeds, July 10-13, 2003).

3. Ciula, Arianna, Spence, Paul, Vieira, José Miguel, Poupeau, Gautier: “Expressing Complex Associations in Medieval Historical Documents: the Henry the III Fine Rolls Project” (paper accepted to the Digital Humanities 2007 conference, Illinois).
4. Dryburgh, Paul: Henry III Fine Rolls Project. Institute of Historical Research, London, 9th of February 2006.
5. Eide, Øyvind. “The Exhibition Problem. A Real Life Example with a Suggested Solution” (paper presented at Digital Humanities 2006, Université Paris-Sorbonne, July 5-9, 2006). Abstract available at <<http://www.allc-ach2006.colloques.paris-sorbonne.fr/DHs.pdf>> (accessed 30 April 2007).
6. Eide, Øyvind and Ore, Christen-Emil. “TEI, CIDOC-CRM and a Possible Interface between the Two” (paper presented at Digital Humanities 2006, Université Paris-Sorbonne, July 5-9, 2006). Abstract available at <<http://www.allc-ach2006.colloques.paris-sorbonne.fr/DHs.pdf>> (accessed 09 March 2007).
7. Poupeau, Gautier. “De l’*index nominum* à l’ontologie. Comment mettre en lumière les réseaux sociaux dans les corpus historiques numériques?” (paper presented at Digital Humanities 2006, Université Paris-Sorbonne, July 5-9, 2006). Abstract available at <<http://www.allc-ach2006.colloques.paris-sorbonne.fr/DHs.pdf>> (accessed 30 April 2007).
8. Spence, Paul. “The Henry III Fine Rolls Project” (paper presented at Digital Humanities 2006, Université Paris-Sorbonne, July 5-9, 2006). Abstract available at <<http://www.allc-ach2006.colloques.paris-sorbonne.fr/DHs.pdf>> (accessed 09 March 2007).