

Expressing complex associations in medieval historical documents: the Henry III Fine Rolls Project

Arianna Ciula, Paul Spence and José Miguel Vieira

Centre for Computing in the Humanities, King's College London,
London, UK

Abstract

This article focuses on the use of technologies traditionally associated with knowledge representation to express complex associations between entities in historical texts that have been marked up in XML, according to the Text Encoding Initiative guidelines. In particular, we describe our exploration of the potential role of an ontology in facilitating the interpretation of implicit and hidden associations in the sources of interest, examining its use, and limits in a digital humanities project in connection with editing tools and delivery issues. We demonstrate our findings based on the Henry III Fine Rolls project, where an ontology—built using the RDF (Resource Description Framework)/OWL (Web Ontology Language) technologies—is being developed to make explicit information about person, place, and subject entities marked up as *instances* in the core texts themselves. For any historian, there is a natural tension between primary sources (as documentary records) and the analysis that produces a context for interpretation. We will argue that the combination of core mark-up (encoded in TEI) and an ontology (in RDF/OWL) provides a powerful model for representing the complexity of this tension and facilitates the necessarily dynamic process of scholarly interpretation.

Correspondence:

Arianna Ciula, Centre for
Computing in the
Humanities, King's College
London, 2nd Floor, 26-29
Drury Lane, London WC2B
5RL, UK.

E-mail:

arianna.ciula@kcl.ac.uk

While the mechanics and rationale for applying text encoding to humanities materials are by now relatively well understood, it is less clear how we might enhance the value of documents marked up in, say TEI,¹ with external frameworks for the scholarly interpretation of entities mentioned within an original text. This article presents and discusses one possible approach, taken on a project to digitize historical sources, and using technologies more traditionally associated with the Semantic Web. For similar approaches in the digital humanities see Boot (2006, 2007), Carracciolo (2006), Eide and Ore (2006), Poupeau (2006).

The Henry III Fine Rolls project,² funded by Arts and Humanities Research Council, is a collaborative venture between King's College London (Department of History and Centre for Computing in the Humanities) and The National Archives of the United Kingdom. It aims to produce both print and digital editions, using as a model the format of the traditional printed *calendar* (an English summary of records, plus a set of indexes), of the parchment rolls (see Fig. 1 for an example of roll membrane) compiled between 1216 and 1248,³ which record mainly (but not only) offers of money made to King Henry III of England in exchange for

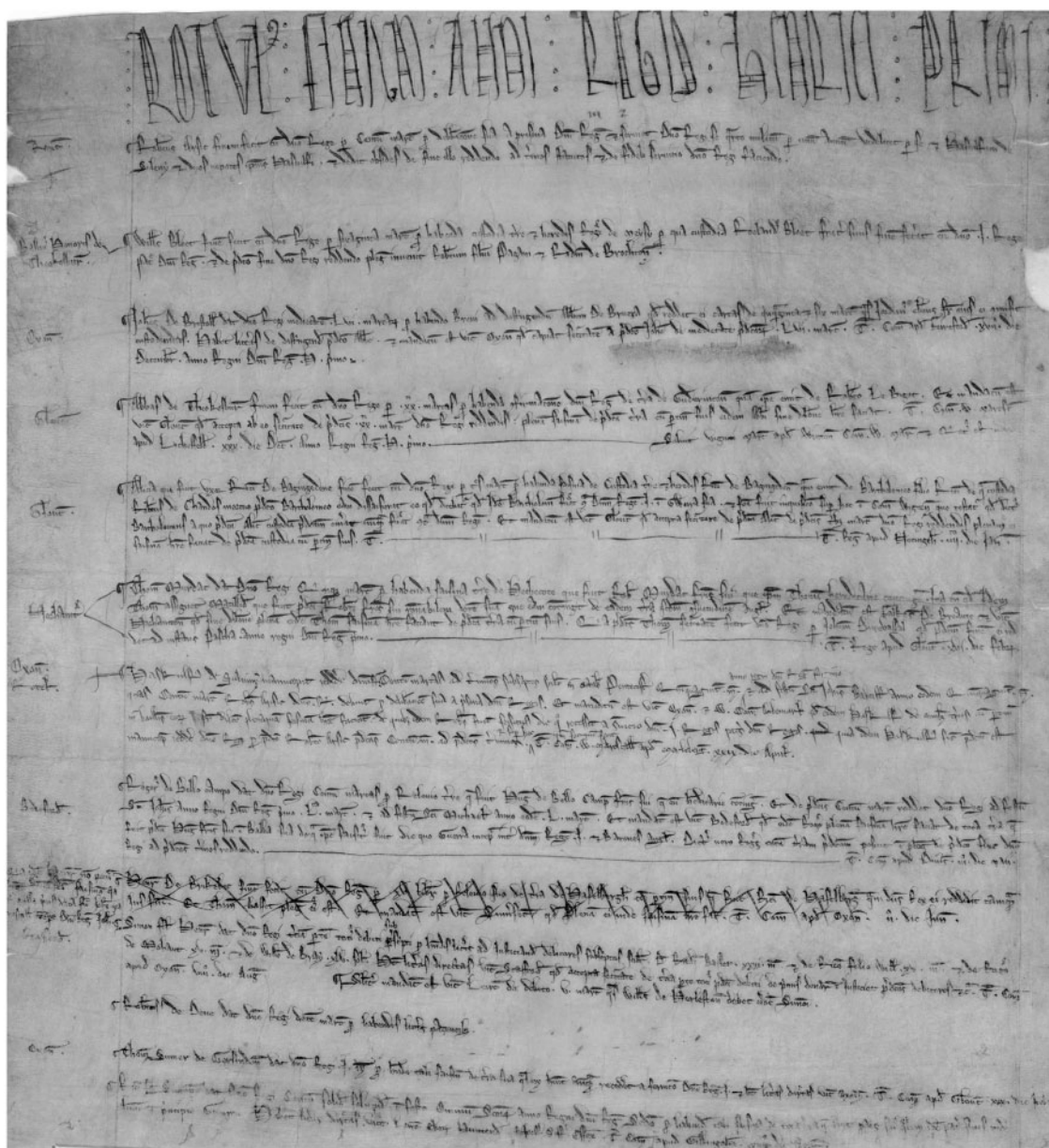


Fig. 1 Sample image of one of the fine rolls membranes (TNA C 60/8, m. 2). © Crown copyright images reproduced by permission of The National Archives, London, England

a wide range of concessions and favours. A total of sixty-four rolls containing around 800 parchment membranes, one for almost all of the fifty-six years of Henry III's reign from 1216–72, survive in The National Archives. Each fine roll was compiled in

Latin by a handful of scribes. Taken as a body of documentary evidence, the rolls are of 'prime importance in the study of political, social, and economic history and of government and administration at a local and national level.' (Dryburgh, 2006)

The first version of the project website was launched on May 2007 and incorporates a wide range of features, which include facsimiles, text translations, image-text connections, indexes of people, places and subjects, and a combined search facility. The edited English summary of the text can therefore be browsed and read as a single unit roll by roll (available in this form for the first time, since the fine rolls have not been published in their entirety before; see Dryburgh, 2006), compared against the images of the rolls membranes or recalled entry by entry through the indexes or the search, which allow unprecedented breadth and ease of access to the material.

1 Sources and Interpretative Layers: TEI Mark-up and Beyond

At the core of the resource are the translated summaries of the fine rolls, which have been encoded using TEI P4 XML,⁴ in order to make explicit:

- the physical structure of the roll—for instance, the fact that it is composed of a series of membranes stitched together;
- the structure of the English calendar, a concise translation of the Latin records, including county and date information concerning the record, body of each entry and witness lists;
- the semantic content of the roll—for instance, names of individuals, names of locations, and key themes mentioned in the text.

TEI offers an excellent framework for representing the structure and semantics of the core texts and the corpus of TEI texts encoded for the Henry III Fine Rolls project is in itself a rich resource. Nevertheless, from a scholarly point of view an extra layer of interpretation is needed in order to fully realize the benefit of encoding the documents. Historical documents such as the fine rolls do not only consist of a textual corpus that documents facts, but can also be considered as sources that give insights into the context of the facts they record. One direct consequence of a socio-historicist approach of this kind towards the text is that the historical

value extends from the text itself to what it reveals about its context of production. The documentary evidence then becomes potentially useful to the scholar as a conceptual framework that can be used to study various aspects of a given society in depth, including its prosopography and genealogy, toponomastics, politics, and economy.

It is no coincidence that the occurrences of names of persons, places and institutions, and their mutual relationships are considered particularly relevant to the study of the fine rolls. They are the means by which historians can get to the heart of the medieval sources as clues to the life and culture of the time. To give some sense of the scale of this, for the period 1216 to 1224—approximately a quarter of the period covered by the first phase of the project—the researchers have identified 3,436 men, 499 women, around 4,070 locations and 1,159 subjects.

Occurrences of people, places, and subjects in the text were marked up to a high level of detail, including such features as the components of a person's name, together with any reference to his or her occupation and relationship(s) with other individuals. However, the process of encoding an entry in the fine rolls privileges the information available in that particular context and lacks the breadth of an overall perspective on the fine rolls as an historical object.

Although, the mark-up is able to capture some aspects of the implicit associations between places, persons, and subjects in the documents—and to a certain extent the new TEI P5 chapter on names and dates further facilitates this type of encoding,⁵ and does so in a systematic manner—this should be complemented with an extra interpretative layer that is both connected to, and independent of, the mark-up itself.

For this reason, the relevant data from the mark-up was extracted and integrated within a set of authorities first to identify the semantic mark-up of persons, subjects, or places as textual references to the correspondent instances of a logical authority, i.e. of a defined entity with characteristic properties; and second to express complex relationships between these authorities and therefore between their instances (see Fig. 2).

After conducting a comparative evaluation of possible standards to create authority list structures—including research into Topic Maps (see Tuohy, 2005) and MADS (Metadata Authority Description Schema)—we opted for RDF/OWL,⁶ a combination of technologies commonly associated with the Semantic Web to represent and exchange knowledge.

The strength of RDF/OWL lies in its ability to model data as a logical representation open to a flexible manipulation of its meanings. In particular, OWL supports a rich expression of the relationships among defined objects or entities, allowing for the capture of knowledge within a given domain of

interest in such a way that complex semantics can be built up from simpler concepts.

In more practical terms, our choice of RDF/OWL was driven by a number of factors, including: our interest in facilitating interconnections with other Digital Humanities projects, particularly those covering similar periods and subject areas, using international standards at the heart of the Semantic Web; the relative maturity of supporting tools; and the fact that RDF/OWL can be expressed in XML format, thus allowing the easy re-purposing of data for web delivery.

2 Creation of the Ontological Framework

We started the process by carrying out a very detailed analysis of what we needed to model, so as to represent the ‘world’⁷ (Thomas, 2004) of the fine rolls in abstract terms with an ontology. Nevertheless, a key principle for our approach was to create a source-based model where the TEI documents themselves were used to populate the ontology.⁸

This source-driven (see Fig. 3) approach meant that—as paradoxical as it may sound—Henry III himself did not in fact formally *exist* in our ontology until his name was actually mentioned in the fifth entry of the first roll. Similarly, the nature of

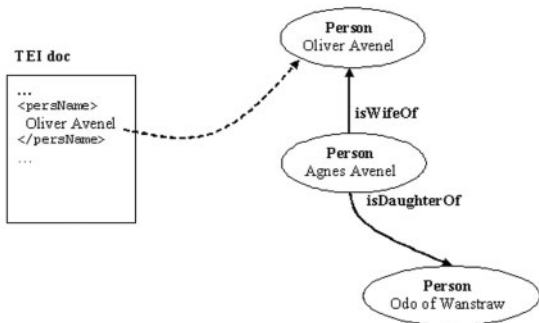


Fig. 2 Example of authority–document and authority–authority associations

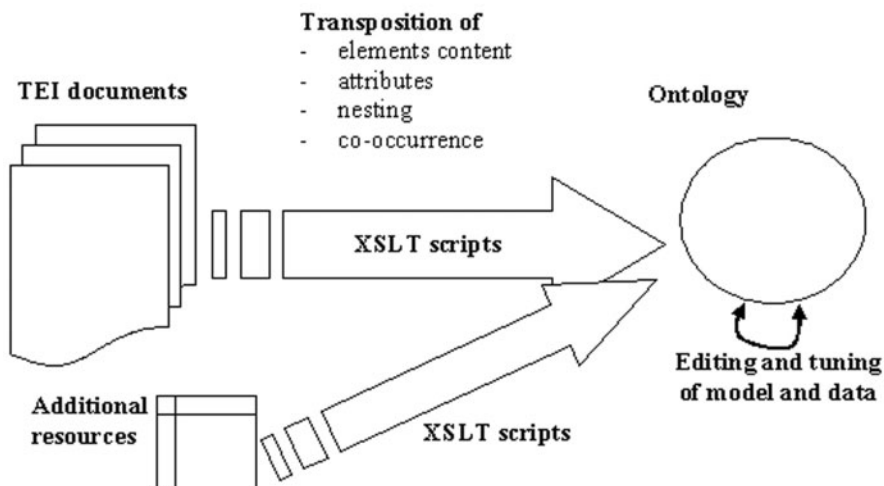


Fig. 3 The process by which the mark-up is used to populate the ontology

his family relationships and the fact that he witnessed some fine rolls writes himself is again information extracted from the XML documents and transposed into the ontology. However, the precise identification of those kinships, and (for instance) the fact that he was King of England from 1216 to 1272, are statements that had to be added directly to his authority record in the ontology by the editors.

It is also interesting to note that we were able to automate the extraction of certain combinations of nested elements from the source XML to feed the ontology so that, for example, a <placeName> (place name element) nested within a <persName> (person name element) was translated into different types of association between the specific person and the place. For example, the nesting of place name inside a person name could imply the occurrence of a toponymic surname (e.g. Henry of London) as well as the occurrence of a location that forms a component of the name of an official (e.g. constable of the Tower of London). This represents only one of the stages in the process of disambiguating the encoding—in this case the hierarchy of nested elements—and translating the co-occurrence of certain combinations of elements into specific types of semantic relationships between concepts.

The XML files not only provided the skeleton of the ontology, but also function as entities within the overall model. The set of entries in the TEI

documents constitute one of the concepts being modelled. Indeed, it could be argued that a faithful historical perspective requires for each reference to a person name, place name, or subject, to be associated with the corresponding textual evidence. In such a model, which aims to include ‘explicit statements of the sources of the assertions exhibited in the text’ (Eide, 2008, p. 36),⁹ references to the relationships between people, as well as references to the different roles they may perform, need also to be anchored to specific *loci* in the documents where they are stated (see Fig. 4).

The RDF/OWL scheme of *classes* and *predicates* was used to define the over-arching knowledge domain which we developed for the core materials of the project. We do not intend to provide a detailed description of this ontological structure here,¹⁰ but it is worth highlighting our main underlying objectives, which were:

- to exploit existing established vocabularies for specific domains such as geography, temporal entities, and cultural heritage representation—for instance, CIDOC CRM, Dublin Core, Geo, Simple Knowledge Organization System, W3C Time—so as to make explicit connections between our work and similar undertakings within the Semantic Web framework;¹¹
- to refine our model and make it reusable for projects that deal specifically with historical sources of a certain period (so, for instance, it

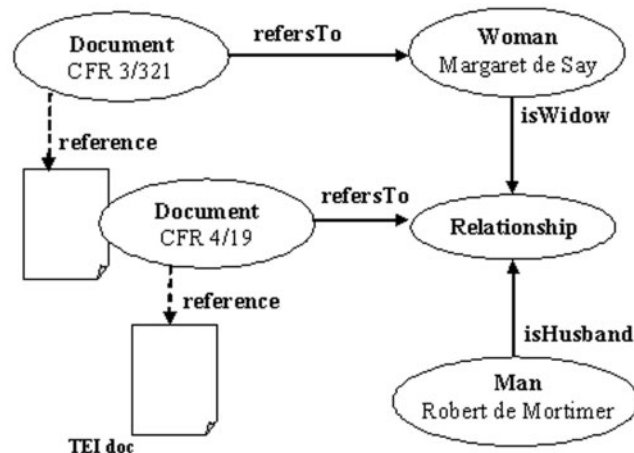


Fig. 4 Explicit associations between the fine rolls source and specific assertions

would be possible for another project to make use of the historical subdivisions of England into so-called Hundreds or Wapentakes,¹² as defined in our model).

In order to achieve the first objective, existing predicates were used as extensively as possible. Amongst the numerous predicates we have borrowed from extant ontology vocabularies, we would particularly like to call attention to the CIDOC Conceptual Reference Model. This vocabulary for knowledge representation provides definitions and a formal structure to describe concepts and relationships relevant to the cultural heritage sector. To give one example, we have used the class *cidoc:E53.Place* to define places in conjunction with some object properties that express the geopolitical inclusion of a Place within another Place, or the fact that a Place was the stage for a certain event (Ore, forthcoming). In our model another class, *cidoc:E21.Person*, has been used to describe supposedly real persons that were mentioned in the fine rolls. In turn, some of the CIDOC object properties that we used allowed us to model the statement that a Person participated in a certain event, or to assert the fact that a Person is identified by a Role or a Relationship.

In addition to making large use of already established vocabularies, we have also created *ad hoc* concepts and project-defined classes and properties where required: for instance, the property *frh3:toponym_name_main*, which allows us to express a person's surname in terms of a toponym (see Fig. 5), or, to give another example, a series of properties to facilitate expressions of uncertainty.

The main classes of the fine rolls ontology (see Fig. 6) are defined as follows:

1. Documents: in our case, these constitute fine rolls entries 'that make propositions about reality' (Crofts et al., 2007), which are subsumed under the class *cidoc:E31.Document*. It is important to note here that we also use some of the Dublin Core Meta Data Terms to assign a title and an identifier to each entry.
2. Any relevant authority belongs to the class *Authority*, which is composed of the three

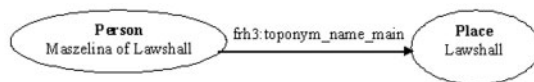


Fig. 5 Example of the property *frh3:toponym_name_main* with the Henry III Fine Rolls project namespace

subclasses of *Person*, *Place*, and *Subject*. The object properties of *Authority* are defined primarily to express relationships between two *Authority* individuals, or between an *Authority* and the documents.

3. 'Something' about an authority as described in the documents is modelled using the class *Factoid*, a concept borrowed from the prosopographical projects developed at the Centre for Computing in the Humanities (Bradley and Short, 2005). It consists of three subclasses: *Role*, *Relationship*, and a hybrid concept between these two that we have called *Role_Relationship* (e.g. the tailor of the King, or the steward of the archbishop). The object properties of *Factoid* relate a *Factoid* to other *Factoids*, to *Authority* individuals and to the documents.

In turn, the subclasses mentioned above have further children. The *Role* class, to name one, itself contains an extensive list of subclasses, each one representing a role (or to be more precise an occupation/status—e.g. *Abbot*, *King*, *Sheriff*). Some of the properties that the class *Role* inherits from its super-class *Factoid* are refined to impose further restrictions; for instance the property *cidoc:P11F.had_participant* for the subclass of *Role* called *Nun* can only apply to individuals belonging to the *Woman* class.

Evidently, the value and granularity of the ontology depends on various factors, which include the richness of the source files themselves (e.g. how well do the fine rolls express the reality of the careers of officials in the thirteenth century?), the specific focus of the researchers (e.g. on the types of fines rather than on the amount of money they keep record of), and the definability of the knowledge domain itself (as evident in the contrast between the identification of individuals and the much more blurred identification of subjects).

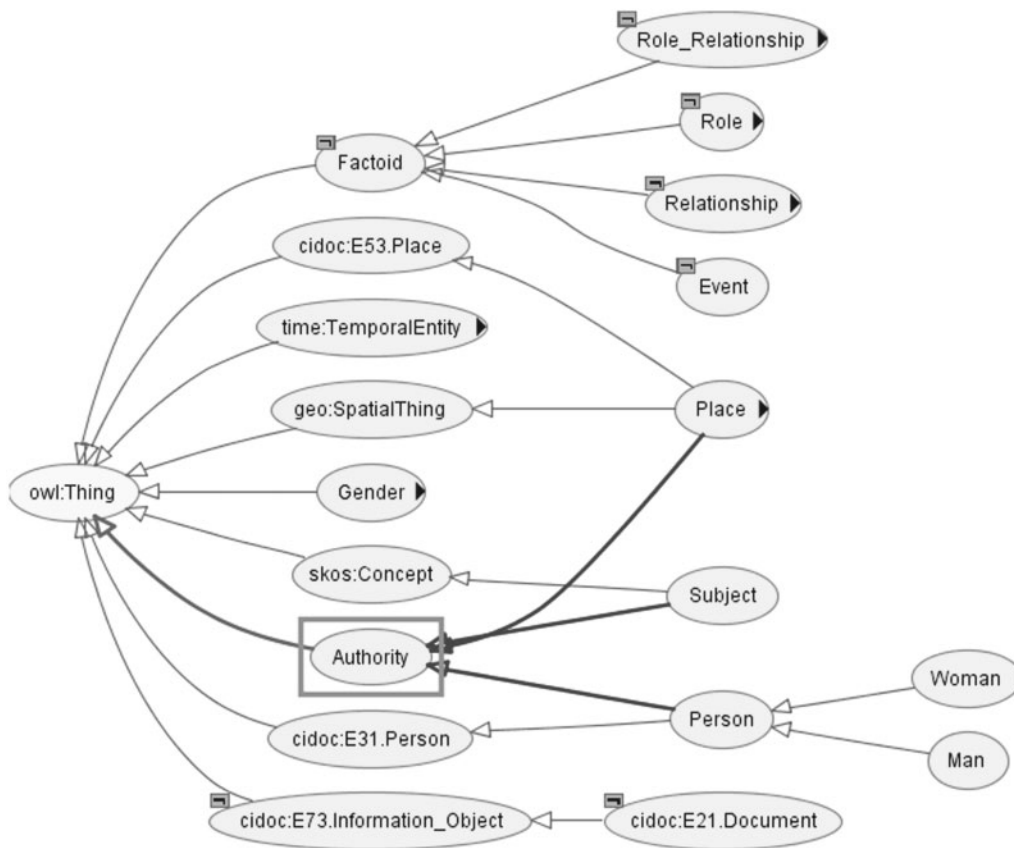


Fig. 6 Structure of the ontology: only the classes closer to the root are shown here

3 Relationship between Mark-up and Ontology

As we have seen, the authorities and their network of associations within the ontology are not only complementary to the mark-up, but also represent, in some cases, alternatives, or corrections. To give a concrete example, references to relevant themes, identified as such by the editors, have been marked up following a flat encoding system, in which each subject is considered to be independent from other themes and is encoded with a *key* attribute in the mark-up linking it to the ontology.

In the ontology, these references are represented as instances of the authority called *Subject*. They are therefore associated with the entries in the documents which contain the references, but, at the

same time, they have been mutually connected following a hierarchical and associative structure, identified by the project editors, of broader and narrower thematic clusters (see Fig. 7).

The current default index of subjects on the website (which also appears in the print volume) uses this structure of connected topics.

To give another example of a situation where the ontology complements the mark-up, there are cases where the scribes wrote something incorrectly or imprecisely. The edition of the fine rolls aims to keep these so-called *lapsus calami*, or errors, highlighting their occurrence with ‘sic.’, as is common editorial practice.¹³ However, the ontology corrects what in the perspective of the historian are imprecise or incomplete assertions and makes alternative statements.

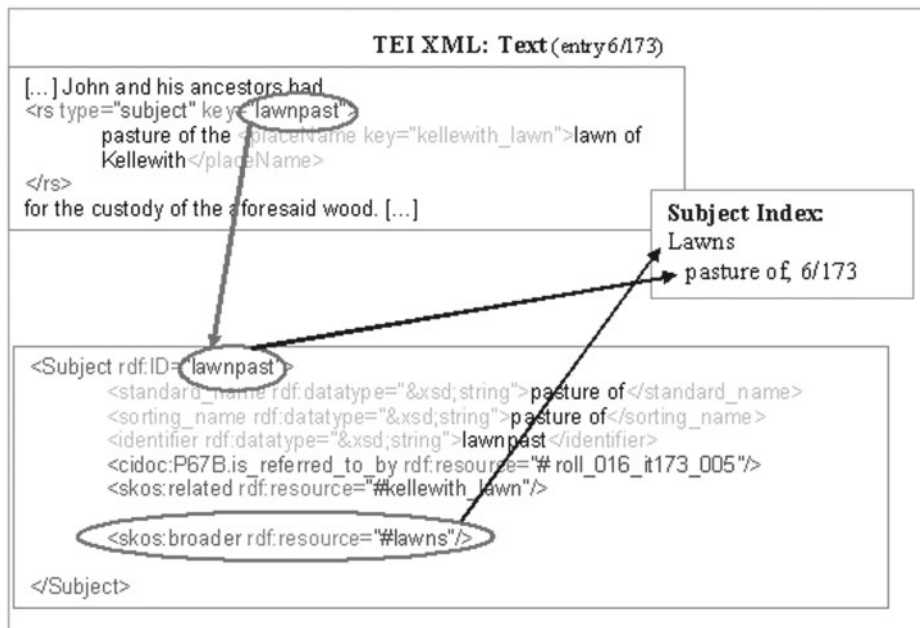


Fig. 7 Connection between the source text (where the reference to a subject is encoded), the ontology (where subjects are organized in thematic clusters), and the output (where the index of subjects is produced)

Take one example where the scribe mistakenly wrote that ‘Alice is sister of Alice’, while in fact it is quite clear from the context that the text should state ‘Alice is sister of Sabina’. So while in the mark-up we only record the assertion that a kinship is mentioned, in the ontology that relationship is specified as being a sisterhood between two differently identified individuals: *alice23* and *sabina02*. Moreover, the ontology also records the fact that Alice is identified by that relationship: in other words, her existence is only manifested by her relationship to Sabina, about whom we have more information (see Figs 8 and 9).

If scholars are to take full advantage of the potential that RDF/OWL provides for formalizing complex associations, it is crucial that the tool kit they use facilitates the editing process by enabling them to perform sophisticated scholarly modelling while concealing some of the technical complexity.

First of all, our own set of editorial checking facilities—built using the XSLT transformation language before importing the data into the ontology—has supported the editors in the transition from the

editing of the text through to its analysis and processing into indexes. In more concrete terms, this intermediate transition from the XML encoding to the editing of the network of authorities has allowed the systematic checking and rationalization of the mark-up especially in the case of identifications.

Second, in building the ontology and managing the data we have used a java-based application called Protégé-OWL,¹⁵ a free, open source ontology editor, and knowledge-base framework. Using its suite of tools we first constructed various prototypes of the ontological model, then tested them with some sample data and finally customized the application to edit the data imported from the XML files. Not surprisingly, the structure of the ontology is subject to continuous refinement and Protégé has responded well to this need.

Once set up, the editing tools have mainly been used by our historian colleagues, who are now able to perform relatively complex operations such as creating subclasses, adding, merging, and splitting instances, as well as being able to perform direct queries on the ontology.

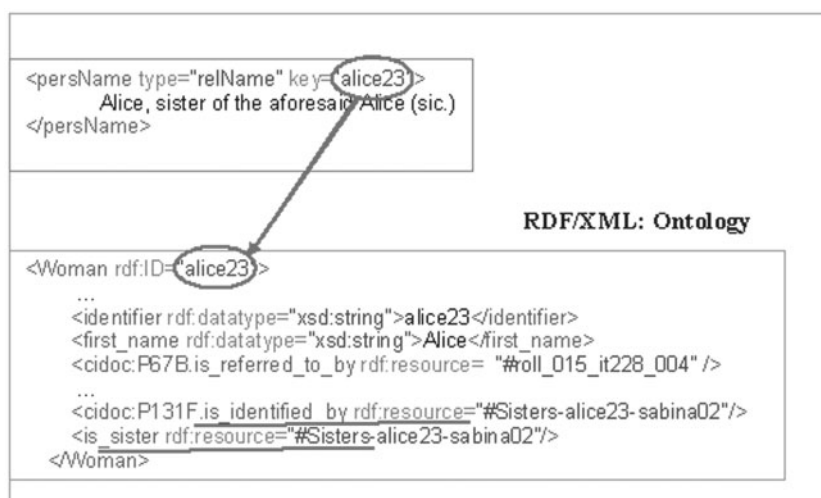


Fig. 8 Occurrence of *alice23* as asserted in the source text in connection with a corresponding authority in the ontology, where, among other things, she is identified by her sister *sabina02*

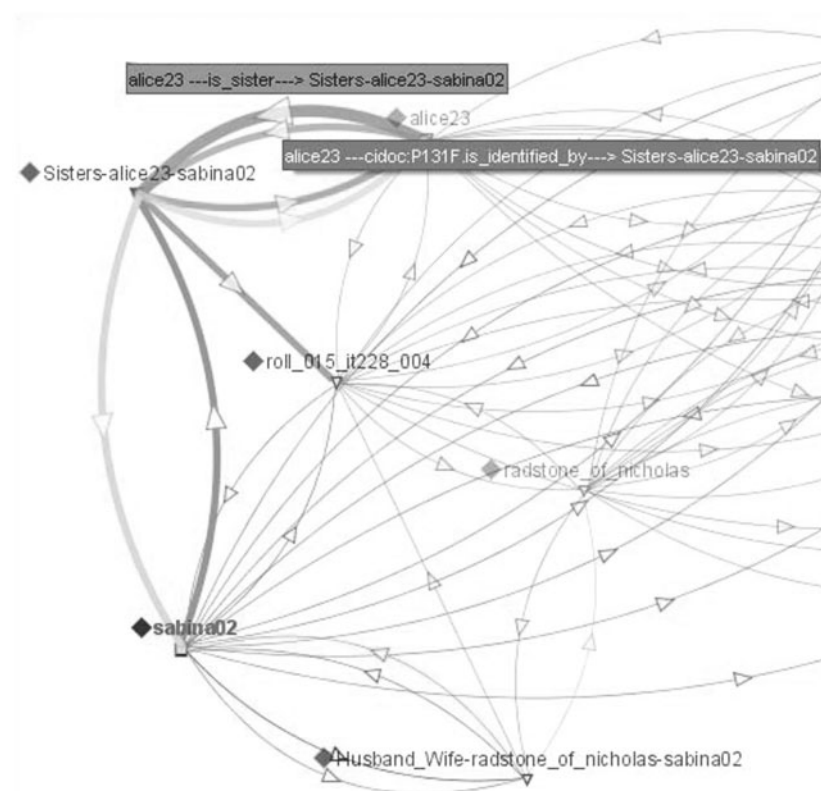


Fig. 9 Jambalaya¹⁴ graph where the instances and properties expressed by the code in Fig. 8 are visualized by rhomboids and coloured arches, respectively

4 Digital and Print Publication

The fact that RDF/OWL is expressed in XML facilitates the re-purposing of data for both web and print delivery. Intermediate RDF/OWL XML files are produced using XSLT transformations on the ontological source data and are used by the print and web publication process at different stages and with different refinements (see Fig. 10).¹⁶ For the website, these files are in turn processed using xMod,¹⁷ the publishing framework developed at the Centre for Computing in the Humanities, running within Cocoon,¹⁸ to produce the website indexes. A parallel print publication process applies XSL-FO to a semantically less rich version of the same material to produce the print indexes.^{19,20} The same intermediate structure in XML format is used to produce the search function (see Fig. 11).

The search facilities for the fine rolls project were created using Ereuna,²¹ a configurable framework that governs the extraction of relevant data from the ontology (Step 1 in Fig. 11) as well as from the original TEI files, which feeds the processed data to the search engine Apache Lucene (Step 3),²² processes search queries (Step 4), and returns search results (Step 5).

The use of OWL also has other more important advantages indirectly related to delivery issues, such as the ability to process the data logically and infer information that is not explicitly asserted in the source-based ontology. So, for example, if Margaret de Beaumont were described as the ‘wife of Saer de Quincy’ in a document, without the use of the ontology we could only search for occurrences of her as the ‘wife of Saer de Quincy’. However, because OWL allows us to define and process associations, we can also say that if Margaret de Beaumont is wife of Saer de Quincy, she is participating in a precise type of relationship—a marriage—where both the participants have identified roles. We can easily deduce that Saer de Quincy is the husband of Margaret de Beaumont, thereby enabling the creation of a much richer search mechanism.

So far we have concentrated our efforts on leveraging the more obvious benefits of the RDF/OWL approach to produce a sophisticated, multi-relational

and searchable resource, but in the future we hope to exploit its potential to the full to generate other outcomes which include:

1. further research into the application of a *reasoner*, which is a tool that can check for consistency and deduce inferences in the current OWL data model by processing assertions;
2. experimentation with more innovative ways of delivering the data, such as two- or three-dimensional visualizations.

5 Conclusions

In general terms, we believe that the approach we have described is perfectly transferable to other historical projects. Besides the broader sharing of models and methods, there is also a more circumscribed level of interchange that this sort of technology encourages. It is very likely that the same people and locations that are mentioned in the fine rolls from 1216 to 1272 are also stated in other related sources, and we hope to explore connections to other projects researching the period in the fullness of time. It is also likely that other sources or other scholars may assert different if not conflicting statements about the same events: the ontological model has the potential to allow effective communication between these propositions by supporting an infrastructure of potentially multiple interpretations and by anchoring them to their evidence in the sources.

While the relationship between the TEI XML files and the ontology has proven to be quite powerful, it has not been without difficulties. Whilst we have successfully populated the ontology with data extracted from the mark-up, it still remains for us to refine the inverse process, so that the mark-up itself can benefit from the process of editing the ontology. A common occurrence is that an editor realizes that two individuals, identified as separate in the mark-up, can in fact be merged into the same authority, using contextual information available only in the ontology that disambiguates the original splitting. While this is a relative easy operation to perform in the ontology, should this knowledge be fed back into the mark-up? Ideally, we would like to

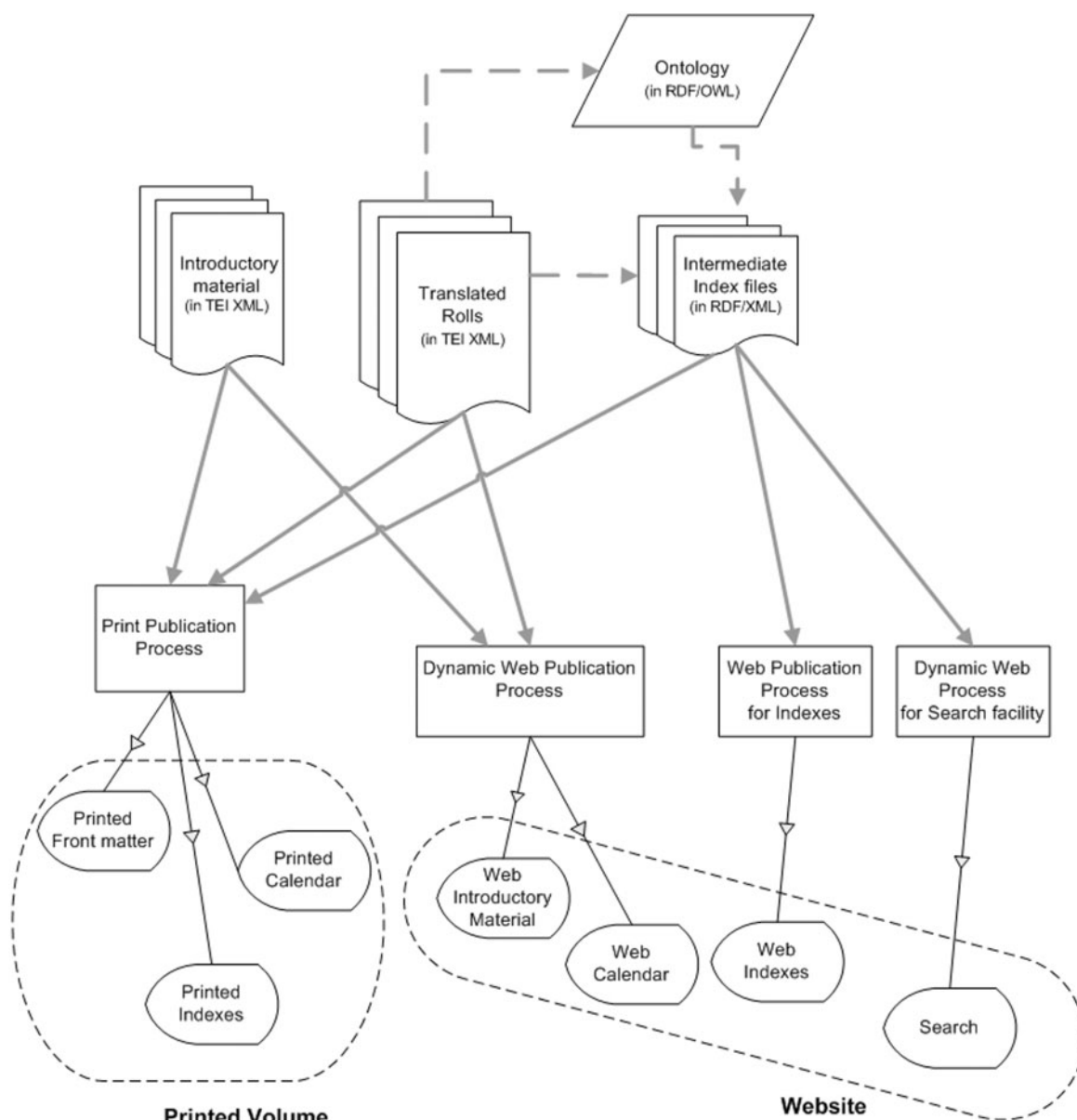


Fig. 10 Diagram of the integration of digital and print publication

keep the two resources synchronized, but the technical implications of achieving this are daunting.

In more abstract terms, the prototyping of the ontology has led to the reification into classes of what we had originally defined as properties, as is the case for institutions and roles/occupations,

which were originally considered to be properties of place and person, respectively, but which were then expanded as entities in their own right. The mark-up then had to be refined accordingly. This is a case where the process of modelling the ontology has affected the editorial strategy of the fine rolls calendar itself.

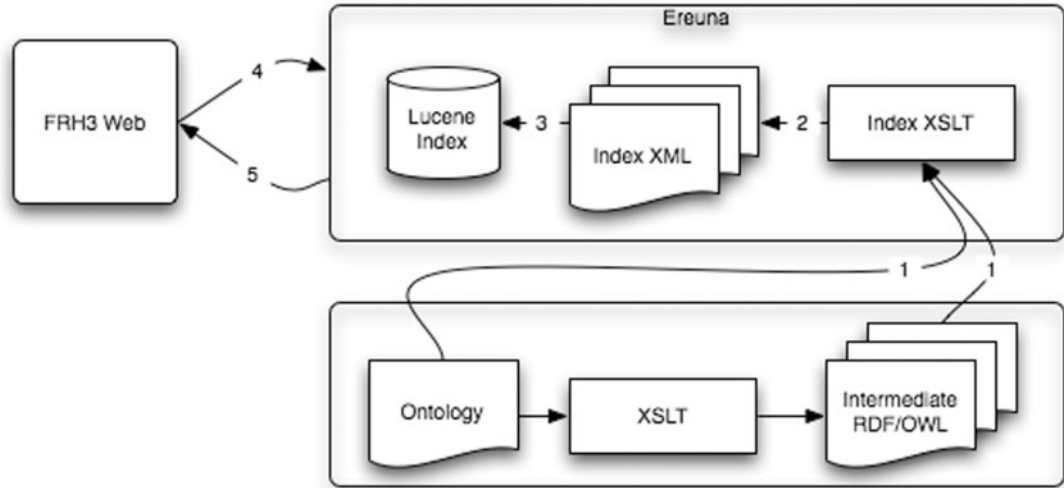


Fig. 11 Search facility

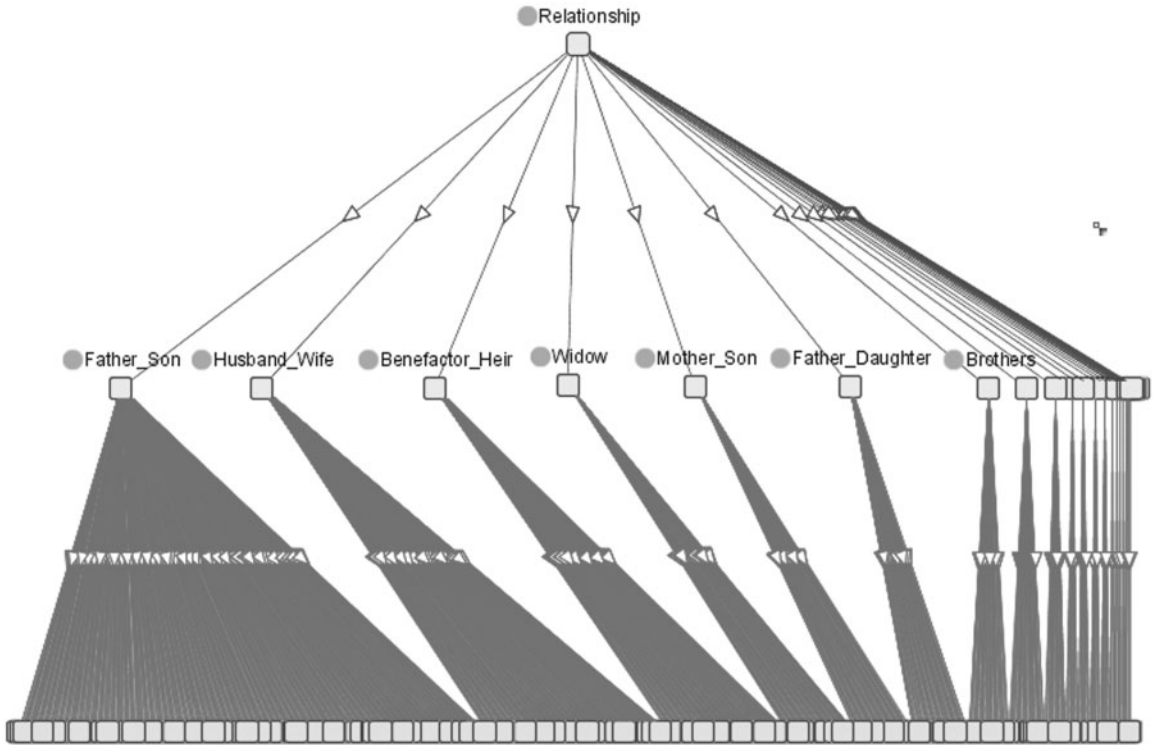


Fig. 12 Jambalaya graph showing the distribution of types of family relationships mentioned in the fine rolls for the period 1216–24

Finally, the ontology not only enables the editors to refine the associations captured at a lower level in the mark-up, but also helps to contextualize these details on a wider scale, which in turn encourages a conscious disengagement from the idiosyncrasies of the text in favour of the bigger picture (see for instance, Fig. 12).

In conclusion, we believe that the use of an ontological approach, in conjunction with TEI XML mark-up for encoding the sources, represents a powerful technical framework for historical projects, which balances respect for the text as unique witness with the need to represent its historical context and interpretation.

Acknowledgements

The Henry III Fine Rolls project is a collaborative project that depends on the contribution of individuals from various institutions (in particular The National Archives and King's College London in the UK). The project website²³ provides a full list of those who have participated, but we would like to acknowledge Gautier Poupeau for his crucial research contribution and work in setting up an initial ontological model with us. We would also like to thank the following people for their advice and contributions in preparing this article: John Bradley, Paul Dryburgh, Beth Hartland, Tamara Lopez, Paul Vetch.

References

- Boot, P.** (2006). Decoding emblem semantics. *Literary and Linguistic Computing*, 21: 15–27.
- Boot, P.** (2007). Mesotext. Framing and exploring annotations. In Stronks, E. and Boot, P. (eds), *Learned Love. Proceedings of the Emblem Project Utrecht Conference on Dutch Love Emblems and the Internet (November 2006)*. Dans: DANS Symposium Publications 2, The Hague, pp. 211–25.
- Bradley, J. and Short, H.** (2005). Texts into databases: the evolving field of new-style prosopography. *Literary and Linguistic Computing*, 20(suppl. 1): 3–24.
- Caracciolo, C.** (2006). Designing and implementing an ontology for logic and linguistics. *Literary and Linguistic Computing*, 21: 29–39.
- Ciula, A.** (2003). *Searching the Fine Rolls: A Demonstration of the Electronic Version*. Paper presented at the International Medieval Congress 2006, University of Leeds, July 10–13.
- Crofts, N., Doerr, M., Gill, T., Stead, S., Stiff, M.** (eds). (2007). Definition of the CIDOC Conceptual Reference Model. Version 4.2.2. August. <http://cidoc.ics.forth.gr> (accessed August 2008).
- Dryburgh, P.** (2006). *Henry III Fine Rolls Project. Paper presented at the Institute of Historical Research, London, February 9.*
- Dryburgh, P. and Hartland, B.** (eds), **Ciula, A. and Vieira, J. M.** (tech. eds) (2007). *Calendar of the Fine Rolls of the Reign of Henry III [1216-1248]. I: 1216-1224*. Woodbridge: Boydell & Brewer.
- Eide, Ø.** (2008). The exhibition problem. A real life example with a suggested solution. *Literary and Linguistic Computing*, 23: 27–37.
- Eide, Ø. and Ore, C.** (2006). TEI, CIDOC-CRM and a possible interface between the two. In *Digital Humanities 2006. The First ADHO International Conference: Conference Abstracts*. Université Paris-Sorbonne. <http://www.allc-ach2006.colloques.paris-sorbonne.fr/DHs.pdf> (accessed August 2008).
- Horridge, M., Knublauch, H., Rector, A., Stevens, R., Wroe, C.** (2004). A practical guide to building OWL ontologies using the Protégé-OWL plugin and CO-ODE tools edition 1.0. <http://www.co-ode.org/resources/tutorials/ProtegeOWLTutorial.pdf> (accessed August 2008).
- Ore, C.** (forthcoming). New digital assets - how to integrate them? In Vogeler G. (ed.), *Digitale Diplomatie. Die historische Arbeit mit Urkunden in der Digitalen Welt. Akten der internationalen Fachtagung in München 28.2.-2.3. 2007*. Archiv für Diplomatik, Beiheft. Köln: Böhlau-Verlag.
- Poupeau, G.** (2006). De l'index nominum à l'ontologie. Comment mettre en lumière les réseaux sociaux dans les corpus historiques numériques? In *Digital Humanities 2006. The First ADHO International Conference: Conference Abstracts*. Université Paris-Sorbonne. <http://www.allc-ach2006.colloques.paris-sorbonne.fr/DHs.pdf> (accessed August 2008).
- Spence, P.** (2006) The Henry III Fine Rolls Project. In *Digital Humanities 2006. The First ADHO International Conference: Conference Abstracts*. Université Paris-Sorbonne. <http://www.allc-ach2006.colloques.paris-sorbonne.fr/DHs.pdf> (accessed August 2008).

- Spence, P. and Vetch, P.** (2004). xMod and XML Publication. Paper presented at Digital Resources for the Humanities (DRH), Newcastle upon Tyne, 5–8 September. <http://drh2004.ncl.ac.uk/abstract.php?abstract=246> (accessed August 2008).
- TEI Consortium, (eds)** (2007). *TEI P5: Guidelines for Electronic Text Encoding and Interchange*. Version 1.0.1. Last updated on 3rd February 2008. TEI Consortium. <http://www.tei-c.org/Guidelines/P5/> (accessed August 2008).
- Thomas, W. G. III.** (2004). Computing and the historical imagination. In Schreibman, S., Siemens, R., and Unsworth, J. (eds), *A Companion to Digital Humanities*. Oxford: Blackwell, pp. 56–68.
- Tuohy, C.** (2005). *Using XML Topic Maps to present TEI*. Paper presented at the 5th TEI Meeting, Sofia, October 28–9.
- Vieira, J. M. and Ciula, A.** (2007). *Implementing an RDF/OWL Ontology on Henry the III Fine Rolls*. Paper presented at OWLED 2007, Innsbruck, June 6–7. http://www.webont.org/owled/2007/PapersPDF/submission_6.pdf (accessed August 2008).

Notes

- 1 See <http://www.tei-c.org> (accessed August 2008).
- 2 The project website is <http://www.frh3.org.uk> (accessed August 2008).
- 3 Further funding was awarded in November 2007 to encode the remaining fine rolls of Henry III up to the year 1272 (the end of his reign). The Department of History and American Studies at Canterbury Christ Church University, Kent, became a partner in the second phase of the project.
- 4 The encoding model developed for the project has been described in Ciula (2003). An experimental framework based on TEI P5 was also developed and is described in Spence (2006).
- 5 See the chapter on *Names, Dates, People, and Places* in TEI Consortium (2007). <http://www.tei-c.org/release/doc/tei-p5-doc/html/ND.html> (accessed August 2008).
- 6 OWL is a language developed on top of RDF by W3C to write ontologies. Indeed, while the latter facilitates the representation of data, it is the former that elevates this representation to a meaningful data model, where relationships between objects are defined in semantic terms and therefore can be *reasoned* upon automatically. Quoting the W3C documentation: ‘The Semantic Web is a vision for the future of the Web in which information is given explicit meaning, making it

easier for machines to automatically process and integrate information available on the Web. The Semantic Web will build on XML’s ability to define customized tagging schemes and RDF’s flexible approach to representing data. The first level above RDF required for the Semantic Web is an ontology language what [sic] can formally describe the meaning of terminology used in Web documents. If machines are expected to perform useful reasoning tasks on these documents, the language must go beyond the basic semantics of RDF Schema.’ <http://www.w3.org/TR/owl-features/> (accessed August 2008). Furthermore, ‘Ontologies are critical for applications that want to search across or merge information from diverse communities. Although, XML DTDs and XML Schemas are sufficient for exchanging data between parties who have agreed to definitions beforehand, their lack of semantics prevent machines from reliably performing this task given new XML vocabularies. The same term may be used with (sometimes subtle) different meaning in different contexts, and different terms may be used for items that have the same meaning. RDF and RDF Schema begin to approach this problem by allowing simple semantics to be associated with identifiers. With RDF Schema, one can define classes that may have multiple subclasses and super classes, and can define properties, which may have sub properties, domains, and ranges. In this sense, RDF Schema is a simple ontology language. However, in order to achieve interoperation between numerous, autonomously developed and managed schemas, richer semantics are needed. For example, RDF Schema cannot specify that the Person and Car classes are disjoint, or that a string quartet has exactly four musicians as members.’ <http://www.w3.org/TR/webont-req/> (accessed August 2008). For W3C specifications on RDF see <http://www.w3.org/RDF/> (accessed August 2008). For an overview and further resources on OWL see <http://www.w3.org/2001/sw/> (accessed August 2008).

7 As Thomas states: ‘The goal for historians working in the new digital medium needs to be to make the computer technology transparent and to allow the reader to focus his or her whole attention on the “world” that the historian has opened up for investigation, interpretation, inquiry, and analysis. Creating these worlds, developing the sequences of evidence and interpretation and balancing the demands and opportunities of the technology will take imagination and perseverance.’ (Thomas, 2004, pp. 66–67). Indeed, the direct experience on this act of modelling has proved very imaginative and particularly challenging when the collaborative aspect of Digital Humanities projects—where meaning

- is under continuous negotiation and partial formalization—is taken into account.
- 8 For a similar approach see Eide and Ore (2006), and Ore (forthcoming).
 - 9 Note that the model does not claim to be a complete formalization of all ‘exhibitions’ and assertions in the text. Indeed, some choices—largely driven not only by the research interests of the project team, but also by pragmatic reasons—had to be made on what to model and what not to model explicitly. Eide (2008) with respect to the modelling of naming events within CIDOC-CRM: ‘In CIDOC-CRM, the modelling of entities we presume to exist without knowing who or what they are is quite possible. And there will always be things we do not know when we read historical texts. What may be a problem is that the naming event we model in this way is an event that is not documented in the text we are basing the model upon. Whether this is acceptable or not is a decision one has to make when building up these kinds of models.’ (Eide, 2008, p. 33).
 - 10 For more details on this see the initial report by Vieira and Ciula (2007).
 - 11 For an overview on these vocabularies see, respectively: <http://cidoc.ics.forth.gr/> (accessed August 2008); <http://dublincore.org/> (accessed August 2008); <http://www.w3.org/2003/01/geo/> (accessed August 2008); <http://www.w3.org/2004/02/skos/> (accessed August 2008); <http://www.w3.org/TR/owl-time/> (accessed August 2008).
 - 12 In England, *hundred* and *wapentake* (in northern counties) are administrative and judicial subdivisions of Saxon origins within each county or shire. In the thirteenth century each had numerous officials and a court where the business of the hundred/wapentake would be done, people tried and judged, and taxes levied and collected.
 - 13 Note that the editors have decided not to encode these occurrences as relevant editorial interventions. Therefore the TEI recommendations to represent primary sources with this respect have not been applied.
 - 14 Jambalaya is a plug-in developed at the University of Victoria, BC, Canada, for visualizing data modelling within Protégé-OWL: <http://www.thechiselgroup.org/jambalaya> (accessed August 2008).
 - 15 Protégé-OWL is downloadable at <http://protege.stanford.edu/> (accessed August 2008). For a tutorial see Horridge *et al.* (2004).
 - 16 It has to be noted that the layout of the web indexes compared to the print indexes does not vary as much as it could have done. Indeed, although at the time of writing the project team had agreed that a more subtle exploitation of the different media should be achieved before the publication of the third volume of the calendar, the switching of editorial paradigms between the two media requires a flexible approach in defining editorial strategies that is certainly favoured by the technologies but which involves some careful re-thinking of well-established editorial practices and conventions.
 - 17 See <http://www.cch.kcl.ac.uk/xmod/> (accessed August 2008); see also Spence and Vetch (2004).
 - 18 See <http://cocoon.apache.org/> (accessed August 2008).
 - 19 The transformation language XSL-FO drives the processor Apache FOP—see <http://xmlgraphics.apache.org/fop/> (accessed August 2008)—to produce the indexes according to the print specifications defined together with the publisher Boydell & Brewer.
 - 20 We are currently in the process of refining the integration between print and web output. For an introduction on how the first print volume of the fine rolls was produced see Ciula, A., Short, H. and Spence, P. *Technical Introduction* (Dryburgh *et al.*, 2007, pp. xlix–lix).
 - 21 The search facility currently accessible on the project website will be enriched further beyond the combination of queries for authorities and date, so as to exploit the complex model embedded in the ontology to the full. See <http://www.cch.kcl.ac.uk/ereuna/> (accessed August 2008).
 - 22 See <http://lucene.apache.org/java/docs/> (accessed August 2008).
 - 23 The personnel and credits page for the project is <http://www.frh3.org.uk/cocoon/frh3/content/about/personnel.html> (accessed August 2008).