Factoid-based Prosopography and Computer Ontologies: towards an integrated approach

Michele Pasin, John Bradley
Department of Digital Humanities
Kings College
London, UK
{michele.pasin, john.bradley} @kcl.ac.uk

Abstract. Structured Prosopography provides a formal model for representing prosopography: a branch of historical research that traditionally has focused on the identification of people that appear in historical sources. Since the 1990s, KCL's Department of Digital Humanities has been involved in the development of structured prosopographical databases using a general "factoid-oriented" model of structure that links people to the information about them via spots in primary sources that assert that information. Recent developments, particularly the WWW, and its related technologies around the Semantic Web have promoted the possibility to both interconnecting dispersed data, and allowing it to be queried semantically. To the purpose making available our prosopographical databases on the semantic web, in this article we review the principles behind our established factoid-based approach and reformulate it using a more transparent approach, based on knowledge representation principles and formal ontologies. In particular, we are going to focus primarily on a high-level semantic analysis of the factoid notion, on its relation to other cultural heritage standards such as CIDOC-CRM, and on the modularity and extensibility of the proposed solutions.

1 Introduction

Structured Prosopography provides a formal model for representing prosopography: a branch of historical research that traditionally has focused on the identification of people that appear in historical sources. Pre-digital print prosopographies presented its materials as narrative articles about the individuals it contains. Since the 1990s, KCL's Department of Digital Humanities has been involved in the development of structured prosopographical databases, and has had direct involvement in Prosopographies of the Byzantine World (PBE and PBW), Anglo- Saxon

England (PASE), Medieval Scotland (PoMS) and now more generally northern Britain ("Breaking of Britain": BoB)[1]. DDH has been involved in the development of a general "factoid-oriented" model of structure that although downplaying or eliminating narratives about people, has to a large extent served the needs of these various projects quite well.

DDH's factoid-oriented prosopographical projects are currently all expressed using the entity-attribute-relationship model of the relational database. The structure formally identifies obvious items of interest: Persons and Sources, and extends to related things like Offices, Places and Possessions. In our prosopographical model the *Factoid* is a central idea and represents the spot in a primary source where something is said about one or more persons. In other words, it links people to the information about them via spots in primary sources that assert that information.

Recent developments, particularly the WWW, and its related technologies around the Semantic Web (W3C 2001) have promoted the possibility to both interconnecting dispersed data, and allowing it to be queried semantically. Central technologies to support this approach are ontology languages such as RDFS (W3C 2004) and OWL (W3C 2009). Modelling the work of prosopography in a framework such as OWL is in many ways similar to relational database modelling: including the idea of classes and slots which correspond quite closely to entities and attributes in RDB modelling, and in the handling of relationships between data.

However there are also some fundamental differences. In the context of the semantic web, ontologies are meant to facilitate the process of data sharing among different (and possibly unknown to each other) parties. For that reason they must possess a degree of generality and reusability that very often cannot be found in relational models. For example, this is of primary importance when the relational schemas being used contain abstractions that are arbitrary (e.g. think of a table for 'people of uncertain origin'), application-specific (e.g. 'document tick-boxes'), or that, more generally, serve purposes which are not compatible with the one of making the semantics of data more transparent, and thus more accessible to other interested scholars.

In this article we review the principles behind our established factoid-based approach to prosopography to the purpose of reformulating it using a more transparent approach, based on knowledge representation principles and formal ontologies. To this aim, we are going to focus primarily on the definition a high-level semantic analysis of the factoid notion, especially in relation to other cultural heritage standards such as CIDOC-CRM (Doerr 2003). Since all of our digital prosopographies rely on the factoid-model, the results presented in this paper will be easily extendable across our projects and will constitute the basis for the construction of the next-generation infrastructure for distributing our prosopographical data over the semantic web.

Secondly, we will discuss the *modularity* of our approach. The way factoids have been reformulated allows for the future development and integration of more vertical historical ontologies e.g. ones that analyze in detail specific historical periods (or aspects of them). As an example of this methodology, we are going to briefly introduce the *feudalism* ontology (which covers aspects of the world of medieval England and Scotland) and

show how it can be used in combination with the suggested ontological rendering of the factoid model.

The rest of the article is organized as follows: section 2 reviews the key characteristics of factoid-based prosopography; section 3 provides a walkthrough of the modeling decisions we have taken in order to create a more interoperable representation of the factoid model; section 4 elaborates on these results, in particular by reformulating them in terms of a well known standard ontology in cultural informatics, the CIDOC-CRM; section 5 shows how the ontology can be extended with more project-specific contents; section 6 outlines a number of unexplored issues that we wish to examine in future work; finally section 7 contains a summary and conclusion.

2 Factoid based prosopography: an overview

Pre-digital print prosopographies, such as Martindale's *Prosopography of the Later Roman Empire* (Martindale 1980), presented its materials as narrative articles about the individuals it contains. For example, in Fig. 1 we can see a typical brief article for someone identified as *Eucharius 4*. In this article Martindale tells us a story about this person, mostly based on information he found on the letters *Sidonius Apollinaris* wrote. Even within the rather terse style of this rather brief article, one can see the complexity and subtlety of narrative playing a role. However, a closer look at these articles reveals a rather formal, structured text with a carefully defined kind of interests: *dates* and *dating*, of course, and information about in which *texts* the person appeared; but also *titles* or *offices* the person held, *places* with which s/he is associated, other *people* with whom s/he had connections, etc. Objects of interest begin to emerge from the text that can provide the basis for a structure to manage them.

Overall, then, our experience has lead us to believe that certain elements of structure represent *key interests* in the mind of many prosopographers as they carry out their work. Furthermore this formal structuring allows us not only to look up a particular person, but to use the computer to approach the data from a whole range of different perspectives i.e. by focusing on a specific facet of persons (e.g. titles), or by combining several of them together so to perform more complex searches that would not be possible otherwise (e.g., we could easily obtains statistics on the titles of people that appeared more than once in imprisonment events).

Narratives, nonetheless, remain a powerful mechanism to present historical research. Multidimensionality, complexity and non-linearity are just a few among the many characteristic features of narratives that could not be easily reduced to the unambiguous abstract language of databases. As a result, downplaying or eliminating narratives about people in virtue of a systematic use of formal structures often causes historians to worry - and rightly so - especially if we take the wrong approach of using the database as the surrogate for the original document.

It is precisely in this context that the 'factoid' based prosopography was first developed. Oddly enough, the word Factoid was coined by Norman Mailer in his 1973 biography of Marilyn Monroe. Mailer described factoids

as "facts which have no existence before appearing in a magazine or newspaper" (Mailer 2012). In a similar fashion, in our prosopographical model the Factoid represents the spot in a primary source where something is said about one or more persons. It links people to the information about them via spots in primary sources that assert that information.

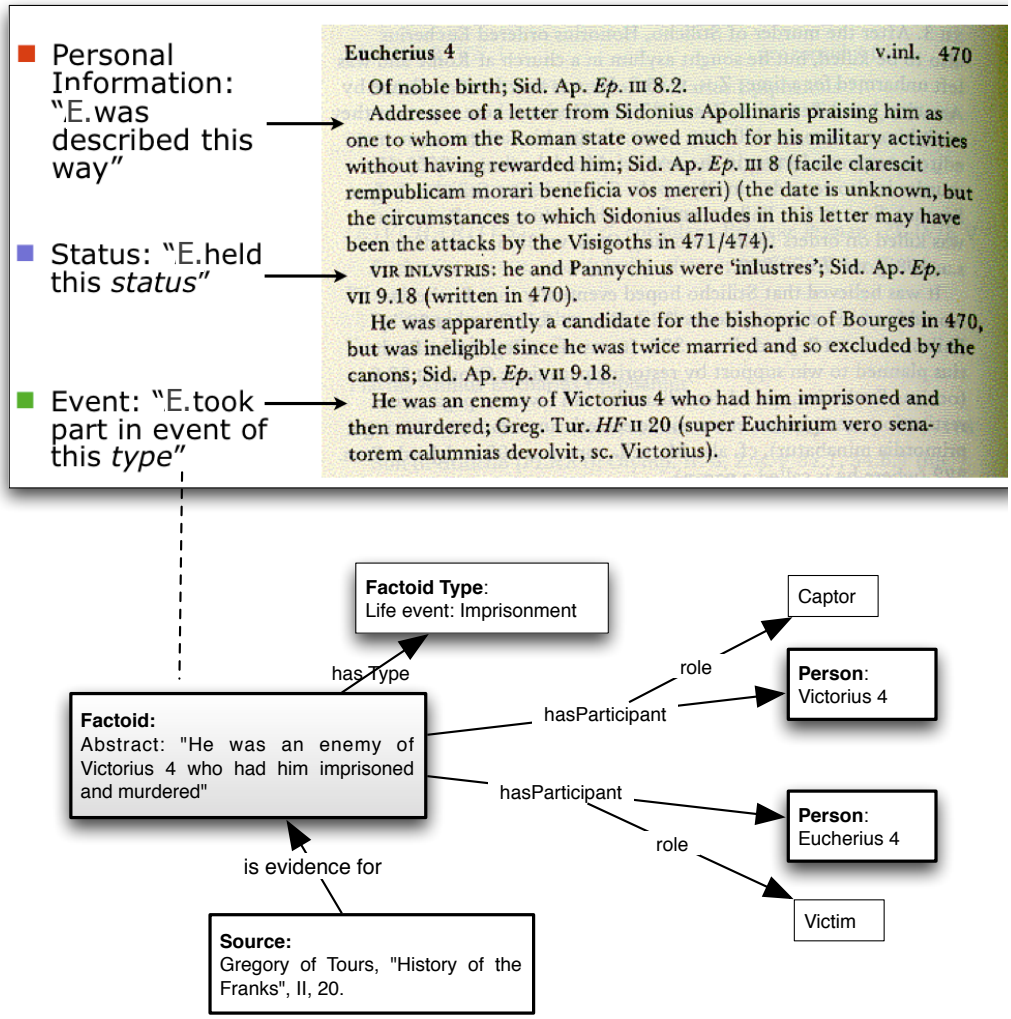

Fig. 1 From textual prosopography to structured prosopography: an example of how Martindale's text could be represented using factoids

The somehow ironic flavor of the name *factoid* is intentional, and reflects a concern that historian often have towards the veracity of sources, since the action of taking materials out of context has to be done with care. By creating factoids which assert things about what the source says about people, the factoid approach prioritizes the sources, rather than our historians' reading of them. Our data about a person is not, then, so much a narrative that presents a summary written by the prosopographer as a collection of information about what the sources say about him/her, and can represent the multiple, perhaps contradictory, voices of the different sources simultaneously: one saying she is a Saxon, but another saying that she was from Northumbria. Perhaps the source is thought to be a forgery, or a statement that arises out what we now consider prejudice. A factoid, however, represents what a particular source

says on its own, rather than what we now believe arises out of our reading of a range of sources. If two sources contradict each other, two contradictory factoids will be recorded. For this reason, we call the factoid model *source driven*.

It is the strength of the factoid approach that, in spite of the very diverse nature of the different kinds of sources that our prosopographical projects have worked with, it is able to provide a useful representation of the prosopographer's task and insights across a range of different periods and cultures and sources. For a more complete overview of the factoid model than what there is room for here, readers should see other publications e.g. (Bradley & Short 2003).

3 Towards a more interoperable definition of factoids

Abstract schemas such as the factoid model never exist in isolation, but rather they are born and live within a specific *context* and for that reason they always have a *purpose* (Allemang & Hendler 2011). In particular, from our experience with prosopographical projects, we can single out four main areas where models (of some sort) normally play a key role, and consequently also have a well defined purpose: data acquisition, data storage, data presentation and data integration (see Table 1).

Context	Purpose	Requirements
Data Acquisition	Extract information from people	Models need to reflect experts' workflow and conceptualization of the domain, so to facilitate the task of entering data into the system.
Data Storage	Implementation of a DB	Models must allow the implementation of practical and effective computing solutions, thus they must comply with principles of efficiency and cost.
Data Presentation	Finding information	Models should facilitate finding information and making sense of it in a way that reflects the discipline's discourse.
Data Integration	Sharing data	Models should facilitate the task of (semi) automatic integration of data streams originating from different digital sources.

Table 1 Main areas of employment of models in digital prosopography

Let us now consider the applicability of the factoid model with respect to these four areas. During the *data acquisition* phase, factoids provided historians with both a guiding metaphor (helping them conceptualise the broader approach being used) and a usable structure for the data entry work. Secondly, within the *data storage* context factoids proved to be a practical, flexible and sustainable schema for designing databases. Thirdly, during the *data presentation* phase the factoid notion has been used with success to the purpose of building user interfaces that are

simple yet rich in the way they combine and organise information about people and make it available to the historians using our online resources.

Lastly, we have the *data integration* context. When attempting to employ factoids to allow cross-database and machine-level interoperability - that is, within a context like the one the semantic web entails - we came across some difficulties. The factoid notion, to non-specialists, may seem rather obscure and under-defined. Sometimes this is due to a mere terminological issue; however, we also recognized that this notion is so tied to our own prosopographical approach that it cannot be easily related to other common vocabularies used on the semantic web.

Overall, we could sum this up by saying that the factoid model embeds a very specific *ontological* view of the world, i.e. a view that relies on an understanding of reality, and thus a conceptualization, that has been developed and refined within a thorough but still relatively circumscribed group of people across a number of projects and years. Consequently, the degree of ontological *commitment* (Gruber 1995) that the factoid model requires is too high, for it to be used successfully in an open context such as the semantic web.

The notion of semantic interoperability reflects this need: to the aim of a more interoperable web of data, it is fundamental that together with programmatic access points to our data-stores we also provide explicit and shared formal descriptions of the *meaning* of our data too (Davis et al. 1993). To this end, *formal ontologies* are one of the core components needed to build a flexible yet solid integration architecture (Doerr et al. 2003). This is due to the fact that the principles of formal ontologies allow us to make categorical distinctions in a systematic and coherent manner - thus allowing the creation of models that can be used like 'semantic glue' among other, more application-specific models.

As a first step towards a more transparent formal representation of the factoid data model, we set out to scrutinise the 'factoid' idea so to deconstruct it into other concepts which are more primitive and context independent. A thorough analysis of the usage of this model across our projects revealed that each time we introduce a factoid we imply the existence of three main entities. An *agent* who creates the factoid (who could be either a person or a collective entity); a *document* that inspires it; finally, a *state of affairs,* that is what the factoid is about. In particular, this last entity is what normally includes information about historical people, places etc. via some kind of specialized *participation* relationship. Furthermore, we also concluded that in all cases the state of affairs described by the document is either a *situation* or an *event*, i.e., more generally, a *temporal entity* - something that in principle can be located at a specific time (or time span) and place (although it is often the case that we lack evidence for this type of information).

The resulting triadic class structure (`agent`, `document` and `temporal-entity`) could give us a quite complete picture of the intrinsic meaning of factoids. However, while these entities appear to be linked together each time we create a factoid, they are not really related to each other, as such. In other words, it is necessary to explicitly represent also the underlying *act* that associates them into a meaningful unit. We achieved this by introducing the concept of a `Document-Interpretation-Act` - defined as another temporal entity that has a characteristic structure - it has an *author*

(who performs the interpretation), a *source* (what is being interpreted) and an *interpretation* (the result of the interpretation act).

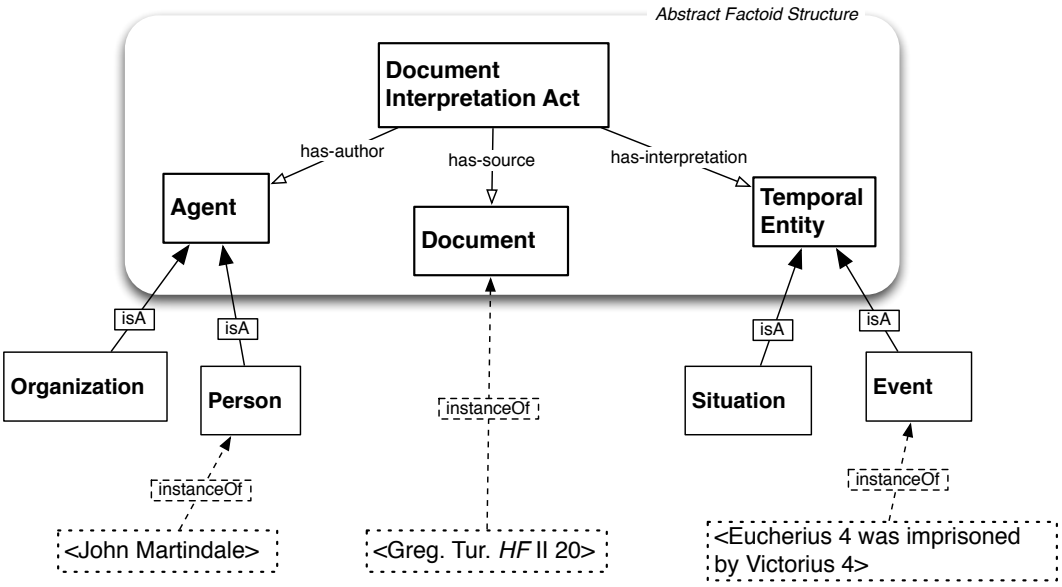

Fig. 2 Conceptual representation of Factoids by mean of Interpretations

We thus obtained a general representation of factoids, depicted in fig. 2. This semi-formal model can easily be implemented in a semantic web language like RDF and instantiated using the data in our prosopographical databases. In practice, this means that each time we come across a factoid structure in our relational database that would be transformed into a `Document-Interpretation-Act` instance; furthermore, the entities attached to the original factoid would serve to fill in the *has-author*, *has-source* and *has-interpretation* slots of the `Document-Interpretation-Act` class.

For example, the gist of the situation described by Martindale at the beginning of this article (see section 2) could be rendered using the following fragment of RDF[2]:

```
@prefix plre: <http://example.com/Prosop_of_the_Later_Roman_Empire#[3]>
@prefix factoid: <http://purl.com/net/factoid#>
@prefix rdf: <http://w3.org/1999/02/22-rdf-syntax-ns#>

plre:john-martindale a factoid:person.
plre:history-of-the-franks a factoid:document. IS A RDFS??
plre:eucherius-imprisonment a factoid:event.

plre:doc-interpretation-001 a factoid:document-interpretation-act;
    factoid:has-author plre:john-martindale;
    factoid:has-source plre:history-of-the-franks;
    factoid:has-interpretation plre:eucherius-imprisonment.
```

4 Improving interoperability via CIDOC-CRM

One of the central approaches to modelling in the Semantic Web world is to develop a schema that contains elements of other, compatible, schemas. Borrowed elements must not only match structurally, but the semantics of the classes and slots in the shared model have to match conceptually as well.

Accordingly, in order to enhance the semantic interoperability of our model we tried to express it by using the formalisms offered by the CIDOC-CRM ontology (Crofts et al. 2011; Doerr 2003). This is a well known ISO standard that includes concepts and relations describing people, organisations, places, events, physical objects and conceptual ones. Quite interestingly, CIDOC also provides classes meant to facilitate the representation of the 'hiatus' between a fact and its readings (i.e., as expressed by the source, or by us in virtue of our editorial role). In particular, CIDOC's `E-13-Attribute-Assignment` appeared to be suitable to express also the intended meaning of our `Document-Interpretation-Act` class (and, in turn, of the underlying semantic structure of factoids).

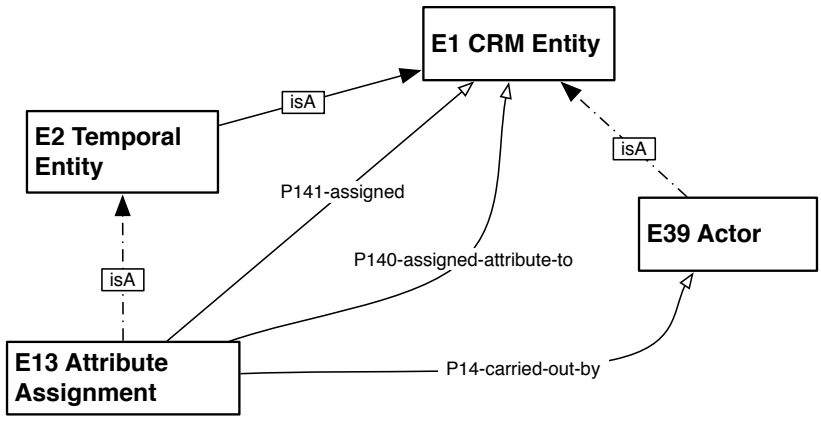

Fig. 3 Structure of the E13 Attribute Assignment class in CIDOC-CRM[4]

An `E13-Attribute-Assignment` is a kind of Temporal entity - that is, something that happened at a particular time and a particular place - and its stated purpose (Crofts et al. 2011) is to represent "*actions of people making propositions and statements during certain museum procedures"* (p. 8). This class allows for the *"documentation of how the respective assignment came about, and whose opinion it was*" (p. xvi), so that the *"properties assigned in such an action can also be seen as directly attached to the respective item or concept, possibly as a collection of contradictory values*" (p. 8).

It is evident that this modelling pattern was created so to be used in a specific context - the museum - consistently with the rationale of CIDOC-CRM. Its semantic structure, nonetheless, is very generic: in fact both the *assigned* and *assigned-attribute-to* properties point to instances of the `E1 CRM-Entity` class (which means, virtually, any possible instance available). So it is possible to describe the `E13-Attribute-Assignment` pattern as a simple reification mechanism, that is, a method by which one can formalize statements of the kind "A says that X is Y". Indeed, this is precisely the kind of semantic structure we wanted to capture with our own

`Document-Interpretation-Act` (see section 3). Consequently, we concluded that it would make sense to use it as the contact point between the two models. In other words, a `factoid:Document-Interpretation-Act` can be seen as a type of Temporal entity that further specialises the `E13-Attribute-Assignment` class. That is to say, a `factoid:Document-Interpretation-Act` encodes the fact that *someone says* that a document *is about* a certain state-of-affairs.

Finally, what remains to be discussed is how the other entities in the factoid model (`agent`, `temporal-entity` and `document`) could be mapped onto CIDOC. For the first two cases, the mapping is straightforward since CIDOC's `E2-Temporal-Entity` and CIDOC's `E39-Actor` classes have the exact same intended meaning as our classes (thus the equivalence can be formalised in RDF via a simple subsumption relationship).

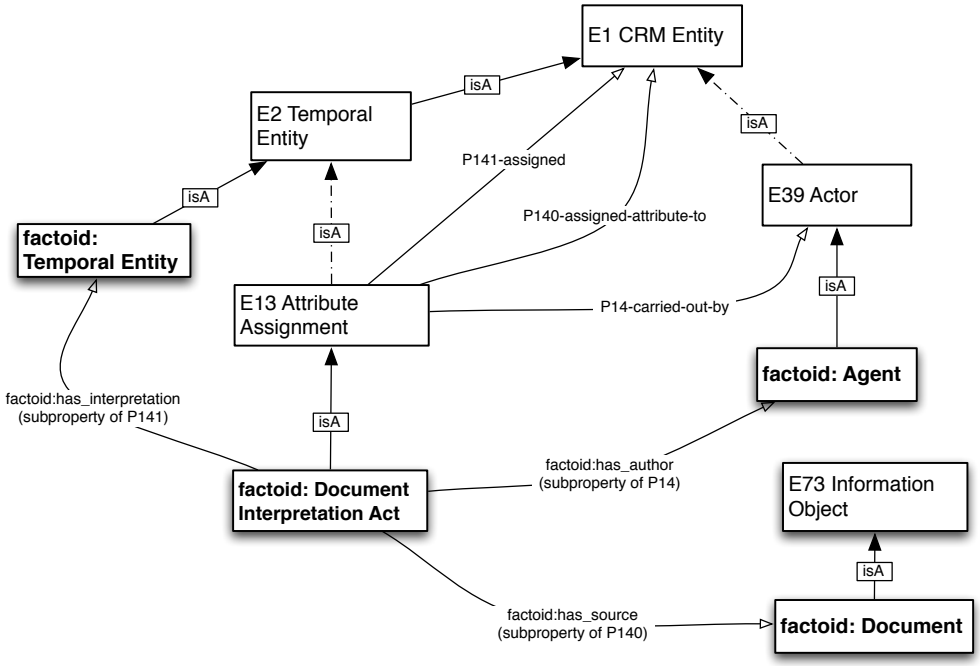

Fig. 4 Mapping the factoid model to CIDOC-CRM primitives

The case of documents is slightly more complicated, since the document we are talking about in a factoid is not a physical entity (e.g. a specific copy of a book, or a manuscript) but rather the abstract contents of it (i.e., its informational content, expressed through a language). This is a key distinction to keep in mind, because only such an approach would allow us to say that the very same factoid could in principle be extracted from identical copies of a source. For example, the factoid in question could be equally derived from a fragment of a medieval manuscript held at the British Library, or from a photograph of that fragment we may find online.

Of course, the modeling issues deriving from a thorough analysis of the tangible and intangible properties of documents are not new to researchers in this area. Indeed, a number of solutions have been proposed, which vary in complexity and expressivity ((Bekiari et al. 2010; IFLA 2007; Renear 2007; Gangemi et al. 2005). In our case, however, we decided to take a 'minimalistic' approach. That is to say, we aimed at

avoiding the inclusion of unnecessary complexity in our model, unless it was strictly required. And in fact, this revealed not to be the case. CIDOC includes in its standard specification a class called `E73-Information-Object`, defined as a conceptual entity that "does not depend on a specific physical carrier" but "it can exist on one or more carriers simultaneously" (p. 29). Since this class definition matches the intended meaning of our own `factoid:document`, we concluded that it was not necessary to include more elaborate structures in our model. As a result, also in this case we were able to formalise the mapping via a subsumption relationship[5].

5 Domain-specific extensions: the *feudalism* ontology

As mentioned at the beginning of this article, the factoid model has been used in a number of prosopographical projects that, taken as a whole, span across almost two thousand years of history. This would not have been possible unless this conceptual framework was general enough to allow this degree of reusability; however it is also true that each single project required a number of extensions to the model. For example, in some projects the main kind of documents available are *charters* and the temporal objects are *transactions*, while in others we have instead *rolls* and events reflecting particular *familiar relationships*.

More generally, the reasons for these differences are to be found in the unique features of an historical period, in the varying kinds of primary sources available, or simply in the specific questions driving the researchers who compile a prosopography. It is thus fundamental for a model to be designed so that this kind of project-specific extensions can be added to it. This would guarantee a base level of interoperability while at the same time also allowing a great degree of flexibility.

One of the advantages of our CIDOC-based representation of the factoid model is that it does offer this kind of modularity. In order to show the reader how this could be done, we can look at a brief example based on the People of Medieval Scotland prosopography (www.poms.ac.uk). In the context of this project we have begun working on an a ontology called *feudalism*[6], aimed at representing key concepts related to the world of medieval Scotland. For example, in this context the notion of *transaction* plays a central role, in fact most of the historical charters being transcribed deal with events where, in general, something is given to someone else. Transactions have therefore been thoroughly categorised, together with the specific participatory relationships actors could have in the context of these events, and the kind of entities (objects, rights, or privileges) that are being given or taken.

The feudalism ontology is currently a work in progress, as it is being enlarged and refined via the progressive inclusion of concepts deriving from other medieval prosopographical projects created in our department (including ones that focus on geographical areas other than Scotland). What is of interest here is that the ontology is being constructed as an

extension of the factoid one (and, indirectly, of CIDOC-CRM). This approach serves to guarantee more interoperability and also to avoid duplication. For example, in fig. 5 we can see how transactions are modeled. In this case a `feudalism:Renewal` becomes simply a specialised type of a `factoid:temporal_entity` object, and used within a `factoid:Document-interpretation-act`. Note also that the various relations involving people (e.g. *witness* or *grantor*) are all defined as sub-properties of `factoid:has_participant`. This means that within a data integration context that operates only at the level of the factoid ontology, we can still benefit from a semantic description of these data.

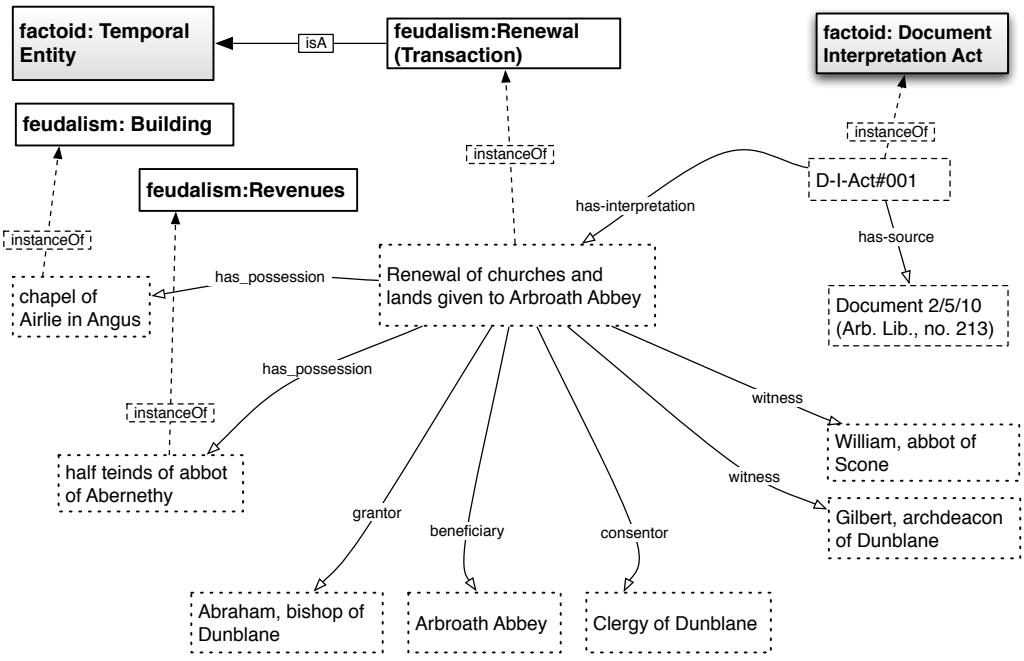

Figure 5 The feudalism ontology built as an extension of the factoid ontology

6. Open issues

The modeling approach we have taken brought to us to a number of unexpected consequences. In this final section we are going to outline three of them, as we believe they can inspire future research in this area.

Interpretations vs *literal* interpretations

The notion of `factoid:Document-Interpretation-Act`, intended as the institution of a 'connection' between two objects (a temporal entity and a document), in reality leaves rather unspecified the nature and purpose of the connection in question. For example, we can think of interpretations that have different degrees of certainty, or interpretations that depend on evidence provided by other background knowledge.

In the original factoid approach, the relationship between a source and the assertions made in that source was (more or less implicitly) meant to

be as *literal* as possible. In other words, historians were always asked to avoid doing any interpretive work, as if they were just 'replicating' the information they found onto the digital medium (despite being conscious of the fact that even the most literal reading cannot escape a fundamental cultural hermeutical horizon).

In the ontology this situation is captured by the `factoid:has-interpretation` predicate, which links a document to a state of affairs. However, the relation in question does not express very well its intended meaning - since in theory any type of interpretation - not just one that wants to reflect the source literally - can be assigned to a document. Ideally, each time an interpretation is made and codified via the factoid model, it would be desirable to encode also some information about the strength or nature of the connection being created. This kind of data could be further organized, e.g. within a continuum, according to their strength, which could result in having N possible *evidence* relations reflecting the degree of certainty historians may have.

So in future versions of the ontology it would be desirable to further elaborate the semantics of this relation.

Modal and Negative statements

Quite interestingly, the previous point opens up a path towards *modal logic*, for it allows creating claims such as "Document X *may* be about the situation Y under conditions Z". We think that historians could make use of this type of claims - the question is then, how to treat them within the database when searching for things, or performing inferences based on existing data? It goes without saying that this opens up a whole range of new issues. It is worth noting though that a very preliminary approach to the implementation of modal statements within a digital prosopography was done in the context of the CCED project (Burns et al. 2007). Here a number of certainty-codes[7] have been used for characterizing the factoids creation process. However, although this type of information is made available to users when they perform a search, it is not used to enhance the search algorithms or perform formal reasoning tasks on the dataset.

Finally, this line of thinking can be even taken further - to the point of deciding to include *negative statements* in our datasets, for example by letting people say that "According to document X, event E never happened". Since a negative statement relies primarily on some type of inference process we are making (based on both the sources and other background knowledge), the most important consequence of this approach is that it forces us to depart from the strictly 'literal', source-based philosophy of factoids. Nonetheless, the speculation is inevitable as long as we decide to move towards an ontological model of interpretations, and we will be able to determine its usefulness or not only through further research work.

7. Conclusion

There has been a stream of argument about the black-and-white nature of assertions made through computer ontologies, implying that this bi-polar nature is a significant flaw when applied to humanities materials. Indeed, we expect that much of the discussion about computer ontologies – centered often on still the relatively simple problems within science and engineering that have been used as examples – have put Humanists (even Digital Humanists) off. Veltman (Veltman 2004) provides this kind of argument when he claims that the preservation of culture requires the dealing of changing meanings over different places and times, and that computer ontologies try to "create data structures that assume a single world-view" (p. 7). Now, this would indeed be a significant concern and, not accidently, fit with, say, Louch's (Louch 1969) reasoning about why narrative, with its subtlety of expression, remains for many historians the main vehicle for research output. However the story need not be as pessimistic as Veltman seems to believe, since ontology modelling need not mean that a single view of the material is an inevitable result.

Indeed, our factoid approach can show that formal structuring if designed correctly need not impose, as Veltman implies, a single perspective on the data it models, but is capable of accommodating a range of views from the different sources. By introducing the source as an intermediary, the model can also accommodate contradictory statements from different sources. Similarly, page 11 of the CIDOC-CRM specification points out that "The CRM does not propose a specific form to support reasoning about possible identity", but we are not asking the ontology for these projects to do that. The identity assertion is the work of our historian partners, and the point of the ontology is not to derive the identity of persons from the ontological-expressed data for us, but merely to express what the historians assert about their materials (including assertion of identity of individuals) in ways that support sophisticated searching and integration.

For these reasons, we are very keen to continue with the exploration of the usage of formal ontologies and semantic web approaches in the field of prosopography, and more broadly within history. In particular, we believe that only once the amount of online resources that are available in RDF/OWL will have reached a suitable critical mass, the less technically-minded research community will become realising the great potential of these data integration approaches.

References

Allemang, D. & Hendler, J., 2011. Chapter 2: Semantic Modeling. In *Semantic Web for the Working Ontologist*. Elsevier.

Bekiari, C., Doerr, M. & Boeuf, P.L.B. eds., 2010. *FRBR: Object-Oriented Definition and Mapping to FRBR-ER (version 1.0.1)*, International Working Group on FRBR and CIDOC CRM Harmonisation: International Working Group on FRBR and CIDOC CRM Harmonisation.

Bradley, J. & Short, H., 2003. Texts into databases: The Evolving Field of New-style Prosopography. *ACH/ALLC conference , Athens, Georgia*, pp.1–14.

Burns, A., Fincham, K. & Taylor, S., 2007. Counting the clergy: the CCEd and the limitations of a prosopographical tool. In K. S. B. Keats-Rohan, ed. *Prosopography approaches and applications. A handbook*. Oxford: P\&G, pp. 257–289.

Crofts, N. et al., 2011. *Definition of the CIDOC Conceptual Reference Model (version 5.0.4)*,

Davis, R., Shrobe, H. & Szolovits, P., 1993. What is a Knowledge Representation? *AI Magazine*, 14(1), pp.17–33.

Doerr, M., 2003. The CIDOC conceptual reference module: an ontological approach to semantic interoperability of metadata. *AI Magazine archive*, 24, pp.75–92.

Doerr, M., Hunter, J. & Lagoze, C., 2003. Towards a Core Ontology for Information Integration. *Journal of Digital Information*, 4.

Gangemi, A., Borgo, S. & Catenacci, C., 2005. *Metokis deliverable D07 - Task Taxonomies for Knowledge Content*,

Gruber, T., 1995. Toward Principles for the Design of Ontologies Used for Knowledge Sharing. *International Journal of Human-Computer Studies - Special issue: the role of formal ontology in the information technology*, 43(5-6).

IFLA, 2007. *Functional Requirements for Authority Data: a conceptual model (draft 2007)*,

Louch, A.R., 1969. History as Narrative. *History and Theory*, 8(1), pp.54–70.

Mailer, N., 2012. *Marilyn: A Biography*, Ebury Publishing.

Martindale, J.R., 1980. *The Prosopography of the Later Roman Empire: Volume 2, AD 395-527*, Cambridge University Press.

Renear, A., 2007. Three of the Four FRBR Group 1 Entity Types are Roles, not Types. In 70th Annual Meeting of the American Society for Information Science and Technology (ASIST).

Veltman, K.H., 2004. Towards a Semantic Web for Culture. *Journal of Digital Information*, 4(4), pp.1–87.

W3C, 2009. OWL 2 Web Ontology Language. *W3C*. Available at: http://www.w3.org/TR/owl2-overview/ [Accessed October 8, 2012].

W3C, 2004. RDF Primer. *W3C*. Available at: http://www.w3.org/TR/2004/REC-rdf-primer-20040210/ [Accessed October 8, 2012].

W3C, 2001. Semantic Web Activity Statement. *W3C*. Available at: http://www.w3.org/2001/sw/Activity [Accessed October 8, 2012].

Notes

1 blog.pbw.cch.kcl.ac.uk; www.pase.ac.uk; www.poms.ac.uk; www.breakingofbritain.ac.uk.

2 For better readibility, we are using here the Turtle serialization of RDF.

3 Note that the :*plre* namespace is purely hyphotetical and introduced here for explanation purposes.

4 For readability purposes, in this and the following figures dashed *isA* arrows indicate that the subsumption hierarchy is partially omitted.

5 If future applications of the factoid model will require a more elaborate formalization of information objects, this could always be obtained by using ad-hoc extensions of CIDOC-CRM (such as FRBR-OO).

6 The ontology URI is http://purl.com/net/feudalism.owl

7 A list is available here: http://www.theclergydatabase.org.uk/jsp/persons/codes.jsp