



King's Research Portal

DOI:
[10.3366/ijhac.2017.0184](https://doi.org/10.3366/ijhac.2017.0184)

Document Version
Peer reviewed version

[Link to publication record in King's Research Portal](#)

Citation for published version (APA):

Bradley, J. D., Rio, A. M. E., Hammond, M. H., & Broun, D. (2019). Exploring a model for the semantics of medieval legal charters. *International Journal of Humanities and Arts Computing*, 13(1-2), 136-154.
<https://doi.org/10.3366/ijhac.2017.0184>

Citing this paper

Please note that where the full-text provided on King's Research Portal is the Author Accepted Manuscript or Post-Print version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version for pagination, volume/issue, and date of publication details. And where the final published version is provided on the Research Portal, if citing you are again advised to check the publisher's website for any subsequent corrections.

General rights

Copyright and moral rights for the publications made accessible in the Research Portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognize and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the Research Portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the Research Portal

Take down policy

If you believe that this document breaches copyright please contact librarypure@kcl.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

Exploring a model for the semantics of medieval legal charters

John Bradley (Department Digital Humanities, King's College London, john.bradley@kcl.ac.uk),

Alice Rio (History, KCL, alice.rio@kcl.ac.uk),

Matthew Hammond, Dauvit Broun (History, University of Glasgow, matthew.hammond |
dauvit.broun@glasgow.ac.uk)

Abstract: This paper describes several aspects of a formal digital semantic model that expresses some issues presented by medieval charters. Surprisingly, perhaps, this model does not deal directly with a charter's text and is not mark-up based. Instead, it draws on the authors' experience with the construction of three highly structured factoid-oriented prosopographical databases that drew heavily on charter sources, and that also did not explicitly contain a digital representation of the charter texts. The paper explains the way in which the structured data model thus derived differs from text-oriented approaches such as TEI/CEI work that has been done so far on charters. It presents a view on why this factoid-based model seems to capture more readily some of the complexity in the apparent meanings of the charters, and suggests that this is because it is also more likely to relate to a richer conception of the broader medieval world in which these charters were created than text-oriented work does. Finally, drawing on recent work on the *ChartEx* project, it explores how a combined approach, that takes the best of both text-markup and structured data modelling techniques, could evolve in the future.

Keywords: prosopography, structured historical data, medieval legal charters, Prosopography of Anglo-Saxon England, People of Medieval Scotland, The Making of Charlemagne's Europe

[This article has been accepted for publication by Edinburgh University Press in the Journal *International Journal of Humanities and Arts Computing* (<https://www.eupublishing.com/loi/ijhac>). Volume 13, No. 1-2. pp. 136-154]

Legal charters are an important class of documents historically because they can offer a particular window into a society's working that otherwise simply is no longer available to us. This significance of charters to medieval history, plus the traditional difficulty of accessing them, has meant that there has been a great deal of activity aimed at producing scholarly editions of their text; at first in print at least as far back as Victorian times, and more recently online. Indeed, the potential of the internet to make charter documents readily available for scholars has been taken up significantly by the *monasterium.net* initiative¹ which has developed an informal collaboration between institutions to support the online publishing of documents held by European monastic archives. The scale of work is evident by the fact that the 125,000 documents available in 2009² has since grown to over 400,000 (according to personal discussion with Georg Vogeler in 2014). As their website claims, *Monasterium's* users are thus set free from a dependence on time and space to study these documents because they are available to everyone at any time.

Charter documents generally were created for the very practical reason of getting something done in their society, and they exhibit textual formalisms that reflect this. Thus, when thinking about how these structures can be made manifest, an obvious approach is to apply textual markup to reveal these formalisms, and to extend the Text Encoding Initiative's (TEI) tagset to express things

the charter texts are about. Thus, we find the formation in April 2004 of a *Charters Encoding Initiative*³. The CEI scheme incorporates the standard TEI manuscript markup elements such as <expan>, and (closer to our interests in this paper) encourages the use of person and place tags such as <persName> or <placeName> to formally identify the name of a person or place in the text. CEI adds further formal abstractions that are relevant to charters such as <issuer>, <recipient>, <scribe> and <addressee>, <issuePlace> and <issueDate>, and there is tagging available that presents transmission information about the charter. As Georg Vogeler wrote in the early days of the initiative, these documents ‘reflect[ed] contemporary attitudes and mindsets as regards legal and representation issues and [...] are tools of diplomatic criticism’⁴. By marking up these texts using the CEI conventions, he believed that one created ‘a platform for seeing the European Middle Ages as they are reflected in their charters.’⁵

Structured Data (Knowledge Representation) for Charters

The authors of this paper have also been involved in projects to make charter materials available to the public and scholarly communities over the WWW. In contrast to *Monasterium* and CEI’s markup approach, however, our projects represent the charter materials in the form not of charter text with markup, but of highly structured data of the kind characteristic of databases or the Semantic Web. We contend here that, as a result, the representation of charters in these projects gives a quite different sense of what is being said about the charters and the historical context in which they exist. What is the nature of this difference? Some of it was observed by one of us (Bradley) in a presentation about the place of structured data in history at the University of Lisbon in 2011 and subsequently published in this journal⁶.

At this workshop Bradley considered what would happen if one had a set of catalogues from exhibitions of photographic prints and wanted to produce a digital resource from them. If one used markup as a way to formally represent the structure, one would produce a something that was clearly a representation of the catalogues. If, instead, one took a highly structured approach, the process that one would go through to produce a model of the material in these catalogues would be more like a representation of these prints ‘in the world’ as it were – in the world of photographers, archives and photographs, and so on. The structured data *derived* from these catalogues causes a shift in focus away from the collection of images on the pages in the catalogue to a representation of the world in which these pictures exist, and ‘although the tagging is truer to the book as an object, the database is closer to the way we deal in our heads with the information it contains.’⁷ Bradley claimed that ‘the better the model we use to hold our material match[es] our understanding of this world, the more useful this representation becomes.’⁸ This phenomenon, which involves the creation of digital surrogates for entities in the world, is not a new idea. See, for example, Davis, Shrobe and Szolovits’s observations in their highly influential 1993 article ‘What is Knowledge Representation?’⁹ (KR) On the first page they lay out five characteristics of KR: the first one characterises it as ‘most fundamentally a *surrogate*, a substitute for the thing itself’, and fifth as ‘a medium of human expression, that is, a language in which we say things about the world.’¹⁰ What happens, then, if instead of looking at the charter *texts* as the primary objects to model, we try to represent something of what the charters were representing in the medieval world? What happens when one conceives the material the charters present in a ‘world’ context rather than a ‘textual’ one?

One worry must be that Knowledge Representation is, by its very nature, a rather reductionist activity, and a reductionist representation might seem to jar with the medievalist’s subtle understanding of medieval society. Furthermore, our view of medieval society is, of necessity,

limited to what materials have survived, and what has survived has fitted with purposes of those institutions or families that preserved them for hundreds of years after they were originally created. Furthermore, some have dubious provenance, or are known to be fakes. In such a complex situation, how can structured data, such as one finds in databases or semantic web technologies, adequately represent the materials of interest – what is this ‘world’ that this representation can express?

It is true that the structuring of this material in our projects was to some degree reductionist and we only represent some of what is there to be seen in our charters. After all, as Willard McCarty reminds us while talking about the act of modelling in the humanities, ‘[t]o render a cultural artefact intellectually tractable, we must ignore some aspects of it and highlight others [...]’¹¹. Our surrogate representation, then, does not claim to accurately capture ‘facts’ about a ‘real’ medieval world. By creating a digital model we of necessity simplified the full subtlety of our understanding of medieval times and the process of transmission of the documents to us. Even then we found that it represented aspects of the charters that fit with our projects’ interests and that were best tractable to KR methods. Our model, then, represents a complex blend of entities that apparently existed in the medieval world with our modern understanding of these entities, mixed together with aspects of how these charters represents views that support the institutions that held them. In the end, we aimed to capture some of what the charters *claim* to represent – including sometimes even highly questionable claims – rather than necessarily what really happened in the ‘real’ medieval world. The ‘world’ we represent in our KR model is drawn from this complex blend of creation in medieval times, transmission through to the present, and our modern-day scholar’s interpretation.

One approach in the projects this paper talks about was to think in terms of the *state of affairs* that these documents actually present to us. The pragmatic purpose of the charters, from the point of view of those who preserved them, was evidently to represent the state of affairs that applied to the things in their world, for example. rights over property,¹² and were a means to an end for these individuals to achieve their goals. Thus, the charters themselves, as objects-in-the-world, also had to be in the model because they provided the evidence of what this state of affairs was, and at least some sense of what the people involved in the creation of these charters, and then their preservation, were aiming to achieve.

We get close to our projects’ intellectual territories by looking at a project with somewhat similar aims: the work of Michael Gervers and others at the University of Toronto, in the *Documents of Essex England Data Set* (DEEDS). As Gervers *et al* remind us in a short article about the project, the DEEDS project had ‘as its objective to provide computerized access to the content of twelfth- and thirteenth-century conveyances concerning the county of Essex, England.’¹³ The DEEDS project is described in the article as a database that presents in a structured, formal manner the ‘often complex patterns of property holding and transmission’ which ‘reflect the exchange of layers and rights and obligations’¹⁴ that were evidently the material of interest in these medieval charters. We are then presented with the set of entities that came out of Gervers’s analysis: Persons, documents, property, lease on the property, roles for the persons, and relationships between persons captures much of what the DEEDS structure is about. The connections between these entities represent some of the meaning that is captured with the transactions described by the charters. Although the DEEDS project has more recently become much more text centred, in 1990 the DEEDS data apparently did not include the actual charter texts in their database. Instead, the actual text had only a kind of indirect presence, and an interpretation of the charter and what it was doing for the people in Essex medieval society was what is being captured.

Charters in Structured Prosopography projects

This paper arises from the work of three projects¹⁵: The *Prosopography of Anglo-Saxon England* project (PASE)¹⁶, the *People of Medieval Scotland Project* (PoMS)¹⁷, and the *Making of Charlemagne's Europe*¹⁸. PASE (Anglo-Saxon Prosopography) began a little before 2000, and continued in several guises through several phases of work up to only a few years ago. PoMS (People of Medieval Scotland) began in 2007 (Hammond's book chapter provides an introduction to it¹⁹), and essentially finished (after two funded phases) a couple of years ago. The *Making of Charlemagne's Europe* (called here *Charlemagne*) was first made publically available in 2015.

The prominence of charters²⁰ increased in each of these projects: 1,445 of PASE's 2,784 documents were charters, and in PASE we began to think about how charters affected the structure the project was already using for its non-charter sources. In contrast, PoMS (about 90% of its 9,261 documents were charters) and *Charlemagne* (about 4,500 charters were expected) both focus almost exclusively on charter sources. Furthermore, whereas PASE is a structured prosopography covering a broad range of kinds of sources, PoMS and *Charlemagne*, with their strong charter orientation, developed a structural approach that is strongly charter centred. All three projects were prosopographies, and prosopography is an activity that by its very nature connects texts to 'objects-in-the-world' outside the charter texts: the historical people. However, all three projects found that as a consequence of developing a formal structure for prosopography, other 'objects-in-the-world' also entered into their thinking. The models for these three projects, then, attempt to achieve a balance between the texts of the charters, and their effect on their society in-the-world. This balance was not the same for the three projects because the interests of the history partners differed, but in all three we found ourselves thinking about entities belonging to the world outside but connected to the charters as well as entities representing aspects of the text of the charters themselves.

Some part of the way to balance the needs of the text with the needs of representing the larger world grew out of the use of the 'factoid' model which was developed by DDH, and described in Bradley and Short's initial article in 2005²¹, and more recently in Bradley and Pasin in 2013²² and again (with an exploration of possible connections between it and the CIDOC-CRM²³) in Pasin and Bradley's article from 2015²⁴. Perhaps a central idea can be found in the recognition of a 'factoid entity' that represents the assertions that the sources are making. This entity is called a *factoid* because one needs to keep in mind that it represents what a document asserts, as best modern historians can establish this, rather than what is necessarily 'true' or 'factual'. Structurally, the point of the factoid as a 'source assertion' is that it represents a *nexus* between something in an historical source, some points or periods in time, a group of one or more people, some geographic places and possibly some possessions, plus various other kinds of assertions made about these historical people, such as offices they held. What is particularly useful about the factoid model for the work here is that it recognises that not only any number of assertions, but also that quite different kinds of assertions, can come from any particular textual source.

The first recognition of this in the context of charters arose from the analysis for PASE. PASE's charters generated three different groups of factoids. The first group categorised people, and recorded offices or occupations or relationships mentioned in the charters. Sometimes something surprising would turn up. In charter *Sawyer 19*, for example, the king *Wihtrud 1* is explicitly identified as being illiterate. This still leaves us with the main business of the charters, which was captured in one or more *transaction* factoids in PASE – the second group of the three. For *Sawyer 19*, for example, PASE has a single transaction factoid that recognises the grant sometime between

697 and 712 of '4 sulungs at Pleghelmestun in Kent' from the king to the king's Royal Minster at Lyminge. We see here the characteristic 'nexus' character of factoids in operation: a date (here date range), people, places and possessions are all brought together by it. This factoid captures the 'business' (as it were) of the charter itself. As we shall see shortly, sometimes the exchanges in a charter are more complex than this, and we shall see how the factoid approach helps to handle this complexity.

What was interesting about *PASE*'s particular approach to charter factoidization was that it separated the business (transactions) from the act of the creation of the charter itself – the third of three groups of factoids associated with charters in *PASE*. *PASE* characterised this as a *charter witnessing* event. The participants in the witnessing of the charters are asserted, and the place where the charter was signed is also attached. By using factoids to separate the transaction that the charter is thought of as being 'about' from the event of the charter signing, the project was better able to capture the different-but-linked nature of the three kinds of assertions made here: one, prosopographical information about people mentioned in the charters; two, that some property was transferred, and three, that a socially-oriented event occurred involving a group of people who the charter says were brought together to witness it.

PoMS and *Charlemagne* were much more focused than *PASE* on charters as exclusive or almost exclusive sources. For both projects, however, factoids about the people themselves, for example, titles, occupations and personal relationships that come out of the charter texts, were still recorded. Also like *PASE*, both projects recognised a transaction factoid as capturing the central act or acts in legal charters. *Charlemagne* added a *place relationship* factoid that did not involve persons at all, but recorded statements about relationships between places, since these relationships were also often unclear and even contradictory.

Many of the charter documents are conceptually quite simple, and involved only one action or transaction. Documents called *Brieves*²⁵ in *PoMS* were often of this kind: for example, where the king forbids anyone from, say, disturbing monks while they are taking wood as fuel from his land. Other charter types in *PoMS* might be more complex, but would still involve only one transaction: a gift of one or more possessions, for example. Here one might find a larger range of people performing different roles: not only a grantor and beneficiary, but often also a person consenting to the gift, the 'pro anima' people (those people for who the grantor claims spiritual benefit from the gift), and the set of witnesses who, as it were, provide 'back up' support for it.

However, many charters could not be properly characterised by a single transaction, but would contain several interconnected ones. Here the factoid approach meant that multiple transactions with complex interconnections between them could be expressed. Matthew Hammond explored some of this kind of complexity in his presentation about *PoMS* for the *Institute for Historical Research* in London²⁶. The *PoMS* team found it useful to establish a classification scheme that they used for different kinds of charters that they found. Figure 1, derived from his slides for that occasion, represents one of the common types of charter that they found; an 'agreement charter' (perhaps a *chirograph*: a document that is split into two so that both parties have a part to hold onto as proof of the existence of the agreement).

Agreement (chirograph)

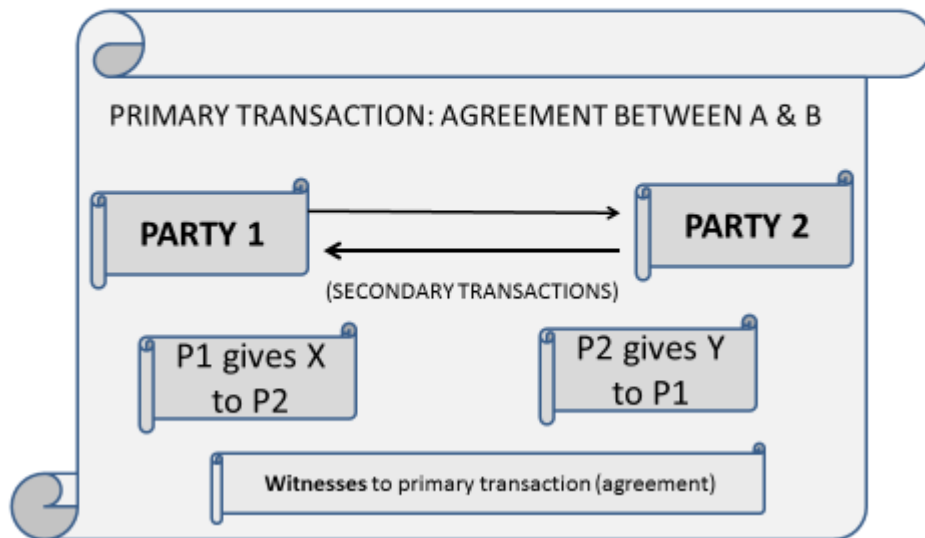


Figure 1: An Agreement charter in *PoMS*

PoMS was able to describe agreements in terms of three related transactions: the agreement between two parties (with the witnesses being linked to the agreement), and the two parts of the agreement: party one gives items to party two, and party two gives items to party one. In the *PoMS*'s structure one was able to identify the three transactions as interconnected factoids, and furthermore say that the agreement transaction was a primary transaction and the other two transactions (the two reciprocal exchanges of property) were related, but secondary, transactions.

Renewal

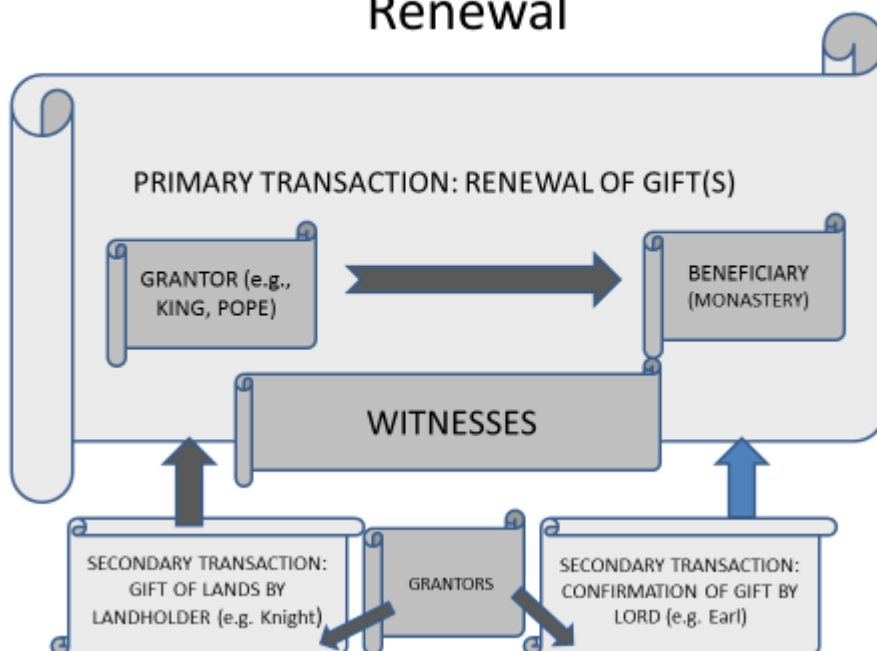


Figure 2: A Renewal charter in *PoMS*

Charters that *PoMS* identified as ‘Renewal charters’ also exploited this characteristic of factoids for multiple tractions, but were somewhat more complex than agreements. Figure 2 (also drawn from Hammond’s 2013 presentation) provides us with a schematic representation where a King or perhaps a Pope renews arrangements that were made under a previous ruler. Like the agreement model just presented, there are three transaction factoids involved but the interaction between them is different. The primary transaction is the renewal, and it still has a grantor and beneficiary attached to it. The witnesses have a role in the renewal itself. However, here the other two other transactions are *previous* grants that are being renewed. Thus, they perhaps involved other people as grantors (although, presumably the same beneficiary). Furthermore, and unlike the agreement, the dates for the various secondary transactions are also going to be different from the renewal itself. Since date or date range is attached to each specific factoid, this is also readily accommodated. Each of these secondary transaction factoids provides, on its own, a separate, but connected, nexus between people, places and property.

One of the important phenomena that arose in the development of the factoid model for the *Charlemagne* project (and, to a lesser extent *PoMS*) was that the project work became proportionally less about prosopography and more about modelling the charters. Although in *Charlemagne* prosopography never entirely went away, more and more of the data structuring began to represent other aspects of what the charters were about and less about the people who were mentioned in them. Since the charters were created to reflect what was, to the people of their day, things that had happened in their world, more and more of the structure constructed for *PoMS* and *Charlemagne* began to represent entities that acted as surrogates for things in these two medieval worlds, and fewer (proportionally, at least) were about the text of the sources themselves. This division between data about the charter texts and data about the medieval world is not fully black and white. However the next section presents an overview of *Charlemagne's* structure in this light, and attempts to explore a little about its significance.

A data model for *Charlemagne*

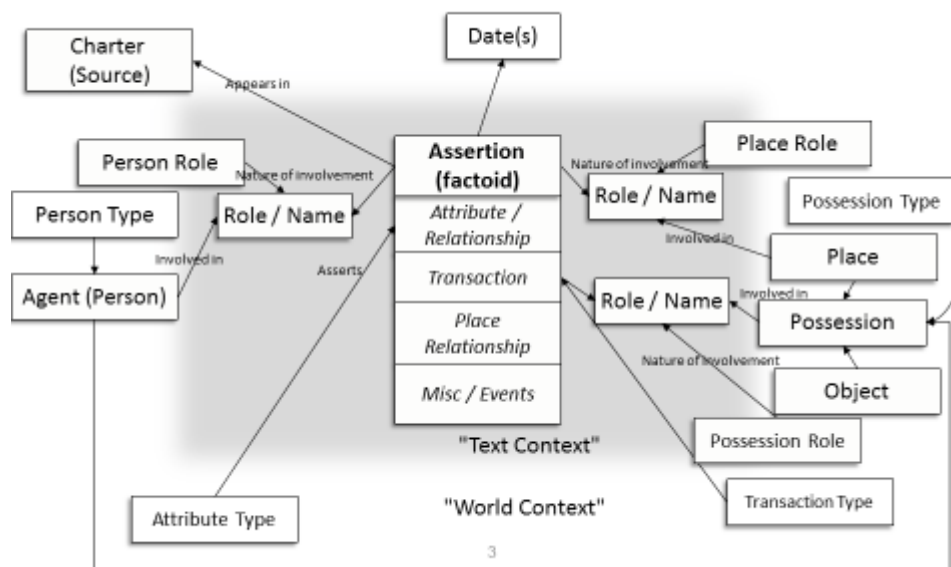


Figure 3: A data model for *Charlemagne* (simplified)

Figure 3 shows a somewhat simplified representation of the *Charlemagne* database's major entities and the connections between them. Before we examine this diagram in more detail it is worthwhile explaining briefly how it should be read. First, although figures 1 and 2 are schematic representations as figure 3 is, the significance of the components is different for figure 3 from the other two in that whereas the diagrams in figures 1 and 2 showed specific entities for two particular kind of charters, the figure 3 diagram describes aspects of the Charlemagne database structure overall. In figure 3 the boxes represent *kind of entities* that the database contains. So, noting the box left centre in the diagram we can see that the database has entities which are called *Agents* or *Persons*. Similarly, to the right we see a box called 'Possession'; meaning that the database has entities called *possessions*. Note, then, that each box does not represent just a single instance of, say, a person or a possession, but a *class* of persons or places – each class representing perhaps hundreds or thousands of particular instances. The lines that connect the boxes, sometimes labelled in figure 3, show that there is a connection between the two entities. Thus, the line between the Charter entity and the Assertion (factoid) entity means that there are connections between individual charters and individual assertions/factoids.

Now that we have briefly introduced how to interpret figure 3, we can begin to examine it in more detail. First, note the grey area labelled 'Text Context' in the middle of the figure. The entities in this area represent material that is closely related to particular charter texts. The objects around this central gray area (labelled 'World Context') are entities that could be argued to exist in a medieval world view and/or our modern understanding of that medieval world independently of their references in the charter texts.

One can see a good number of entity types in Figure 3, and the full formal structure for *Charlemagne* is, in fact, still more complex than this. However, one can see several obvious historical entities. *Agents* (*Charlemagne's* name for persons) are historical entities that one could view as having an existence outside of the charters themselves. Similarly, places, and possessions can arguably be usefully thought of as having an independent historical existence. The *Place* entity represents geographic places in our data, and so of course also exists in the 'world context'. About half of the places in *Charlemagne* charters have known geographic coordinates, and well more than half can at least be located in larger modern geographic regions such as modern day region and country. Furthermore, the charters themselves, as physical document-objects, also exist as historical entities outside of their texts.

Although these entities have a physicality that makes it relatively straightforward to see them as objects with an existence 'in the world', one can place other entities there too. Several of these are entities that, although without a physical character, act within a societal context which also has an existence in our thinking about the world.

- 'Attribute Type', for example, is connected to people through *Charlemagne's* Attribute/Relationship factoid and is the place where ideas such as 'Duke', 'Lord', 'King' where various kinds of relationship ('Son', 'Mother') are identified. These ideas of how individuals are organised in society, although without a physical character, have an existence outside of the charter texts themselves.
- Similarly, for *Charlemagne's* 'Person Type'. Each person/agent is assigned a type. For human individuals this is their sex, and for 'legal persons' that are groups of persons the type reflects something of how historians believe their society categorised them: monasteries as Female or Male institutions, for example.

- Possession types offer categories like ‘Land’, ‘Goods’, ‘Animals’, ‘Money’, and ‘Rights’. Like types for persons, they provide a way that our historians could organise the large range of possessions into categories that are meant to represent one aspect of how they were thought of in medieval society, and thus they had a sociological existence outside of the charters themselves.
- Finally, there are dates. Dates are attached to factoids, and are – as one would expect for historical dating from medieval times – a more complex structure than the simple box in Figure III suggests. We cannot discuss them here (although their structure is similar to the TEI date tags²⁷ in conception) – but dates also exist in the medieval world outside of any particular charter text.

These ‘in the world’ concepts are important because they allow the users of these prosopographical databases to group material by the societal structures they represent: ‘find me all charters that involve animal possessions and female institutions’, for example.

With so many entities having an existence in the medieval world, what ones are in fact specific to the actual charter textual? Of course the factoids, those assertions that the sources make, are closely related to the texts from which they emerge, and one can see in Figure 3, under the Assertion/factoid box, the four kinds of factoids that are associated with *Charlemagne*. Although they *connect* material in the text to the world entities of people, places, dates, and so on, they have an existence only in the context of the text of a particular charter.

So, what then is the nature of each connection between the factoids and these world objects? First, notice the existence of an entity in Figure 3, rather generically called ‘Role/name’, that sits between a factoid and a person. This entity captures the way in which a particular person is involved in any particular factoid.

- First, there is the role of the person in the particular factoid. For example, in a transaction a person can have the role of Grantor, Recipient, Witness, Spiritual Beneficiary, and many more.
- Also, this particular entity provides a place where the text that the source at this particular spot uses to identify the person can be recorded.

In a way similar to the one for Persons, then, the Place Role/Name entity between factoid and Place provides a place where a role (for example, *location of the transaction*, *location of a possession*, or the *location of residence*) for a particular place in a particular transaction can be recorded. Also, it provides a place where the place’s name, as actually found in the source, can be recorded.

The possession structures are somewhat more complex and reflect the more complex nature of possessions in our data.

- Possessions can be typed into categories such as land, money, goods, buildings, olive trees. *Charlemagne* is developing a rather rich hierarchical (and therefore thesaurus-like) classification scheme for the possessions. Some are non-physical in nature – such as what *Charlemagne* (and *PoMS* before it) called ‘Spiritual Benefits’: prayers for someone’s soul.
- Most possessions are of three broad kinds: land, valuable objects and persons. Possessions which are land also therefore have connections to places and will be linked to an instance of the Place entity. Possessions which are people – unfree individuals – will be linked to an instance of the Agent/Person entity.
- In the case of valuable objects as possessions, most of them are transient objects that as physical entities actually appear only in a single particular charter, and they are often quantity-based or collective objects – a sum of money for example, or a herd of horses.

Charlemagne's possession structure provides a place to describe these kind of possessions and to categorize them, but doesn't require the creation of a unique object.

- Other objects are specific things that might appear in more than one charter – relics for example. Here an instance of an Object is created, and in this way more than one factoid can refer to it.

As in the other objects, each one can be given a role in the transaction in which it appears: as the object being transferred, as a basis for equivalent value, as a price, and so on.

The structure just presented, although in fact a rather simplified representation of the actual structure of the *Charlemagne* data, is accurate and complex enough to give one a sense of the ways in which it can represent rich data about the objects with which the *Charlemagne* charters concern themselves with. The complex place of people in *Charlemagne* is a good example of this. Not only can *Charlemagne*'s 'people' be human beings here, with the normal division into Male and Female, but can also represent entities that apparently act like persons in these documents: church institutions such as abbeys, monasteries, and so on., and *Charlemagne* even allows for the distinction between 'male' and 'female' institutions to be expressed since, as our historian colleagues tell us, this was an important aspect of the thinking of medieval people about their church institutions. A main focus here is people and their role in the transactions as agents, but the model also allows us to record personal relationships between individuals, and offices, titles, and other attributes that they hold and develop during their lives. Furthermore, this particular approach allows persons not only to be actors or agents in transactions, but as objects to be transacted as well.

Fusing Structured Data with Markup for Charters?

We have presented here a somewhat simplified representation of the structure of our *Charlemagne* dataset. We have a similar (although not the same) structure for the *PoMS* project as well, where, like *Charlemagne*, we could categorise those entities that we captured in its DB structure quite richly.

What does this highly structured interpretive model of what charters talk about bring to the act of charter interpretation that markup, such as the TEI-based CEI markup, does not? CEI does indeed provide tags to identify some things that are similar to what we are capturing in our *PoMS* or *Charlemagne* data structures. Examples of CEI markup often contain tags that, for instance, identify the name of a person as the issuer, a place as property involved in a transaction, or a date and place of issue of a charter. And, indeed, it is true that textual markup *could* provide a vehicle for representing data similar to what we are building in our highly structured *Charlemagne* and *PoMS* data sets. However, the 'markup' approach, which tends to focus the analysis on the text of the charters also tends to encourage one to only develop a formal structure that is 'close' to the charter text – identifying materials that belong in the area that we show in figure 3 as the 'textual context'. So, for instance, CEI provides a tag to identify a *reference* to the person who is the issuer (the kind of material we have identified as textual context in figure 3), it does not actually formally *identify the person* (world context) being referenced. Indeed, in all the CIL examples we have examined, the reference to the person was as far as the tagging went. We saw no mechanisms in these CIL examples that showed how one could give these referenced individuals their own identity with attributes such as whether, for example, each person was male or female. The same thing happened in the tagging of places or dates: the spots in the text that make *reference* to a place or a date might be identified by CEI markup but they are not then turned into an interpretive representation of that date or the place. Indeed the 'world context' structures that our entities are

able to represent (as identified in figure 3) are not in fact recognised as having an independent place in the CEI markup approach. Substantial new XML/TEI structures outside of the charter text could have been added to the markup to handle them, but this does not appear to be done. Perhaps this is because the whole approach of looking at getting structure out of text in terms of markup simply does not encourage the markup designer to think that ‘far away’ from the texts. It is not that a markup approach could not accommodate the representation of ‘world structures’ that are more tangentially represented by the actual charter text; it is more that markup does not encourage one to think in that way.

Here, then, is the issue of how the *approach* to structuring affects how we think about what we are structuring – similar to what we described earlier between the markup of a print catalogue, and the creation of structured data to represent its contents. By designing database structures for holding data derived *from* the charters the *PASE*, *PoMS* and *Charlemagne* teams were encouraged to think about the objects-in-the-world that these charter texts claim to represent. A markup project based on charters thus would have been a significantly different thing than what *PoMS* or *Charlemagne* ended up being about.

So, is this paper building a case that one is actually better off if one does not include the actual charter text at all in our representation? By no means. Our earlier projects had to be built without the texts because they were, in general, not available in digital form – indeed, most of the texts from Scottish Medieval charters are still not available online, although one of the authors of this paper (Broun) has been working on a project which will be making some of them available²⁸, and which will be linked to *PoMS*. A good number of *Charlemagne*’s texts are already available over the WWW, so we made provision there to store a hyperlink from our charter data to an online text when the project team knew of one that was available. This link, however is at the level of the entire document. Surely it would be better if all our ‘world entities’, persons, places, and so on., could have been more intimately linked with the references to them in the charter texts. To achieve this would require a more complex structure than the charter-text-plus-markup that CEI provides by itself, since the world entities exist outside of any particular piece of charter text. What, then, could happen if we tried to do this: to directly connect the texts through markup or something like it with a representation of the semantics of what the charter is about? It turns out that this has been attempted in the *ChartEx*²⁹ project.

ChartEx was funded out of the *Digging into Data* initiative³⁰ and was thus primarily focused on the challenge of applying big data techniques to digital text representations of charters. One of the big data techniques *ChartEx* used was a Natural Language Processing (NLP) mechanism called ‘entity recognition’. Entity Recognition provides techniques that allow the computer to automatically identify references to things like places, people and events in plain digital texts. So, *ChartEx* explored how these automatic *entity recognition* strategies could be used to locate references to entities like people or places in these medieval charter texts – definitely a useful venture in situations where perhaps the number of charters is too large for human manual processing. Of course, as we hope it is clear by now, entity recognition alone is only a part of what our charter data model is about – not only are references to historical entities identified in *PoMS*, *Charlemagne*, and *PASE*, these entities are also connected to structural elements that identify their roles and functions in the events the sources are talking about.

In a similar vein, a 2013 presentation³¹ about *ChartEx* showed that it had worked to go beyond merely entity recognition to try to automatically locate additional structure in these charter texts, and to put the entities it found into that structure. Indeed, one sees a BRAT³² representation

in a *ChartEx* presentation that shows of the semantic structure of a charter (p. 17) and that echoes many of the concerns we have also been talking about. Persons or agents, transactions with their types, property and place, as well as a source are all present here (page 15), and linked together with predicates that establish the relationships between them that are very similar to what we have been recording in our projects too. The BRAT representation shows the information extracted from a charter text (or, at least, a modern language rendition of it) as a hierarchy imposed over top of it.

The potential of the software they have used to detect and tag this information semi-automatically is impressive, but one must add a couple of caveats. First, the text shown there and in other examples *ChartEx* shows in its presentations is a modern English rendering of the charter, and indeed modern English is the language most effectively supported by existing entity recognition and other NLP software. One would expect that any automatic extraction would work less well or fail altogether if the text was, say, in medieval document Latin. Second, the structures that are shown in *ChartEx* examples are seemingly rather simple whereas, as we have illustrated above, many of the *PoMS* and *Charlemagne* charters exhibit a truly complex multi-transaction structure. Finally, the task of personification (turning the appearance of a person's name into a pointer to a record representing the corresponding historical person), and the parallel activity of identifying land, is not actually shown in the examples we have found, although it apparently was undertaken by *ChartEx*. Thus, this particular parsing process which identifies things in the charter's text, as impressive as it is, still might well leave one with the task of subsequently connecting the things found here to the historical world of people and places. Furthermore, the BRAT notation itself, as wonderful as it to show structures within the text, by focusing on the text, might suffer a bit from the distinction between markup and its expressive nature and structured data that was just described in this paper.

Nonetheless, the connection between the text and the structure is pretty clear here. Perhaps *ChartEx* and our factoid-based *PoMS* and *Charlemagne* projects would benefit by a sharing of their overlapping, but somewhat different, insights into the nature of charters.

Summary and Future Work

In this article we have made the argument that a structured data approach to the representation of materials from medieval charters encourages a view of this material that more readily incorporates entities arising from a historical sense of the medieval world than charter text markup generally does by itself. We believe that through the use of formal structure we show a way towards a perspective on charters that moves the focus from a charter's actual text to an *historical interpretation* of what that charter was doing in its society. Our Factoid approach, which allows multiple things of different kinds to be asserted from charters, provides a rich basis for supporting the complex set of things often going on in the charter text. Indeed, projects like *Chartex*, which aim to use semi-automatic NLP techniques to draw semantic structure from the text, generate models for the data they create that are similar to those we have developed, and an integrated environment that combines highly structured data, of the kind we used in our projects, with links into the charter text would surely provide the best and richest result.

Two directions for further work in this area suggest themselves. First, the model shown in Figure 3 blends assertions from the sources with entities that represent a view of the state of affairs in their medieval societies, and presents a perspective on what is going on based directly on charter assertions. The model could, in fact, be taken instead further towards a 'state of affairs' representation by enriching the representation of ownership over time. A transaction would, in this model, have at least two ownership components: before and after the transaction, and the data

could be looked at perhaps more clearly as a representation of how people in the medieval society viewed what the charters were doing for or to them.

A second, related, initiative could be to explore the development of a formal ontology for charters, similar in spirit to CEI, but growing out of the structured data context of the semantic web. Such an ontology would perhaps share elements with one of the authors' proposal for an *Ontology for Historical Persons*³³, but would expand upon a blending of conventional diplomatic interpretations with the state of affairs perspective explored here.

Medieval charters provide a rich source of information about the societies in which they were created. We can be sure that a sophisticated structured data approach representing what they are about will enable some new ways of exploring, and understanding, those medieval societies.

Acknowledgements

The project work that supported the research in this paper was funded by the UK's *Arts and Humanities Research Council*, and by the *Leverhulme Foundation*. This paper grew out of a preliminary version given at the DH2014 conference in Lausanne Switzerland by one of the authors.

¹ 'Monasterium.net', accessed 6 January, 2016, <http://www.monasterium.net/>, last accessed 14 January 2016.

² G. Vogeler, 'Charters Encoding Initiative overview', *Digital Proceedings of the Lawrence J. Schoenberg Symposium on Manuscript Studies in the Digital Age*: Vol. 2, 1 (2009), Article 8. <http://repository.upenn.edu/ljsproceedings/vol2/iss1/8/>, last accessed 14 January 2016, Cited here at slide 9.

³ 'CEI - Charters Encoding Initiative', accessed 6 January, 2016, <http://www.cei.lmu.de/index.php>, last accessed 14 January 2016.

⁴ G. Vogeler, 'Towards a standard of encoding medieval charters with XML', *Literary and Linguistic Computing* 20, 3 (2005), 269-280. Cited here at 276.

⁵ G. Vogeler, 'Towards a standard for encoding medieval charters with XML', 279.

⁶ J. Bradley, 'Silk purses and sow's Ears: can structured data deal with historical sources?', *International Journal of Humanities and Arts Computing*. 8,1, (2014), 13-27, doi: 10.3366/ijhac.2014.0117.

⁷ J. Bradley, 'Silk purses and sow's Ears', 19.

⁸ J. Bradley, 'Silk purses and sow's Ears', 20.

⁹ Knowledge Representation (KR) has, as a part of its conception, the representation of information as highly structured data.

¹⁰ R. Davis, H. Shrobe and P. Szolovits, (1993). What is a Knowledge Representation? *AI Magazine* 14, 1 (1993), 17-33. Cited at 17.

¹¹ W. McCarty, 'What's going on?', *Literary and Linguistic Computing* 23, 3 (2008), 253-61. Cited here at 254.

¹² Of course, the aim even of forged charters is the same! Our projects therefore often took evidently forged documents as part of the canon of charters they were interested in, although the system provided ways for the team to indicate that they thought they were forgeries.

¹³ M. Gervers, G. Long, G. and M. McCulloch, 'The DEEDS Database of Mediaeval Charters: Design and Coding for the RDBMS Oracle 5', *History & Computing*, 2, 1 (1990), Cited here at 1.

¹⁴ M. Gervers *et al*, 'The DEEDS Database'.

¹⁵ all three undertaken by the Department of Digital Humanities (DDH) at King's College London with historian partners from the University of Cambridge, Glasgow, and King's.

¹⁶ "PASE: Prosopography of Anglo-Saxon England", <http://www.pase.ac.uk>, last accessed 14 January 2016.

¹⁷ 'People of Medieval Scotland: 1093-1314', <http://www.poms.ac.uk>, last accessed 14 January 2016.

¹⁸ 'The making of Charlemagne's Europe', <http://www.charlemagneurope.ac.uk/>, last accessed 14 January 2016.

¹⁹ M. Hammond, 'Introduction: The paradox of medieval Scotland, 1096-1286' in *New Perspectives on Medieval Scotland, 1093-1286* (the Boydell Press, 2013), edited by H. Matthew, 1-52.

²⁰ The meaning of the word *charter* in *Charlemagne* and *PoMS* included any kind of legal document that disposed specific rights over property – thus including some things that are not technically charters such as royal diplomas or testaments.

²¹ J. Bradley and H. Short, 'Texts into databases: the Evolving Field of New-style Prosopography', *Literary and Linguistic Computing* 20, Suppl. 1, (2005), 3-24.

²² J. Bradley and M. Pasin (2013). 'Structuring that which cannot be structured: A role of formal models in representing aspects of Medieval Scotland,' in *New Perspectives on Medieval Scotland, 1093-1286*, edited by H. Matthew (the Boydell Press, 2013), 203-214 (2013).

²³ 'The CIDOC Conceptual Reference model', <http://www.cidoc-crm.org/>

²⁴ M. Pasin and J. Bradley, 'Factoid-based prosopography and computer ontologies: Towards an integrated approach,' *Digital Scholarship in the Humanities*. 30, 1 (2015), 86-97, doi:10.1093/lc/fqt037.

²⁵ The online *Law dictionary* tells us that a *Brieve* is the name for a writ in Scotch law. <http://thelawdictionary.org/brieve/>, last accessed 14 January 2016.

²⁶ M. Hammond, 'The People of Medieval Scotland database' (paper presented at the *Institute for Historical Research* (London) Digital Series, 9 May, 2013).

²⁷ Text Encoding Initiative, *P5: Guidelines for Electronic Text Encoding and Interchange* section "Names, Dates, People, and Places", sub-section 13.1.2. <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ND.html>, last accessed 3 August 2016.

²⁸ *Models of Authority: Scottish Charters and the Emergence of Government 1100-1250*. Project materials available online at <http://www.modelsofauthority.ac.uk/>.

²⁹ 'ChartEx', <http://www.chartex.org/>, last accessed 25 June 2016.

³⁰ The *Digging into Data* initiative is a research granting initiative supported by a number of national research funding bodies. According to its website, its purpose is 'address how "big data" changes the research landscape for the humanities and social sciences'. See <http://diggingintodata.org/about>, last accessed 25 June 2016.

³¹ R. Sutherland-Harris, and R. Evans, 'People, places and events in charters: exploring the language of charters within ChartEx' (paper presented at the *Digital Diplomatics* conference in Paris November 2013). Slides available at <http://www.chartex.org/docs/Chartex-Paris-14112013.pdf>, last accessed 14 January 2016.

³² 'brat rapid annotation tool', <http://brat.nlplab.org/>

³³ J. Bradley, (2013), 'Towards an Ontology for Historical Persons' (paper presented at *Culturecloud, Co-reference, Archive workshop*. Swedish National Archives (Riksarkivet), Stockholm, Sweden, 4 June, 2013). Slides available at <http://www.slideshare.net/johnBradley/towards-an-ontology-for-historical-persons>, last accessed 14 January 2016.