DIGITAL CLASSICS ONLINE

Building a Digital Prosopography: Towards a Scholarly / Technical Partnership

John Bradley

Abstract: The paper explores how digital resources can be successfully built for historical projects. It focuses on projects that use highly structured data as the paradigm for representation, and it explores the basis for a collaboration between technically-oriented partners who understand and can work within the formalisms of highly structured digital data and historians who must ensure that the formal structures that emerge properly represent the historical issues. The author has experience with over 20 such projects and believes that a rich, understanding of the model for the structured data that is shared between technical staff and historians is the key component in achieving success. Through looking at examples from these projects the paper shows how this shared understanding can positively affect several aspects of the project work and the resulting digital resource.

Introduction

I recently came across an article (which must remain anonymous here) written by historians who had obviously not enjoyed, or even benefited from, their development of a digital historical resource. In the article it became evident that the work had presented the historians with a number of serious academic problems that were made worse by their sense that they were not properly in control of their project. Clearly, they had problems working with the part of the project they called the 'computer engineers' who were meant to be working with them to transform their ideas into a digital form. The historians felt that they were not properly in control of the project and that, as a consequence, their historical concerns were not being addressed.

There is indeed a challenge in creating academic digital resources. On one hand, the digital representation has to be expressed in forms that are amenable to digital expression through a number of highly disciplined and highly formal representations. To make effective use of these formal means of expression to produce something useful does require highly technical digital expertise which can usually only be found with technically trained personnel. However, on the other hand, what needs to be expressed in this highly structured, perhaps even rigid, set of formalisms must satisfactorily capture a complex historical conception which must come out of the expertise of academic historians. Furthermore, if the material that is developed is to be shared by an academic public, it needs to be understandable to them; a community who, because they are outside the development work and are mostly unfamiliar with the nature of highly structured data, are less likely to be able to grasp what the resource is all about than the historians in the project are.

An Analyst's Perspective

I am not an historian, and, indeed, although I have worked almost my entire professional career of more than 40 years in academia, was only actually even made an academic by King's College London in 2011 – 4 years before my retirement. Nonetheless, the reason why I think I have something to say here is based on my experience; mostly in what was classified as an 'academic related' post first at the University of Toronto starting in 1977, and then at King's College London since 1997 in what was called a 'Senior Analyst' post.[1] During my time at King's I worked, with varying levels of involvement, with academic partners on the building of a good number of digital resources: my CV lists more than twenty. The degree of success, as judged by the project's historian academic partners, did vary – but I think it is fair to say that almost all of my historian colleagues have expressed a large degree of satisfaction with what was produced. Several of them were of sufficient research quality that they were submitted by our historian partners as evidence of high quality research to the UK's Research Assessment Exercise (RAE) and Research Excellence Framework (REF)[2] initiatives. All were projects that involved the creation of highly structured data (the kind of data stored in formal databases) to represent scholarly research (mostly history), and almost all of them resulted in the provision of public scholarly access for the resources that were developed.

Some of the projects in which I have been involved include:

- *Prosopography of Anglo-Saxon England*[3] (Janet L. Nelson [KCL], Simon Keynes [Cambridge] and Stephen Baxter [Oxford], Harold Short [KCL] as technical director; I had a relatively junior role here as principal analyst): recorded information about individuals that appear in historical sources related to Anglo-Saxon England. *PASE*'s sources are of all kinds from, for example, Saint's Lives, through chronicles, letters, legal charters and even Domesday book.

- *People of Medieval Scotland*[4] (Dauvit Broun [Glasgow], David Carpenter [KCL] and Matthew Hammond [initially from Glasgow]; I was co-investigator and technical director here): another prosopography based almost entirely on Scottish medieval charters in the period from 1093 to 1314.

- *Digital Prosopography of the Roman Republic*[5] (Henrik Mouritsen and Dominic Rathbone and Maggie Robb [all KCL]; again I was technical director and co-investigator): unlike the other prosopographies in which I was involved which worked directly with primary sources, *DPRR* aimed to assemble and combine together a set of about 20 pre-existing mostly 20th century prosopographies, including T. R. S. Broughton's masterful *Magistrates of the Roman Republic*, first published in the 1950s, which forms the backbone of the database.

- *Art of Making in Antiquity*[6] (William Wootton [KCL] and sculptor the late Peter Rockwell; here I was co-investigator): about a third of my projects were not prosopographies. Rockwell, as an active artist and sculptor living in Rome, became interested in how historic sculptors created their works – what tools and processes did they use – and he looked at evidence for these things on the pieces themselves. He created an archive of thousands of

1 See some discussion of this in Bradley (2023).

2 https://www.ref.ac.uk (consulted 12.03.2024).

3 https://pase.ac.uk/ (consulted 12.03.2024).

4 Broun et al. (2014); https://www.poms.ac.uk/ (consulted 12.03.2024).

5 Mouritsen et al. (2016); https://romanrepublic.ac.uk/ (consulted 12.03.2024).

6 https://artofmaking.ac.uk/ (consulted 12.03.2024).

images that showed what evidence he found. We worked with him and King's historian William Wootton to present a subset of his images and to identify the evidence of how the objects they showed revealed the processes and tools used in their making.

No project of this kind is perfect, of course, but almost all of the ones in which I was involved were completed and digitally published within the funded period, and with the funds provided by the funding body. They created resources that have been, broadly speaking, thought of as historically interesting and useful by our partners and by the broader research community. Furthermore, there was a sense of ownership of the online resource by all the partners, technical and scholarly alike. And the happiest result came about when historian partners felt themselves to be fully owners of the data, its structure, and presentation. One way or another, there was a partnership involved. Figuring out how to make it work well is key.

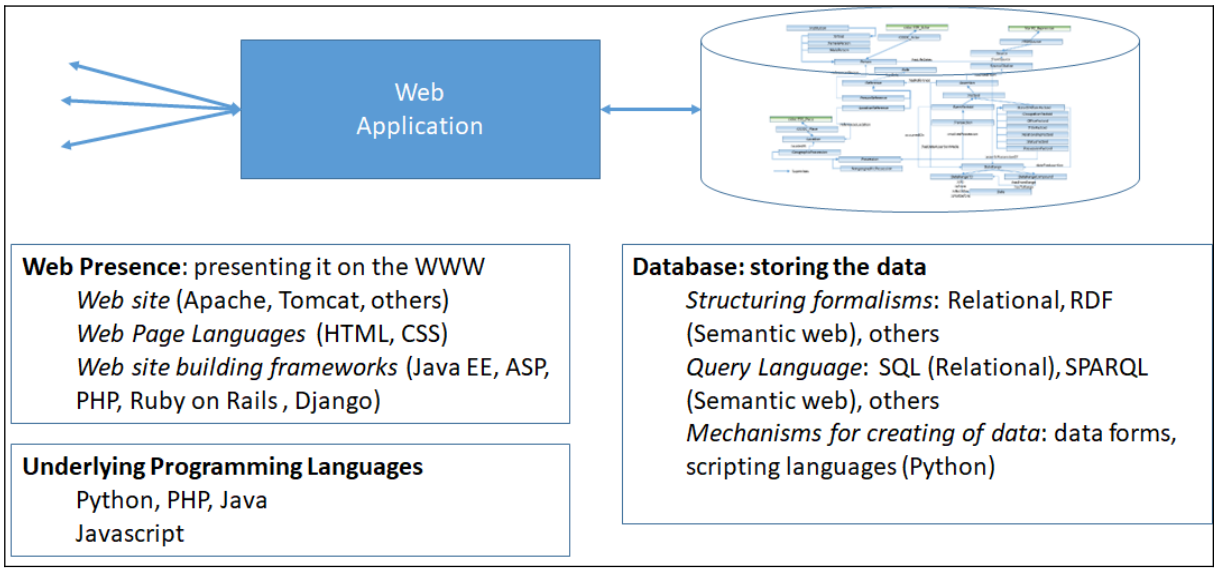


Fig. 1: Technical components in a structured data web application.

Technical Challenges

The kind of project we shall consider here is one that represents its materials in the form of highly structured data, and shortly in this paper will be a description what this means in a bit more detail. For the present, however, it is sufficient to understand that this paper considers projects that involve first the creation of a database (or similar technology such as, say, the semantic web) to hold the results of the scholarship, followed by its subsequent publication on the World Wide Web.

Fig. 1 provides a high-level overview of the technologies involved in the creation of such a project. On the right is a representation of the database which is holding the highly structured data.[7] On the left is the piece of software called the 'Web Application'. Someone can use it via the World Wide Web to access and interact with the data in the database. The challenge that I am exploring here comes out of the nature of the two groups of technologies (for the database, and for the web application) that are represented in the boxes at the bottom of the figure.

Focusing on the right side of fig. 1 for a moment, I believe that one needs to understand that although the formalisms of all database technologies now being used (relational, semantic web, NoSQL etc.) have many times over proven to be quite expressive in the situations for which they were developed,

7 The structure in the database shown in fig. 1 upper right corner will be too small to be examined in detail here. A full-size version can be found at https://www.kcl.ac.uk/factoid-prosopography/overall-concepts (consulted 12.03.2024) and similar diagrams for other projects will appear later in this article.

they are also quite rigid. Our experience has shown that the relational database model – far and away the most widely used data paradigm – can work well with a broad range of materials of interest to historians, and other humanities scholars too. But it is not trivial work to find out how. Furthermore, the tools in which databases are created (shown in fig. 1 in the box below the representation of the database) are meant to be managed by technical professionals. Thus the mechanisms to get the data into the database – via, say, web data forms, or perhaps extracted from, say, spreadsheets – require software created by professional developers to be built to request the data and transform it into a form suitable for deposit. All this work is built using software platforms that are designed in such a way to assume that technical professionals are using them. Some of the technologies are named in the right-hand box in fig. 1. A person initially without this technical training could, of course, acquire enough to do some of it, but to do it well requires extensive technical training and experience.

Furthermore, this is only a part of the whole story of the role of technology because although data structures such as those in a database can reflect well the expressive needs of many historical projects, including prosopography, the data in the database becomes largely inaccessible to an historian user. The mechanisms to interact with it – through the established query languages – are, again, aimed at professional technical folk, and are not appropriate to historians unless they invest considerable time in learning their formalisms and means of expression.

Hence the need to create a presentation of the data in ways that historians can access. Almost inevitably this involves the World Wide Web as the medium, with the historian user using her browser as the access software. In almost all of the projects I have been involved in this was managed by creating as a part of the project work an elaborate piece of software called a web application in fig. 1 that sits between the user's browser and the data. Here, again, complex technical issues must be managed, and they are different from the formalisms involved with the data and data storage. All the technologies (summarised in the box below the 'Web Application' box) are, again, meant to be used by highly professional developers with possibly years of training behind them.

There are other approaches than web applications that can be used to create materials that someone could access from the World Wide Web (see the strategies of 'minimal computing' which does away with a web server, but puts specialised functionalities into web pages, for examples), but in all mechanisms that I know of, the building of objects which allow the public through their browser to have access to a databases is so complex, with many different specialisms and involving different technical people, that their development still requires a substantial involvement of the technically trained. It is not surprising that unless one thinks through the issues with care, the historians can feel that they have lost control of the project: exactly as I mentioned at the very beginning of this paper. So, if this is to be avoided, regular and substantial collaboration between the historians and the technical team is necessary. Historians cannot expect to just leave the technical folk to 'get on with it'. Instead, in our experience, a kind of enduring partnership must be created and maintained, in which information is exchanged in both directions. What is the nature of these interactions, and partnership?

Structure and Semantics

There is, of course, more than one area where significant and ongoing collaboration is most useful. However, undoubtedly the most important, indeed the *key* domain, is in the area of the way the data is formally structured in the database that will be holding it. The formalisms of data design are quite rigid and significant complexity can lie in the details. Nonetheless, with some effort, it is possible to have a shared rich and deep understanding of the data with its structures. From this it then becomes possible to retain scholarly control of first the mechanisms that are created to put the materials into the database and then the public web interface – what a user sees of the data and how they can interact and

move through it – that will represent the material in a historically responsible manner. In spite of the need for much of the development work to be done in highly technical frameworks controlled by highly technical people, by fully grasping the formal design of the data, and understanding how to work with it to define and manage the public presentation of it, historians can be in control of this material.

One important thing to understand is that the formalisms are not merely rules for how the data must be organised; they also provide a framework into which the semantics of the material being represented can be expressed. A database structure must encapsulate part of the semantics of the material being studied and its scholarly interpretation. I use the word 'semantics' here, but it is worthwhile clarifying a little more what I mean by it.

Jerry Fodor, the well-known philosopher spoke about semantics in 2007:

> "Semantics [...] is part of a grammar of [a] language. In particular, it's the part of a grammar that is concerned with the relations between symbols in the language and the things in the world that they refer to or are true of. […] The intuition is that […] semantics is about how they relate to their referents in the nonlinguistic world."[8]

Fodor connects semantics to grammar and thus makes them both linguistic ideas: for him semantics become a key component both of thinking about the world, and also language that one uses to describe it. The nature of semantics for highly structured data – databases, XML, Semantic Web, etc. – are not so clearly related to natural language with its grammatical structures. Words are used of course – to name things that are represented – but the relationships between these words-as-names is incorporated into structures that appear to be quite different from natural languages, even though these structures still capture something about the significance of Fodor's 'nonlinguistic world'. So, how does this work? If the meaning between the names is not captured by a language, how is it captured? We will look at a simplified version of the process of building the semantic structure that was used in the prosopographical project *Peoples of Medieval Scotland* (*PoMS*).

8 Fodor (2007), 1.

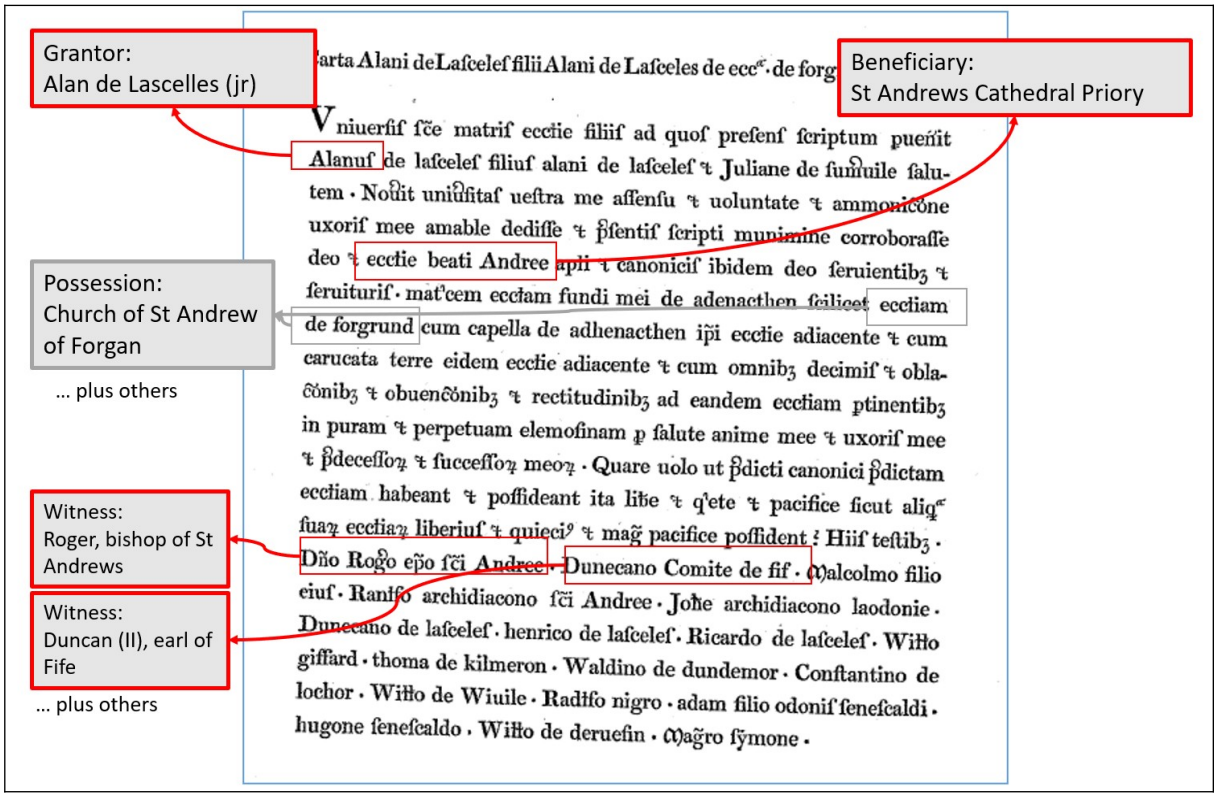


Fig. 2: A transaction based legal charter in *St Andrews Liber*, 260 (3/337/1).

PoMS's primary sources were medieval charters, and fig. 2 shows an example of one of them, from Thomas Thomson's May 1841 edition of the Cartulary of the Priory of St Andrew's.[9] This charter has Hammond Number 3/337/1.

What kind of historical, prosopographical, assertions could be taken from a document like this – what, in some sense is this document 'about'? As a legal charter, it is about a transaction of some land. People are, of course, involved, as are possessions. Fig. 2 annotates the reference to Alan de Lascelles, the grantor, and the Priory of St Andrew's as the beneficiary in an exchange of the Church of St Andrew in Forgan, plus some other, named, pieces of property. A list of people are identified as witnesses, including Roger, Bishop of St Andrew's, and Dunan, Earl of Fife.

There is more prosopographical information in this document than just this. A woman is identified as Alan's wife, and is presented as a person who consents to this transaction. Furthermore, the text of the charter followed certain legal conventions that were also of interest to the *PoMS* historians. In the interest of simplicity, how these other items were handled will not be discussed at present.

PoMS did not transcribe the text of the charter itself. Instead, it developed a data structure that could record what the document was about that was of interest to the project. Fig. 3 shows the kind of diagrams that was useful in the early planning phases of the project. Later diagrams were more complex as the project team gradually identified what formal structure would be needed to adequately capture what needed (historically) to be formally represented.

9 Thompson (1841), 260.

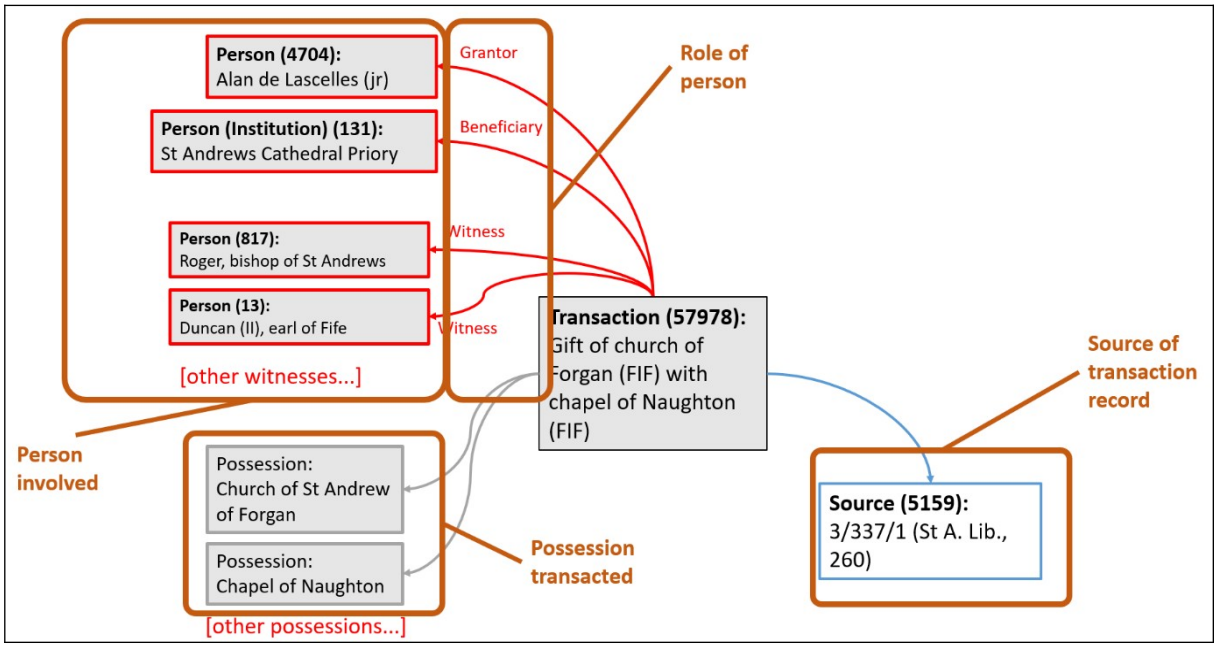


Fig. 3: A model for a transaction in *PoMS*.

Fig. 2 identified several of the spots in the text where persons were mentioned. In fig. 3 some of the persons that appeared in the text are displayed top left as formal digital entities that represent the corresponding historical persons: Alan de Lascelles, St Andrew's Priory, Roger the bishop and the other people mentioned as witnesses. The bracketed numbers are IDs for these particular persons. The possessions have also become digital entities.

In addition to the persons, the transaction itself (showing in the middle of fig. 3) became a formal entity in *PoMS*'s data structure with its own ID number: 57978. The roles of the people in this transaction are shown as lines that connect the transaction to the various people involved: Alan de Lascelles (historical person with ID 4704) is identified as the grantor for this transaction, etc. (in the actual data, Lascelles' wife is also formally identified as a consenter to the transaction).

Finally, the historical source for the transaction is given as an instance of *PoMS*'s source entity. Its ID is 5159, and the transaction is identified as found in it by the line that connects the transaction to its source.

Fig. 3 has focused on a particular transaction – but a good part of the work in capturing the semantics for representation in a database requires the step of generalising from particular instances such as this particular transaction into concepts that will apply across the data as a whole. For *PoMS* it was evident that the project would be interested in all transactions, to varying degrees like this one, along with a selection of other assertions that the sources made about persons, such as personal relationships and offices they held. After looking at a number of surviving legal charters the *PoMS* team concluded that all transactions would involve specific classes of entities:

- First, historical people who were referenced as being involved in each transaction.
- Second, these people would have specific roles in any particular transaction: as grantor, as beneficiary, as witness, as consenter, and perhaps other roles too.
- Third: all transactions were going to involve one or more possessions.
- Finally, there would always be a source in which the transaction appears. *PoMS* did not choose to transcribe the sources into, say, TEI. However, it did record bibliographic information about each source that it used in its database.

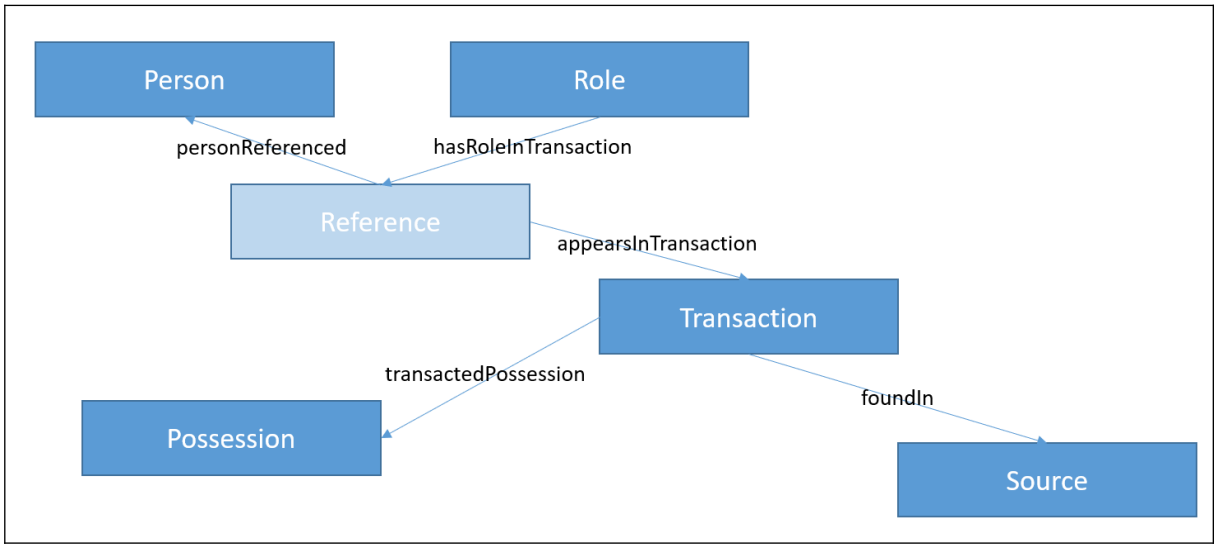


Fig. 4: Entities and relationships for transactions in *PoMS*.

Database representation becomes most useful when you have many instances of things that exhibit the same or similar characteristics. Fig. 4 shows how the analysis for specific transactions was generalised into a set of overarching concepts that expressed what *PoMS* intended to record about transactions.[10] Here can be seen the types of entities that were identified in the previous figure: Person, Role, Possession, Transaction and Source (plus a new one: Reference, which is necessitated by the need to connect roles to people in transactions, but also turned out to have semantic significance itself which was discovered during the analysis). The lines between the entities are named in ways that suggest the nature of the connection. In the end our actual structure to deal with transactions was more complex than this, and was the result of a more complex and time consuming process that took a number of months to complete and looked at a range of sources. The structure that emerged became the primary structure that was expressed in a database. Even after this amount of planning work, the resulting structure for transactions turned out to be incomplete, and had to be modified somewhat on a few occasions during the project work.

Prosopographies like *PoMS* are interested in transactions because they express significant and interesting things about historic people. Right at the start of his 1970 article on the idea of Prosopography Lawrence Stone suggests a range of other kinds of assertions that are interesting too:

> "Prosopography is the investigation of the common background characteristics of a group of actors in history by means of a collective study of their lives. The method employed is to establish a universe to be studied, and then to ask a set of uniform questions – about birth and death, marriage and family, social origins and inherited economic position, place of residence, education, amount and source of personal wealth, occupation, religion, experience of office, and so on. The various types of information about the individuals in the universe are then juxtaposed and combined, and are examined for significant variables."[11]

Indeed, one of Stone's 'uniform questions', marriage and family, appears as an assertion in the charter we've been looking at, since Alan's wife is explicitly identified as his wife there. Other kinds of information found in charters are people's offices and sometimes their occupations.

10 Well, in truth, only a part of what *PoMS* chose to record about transactions in fig. 4: this is in fact a simplified diagram since, as we will see later, *PoMS* chose to record other characteristics of the transactions that are not shown here.

11 Stone (1971), 46.

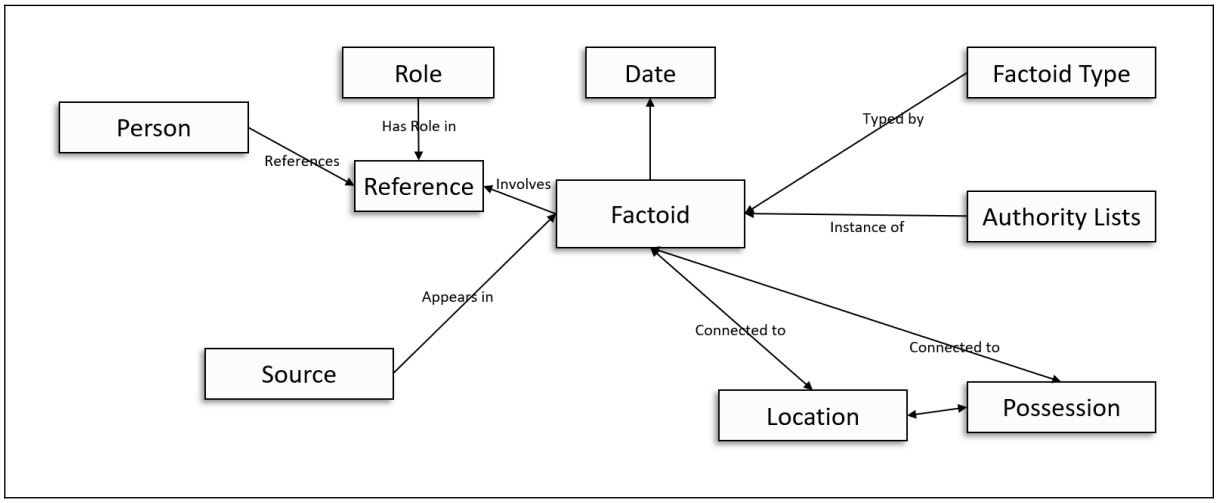


Fig. 5: A simple representation of the prosopographical factoid model.

Thus, it is useful to perform a further generalisation beyond the transaction-focused one that we have just seen by turning the kind of thinking represented by our analysis of legal transactions into a more general approach that will cover other kinds of assertions found in the sources too. It was from this kind of thinking in earlier prosopographies developed at King's that the idea of the factoid arose, and this idea was found to work well with *PoMS* as well. This chapter is not the place to expand upon the idea of the prosopographical factoid in great detail, or to explain why it has been given this slightly ironic name. There is more information online at the factoid website,[12] and, more specifically in terms of legal charters at a recent article.[13] A simple schematic of the factoid model for data can be seen in fig. 5.

If we compare fig. 4 – which is specific to legal transactions – with the factoid-oriented one in fig. 5 we can see many similar elements. However, the central change is that the central box that was labelled 'transaction' in fig. 4 has become 'factoid' in fig. 5. In Bradley 2018 the factoid is described as "a spot in a source that says something about a person or persons." It is thus an assertion. The transaction becomes one type of assertion/factoid and other kinds of assertions can be included as well.

- As has just been stated, in *PoMS* the legal transaction is one type of factoid.
- Relationships between people is another (and indeed, we found this second kind of factoid in the husband-wife relationship provided inside what was primarily a legal document),
- and titles and occupations are a third.
- Other kinds of assertions might be of interest too.

For all these types of factoids:

- We have people referenced, and a historical source, and place (or 'spot') in the source identified.
- Some factoids can involve references to associated geographic locations and/or (as we have seen) possessions: land and other kinds of things.
- The *Factoid Typ* entity is referenced to specify which kind of factoid we are dealing with: transactions, relationship, title, occupation, etc.

12 Bradley (2018): https://www.kcl.ac.uk/factoid-prosopography (consulted 12.03.2024).

13 Bradley / Rio / Hammond / Broun (2019).

- *Authority Lists* are used for a broad range of semantic purposes centred on Prosopography's frequent dependency on classification of people. One kind of authority list would be associated with relationships and specify what kind of relationship was specified: son, mother, aunt, etc. Another would be the list of titles or occupations so that each title or occupation factoids can point to one of them.

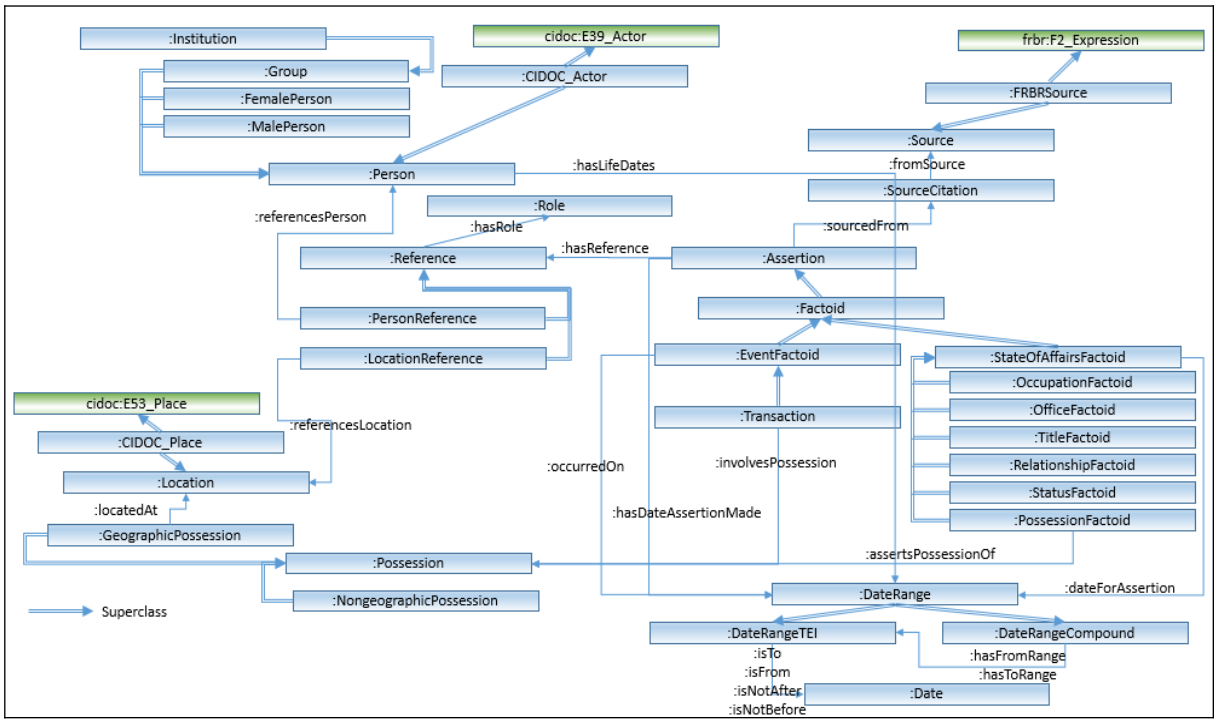


Fig. 6: A diagram representing the *Factoid Prosopography Ontology*.

A further generalisation and enrichment of the factoid approach is given in its Semantic Web representation through the *Factoid Prosopography Ontology* described in Bradley 2018. An Ontology has been described by Thomas Gruber as a conceptualization. The objects, concepts, and other entities that are assumed to exist in some area of interest (called *Classes* in Semantic Web terminology) and the relationships that hold among them form such a conceptualization, and in the Semantic Web an ontology is "an explicit specification of [such] a conceptualization."[14]

Fig. 6 shows the overall structure of the *Factoid Prosopography Ontology*, focusing on the classes and relationships between them. It is richer and more complex than the simple representation of the Factoid shown in fig. 5, but one can still see many of the same components in it: Person, Factoid, Source, Reference, Role, etc. One thing an ontological description can do is present different 'kinds' of classes. People, for example, come in different kinds, and the classification by sex and the recognition of legal corporate entities that act as people in legal documents are shown as 'kind of' relationships (with a double line) in this diagram. We've already mentioned that factoids come in various kinds, and the common kinds are shown here. Dating can be a complex historical process and some of that complexity is dealt with by the classes shown in the bottom right corner which adapt the Text Encoding Initiative's dating formalisms[15]. One can find an extensive description of the whole ontology at Bradley 2018.

One of the ways that ontologies work is as a *basis* for a data structure for a specific project. By incorporating ontologies such as the *Factoid Prosopography* in the structure developed for a particular project one explicitly associates the semantics attached to it to the structure developed by the project as

14 Gruber (1993).

15 TEI 2022, Section 3.6.4.a.

well. In the Semantic Web thinking, one or more external ontologies can provide common core ideas upon which a project can build their own specialised ontology for their specific needs.

- Not all projects may need all the kinds of things here – others will need new kinds of things that are not here, but perhaps connect to these core elements.

- Sometimes entities shown here will need to be extended further to serve the needs of a particular project. One can see an example of this where the Possession class by itself was insufficient for *PoMS*'s needs.

The *PoMS*-specific ontology was based on this general factoid one, but the single Possession class that the *Factoid Prosopography Ontology* includes was insufficient of its needs. Thus, *PoMS*'s ontology enriched the single possession class the *Factoid Prosopography Ontology* gave by defining an entire set of kinds of possessions. You can see this extension, plus others, in the diagram shown in fig. 7, which in turn comes from the online description of *PoMS*'s ontology.[16]

Fig. 7 reveals a relatively complex structure when compared to the previous figures discussed here. However, you can find types of entities and relationships between them that correspond with both the *Factoid Prosopography* approach, and *PoMS*'s analysis of the sources and the transactions found within them that has been described here earlier.

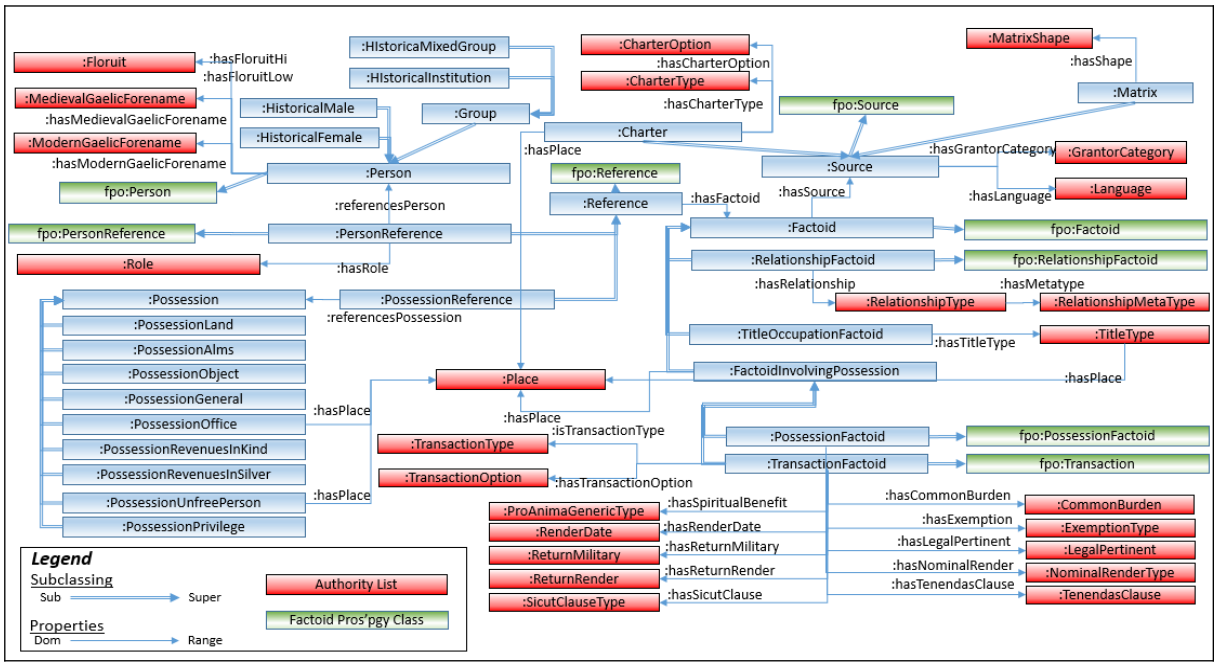


Fig. 7: A diagram showing *PoMS*'s ontology.

There are clearly areas where the entities described in the *Factoid Ontology* shown in fig. 6 have been extended. In the bottom left area we can see the significant different types of Possessions that were identified by the *PoMS* team during the analysis phase of the work and that appeared as possessions in Scottish Medieval Charters: kinds of things ranging from property as land, to property as unfree persons or possessions as offices, or specific objects such as horses or golden statues. Elsewhere, the person structure in *PoMS* followed closely that of the generalised *Factoid Prosopography Ontology*, but it was extended to deal with certain aspects of Gaelic naming practices. The 'Source' corner of the ontology was enriched by a recognition that not only the charters, but the attached seals and the matrices that created them provided prosopographical information. In addition, charters were classified into types by the *PoMS* team, and each charter was then classified according to this scheme.

16 https://www.poms.ac.uk/rdf/doc/ontology.html (consulted 12.03.2024).

Finally, in a project so focused on legal charters, it is perhaps not surprising that the rudimentary model for transactions that was described earlier in this paper was enriched substantially by further pieces of information reflecting Medieval Scottish interests, such as a transaction's Render Date, or associated Common Burdens.

What the data model is like will affect the semantics associated with the database, and will, as a consequence, control what kinds of assertions the database can make about the historical sources. Thus, the organisation of *PoMS*'s data demanded a complex structure for the complex historical task that *PoMS* represented. Indeed, one should, surely, expect a substantial historical project drawn from rich materials to require a complex and subtle analysis, and this figure is simply a result of the complexity that the historians uncovered and formally expressed in their sources.

As a consequence, it should be clear that the creation of this model is a task that requires the detailed understanding of the historians who can judge to what extent the formal model represents adequately what is needed. They must be closely involved in the formal expression of the data that the model represents if it is to capture properly what they feel should be said about their materials. For *PoMS* the preliminary model took many months to create, and was based on a close analysis of a good number of charters. In any project of this kind, and certainly for *PoMS* too, the modelling work will to some extent continue throughout most of the building of the data, and thus requires a continuous partnership between the historians, who might, from time to time, spot problems in the expressiveness of the current model, and can then work with the technology team who can implement changes to address them.

The Model and Data Preparation

This paper has illustrated how the data model is connected to an idea of semantics and thus can explicitly represent some significant part of the scholarly meaning of the material. The data model will affect the semantics associated with the database, and will, as a consequence, control what kinds of assertions the database can make about the historical sources. I have tried, through this extended illustration, to make it clear that the creation of the model is a task that requires the understanding of the historians who can judge to what extent the formal model adequately represents what is needed. Ideally, then, its development can be a significant partnership between the historians and the technical team; an arrangement that is likely to continue throughout the project.

Having developed a repository for the data, the project will then be ready to move on to the phase of putting materials into it. Here is one place where the model is critical since it will affect in many ways the mechanisms that can be created by the technical team to collect the data.

Data forms are the classic way to get data into a database, and they have been widely used in projects in which I have been involved. Fig. 8 shows portions of one of about 20 different forms for data entry that was created for the *Prosopography of Medieval Scotland*, and this particular one was used to enter the information for its transaction factoid. *PoMS*'s researchers would read the source text, locate the aspects of interest to the *PoMS* project, and record them in this and other forms. Like most of the data entry mechanisms that *PoMS* used, this particular form is generated semi-automatically by the *Django* framework[17] which we use to build our data-oriented projects. Its ability to create and adjust these forms relatively painlessly as the data structure is first created, then modified, is a major advantage we have found for the *Django* platform over other frameworks with which we are familiar.

Perhaps the first thing to note is that even though the project is a prosopography, and one might think that a form for persons would be front and centre, that is not so for *PoMS*. Most of the time the data entry operation was centred on the sources being read, and thus it was centred on factoid-based forms,

17 https://www.djangoproject.com/ (consulted 14.03.2024).

like the transaction shown in fig. 8. Records for individual historic people were created as needed as the source was read but, surprisingly perhaps, this often was carried out as a seemingly secondary operation – driven by their appearance in the sources.

Fig. 8 shows the transaction form as it was filled in by the *PoMS* researcher about the transaction from source 3/337/1 that formed the basis for the discussion here earlier. The entire form is rather long and complex, and fig. 8 shows only three fragments of it. The top part, which is also the top of the form itself, begins with an area for specifying the historic source; the legal charter; for this transaction. Within the database the Source is a separate entity, so the first form element, labelled 'Document', is actually used not just to hold a bit of text that identifies the document, but instead to point to the appropriate instance of the Source entity that describes it bibliographically.

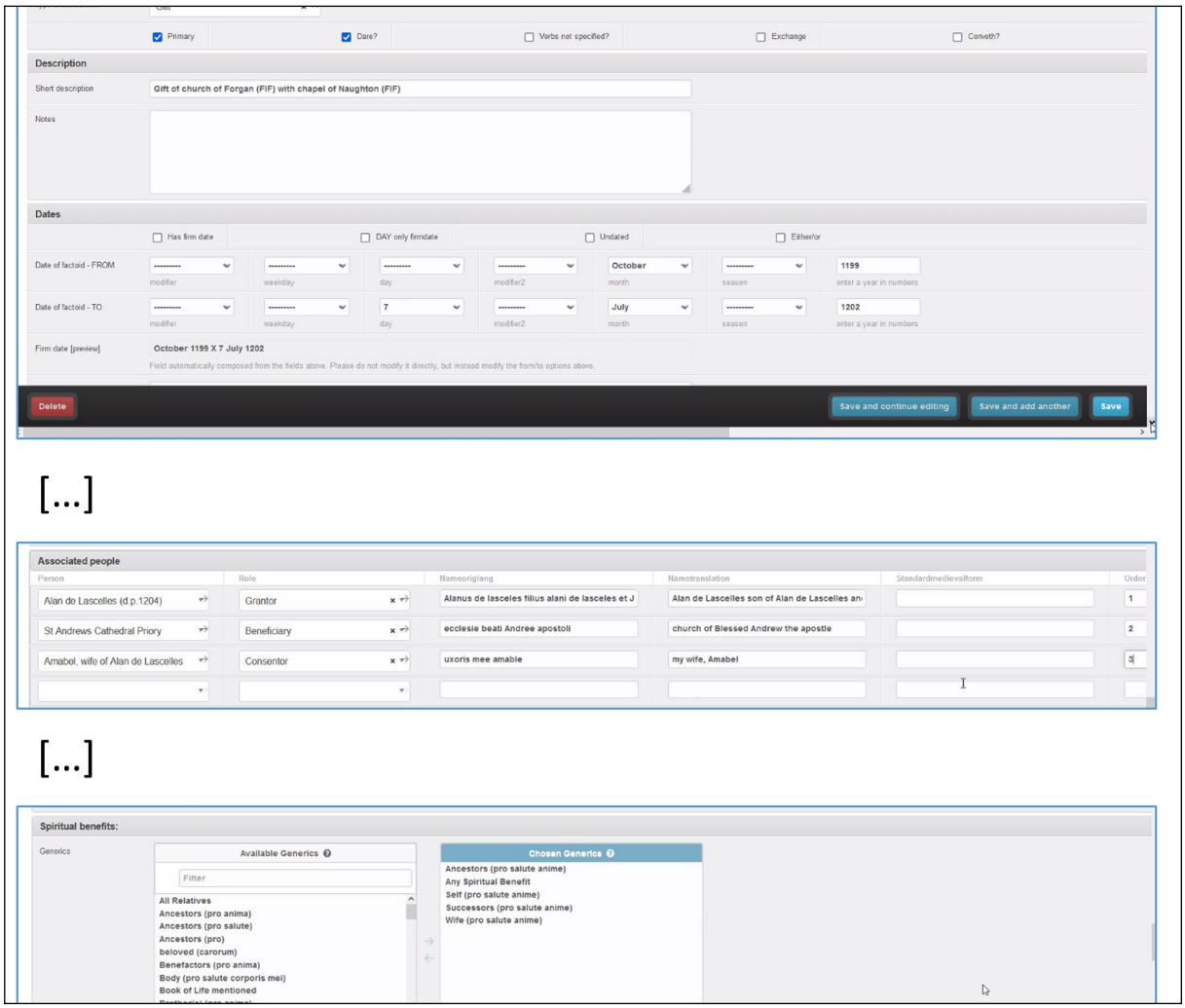


Fig. 8: Three fragments from *PoMS*'s transaction data entry form.

Other source-related information that has not been mentioned here before appears here too. The transaction is specified as a 'Gift', and the person entering the data can also assert that this transaction is the primary thing presented in this particular source. Below the source area on the form is a field for a short verbal description of this transaction. The system can then display this to the user whenever that is appropriate. Then, below that is the dates area, which is revealed to be a rather complex structure: dates being, of course, in historical sources like this, a complex matter to handle with the flexibility that is required.

The second fragment of fig. 8 shows that part of the form which focuses on the people that are connected to the transaction. Although this area is part of a form about the transaction, it actually specifies what references are associated with the transaction. Each reference, in turn, will point to the person

rather than simply name them, and when the form is filled in the researcher is presented with a mechanism that allows him/her to either select or (if an entity for them did not yet exist) create a new person entity. The role of the person in the transaction is indicated in the 2nd column, and two data attributes are also attached to the person reference that record what the source actually says to identify the person, and how this bit of Latin could be translated into modern English.

As was mentioned earlier, the *PoMS* team wished to record a rich interpretation of their transactions, and the third segment of fig. 8 provides an example of one of the other kinds of information that *PoMS* collected: what the *PoMS* project called 'generic Spiritual benefits'. Often charters of this period used formulaic language to say that something was being done for the good of someone's soul, and here the project team could choose from the list of formulae found (on the left) to indicate which ones actually appeared in this document.

It is probably obvious by now that the form is quite complex reflecting the complex nature of what the *PoMS* team wanted to say about transactions. It is useful to note two things about it.

First of all, the form makes it evident that the structure diagram shown in fig. 7 does not tell the whole story about the structure of the *PoMS* data. It emphasized the kinds of entities that were involved in the dataset, and how they were interconnected. What it did *not* show was the kinds of information that are *attached* to each type of entity: for Person Entities, things like a head name, and perhaps other data such as floruit dates, for Transactions: dates, or short bits of narrative attached to a transaction that records something the team wants to say to the user audience. These kind of attributes are a kind that don't point to other entities, but instead hold relatively simple values; numbers or bits of text. There is often many of these for each kind of entity.

Second, the things a form shows and how it works *must* align with the data structure: it cannot be made to work otherwise. If what the form allows to be entered cannot be made to match the historical needs of the project, this is an indication that the model behind the form is inadequate or inaccurate. If this problem occurs, the data model must be changed, and then the form changed too to reflect the new model. In that the semantics of the data structures must be in the domain of the historian, the contents and order of the data entry form can, and should, be sorted out between the historian and technical team as partners: the technical team is responsible for what the tool generating the form can accomplish, and the historian for what it does express and how it does so in ways that best facilitate the entry of the material by the historical team.

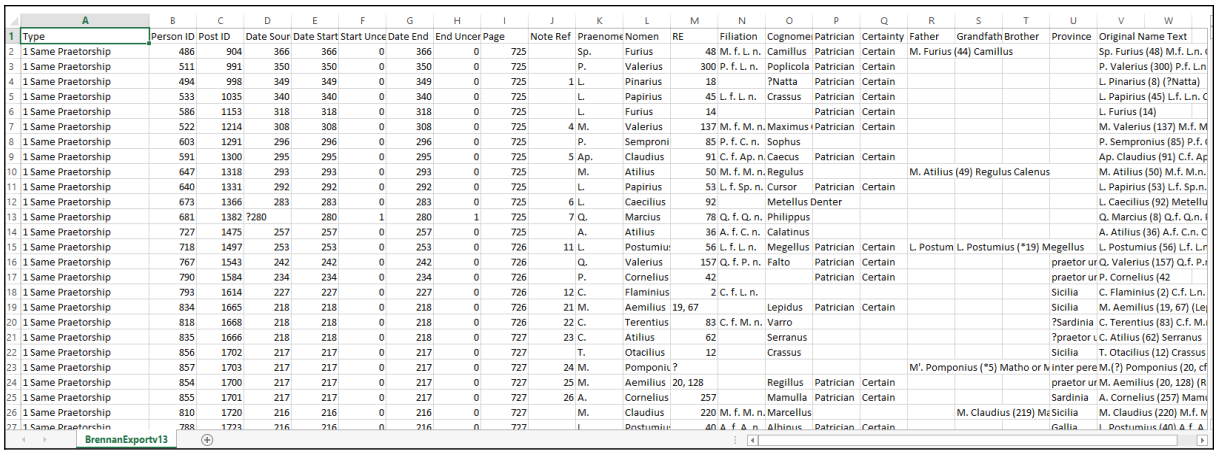
Although we have found forms to be the usual mechanism for inputting data into databases, it is not unusual to sometimes have other mechanisms too. Academic project partners sometimes find it more convenient to create materials in, say, a spreadsheet or perhaps an XML document. Indeed, a spreadsheet is often proposed as a method of data collection since it would appear to be a kind of half-way house towards structured data, and historians are relatively comfortable dealing with it.

Fig. 9 is an example of this alternate, and in some circumstances better, approach to entry of the data. It deals with material which the *Digital Prosopography of the Roman Republic project* (*DPRR*)[18] wanted to include that came from T.C. Brennan's *The Praetorship in the Roman Republic*[19]. The principal research associate in *DPRR*, Dr Maggie Robb, collected the material from Brennan and put it into a spreadsheet. The technical team then wrote a 298 line Python script – a little software programme – to take the material in the spreadsheet and transform it into data loaded into *DPRR*'s database. We will not consider the details of this process here, but you can see that this work which allowed this approach to be taken was a part of the ongoing partnership between the historians involved in *DPRR* and the technical team well after the database was up and running. The script was what is often called 'throwaway' code in that it did precisely and only the job of handling the Brennan spread-

18 Mouritsen et al. (2016); https://romanrepublic.ac.uk (consulted 12.03.2024).

19 Brennan (2000).

sheet, and was not needed again for anything else. However, although it did require a professional technical understanding to create it, it was not a large technical challenge to create, and was written in a day or two.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
1	Type	Person ID	Post ID	Date Sour	Date Start	Start Unce	Date End	End Uncer	Page	Note Ref	Praenome	Nomen	RE	Filiation	Cognome	Patrician	Certainty	Father	Grandfath	Brother	Province	Original Name Text	
2	1 Same Praetorship	486	904	366	366	0	366	0	725		Sp.	Furius	48	M. f. L. n.	Camillus	Patrician	Certain	M. Furius (44)	Camillus			Sp. Furius (48) M.f. L.n. (	
3	1 Same Praetorship	511	991	350	350	0	350	0	725		P.	Valerius	300	P. f. L. n.	Poplicola	Patrician	Certain					P. Valerius (300) P.f. L.n	
4	1 Same Praetorship	494	998	349	349	0	349	0	725	1	L.	Pinarius	18		?Natta	Patrician	Certain					L. Pinarius (8) (?Natta)	
5	1 Same Praetorship	533	1035	340	340	0	340	0	725		L.	Papirius	45	L. f. L. n.	Crassus	Patrician	Certain					L. Papirius (45) L.f. L.n. C	
6	1 Same Praetorship	586	1153	318	318	0	318	0	725		L.	Furius	14			Patrician	Certain					L. Furius (14)	
7	1 Same Praetorship	522	1214	308	308	0	308	0	725	4	M.	Valerius	137	M. f. M. n.	Maximus	Patrician	Certain					M. Valerius (137) M.f. M	
8	1 Same Praetorship	603	1291	296	296	0	296	0	725		P.	Semproni	85	P. f. C. n.	Sophus							P. Sempronius (85) P.f. (	
9	1 Same Praetorship	591	1300	295	295	0	295	0	725	5	Ap.	Claudius	91	C. f. Ap. n.	Caecus	Patrician	Certain					Ap. Claudius (91) C.f. Ap	
10	1 Same Praetorship	647	1318	293	293	0	293	0	725		M.	Atilius	50	M. f. M. n.	Regulus			M. Atilius (49)	Regulus	Calenus		M. Atilius (50) M.f. M.n.	
11	1 Same Praetorship	640	1331	292	292	0	292	0	725		L.	Papirius	53	L. f. Sp. n.	Cursor	Patrician	Certain					L. Papirius (53) L.f. Sp.n.	
12	1 Same Praetorship	673	1366	283	283	0	283	0	725	6	L.	Caecilius	92		Metellus	Denter						L. Caecilius (92) Metellu	
13	1 Same Praetorship	681	1382	?280	280	1	280	1	725	7	Q.	Marcius	78	Q. f. Q. n.	Philippus							Q. Marcius (8) Q.f. Q.n. I	
14	1 Same Praetorship	727	1475	257	257	0	257	0	725		A.	Atilius	36	A. f. C. n.	Calatinus							A. Atilius (36) A.f. C.n. C	
15	1 Same Praetorship	718	1497	253	253	0	253	0	726	11	L.	Postumiu	56	L. f. L. n.	Megellus	Patrician	Certain	L. Postum	L. Postumius (*19)	Megellus		L. Postumius (56) L.f. L.n	
16	1 Same Praetorship	767	1543	242	242	0	242	0	726		Q.	Valerius	157	Q. f. P. n.	Falto	Patrician	Certain				praetor ur	Q. Valerius (157) Q.f. P.(	
17	1 Same Praetorship	790	1584	234	234	0	234	0	726		P.	Cornelius	42			Patrician	Certain				praetor ur	P. Cornelius (42	
18	1 Same Praetorship	793	1614	227	227	0	227	0	726	12	C.	Flaminius	2	C. f. L. n.							Sicilia	C. Flaminius (2) C.f. L.n.	
19	1 Same Praetorship	834	1665	218	218	0	218	0	726	21	M.	Aemilius	19, 67		Lepidus	Patrician	Certain				Sicilia	M. Aemilius (19, 67) (Le	
20	1 Same Praetorship	818	1668	218	218	0	218	0	726	22	C.	Terentius	83	C. f. M. n.	Varro						?Sardinia	C. Terentius (83) C.f. M.(	
21	1 Same Praetorship	835	1666	218	218	0	218	0	727	23	C.	Atilius	62		Serranus						?praetor u	C. Atilius (62) Serranus	
22	1 Same Praetorship	856	1702	217	217	0	217	0	727		T.	Otacilius	12		Crassus						Sicilia	T. Otacilius (12) Crassus	
23	1 Same Praetorship	857	1703	217	217	0	217	0	727	24	M.	Pomponi	?					M'. Pomponius (*5)	Matho or N		inter pere	M.(?) Pomponius (20, cf	
24	1 Same Praetorship	854	1700	217	217	0	217	0	727	25	M.	Aemilius	20, 128		Regillus	Patrician	Certain				praetor ur	M. Aemilius (20, 128) (R	
25	1 Same Praetorship	855	1701	217	217	0	217	0	727	26	A.	Cornelius	257		Mamulla	Patrician	Certain				Sardinia	A. Cornelius (257) Mam	
26	1 Same Praetorship	810	1720	216	216	0	216	0	727		M.	Claudius	220	M. f. M. n.	Marcellus				M. Claudius (219)	Ma	Sicilia	M. Claudius (220) M.f. M	
27	1 Same Praetorship	788	1723	216	216	0	216	0	727		L.	Postumiu	40	A. f. A. n.	Albinus	Patrician	Certain				Gallia	L. Postumius (40) A.f. A.	

BrennanExportv13

Fig. 9: Some of Brennan's *The Praetorship of the Roman Republic* expressed as a spreadsheet.

The Model and the Public View

So far this paper has considered tasks involved in the creation of the database. The contention here is that although the creation of the structure of the database is a highly formal and technical task, and requires technical expertise to achieve it successfully, it must grow out of the historian's understanding, interests and aims. Although few historians will be familiar with the formal issues involved in highly structured data, it is possible, with some effort, to understand in some detail how the design works. In this way, control of the database becomes not purely to domain of the technologist, but also something that is properly and fully jointly owned. By involving him or herself substantially in the design, the historian can also then direct the creation of mechanisms that put the materials into the database in ways that work best for his or her needs.

However, this is not the end of the engagement between historians and technical staff. As fig. 1 shows, the database, by itself, will be unusable by the great majority of historians. These days, the common way to give access to a historian community is through the World Wide Web, and conventionally this means that a web application has to be created that sits between the database and the World Wide Web so that an historian, using a web browser, can make use of the data.

As discussed briefly earlier, the creation of this web front end involves quite a different set of skills from those needed for database creation, often requiring different technical folk. Furthermore, in pretty well all of the projects in which I have experience, it has proven to be the most complex piece of technical work of all. Even then, however, historians must keep engaged with the work of building the web application so that the resulting software server presents their material properly and accessibly. Given the highly technical nature of the work needed to build this web application, how can historians retain control of the result so that they are also happy with what appears? The issue comes more into focus when one sees what the data in a database looks like when viewed without a web application to provide some interpretive packaging.

As it happens, one can see the data from both the *PoMS* and *DPRR* projects essentially directly as they appear in their databases. Because of the significant interest in the digital humanities at a time in the idea of 'open data' (where raw data are made freely available directly for others to use) we decided to provide, in addition to our presentation-oriented web application, two data portals which gave direct access to the *PoMS* data and the *DPRR* data in their raw Semantic Web RDF expression. Later in this

paper there will be an argument about how this might be useful, but for now it is useful to see what the raw data for *PoMS* or *DPRR* is like.

To build these portals, we chose to follow the Open Data principles developed by Tim Berners-Lee (the inventor of the WWW), and described in a highly influential paper by him and some colleagues in 2009[20]. Of the set of principles described there for Open Data, there are two pieces necessary for our discussion here.

- The first one is that all entities represented in the data need to have access mechanisms that follow open and public standards. Berners-Lee proposes using URIs (which are almost identical to the WWW's URLs) as public identifies for these entities.

- Then, second, the resulting data that is returned should be provided in a form that is compatible with further processing by the person that fetched it, and is in an open format. Berners-Lee proposes the W3C's RDF format[21] for this.

This is exactly what was done with the *PoMS* and *DPRR* data. For both projects all their database materials were expressed in RDF, and all of it was made accessible world-wide through URI's which anyone can fetch by giving the URI to their browser. For those interested in information on a classically-oriented project, there is extensive discussion of the nature and implications of this for *DPRR* in an article in the *Digital Humanities Quarterly*[22]. Here, we shall continue to centre our material on the *PoMS* example of a transaction started earlier.

The first principle from Berners-Lee requires that all entities have a public identifier that gives access to the data that is attached to it[23].

In true Open Data fashion, one can simply give this URL/URI to your browser to see the data that is attached to it displayed as a list of RDF assertions. Fig. 10 shows the data for this transaction starting a little way into that list of assertions. It is not straightforward to see one's way around this list of RDF statements, but it is useful to see how the semantics attached to this transaction have been represented.

As fig. 10 implies, RDF statements have three parts. The first line in fig. 10, for example, formally asserts that this entity is of type 'Transaction Factoid'[24].

Continuing the exploration; the 15th RDF statement in the list in fig. 10 identifies the source in which this transaction can be found. Attribute *hasSource* identifies the source using the Source's public URI/URL[25] that RDF and Open Data requires. What source this actually is can be found by going to look at this URL for the source itself.

Elsewhere one can see RDF assertions related to the dating of the Transaction. This paper has not discussed how dating is handled in *PoMS* in any detail, but one can find groups of assertions about dates that allow for complex dating to be represented. The charter source that was available to *PoMS* for this transaction was undated in any explicit form, but from examining the witness list and knowing when the various people were in office, it was possible to provide a plausible date range.

---

20 Bizer / Heath / Berners-Lee (2009).

21 Cyganiak et al. (2014).

22 Bradley (2020).

23 For the *PoMS* transaction described earlier this is https://www.poms.ac.uk/rdf/entity/Factoid/57978 (consulted 12.03.2024).

24 https://www.poms.ac.uk/rdf/entity/Factoid/57978 (consulted 12.03.2024).

25 https://www.poms.ac.uk/rdf/entity/Source/5159 (consulted 12.03.2024).

<https://www.poms.ac.uk/rdf/entity/Factoid/57978>	rdf:type	vocab:TransactionFactoid
<https://www.poms.ac.uk/rdf/entity/Factoid/57978>	rdfs:label	"Factoid 57978 (transaction): Gift of church of Forgan (FIF) with chapel of Naughton (FIF) St Andrews Liber, 260 57978 "
<https://www.poms.ac.uk/rdf/entity/Factoid/57978>	vocab:FactTransactionTenendasClauseOrigLang	"in puram et perpetuam elemosinam"
<https://www.poms.ac.uk/rdf/entity/Factoid/57978>	vocab:hasAssociatedWebpage	<https://www.poms.ac.uk/record/factoid/57978/>
<https://www.poms.ac.uk/rdf/entity/Factoid/57978>	vocab:hasDatingNotes	"Election of William Malveisin, archdeacon of Lothian to see of Glasgow × death of Bishop Roger of St Andrews. "
<https://www.poms.ac.uk/rdf/entity/Factoid/57978>	vocab:hasFirmDate	"October 1199 X 7 July 1202 "
<https://www.poms.ac.uk/rdf/entity/Factoid/57978>	vocab:hasFromMonth	10
<https://www.poms.ac.uk/rdf/entity/Factoid/57978>	vocab:hasFromYear	1199
<https://www.poms.ac.uk/rdf/entity/Factoid/57978>	vocab:hasID	57978
<https://www.poms.ac.uk/rdf/entity/Factoid/57978>	vocab:hasPlace	<https://www.poms.ac.uk/rdf/entity/Place/136>
<https://www.poms.ac.uk/rdf/entity/Factoid/57978>	vocab:hasPlace	<https://www.poms.ac.uk/rdf/entity/Place/1748>
<https://www.poms.ac.uk/rdf/entity/Factoid/57978>	vocab:hasPlace	<https://www.poms.ac.uk/rdf/entity/Place/412>
<https://www.poms.ac.uk/rdf/entity/Factoid/57978>	vocab:hasShortDesc	"Gift of church of Forgan (FIF) with chapel of Naughton (FIF)"
<https://www.poms.ac.uk/rdf/entity/Factoid/57978>	vocab:hasSicutClause	<https://www.poms.ac.uk/rdf/entity/SicutClauseType/1>
<https://www.poms.ac.uk/rdf/entity/Factoid/57978>	vocab:hasSource	<https://www.poms.ac.uk/rdf/entity/Source/5159>
<https://www.poms.ac.uk/rdf/entity/Factoid/57978>	vocab:hasSpiritualBenefit	<https://www.poms.ac.uk/rdf/entity/ProAnimaGenericType/12>
<https://www.poms.ac.uk/rdf/entity/Factoid/57978>	vocab:hasSpiritualBenefit	<https://www.poms.ac.uk/rdf/entity/ProAnimaGenericType/4>
<https://www.poms.ac.uk/rdf/entity/Factoid/57978>	vocab:hasSpiritualBenefit	<https://www.poms.ac.uk/rdf/entity/ProAnimaGenericType/53>
<https://www.poms.ac.uk/rdf/entity/Factoid/57978>	vocab:hasSpiritualBenefit	<https://www.poms.ac.uk/rdf/entity/ProAnimaGenericType/75>
<https://www.poms.ac.uk/rdf/entity/Factoid/57978>	vocab:hasSpiritualBenefit	<https://www.poms.ac.uk/rdf/entity/ProAnimaGenericType/8>
<https://www.poms.ac.uk/rdf/entity/Factoid/57978>	vocab:hasTenendasClause	<https://www.poms.ac.uk/rdf/entity/TenendasClause/5>
<https://www.poms.ac.uk/rdf/entity/Factoid/57978>	vocab:hasToDay	7
<https://www.poms.ac.uk/rdf/entity/Factoid/57978>	vocab:hasToMonth	7
<https://www.poms.ac.uk/rdf/entity/Factoid/57978>	vocab:hasToYear	1202
<https://www.poms.ac.uk/rdf/entity/Factoid/57978>	vocab:hasTransactionOption	<https://www.poms.ac.uk/rdf/entity/TransactionOption/1>
<https://www.poms.ac.uk/rdf/entity/Factoid/57978>	vocab:hasTransactionOption	<https://www.poms.ac.uk/rdf/entity/TransactionOption/2>
<https://www.poms.ac.uk/rdf/entity/Factoid/57978>	vocab:isDare	true
<https://www.poms.ac.uk/rdf/entity/Factoid/57978>	vocab:isPrimary	true
<https://www.poms.ac.uk/rdf/entity/Factoid/57978>	vocab:isTransactionType	<https://www.poms.ac.uk/rdf/entity/TransactionType/2>
<https://www.poms.ac.uk/rdf/entity/PersonReference/109854>	vocab:hasFactoid	<https://www.poms.ac.uk/rdf/entity/Factoid/57978>
<https://www.poms.ac.uk/rdf/entity/PersonReference/113765>	vocab:hasFactoid	<https://www.poms.ac.uk/rdf/entity/Factoid/57978>
<https://www.poms.ac.uk/rdf/entity/PersonReference/117408>	vocab:hasFactoid	<https://www.poms.ac.uk/rdf/entity/Factoid/57978>
<https://www.poms.ac.uk/rdf/entity/PersonReference/120901>	vocab:hasFactoid	<https://www.poms.ac.uk/rdf/entity/Factoid/57978>
<https://www.poms.ac.uk/rdf/entity/PersonReference/124171>	vocab:hasFactoid	<https://www.poms.ac.uk/rdf/entity/Factoid/57978>
<https://www.poms.ac.uk/rdf/entity/PersonReference/127223>	vocab:hasFactoid	<https://www.poms.ac.uk/rdf/entity/Factoid/57978>
<https://www.poms.ac.uk/rdf/entity/PersonReference/129970>	vocab:hasFactoid	<https://www.poms.ac.uk/rdf/entity/Factoid/57978>
<https://www.poms.ac.uk/rdf/entity/PersonReference/132401>	vocab:hasFactoid	<https://www.poms.ac.uk/rdf/entity/Factoid/57978>
<https://www.poms.ac.uk/rdf/entity/PersonReference/134423>	vocab:hasFactoid	<https://www.poms.ac.uk/rdf/entity/Factoid/57978>
<https://www.poms.ac.uk/rdf/entity/PersonReference/136084>	vocab:hasFactoid	<https://www.poms.ac.uk/rdf/entity/Factoid/57978>

Fig. 10: RDF data attached to a *PoMS* transaction factoid.

- The resulting data range is shown in a human-readable form as the value of the *hasFirmData* property (RDF assertion 6 in fig. 10), and the property *hasDatingNotes* (assertion 5) explains why these dates are proposed.

- As well as showing the date in a human friendly way, the date is broken down into its constituent parts to make it machine processible by means of a set of *hasFrom* and *hasTo* attributes. You can see the year and months for the range given in machine-friendly numbers elsewhere in the list.

Associated places involved in the transaction are given through three *hasPlace* attributes, which provide URLs that identify the three places we mentioned earlier.

PoMS's historians were interested in certain kinds of formulaic language that appeared in the charter text. One can find assertions related to the Spiritual Benefits associated with the transaction, for example. The use of other formulaic language was also captured: the use of the 'Dare' phrase, the kind of 'Sicut Clause' and the 'Pro Anima' and 'Tenendas' clauses that appeared.

Finally, one can see the linking to information about people associated with this transaction at the bottom of fig. 10. There are ten links to persons visible here, but there are more links below this part of the display. In fact, twenty-one people are connected to this transaction, most of them as witnesses. For each reference one can see who they are, and what their role is in the transaction by following the URL link to their Person Reference entities.

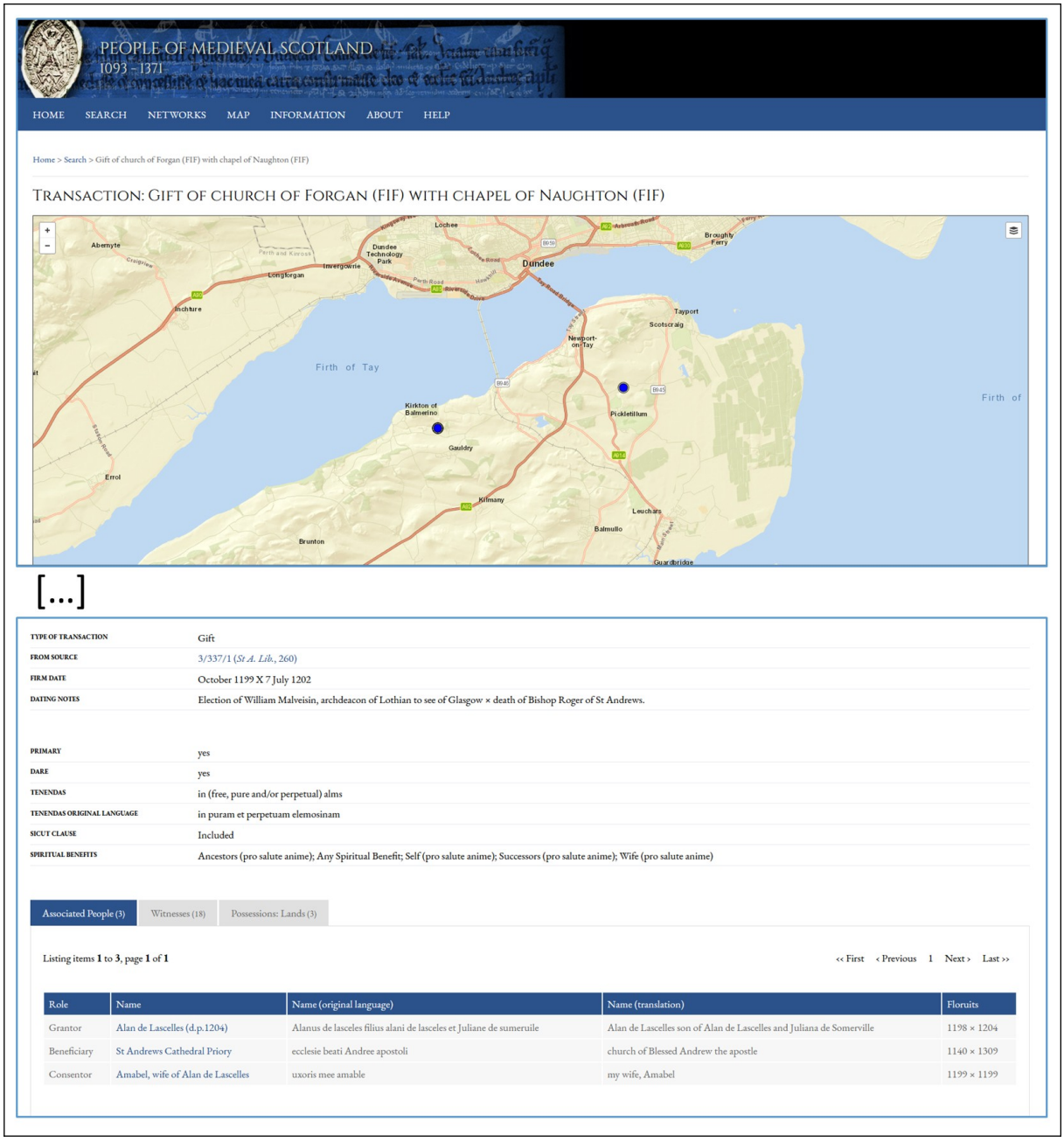


Fig. 11: *PoMS*'s web application's display of a transaction factoid.

One can see that this raw-data approach does give one access to the materials that *PoMS* holds about this transaction. However, it is difficult to see things in it. It is not, to use the well-worn phrase, user friendly. This is where a web application comes in. It packages up the data like this in a way that allows a human user to more readily grasp what is going on. Fig. 11 shows the same material – this particular transaction – as *PoMS*'s web application presents it. One can see this display live with the following URL: https://www.poms.ac.uk/record/factoid/57978 (consulted 12.03.2024).

PoMS's web application has been programmed to start with a heading as a label for the rest of the display (indeed the reader might recall the place where this was entered in fig. 8). This heading is intended to ground the human viewer, and the web application has programming in it to assemble this label from a couple of attributes from the transaction itself and display it here for that purpose. The next thing one sees is a map showing the main geographic locations that are coded as places in the attributes connected to the transaction. We can immediately see a benefit in the web application approach over the simple data access we saw in fig. 10: the application can be coded so that, as it does here, it automatically generates this map display directly from the place data we had in our database.

The second part of fig. 11 shows the contents of the display below the map: a more prosaic display of the rest of the data for our transaction. This may be less technically exciting than the map was, but it has been designed to serve the human user more readily than the display of the RDF data does. The firm date is shown more or less exactly as it appeared in the RDF display, but other items are more significantly transformed. In particular, attributes which pointed to what we call authority list items such as value of the Tenendas Clause or Sicut Clause are shown with their corresponding names that immediately mean something to the human user rather than their associated URI identifiers. Even more significantly, the references to persons have been resolved into actual names, to the roles they have been assigned, and to the floruits in which they appear in the historical record.

Note as well horizontal clickable headings for 'Associated People', 'Witnesses' and 'Possessions' which groups these associated entities into semantically useful groups. All this material could simply have been listed one after the other in the display. Instead, the reader can choose one of the categories here to focus on it. This was a decision of the historians to present it this way.

Overall, the screen layout, with its use of highlighting and positioning has made the data much more accessible than it was in the dump of raw data we saw earlier. This was a consequence of some thinking both by the technical graphic design specialists that worked with the *PoMS* team and by the historians who considered how the layout could best be designed to be effective for other historians. Thus, the layout of the page was not purely a technical issue. It had to be dictated by the historical demands of the project. The page needs to speak to historians.

The software to slot the data into the various places required to display the transaction in this way required sophisticated technical work with the *Django* software framework. The mapping, of course, required even more effort. All this work had to be carried out by a technical team. However, these technical programming tasks that displayed this data for this part of the web application was guided by both graphic design and historical principles. Since the historians had a good understanding of *PoMS*'s data structure they were in the best place to do this, and the layout, as a consequence, reflected their interests and they felt in control of it.

Of the various projects in which I have been involved, the design details of how their material was to be presented varied – based on the interests of the historians, and the content of their project's respective data sets. This functional range in design could be quite significant. Fig. 12 shows how person data is shown in the *Prosopography of the Byzantine World* (*PBW*)[26] and the *Clergy of the Church of England Database* (*CCEd*)[27]. It shows only the beginning of each of the full displays.

In the top area we see the display for *PBW*'s historical person Aaron 101. The UR[28] will generate the same display. In *PBW* the header acts as an historical display name for the person, and is followed by a floruit date 'E / M XI': early- or mid-ninth century. There is a limited other general attributes stored in the *PBW* database for its persons. However, at the bottom one can see the area reserved for the factoids associated with Aaron 101: there are 23 factoids that assert that this person held a dignity or office, etc. Details about them can be seen by unfolding the area by clicking on it. The figure has the Narrative factoid area unfolded, and you can see the beginning of list of 17 'Narrative' factoids associated with Aaron 101.

In contrast, the *CCEd* display for its historical persons prioritises a person's career path. This is a consequence of the nature of the sources from which *CCEd* was built, and the interest of the historians who wanted *CCEd* to focus on the careers of their clergy.

26 Jeffreys et al (2016); https://pbw2016.kdl.kcl.ac.uk/ (consulted 12.03.2024).

27 Burns et al. (2017); https://theclergydatabase.org.uk/ (consulted 12.03.2024).

28 https://pbw2016.kdl.kcl.ac.uk/person/Aaron/101/ (consulted 12.03.2024).

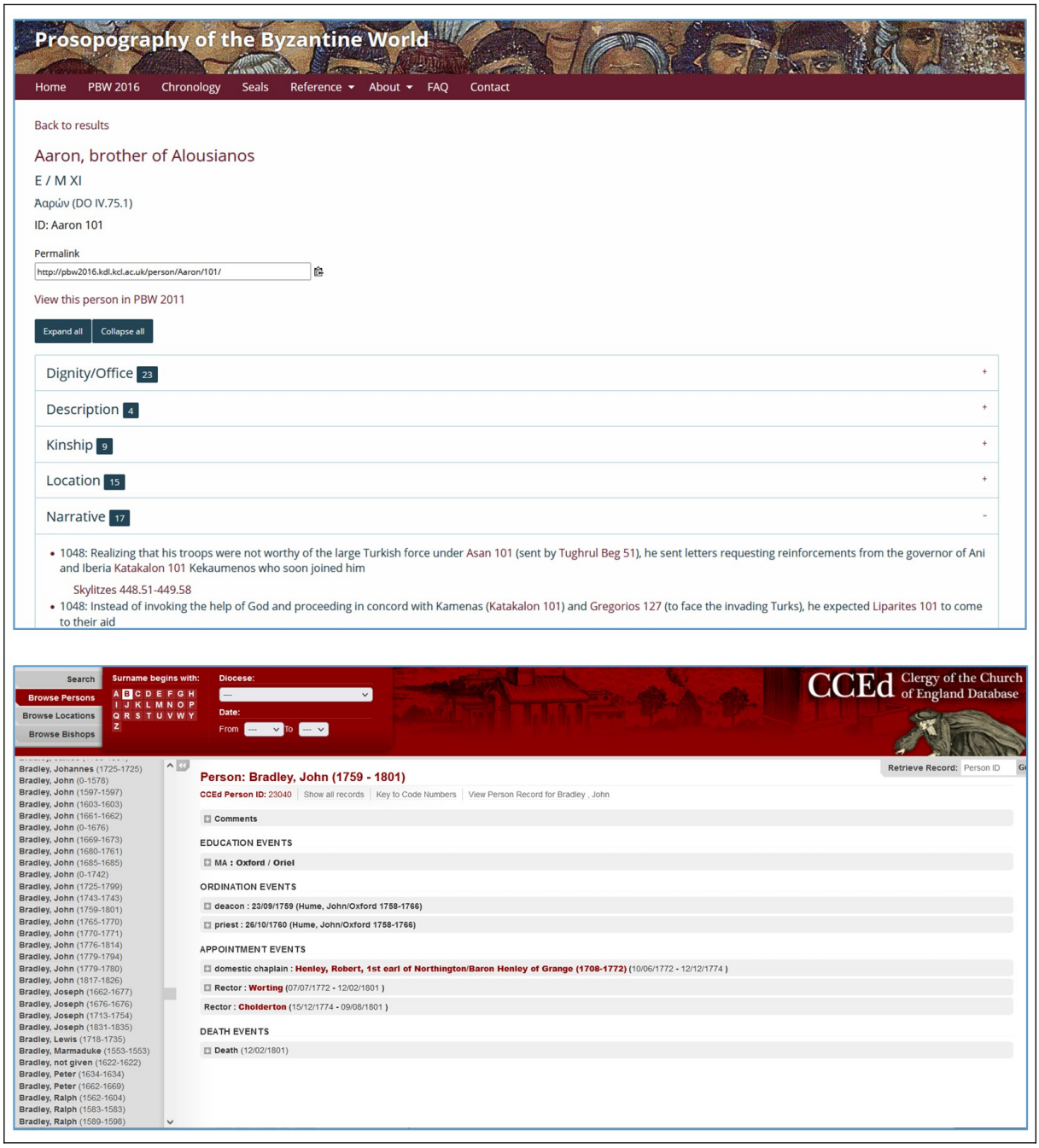


Fig. 12: Displaying person data in *PBW* and *CCEd*.

The difference between the two displays for historical persons was thus caused not only by the difference in the semantic nature of the structure of the database, but also by the interests of the historians: who for *PBW* could not organise the data they had usefully into a career path and those for *CCEd* who could. Both these displays were based on historical considerations, rather than technical ones.

There need not be a 'one style fits all' approach when it comes to how the web application can be created to display information drawn from the database. Although the coding work to make the displays operate will have to be done entirely by the technical team, it is clear that it requires an understanding of the meaning (or semantics) of the dataset, as well as an understanding of how things should be presented to most benefit historian users. Thus the decisions about how to present the data lies, at least in good part, in the domain of the historian project partners, and as long as the historians have a strong understanding of the data model so that they understand exactly what is possible to expect from the data, those partners can, and indeed must, have a prominent role in what gets built.

Navigation and Data

The presentation of the data is not the only task for a web application. Users also have to have ways to find the things that interest them in these often large and complex datasets – a task often described as navigation. What strategy can be employed to help readers find items that are of interest to them?

When one thinks about the situation in which users find themselves while using databases like these, it becomes clearer what overall paradigm should be. *PASE* has about 20,000 people in it, *CCEd* has nearly 200,000. Simply reading, or even skimming them, is not an option. Instead, researchers will usually find themselves thinking of criteria that will focus in on the people they might be interested in. The web interface should, then, model this criteria-driven approach. Fortunately, this approach fits well with how web applications for databases actually work. In almost all of them all actions involved in getting material from the database – including supporting finding things in it – involve what is called database *querying*. The paradigm is not so much one of exploring or browsing the data; one must think instead of asking questions of it. And all the standard database paradigms provide their own highly formal *query language* that a web application must use to do exactly this.

It is possible to see what interaction through a formal query language with a collection of data is like, since the *Digital Prosopography of the Roman Republic (DPRR*) data is, like *PoMS* shown earlier, available as RDF through its RDF server[29]. Since the data is in RDF format, RDF's query language SPARQL[30] is made available for *DPRR*'s RDF portal. Here is a simple query in SPARQL that works with the *DPRR* RDF server:

```
PREFIX: <http://romanrepublic.ac.uk/rdf/ontology#>
SELECT ?Source (COUNT( ?assertions ) AS ?NumbAssertions )
WHERE { ?assertions a :Assertion ;
:hasSecondarySource ?secondarysource .
?secondarysource a :SecondarySource ;
:hasAbbreviation ?Source . }
GROUP BY ?Source
ORDER BY DESC( ?NumbAssertions )
```

There will be more to say about RDF and its SPARQL query language later, but for now the main thing to understand is that it is not a straightforward thing to write. Like all the mainstream query languages, SPARQL is rigid and highly formulaic. And its cousin query languages: SQL for relational databases, XQUERY for XML-based databases, etc., are all similarly formal and not trivial to master. It takes significant intellectual effort to sort out how SPARQL works in general, and after that one must then understand the data structure you are querying well enough to be effective. As we saw earlier, just as simply showing the data exactly as it is is insufficient to allow the data to be used readily in historical research; exploring the data through SPARQL queries is also outside the range of most historians.

As a consequence of this, it is common practice to put mechanisms into the web application that allow users to specify criteria in ways closer to that they can readily understand. The application's job then becomes taking these criteria, translating them into a formal data query and translating the results returned from the database into an understandable HTML display that a user can see on his/her browser. Fig. 13 shows this process underway in the *Clergy of the Church of England Database* . It is described there as a *search* , and presents the user with a form containing specific fields that can be used to specify criteria for selection.

29 https://romanrepublic.ac.uk/rdf (consulted 12.03.2024).

30 Harris / Seaborn (2013); https://www.w3.org/TR/sparql11-query/ (consulted 12.03.2024).

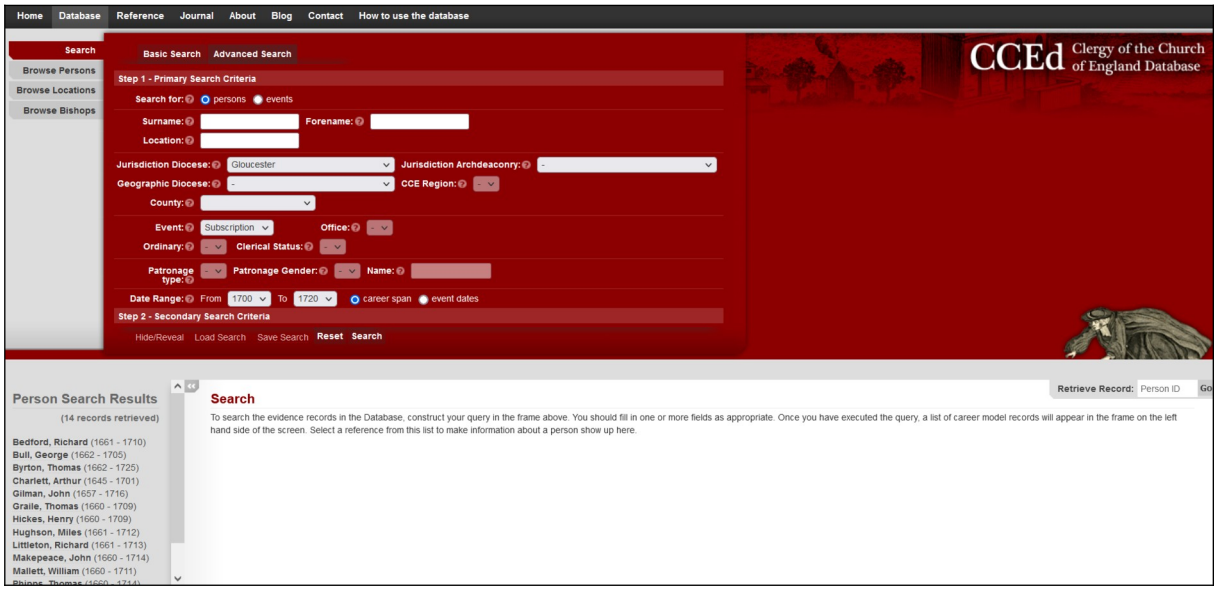


Fig. 13: A query and results for the *Clergy of the Church of England Database*.

The form allows you to specify names (or just name beginnings), specific locations, a diocese or archdeaconry, and various other criteria for filtering. The *CCEd*'s web application then uses the specified criteria to filter the persons and select ones that match all the criteria that were given. Fig. 13 shows a search that asks for people who served in the Diocese of Gloucester, who have subscription information associated with them, and who had careers under way during the period 1700–1720. When this Search button is pushed the criteria are turned into a query in SQL and it is run against the database. The bottom left area shows the first twelve of the fourteen people that match the criteria.

There is a basic and advanced search (with more options for criteria). What options were chosen to be included in the form as criteria were selected by *CCEd*'s historians, and to do this they needed an understanding of what the structure was in the database, and from that, some sense of how these criteria could be combined together into a single query so that historically valid and useful results would be generated.

This kind of 'select by form' approach is widely used as a search-oriented navigation mechanism. However, one of the problems with this relatively simple technical approach is that it is possible, indeed sometimes far too easy, for a searcher to create combinations of criteria for which none of the data matches the criteria and as a consequence nothing is returned.

My technical team considered the problem in the early 2000s and adopted a strategy in 2006 for search/navigation called *faceted browsing*. The first project in which it was used was *British Printed Images to 1700H*[31], and we have continued to use it for data-oriented projects such as our prosopographies ever since. By using the faceted approach one avoids 'dead end' criteria combinations – which result in nothing being selected – from arising. The most significant cost to using this approach is that the software in the web application must be an order of magnitude more complex than the basic 'criteria from a form' approach shown in fig. 13 for *CCEd*.

The ideas in faceted browsing grew out of the closely related idea of *faceted classification*: an approach to the library classification of books that was invented by the great Indian library scientist Shriyali Ramamrite Ranganathan. The central idea here was that one classified books along multiple paths or facets (personality, matter, energy, space, and time) in a single multi-component classification entry.[32] By 2003 people were thinking of how the principles of faceted classification could be applied

31 Hunter (2007). This site has been moved to a new host, and no longer uses the faceted approach.

32 Ranganathan (1962).

to the World Wide Web. We were much influenced by William Denton's article[33], in which he presented an argument about why a multi-faceted approach could be useful, and put the ideas from faceted browsing into the context of designing a faceted scheme for a collection of data that he claimed kept users properly in mind.

The faceted approach resonated particularly strongly with us when we realised that the elements in the search form for *CCEd*, for example, could be viewed from a multi-faceted perspective: the facets being the clergyman's name, the places where he had a living, the diocese to which he was responsible, the dates in which he held posts, and several others.

Several of our faceted search mechanism take on the form of this kind of *faceted browsing*. *PoMS* provides an example of this.[34] The facets for selecting people – such as historical persons' names, historical sources in which they appear, titles they held, roles and personal relationships, kinds of property, and many more – are listed down the left-hand side. A user selects one of these, and two things happen: the list of people who are connected to that facet appear on the right, and the list of facets is adjusted so that the list contains only those still available among the historical people selected by the first facet selection – in this way one avoids allowing a user to select two or more criteria that, when combined, apply to none of the historical persons. They can now select a second selection criteria knowing that at least some historical figures have both their first and second criteria applying to them. The experience of the user is that they select people by a sequence of filtering steps, hence the term *browsing*.

An alternative approach to faceted searching was taken on for *DPRR* (*Digital Prosopography of the Roman Republic*). You can see two screen shots in fig. 14 suggesting how it works by showing the first two steps in an exploration. The readers can, alternatively, go to the *DPRR*[35] search facility itself and try it out for themselves.

The top panel in fig. 14 shows the top part of the search screen as it appears when one first sees it. At a quick glance it appears similar in style to the 'criteria by form' mechanism used for *CCEd* and illustrated in fig. 13. However, something that is relevant to the faceted search mechanism has been added: there are the numbers beside many of the items which specify how many historical people have that particular attribute assigned to them. For example, beside the status of *nobilis* is the number 264, meaning that 264 individuals have been recorded as having the status of *nobilis* in *DPRR*. We can only see a part of the full search form here, but from it one does get a sense of the range of data recorded in *DPRR* for historical persons from this display, and consequently the range of facets they can specify. Filtering by names is possible in the top area, filtering by a range of kinds of personal data (including life dates) from the second, and by career in the third.

Suppose a user wishes to focus on adopted people. She can see under 'Life Events' in the Personal Data box that *DPRR* has 64 people recorded as adopted. If she clicks on this the search engine will then focus on only those people who are adopted, and the screen changes to that shown in the bottom panel of fig. 14. It looks quite similar to the panel above, but there are changes. First, almost all the numbers have changed. Whereas, when the full set of people were included there were 264 classified as *nobilis*, now, among the adopted people there are only 15. Furthermore, the Female group has entirely disappeared because *DPRR* has no adopted women recorded. Similarly, the Life Events-Restored category has disappeared because there are no records of adopted people being restored.

A major theme of this paper is that historians need to hold on to the structure and the semantics it implies if the work that is produced is to be satisfactory as a piece of digital publication for historians.

---

33 Denton (2003).

34 https://www.poms.ac.uk/search/ (consulted 12.03.2024).

35 https://romanrepublic.ac.uk/person/ (consulted 12.03.2024).

This paper has shown several examples of why this is necessary and how it works, and this same need extends to the selection of the facets, and the semantics behind them too. If the web application is to support a particular facet the developers need to know how to query the data in the backing data structure to make it happen. Thus, a direct connection between all the facets and the data structure can, and indeed must, be identified if they are to work properly from a historiographic point of view. Historians have to be engaged in establishing what facets should be made available to the user, and they must ensure that the data structures enable the facets they want to provide. Although the technical work to set up facets is a complex technical one, it cannot be done satisfactorily by the technical team alone.

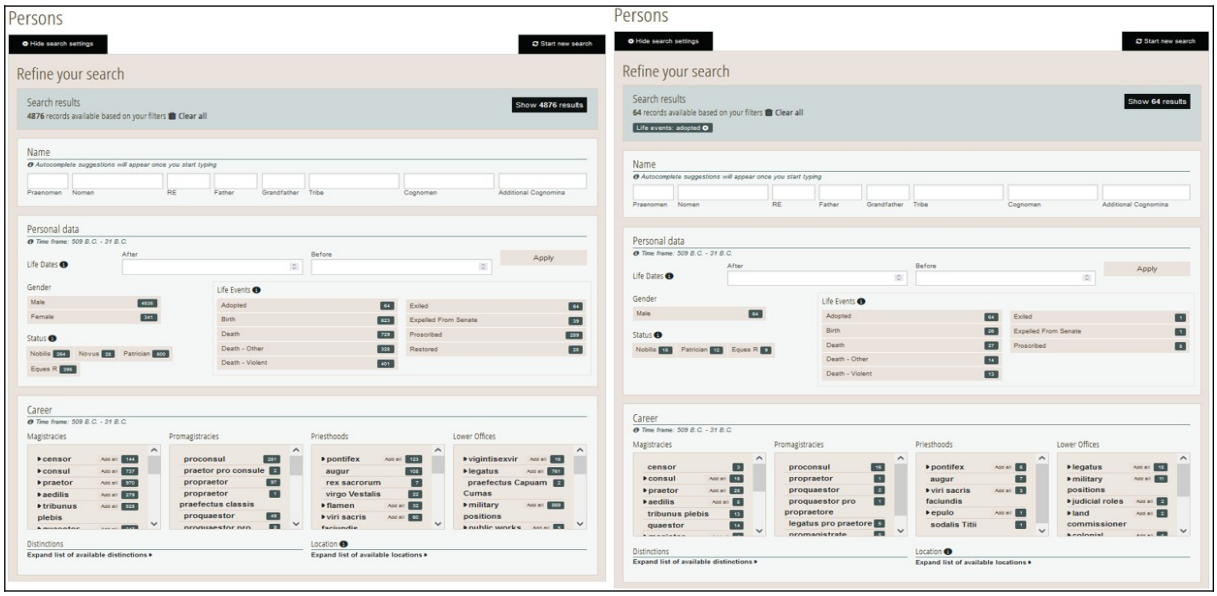


Fig. 14: Faceted searching in *DPRR*.

There is a final point that I can make about searching through data with faceted browsing, and it is this: although faceting browsing has worked well in almost all of our projects' web apps as a way to help historian users engage reasonably effectively with our projects' data, the faceted approach has limitations. The thinking involved in creating faceted browsing results in a mechanism which enables only certain pathways through the data to be made available: they only allow certain types of queries to be asked. In the end, there will always be questions that could be legitimately asked (through query languages such as SQL or SPARQL) of the data that any manageable facet approach will not be able to deal with.

An example can be found in the SPARQL query that was shown earlier in this paper. The point then was that using the query language against the data directly was a complex thing, beyond the capability of most historians. However, this particular query, written in the SPARQL query language, asks the *Roman Republic Prosopography* (*DPRR*)'s RDF server to count the number of assertions made for each of the sources that *DPRR* draws on. From which sources did the project team draw on heavily, and which were used quite a bit less? Although *DPRR* has a rich set of facets that support faceted browsing, it does not provide a way to ask this question. If this query is given to *DPRR*'s RDF server it sends back a list of the sources with the count of the number of assertions that came from each one. Furthermore, unlike the web-browser orientation that has been seen above, the tabular result is available in a form that allows for further independent processing.

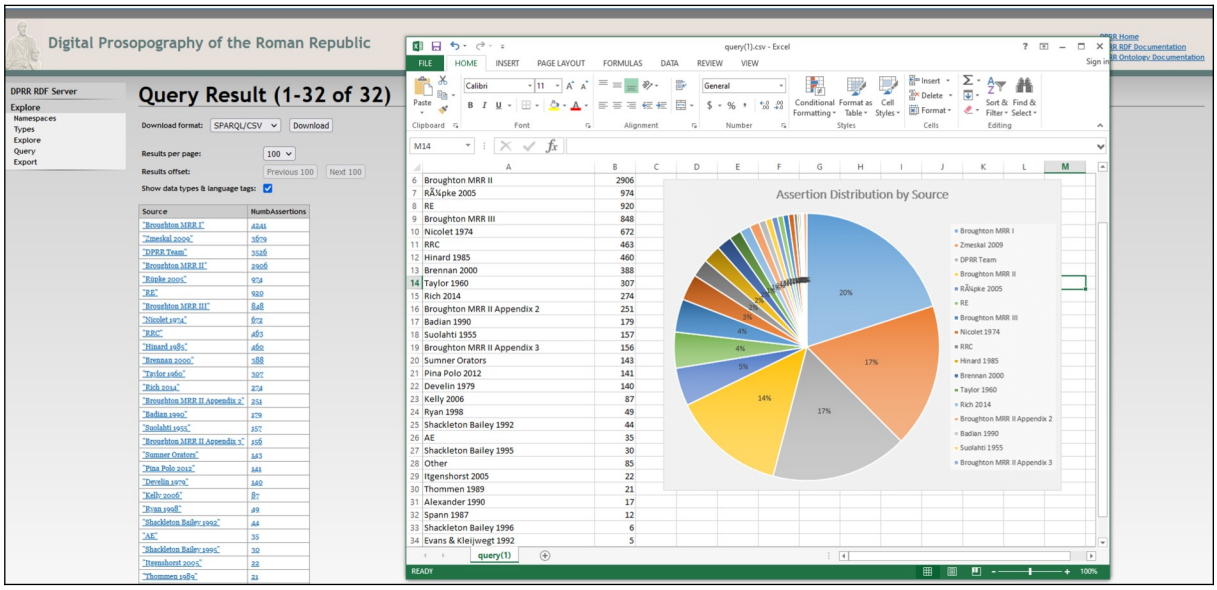


Fig. 15: *DPRR* sources and their contribution.

Fig. 15 shows a simple example of this. The tabular result of the query (shown on the left) has been imported directly into an Excel spreadsheet, and from it a pie chart has been generated; providing a visualisation of the results.

This simple example illustrates that although we have found that the web app approach with faceted browsing creates a web experience that is understandable and usable by most historians, it can only give specific kinds of access to the data: along paths that are predefined in the facets. It also limits what historians can readily do with the results: look at them. In contrast, direct access to the data through one of the standard query engines allows one to ask any kind of question that the data structure is capable of answering. Furthermore, unlike the web application approach which focuses on presenting results for human viewing on the browser, standard query systems such as SPARQL return results in the form of data that allows for further processing. In experiments with our RDF servers we've gone beyond relatively straightforward processing such as Pie Charts to including passing data to geographic mapping systems, and creating data that could be given to more sophisticated analysis tools, such as the Social Network Analysis tool *Gephi*.[36] Anyone who is prepared to learn a significant amount of technical knowledge can do this with our RDF servers. The raw data for projects such as ours can support kinds of research that cannot be carried out with our given web application mechanisms.

Summary and Conclusions

In my own experience over many projects I have found that the execution of a data-oriented historical project is a complex venture best viewed as an extended partnership between the historians and a technically oriented team. Although, of course, the product of such a project must be something valid and useful to historians, who thus must of necessity lead on issues involving the meaning of results, the nature of digital data representation means these projects cannot be carried out without engaging technologists in a complex and continuous way.

In the approach to digital resource building we have used across many projects, project work begins with the definition of data structures that are meant to represent the ideas the project wishes to represent, and developing this, in and of itself, requires a long-term partnership between the historians and

36 https://gephi.org/ (consulted 14.04.2025).

the technical team who must jointly understand how to use the forms and structure of data engines such as databases to represent materials.

To achieve this one must be sure that historians engage closely and carefully with the database structure. The formalisms of a database can be quite expressive, but their semantics are also quite rigid, and historians need to be sure that they fully understand what is being represented, and how. This does require a focused dipping into the technical concepts behind data structuring on the part of the historians, but the formal conceptions of structured data are not so complex as to make this impossible. But they can be subtle – so historians should invest some time to understand fully how they work.

Furthermore, one should not underestimate the amount of initial work needed to sort out how the database should be structured even before building anything digital: the more one gets right from the beginning the better off one will be, because retrofitting in changes, once the web application is being developed in particular, can often require quite a bit of technical effort and delay. One should expect a complex model to be evident before any significant coding work is undertaken. In addition, whatever design one come up with, one should expect to have to change it perhaps more than once, and perhaps even quite late in the project. A supportive, but intellectually critical engagement between the technology and the history partners is essential to make this practical. Overall, one needs technical people with a deep understanding of how to do this on-board and onside throughout the project, and the historians will need to respect what their technical partners know and contribute.

After a database is built, and mechanisms are in place to allow material to be put into it, one will need to make it available to the general community. There are various approaches to this, but for the projects in which I have been involved, this has always meant the construction of a web application that sits between the database and the outside world who will access it using browsers and the World Wide Web. The technical work for creating the web application requires expertise in quite different technologies than what are needed for database design – expect to need more than one person on the technical team. Furthermore, the work to create a web application is often more complex and subject to error than the database work is.

- The database structure will drive what the web application can do. While it is being built, you might expect to discover that changes to the underlying structures are necessary, perhaps even quite late in the day. Expect this in the project's planning.

- One needs to think about how materials in the database are to be represented as WWW HTML pages. This requires historian expertise and active critical participation, but will need to be filtered through technical understanding of web page design.

- One also needs to think about how a user community will navigate: how they will find things of interest to themselves. We have taken on the faceted browsing approach for this, but it is both conceptually and technically complex to implement, and you need technically sophisticated team members to make it work. Overall, the amount of work to do to build it should not be underestimated.

As mentioned briefly earlier, recent thinking has developed an approach to web-oriented presentation that removes the server-based web application and instead puts all the user interaction into the HTML pages using Javascript. However, even when a complex web server is replaced by complex Javascript run in the browser, similar complex development work still needs to be done. Overall, then, whether a web application is used or not, to make results available over the WWW projects require an enduring partnership between a technical team of probably more than one person (with different expertise), and the historians. This must be funded and planned for from the beginning.

The idea of an intellectual partnership between scholars and non-scholars in the academy is not a particularly new one. See, for example, the comments of the historian of religion Jaroslav Pelikan from his book *The Idea of the University*. About those "who usually stand outside the classroom but without whom research would halt", he wrote as follows:

> "Indeed, even such a term as 'providers of support services' is becoming far too limited to describe both the skills and the knowledge required of those who hold such positions. Scholars and scientists in all fields have found that the older configurations of such services, according to which the principal investigator has the questions and the staff person provides the answers, are no longer valid, if they ever were; as both the technological expertise and as the scholarly range necessary for research grows, it is also for the formulation and refinement of the questions themselves that principal investigators have to turn to 'staff', whom it is increasingly necessary – not a matter of courtesy, much less as a matter of condescension, but as a matter of justice and of accuracy-to identify instead as colleagues in the research enterprise."[37]

I was invited to say a few words at the launch of *The Making of Charlemagne's Europe* project[38], and I chose to comment on the intellectual partnership that I had found there:

> "To me, it has been striking that the historical team was evidently so highly committed to exploring and developing sophisticated models for their data – particularly around the some of the complex aspects of historical charters of charters – say, for example, possessions. In my view this has been one of the real strengths of the partnership between the history and the digital humanities aspects of this project. Overall, then, the work on Charlemagne can be characterised by a strong sense of commitment, a high degree of intellectual engagement, often a strong sense of excitement, sometimes a sense of adventure (!), and I hope we are now finishing this project with a strong sense of satisfaction at the quality of work that we have achieved."

37 Pelikan (1992), 62.

38 https://charlemagneseurope.ac.uk/ (consulted 12.03.2024).

Sources

Online References

Broun et al. (2014): D. Broun / D. Carpenter / M. Hammond / J. Bradley, People of Medieval Scotland, online 2014, https://www.poms.ac.uk/ (consulted 12.03.2024).

Burns et al. (2017): A. Burns / K. Fincham / Stephen Taylor, CCEd: Clergy of the Church of England Database, online 2017, https://theclergydatabase.org.uk/ (consulted 12.03.2024).

Bradley (2018): J. Bradley, Factoid Prosopography, online 2018, https://www.kcl.ac.uk/factoid-prosopography (consulted 12.03.2024).

Cyganiak et al. (2014): R. Cyganiak / D. Wood / M. Lanthaler, RDF 1.1 Concepts and Abstract Syntax, online 2014, https://www.w3.org/TR/rdf11-concepts/ (consulted 12.03.2024).

Harris / Seaborne (2013): S. Harris / A. Seaborne, SPARQL 1.1 Query Language, online 2013, https://www.w3.org/TR/sparql11-query/ (consulted 12.03.2024).

Hunter (2007): M. Hunter, British Printed Images to 1700, online 2007, https://bpi1700.org.uk/ (consulted 12.12.2024).

Jeffreys et al. (2016): M. Jeffreys et al., Prosopography of the Byzantine World, online 2016–, http://pbw2016.kdl.kcl.ac.uk (consulted 12.03.2024).

Mouritsen et al. (2016): H. Mouritsen / D. Rathbone / J. Bradley, Digital Prosopography of the Roman Republic, online 2016–, https://romanrepublic.ac.uk (consulted 12.03.2024).

TEI (2022): The Text Encoding Initiative, P5: Guidelines for Electronic Text Encoding and Interchange, online 2022, https://tei-c.org/release/doc/tei-p5-doc/en/html/index.html (consulted 12.03.2024).

References

Bizer / Heath / Berners-Lee (2009): C. Bizer / T. Heath / T. Berners-Lee, Linked Data – The Story So Far, International Journal on Semantic Web and Information Systems, 5/3 (2009), 1–22, DOI: https://doi.org/10.4018/jswis.2009081901 , http://tomheath.com/papers/bizer-heath-berners-lee-ijswis-linked-data.pdf (consulted 12.03.2024).

Bradley (2020): J. Bradley, A Prosopography as Linked Open Data: Some Implications from DPRR, Digital Humanities Quarterly, 14/2 (2020), https://digitalhumanities.org/dhq/vol/14/2/000475/000475.html (consulted 12.03.2024).

Bradley (2023): J. Bradley, Four Corners of the Big Tent: A Personal Journey through the Digital Humanities, in: J. Nyhan / A. Ortolja-Baird / G. Rockwell (eds.), On Making in the Digital Humanities. The Scholarship of Digital Humanities Development in Honour of John Bradley, London, online 2023, https://www.uclpress.co.uk/products/211149 (consulted 12.03.2024).

Bradley / Rio / Hammond / Broun (2019): J. Bradley / A. Rio / M. Hammond / D. Broun, Exploring a Model for the Semantics of Medieval Legal Charters, International Journal of Humanities and Arts Computing, 13/1 (2019), 136–154, DOI: https://doi.org/10.3366/ijhac.2017.0184.

Brennan (2000): T. C. Brennan, The Praetorship in the Roman Republic, Oxford 2000.

Denton (2003): W. Denton, How to Make a Facet Classification and Put it on the Web, Library Trends, 48/1 (2003), 22–47, https://www.miskatonic.org/library/facet-web-howto.html (consulted 12.03.2024).

Fodor (2007): J. Fodor, Semantics: An Interview with Jerry Fodor, ReVEL, 5-/8 (2007).

Gruber (1993): T. R. Gruber, Toward Principles for the Design of Ontologies used for Knowledge Sharing, Technical Report KSL 93–04, Knowledge Systems Laboratory, Stanford University, 1993.

Pelikan (1992): J. Pelikan, Jaroslav, The Idea of the University. A Reexamination, New Haven 1992.

Ranganathan (1962): S. R. Ranganathan, E lements of Library Classification. New York 1962.

Stone (1971): L. Stone, Prosopography, Daedalus, 100(1) 1971, 46–79.

Thompson (1841): T. Thomson (ed.), L iber Cartarum Prioratus Sancti Andree in Scocia, Edinburgh 1841.

Figure References

Fig 1: Figure created by author.

Fig 2: Underlying text from Thompson 1841. Labels added by author.

Fig 3: Figure created by author.

Fig 4: Figure created by author.

Fig 5: Figure created by author. An extension to figure in https://www.kcl.ac.uk/factoid-prosopography/about (consulted 12.03.2024).

Fig 6: From https://www.kcl.ac.uk/factoid-prosopography/overall-concepts (consulted 12.03.2024).

Fig 7: From https://poms.ac.uk/rdf/doc/ontology.html (consulted 12.03.2024).

Fig 8: Screenshots of administrative interface for the *PoMS* project (Broun et al. [2014]).

Fig 9: Screenshot of spreadsheet created for the *DPRR* project (Mouritsen et al. [2016]).

Fig 10: Screenshot of output generated by query from https://poms.ac.uk/rdf (consulted 12.03.2024).

Fig 11: Screenshot from *PoMS* website: https://poms.ac.uk/record/factoid/58027/ (consulted 12.03.2024)

Fig 12: Combined screenshots from *PBW* (Jeffreys et al. (2016)) and *CCEd* (Burns et al. [2017]) projects.

Fig 13: Screenshot from *CCEd* (Burns et al. [2017]) search engine page.

Fig 14: Screenshots from *DPRR* (Mouritsen et al. [2016]) persons search page.

Fig 15: Screenshot created by author.

Author Contact Information[39]

John Bradley
Honorary Senior Research Fellow
King's College London
Strand, London UK WC2R 2LS
E-mail: john.bradley@kcl.ac.uk

39 The rights pertaining to content, text, graphics, and images, unless otherwise noted, are reserved by the author. This contribution is licensed under CC BY-SA 4.0.