DHQ: Digital Humanities Quarterly
2020
Volume 14 Number 2

A Prosopography as Linked Open Data: Some Implications from DPRR

John Douglas Bradley <john_dot_bradley_at_kcl_dot_ac_dot_uk>, King's Digital Lab, King's College London

Abstract

The Digital Prosopography of the Roman Republic (DPRR) project has created a freely available structured prosopography of people from the Roman Republic. As a part of this work the materials that were produced by the project have been made available as Linked Open Data (LOD): translated into RDF, and served through an RDF Server. This article explains what it means to present the material as Linked Open Data by means of working, interactive examples. DPRR didn't do some of the work which has been conventionally associated with Linked Open Data. However, by considering the two conceptions of the Semantic Web and Linked Open Data as proposed by Tim Berners-Lee one can see how DPRR's RDF Server fits best into the LOD picture, including how it might serve to facilitate new ways to explore its material. The article gives several examples of ways of exploiting DPRR's RDF dataset, and other similarly structured materials, to enable new research approaches.

In a TED conference in 2009 Tim Berners-Lee gave a presentation entitled *The Year open data went worldwide* [Berners-Lee 2010]. In it he gave some examples of how open data from “governments, scientists and institutions” could be used to make significant statements about the state of affairs in society. He then asked governments, scientists and institutions to support this kind of work by making more of their data freely available in a form where it could be further processed rather than just looked at. This was a part of what he called the “Linked Data” Initiative, and which has more recently often been called “Linked Open Data” (LOD). 1

Berners-Lee has a scientific background, so perhaps it was not surprising that he didn't seem to think about LOD data from the humanities. Nonetheless, there is no technical reason why those Digital Humanists who have suitable data should not be be making their material openly available too. Indeed, the desirability of Linked Open Data from and for the humanities has been expressed by others in the digital humanities community for some time. One can, for example, see a similar motivation in the premise behind the workshop *Linked Data for Digital Humanities* that was held in the 2016 DHOxSS Summer School [Nurmikko-Fuller 2016], James Smith's *RDF and Linked Open Data* [Smith 2017] at University of Victoria's Digital Humanities Summer Institute, and in other similar workshops that explore the ideas of applying Linked Open Data technologies to humanities-oriented materials. 2

This paper introduces recent project work done at King's College London which makes one of its many online web resources available as Linked Open Data. The project is the recently completed *Digital Prosopography of the Roman Republic* (DPRR), and this part of its work was a response to Berners-Lee's TED talk challenge mentioned above in which he asked people to deliver their data as LOD. This paper is also a part of this response. Here we will consider why DPRR was the first humanities project by KCL to have its full set of data published as LOD data, and what in DPRR's characteristics made it particularly suitable for this. The paper will then explore what DPRR's LOD server looks like to a user, what kind of interactions with the data it makes possible, and how this connects with Berners-Lee's view of how LOD should be expressed. Having done this, it will then consider what might come from allowing anyone to get at this historical material directly as pure data rather than exclusively through a browser oriented front end which, as we will see below, acts as a focusing filter between the material-as-data and the user's browser. Does this direct access truly empower people to explore our material in the way Berners-Lee and other people who have taken up the LOD cause intend? Does it allow for new kinds of analysis and research to be carried out, hopefully revealing new insights for 3

the materials that are not visible through even our rather sophisticated browser-oriented front end? Some part of this issue arises out of research about the nature of querying that has been carried out in the context of the Semantic Web, and we will briefly describe this here; contrasting the original AI-related vision of the Semantic Web in the late 1990s, with the more pragmatic Linked Data vision that emerged a few years later. And finally, how does DPRR's RDF server fit with one of the major interests from the Digital Humanities that have come out of LOD thinking: an interest in adding links from digital resources to standard authority lists such as *VIAF* [VIAF 2010-16] or, say, *Pelagios* [Pelagios nd] to aid in the aggregation of data between different data sets?

Why have DPRR as Linked Open Data?

I started work at King's College London (KCL) in 1997, first at its Centre for Computing in the Humanities which was subsequently renamed the Department of Digital Humanities (CCH/DDH). Most recently I have also become associated with at a new unit at KCL called the King's Digital Lab (KDL). During this period CCH, DDH and KDL have built many academic resources in collaboration with humanities academic partners. In all these projects we have championed the concept of openness and accessibility. As a consequence, since the late 1990s we have, as conscious policy, made sophisticated online digital resources available for free over the WWW. 4

One of the challenges for the work in which I particularly was involved arose out of the fact that almost all my collaborative academic projects took a strong “data” perspective to their materials. This is an approach which can seem very foreign to the text-orientation of much of the humanities and the digital humanities too. Having created a highly structured and complex set of data as a product of the scholarship, how could it be made accessible to the rather non-technical scholarly audience? Out of this issue came much work on how to present this complex data adequately through web applications. Indeed, the development of these delivery apps became a large part of CCH/DDH's standard project practice, and the focus was always (indeed, had to be, due to the intended audience) on making that access as non-technical as possible by creating a web application that mapped each project's data into dynamic web pages that could be displayed by any browser. Thus, although almost all browser-mediated resources created at KCL have been open and freely available, they have not really been conceived as providing direct access to the data *behind* the web application in the way that Berners-Lee meant for LOD in his TED talk. 5

Recently, however, the *Digital Prosopography of the Roman Republic* [Mouritsen et al 2017] was completed, and it presented us with an opportunity to publish the same material in two different forms. First, like all pre-existing data-oriented resources that we had created, DPRR has its web application that made access possible for nontechnical users. Second, however, and at very little additional development cost, DPRR's data has also been made available as pure data, in a form suitable for LOD. 6

Why was DPRR the target for this work? From a fully pragmatic perspective, DPRR came to be expressed as LOD because in its AHRC funded research proposal we actually proposed offering direct data access, in the spirit of LOD, as one of the project's outcomes. Furthermore, RDF and related technologies had been in the mix within the *Sharing Ancient Wisdoms project* (SAWS) project which had been carried out with DDH as a partner, so there was some significant experience of RDF to draw on in previous work. However, we did not do the work of expressing DPRR materials in LOD-compatible ways only because we had promised it in the proposal, or because of our experience with the SAWS project, but because we believed that DPRR connected in particularly useful ways to the three components of the idea of LOD: *openness*, *linked*, and *data*, and we thought it plausible that by opening up DPRR in this way we would allow others to explore more richly what DPRR contains than what our conventional browser-oriented mechanisms, as sophisticated as they are, would enable. 7

First, *Openness*: DPRR is a published prosopography, and we believe that, as such, it offers a highly suitable source for open data. A published prosopography is consciously intended by its creators for a global audience and for this reason it is ideally an open publication and compatible to many of the ideas of open data. This is particularly true for a prosopography which is free to all online, as DPRR is. 8

Linked: DPRR is also a good example of scholarship that invokes the essential spirit and principles of linked data. Of course, DPRR hopes that modern Roman Republic scholars will explicitly link to it by referencing the historical entities 9

— presumably primarily historical people — that it defines. If this happens, DPRR will become integrated into the global scholarship around the Roman Republic. However, DPRR has more significance to linking than just this. DPRR, like any prosopography, establishes formal identities for their historical persons out of the appearance of them in a range of sources, and it thus links these sources together through their shared historical people. However, DPRR takes a different approach to its prosopography than the other, factoid-based (defined in [Bradley 2017a]), digital prosopographies in which DDH/CCH has been involved. Unlike these other prosopographies, such as *the People of Medieval Scotland* [PoMS 2014], or *Prosopography of Anglo-Saxon England* [PASE 2016], which draw almost exclusively on their projects' interpretation of their primary sources, DPRR has assembled and aligned work done by a range of already existing nineteenth, twentieth and twenty-first century prosopographies, and could thus be said, in itself, to represent a multi-source “global graph” (to use RDF terminology) of recent Roman Republic prosopographical scholarship. It, and many of the works upon which it draws, has been built on the work of T. Robert Broughton's study of office-holders [Broughton 1951-2, 1986] which remains to this day a standard reference work. Furthermore, underpinning all these other prosopographies, including Broughton, is the monumental 83 volume nineteenth century *Real-Encyclopaedie der classischen Altertumswissenschaft* [Pauly et al 1893-] — referred to as RE and once called by a DPRR project member the "grandfather" of all DPRR's prosopographical sources. RE continues to provide the basis against which historical identity of individuals is argued even today. A full list of sources that provided data for DPRR can be found on their Bibiography page. [Mouritsen et al 2017] at web page Bibliography.

Data: Finally, DPRR is like DDH/CCH's many other prosopographical projects in that it is data-oriented rather than being, as traditional published prosopography has been, article oriented. Like PoMS or PASE, DPRR represents its materials in the form of highly structured data, and, like DDH/CCH's other structured prosopographies, is built on top of that quintessential highly structured paradigm: the relational database. Thus, DPRR's historical research work has been expressed in terms of the semantic concepts of “entities”, “attributes” and “relationships” as they are thought of in the relational data model. 10

Overall, then, DPRR can be thought of as an ideal candidate for all three aspects of the LOD model: linked, open, and data oriented. 11

To take DPRR's materials in its relational database and to turn it into LOD presented in the forms apparently meant by Tim Berners-Lee requires taking up technologies developed for LOD. Thus, we followed the thinking of the original developers of the Linked Data (LD) concept, and of the rather broader Semantic Web too, in adapting the Resource Description Framework (RDF) [RDF 2014] and its related components as fundamental technologies for expressing DPRR as LOD. RDF links have been described as “the glue of the data web” [Bizer 2008, 1265], and RDF has been given by LD's original thinkers as a key part of Berners-Lee's “four rules” to allow published data to become “part of a single global data space” [Bizer et al 2009, 2]. Furthermore, relational data structures (the paradigm used for organising data in DPRR) generally map particularly well onto RDF. As Berners-Lee remarks about RDF and the Semantic Web: “[O]ne of the main driving forces for the Semantic web has always been the expression, on the Web, of the vast amount of relational database information” [Berners-Lee 1998]. Indeed, exactly because of this thinking within the fundamental design of RDF, the task of mapping DPRR's materials into RDF -- turned out to be conceptually relatively straightforward. 12

In the online descriptive material I have provided about the DPRR RDF server [Bradley 2017b], I describe how I used the *d2rq* tool [D2RQ nd] to map DPRR's database structures into RDF, how I created a basic semantic web ontology to supplement the DPRR data, and how I created an RDF server as a mostly stripped down, but in a couple of areas somewhat extended, version of the *rdf4j* [RDF4J 2017] workbench. I based much of the RDF server on rdfj4's workbench because I believed that it produced quite an elegant thin HTML-based wrapping around the RDF data that allowed a browser user to explore and better understand the data without having its HTML wrapping mask or hide the nature of the RDF. The server's functions are documented at [Bradley 2017b] on web page Using DPRR's RDF server. 13

Exploring DPRR's RDF Server

Where on the WWW, then, does one find DPRR's LOD data representation? One finds its RDF server at 14

http://romanrepublic.ac.uk/rdf/. All URIs and URLs that start in this way are delivered to DPRR's RDF server, and processed by it.

We have built the server to provide support for what we believed to be the main characteristics of RDF-oriented Linked Open Data. What are these characteristics? 15

- The server meets the Linked Data requirements outlined in [Bizer et al 2009]. In particular, it is designed so that, first, all of DPRR's data are given public URIs (although DPRR is a prosopography, not just historical Persons are formally identified with URIs), and second that if any of these URIs is given to the WWW, they will find their way to DPRR's RDF server, which will deliver data it has which is connected to this entity.
- Furthermore, the server supports querying via RDF's standard query language SPARQL (see SPARQL 2013).
- Materials can be fetched as pure RDF data, suitable for further processing, or filtered through a light-weight browser oriented HTML presentation to facilitate human browsing of the data.
- The interconnections between the different entity types that makes up the DPRR data is made evident through the provision of a basic OWL ontology.

The attentive reader may have noticed that I claim here the RDF server is capable of delivering the DPRR RDF data in a browser-friendly manner, and may have remembered that DPRR's *other*, more conventional, browser-friendly interface *also* delivers DPRR data in a browser-friendly manner. What, then, is the difference between the two? 16

Although both DPRR's RDF server and browser-oriented search engine interact with the same data, they present quite a different face to their users. As a point of comparison, Figure 1 shows the top of the front “Person Search” page of DPRR's conventional browser-oriented site: 17

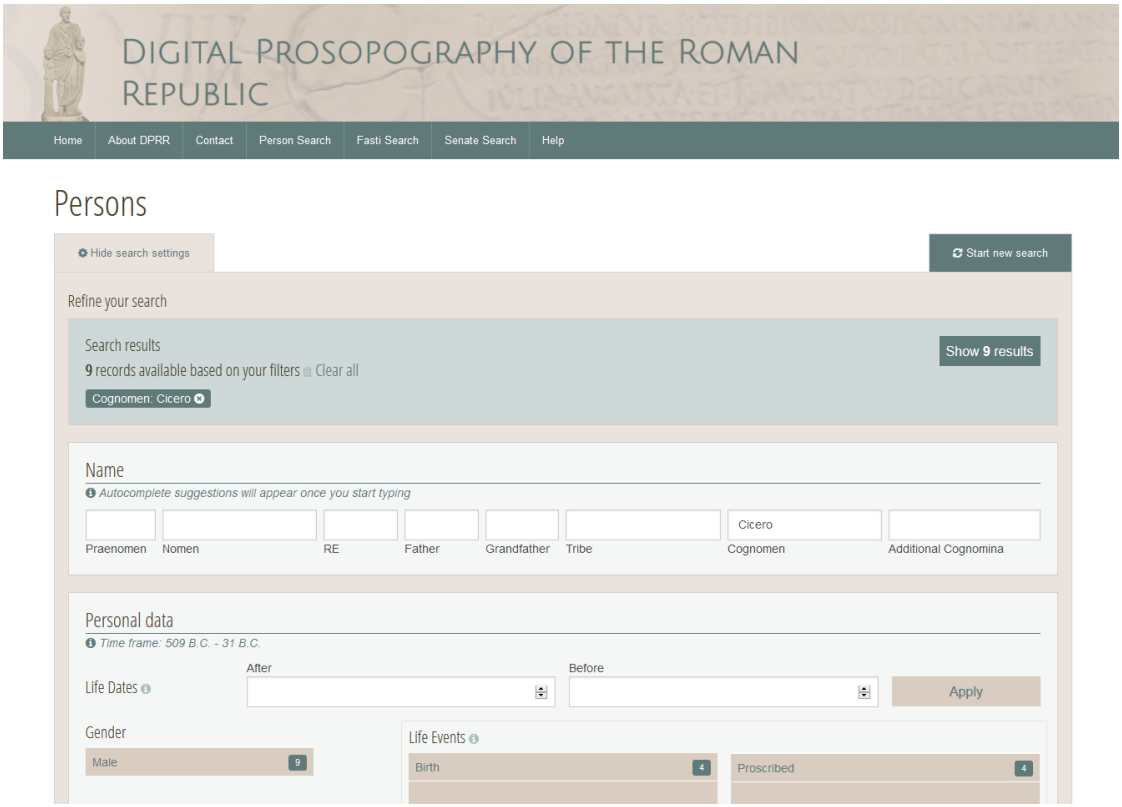


Figure 1. DPRR's bowser-oriented site: the Person Search

This figure shows what someone sees if they enter “Cicero” as the Cognomen for a person. We can see there that there are 9 records (persons) who have *Cicero* as their cognomen (and they are actually listed on the page, but below the displayed area in this figure). If one focuses for a moment on the form area in the bottom half of the figure, one sees a 18

good number of labelled boxes that can be filled in to filter the selection of persons. Note that the boxes and their labels immediately tell the user what kind of data the DPRR dataset holds that can be used for filtering (and there are even more filtering items off the bottom of this screen shot that are also available).

This web page uses a user interface strategy called “facetted search” (see Wikipedia's “faceted search” entry) to steer its user towards materials relevant to them in the dataset. This facetted search approach implements interface strategies used in other commonly used sites such as Amazon's, and is designed to help users with a limited knowledge of a field to find things that they want. Thus, the facetted approach for this DPRR selection page is intended to help novice users (although, of course, perhaps not so much novices to the study of Roman Republic society, because they are expected to know, for example, what a “Praenomen” is) to find things that will interest them. The intent of the design is to allow historians of the Roman Republic to use this page effectively with what is only now-a-days conventional web-access skills. 19

Contrast this with the front screen one sees (shown in Figure 2) when one fetches the front page of the DPRR RDF server. It allows the user to explore and select DPRR's data using RDF's SPARQL query language (which one provides in the large text box labelled “query”. 20

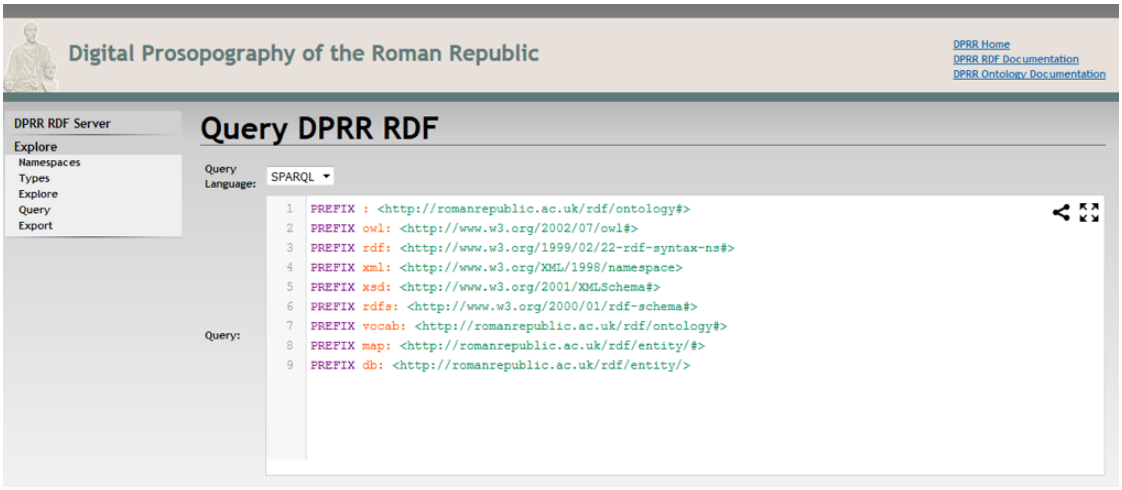


Figure 2. DPRR's RDF Server's front screen

Of course, there is (not surprisingly) more to the RDF server's web-oriented interface than this page alone, so only so much can be learned by examining it critically by itself. Nonetheless, even though we are only looking here at one of the pages that the RDF Server can show to us, one can quickly see that the RDF's server's web interface is built under a very different set of assumptions about the kind of user who will be working with it. Indeed, although there is a banner at the top of the screen that identifies it with DPRR, the DPRR RDF server's public interface is not specific to DPRR in the way that the facetted search browser presented earlier is, but represents instead a general kind of interface that could be used with any collection of RDF data on any subject. The web-browser interface for DPRR we looked at a moment ago has been tailored specifically to make front and centre how DPRR's materials are organised and to show under what semantic issues they operate. Here, in contrast, other than that obvious DPRR banner, this RDF server could look virtually the same if it was giving access to an entirely different set of RDF data. 21

In fact, this screen is part of *rdf4j*'s RDF workbench interface which has been specifically designed by the *rdf4j*'s developers to work usefully with *any* kind of RDF data. Indeed, DPRR's RDF server's browser interface focuses primarily on providing an interface that fits with RDF and related technologies like SPARQL rather than being an interface that is tailored specifically to express DPRR's concepts and materials. For someone to use the RDF server they need to know not only about DPRR's data and how it is represented in RDF, but also how RDF works and (for this particular web page) how to express queries in the SPARQL language that will be able to fetch data for the user's particular needs. We'll see examples of SPARQL being used in this way later in this article. The important point at this moment is that this browser interface says something about its intended audience: to use it one needs to have a solid 22

technical familiarity with RDF and its technologies, and to be capable of exploiting materials presented in this way. This article will look briefly at some of the other parts of its interface that is derived from the *rdf4j* workbench later.

23 I have chosen to include in this article HTML links and forms that actually invoke the DPRR RDF Server, based on the principle that by actually sending readers to the server, they will be better enabled to explore for themselves what the server is doing. Therefore, I recommend that you, the reader, click on the provided links and thus directly engage with the server yourself. The links are set up to cause your browser to open them in a new tab or window. Thus, to return to this article, you can simple close the display the link or form created when you are done with it. Furthermore, if for some reason you are unable to make the links work you can instead find screen captures in the appendix which show what appears in my Firefox browser when I click on the links. Each figure in the article is linked to the the spot in the article where it is needed.

24 Now let us turn our attention to how DPRR's RDF server addresses the basic requirement of Christian Bizer, Tom Heath and Tim Berners-Lee's conception of Linked Open Data as they describe it in their 2009 article that was mentioned earlier: [Bizer et al 2009].

25 The first point to notice is that the server supports the fundamental principle stated by Berners-Lee and others about open data: that all entities in the data have globally defined URIs for them, and if one gives the URI for any one of these entities to the web as a URL, one gets data back from the server about it. Thus, all of DPRR's data (as we will see shortly, not just DPRR persons) are globally accessible in this way, since all entities in the DPRR dataset are assigned global URIs and can be directly referenced by anyone with web access who wishes to do so.

26 For example, http://romanrepublic.ac.uk/rdf/entity/Person/2072 refers to one of the historical persons in the dataset: in this case the famous Roman author Cicero. (A screen shot showing what my browser gives me in response is given in the appendix as Figure 3.) If you give your browser this URI (and if you are reading this article online you can readily do this by simply clicking on the URI-as link showing here) it will find its way to the DPRR RDF server. There, the server will fetch the data about the person identified by this URI (Cicero) and will return to your browser all the data it has about him, delivering it to you through the *rdf4j* workbench “wrapper” which presents all these RDF statements wrapped in lightweight HTML so your browser can effectively display them. The tabular part of the display shows the RDF statements that reference DPRR's URI for Cicero. RDF statements have three parts in the order “<subject> <predicate> <object>”. Thus, one of the triples part way down the list in the table can be read “The entity with URI `http://romanrepublic.ac.uk/rdf/entity/Person/2072` has Cognomen 'Cicero'”. Cicero's URI is likely to appear as Subject or Object part of the RDF statements (and is allowed as a Predicate, although because of the way DPRR's RDF works, Cicero's URI does not in fact occur there), and this display shows all the RDF statements that reference Cicero's URI for all three possible types of reference.

27 It is important to grasp the fact that DPRR's other non-RDF “browser oriented” web interface can *also* present similar data about Cicero, and this function is invoked through a URL that looks somewhat similar to the RDF URI used to identify Cicero: http://www.romanrepublic.ac.uk/person/2072/ (A screen shot is shown in the appendix as Figure 4. The data about the same historical person, Cicero, is all included in the web page returned to the browser too, but it is wrapped in rather more complex HTML which has been tailored specifically to represent DPRR Person data, and which is designed to present visually well in a conventional browser for a human reader. Although both the “browser friendly” URL and the RDF-oriented URI for Cicero are based on the same underlying data and return similar results the differences between them are similar to the differences described earlier about the two front pages: Cicero's RDF URI is presented in terms of its RDF representation, whereas his browser-oriented URL immediately presents its material in terms focused on how DPRR data about Cicero is organised in a format which is calculated to be more immediately accessible to a less technical reader.

28 Although the RDF URI for Cicero's data caused the RDF server to respond with the RDF statements it holds still wrapped in a little presentation HTML you can in fact ask the RDF server to deliver its result in pure RDF — immediately suitable for further machine processing. There are two ways to do this. One can use the mechanisms recommended in the W3C's specification for RDF servers [Speicher et al 2015]: to ask the server to create the result in a particular RDF

format by identifying the type you want with a suitable RDF mime-type (such as “application/rdf+xml”, which requests RDF expressed in XML) in the *HTTP request header*. This approach can be relatively readily done if you use the http support in most programming languages such as Python or Java. However, if you are trying to use a web browser to fetch data as simple RDF it is difficult to follow these W3C guidelines and to control the mime-type the browser will specify in the HTTP request it generates for you. So, for browser users who actually want the plain RDF rather than an HTML representation of the RDF this W3C recommended method is difficult to carry out. For this reason, DPRR's RDF server has been extended beyond the W3C specification to support a parameter “format”. Specifying one of the standard mime-types for RDF (or more simply “rdf”) with it will cause the DPRR server to deliver the RDF data directly in the corresponding standard representations of RDF: http://romanrepublic.ac.uk/rdf/entity/Person/2072?format=rdf (A screen shot of what a browser shows for this is shown in the appendix as Figure 5.)

This pure RDF is perhaps even more difficult for a human reader to read (especially those not familiar with RDF), but it presents the data in RDF's standard Turtle format [Beckett et al 2014] that can be readily processed by RDF software in programming languages like Python or Java. 29

We have now seen the data DPRR has about Cicero in both the browser-friendly and RDF data-oriented views. The packaged presentation of DPRR's reader-friendly view is clearly more straightforward for a non-technical DPRR user to understand: that is the intent of its design. Furthermore, the DPRR development team worked to combine together data from various related parts of the DPRR dataset to create a unified and concise presentation that appears assembled together on a single web page. In contrast, to get all the data shown on this one screen through the RDF display requires the user to, themselves, follow links given as URIs in the RDF statements and thus to look at other related parts of the DPRR RDF dataset. Since, as we have seen, the browser-oriented interface delivers information about Cicero in a way that is more user-friendly, who would want to use the RDF Server's representation when arguably the browser-oriented presentation is easier for us to read? 30

This question takes us to the point of the Semantic Web and Linked Open Data too: that it expresses its materials in a highly structured form (RDF) that is suitable for *further* processing rather than just human viewing. Whereas arguably the browser-oriented presentation is easier for a person to interpret, it is not as straightforward to use when the purpose is to gather data from it for further processing. Techniques called “screen scraping” or, more specifically, “web scraping” (see Wikipedia's definition of “web scraping” for a good introduction) have been developed to get data out of human-oriented web pages such as DPRR's reader-oriented presentation — but screen scraping techniques are notoriously unreliable for getting at the underlying data which is presented for human eyes through the web page. In contrast, RDF has straightforward and consistent structures that are easy to process in a programming language such as Python or Java. If your aim is to further process the DPRR materials you fetch, using the RDF formats as the delivery mechanisms from DPRR are most definitely the better bet. Furthermore, as we shall see when we look at the server's query (SPARQL) mechanisms, the data there can also be delivered both in formats not only suitable for further processing in Python or Java, but also in spreadsheet-friendly formats such as Comma-separated values (CSV) (See Wikipedia's definition of “Comma-separated values” for a brief introduction). 31

We have now seen how the RDF server delivers LOD data about persons held in DPRR. However, as mentioned earlier, one of the important characteristics of the RDF server is that *all* kinds of DPRR data — not just persons — have open and public URIs assigned to them, so that a user can fetch DPRR's data not only about persons but also about any other kind of information that DPRR holds. 32

For example, Cicero is recorded in DPRR as having held a post of consul in the year 66 BCE. This kind of assertion is what in DPRR is called a “Post Assertion” and the one about Cicero being a consul is one of the many Post Assertions recorded in DPRR. This particular Post Assertion is expressed as a set of RDF statements, and has its own global URI: http://romanrepublic.ac.uk/rdf/entity/PostAssertion/5439 (A screen shot of what a browser shows for this is shown in the appendix as Figure 6.) Giving this URI directly to the WWW will fetch the RDF statements that are associated with this particular Post Assertion about Cicero's consulship. Indeed, we can continue in the same line and, following the principle that *all* DPRR data has a public URI attached to it, note that the concept of consulship (which is referenced in this Post Assertion) also has its own global URI: http://romanrepublic.ac.uk/rdf/entity/Office/3 (A screen shot of what a 33

browser shows for this is shown in the appendix as Figure 7.) and all the data linked to the office of Consul, as identified by this URI will be returned — including all the Post Assertions that state that someone was a consul since they will all refer to this “Consulship” URI through their “hasOffice” predicate.

Why does having kinds of data other than just persons directly addressable via the WWW matter in what is, after all, a prosopography? Because, as we discuss later in this article, being able to start anywhere (from any kind of data) rather than just one or two kinds of “entry points” (such as, for a prosopography, “historical person”) is a key reason why structured, interconnected, data (such as that represented using the relational paradigm or by graph representations such as RDF) is likely to be most useful. Being able to enter DPRR's data structures in any number of different ways makes possible fresh ways of looking at the data, something that would difficult to achieve if you could only enter the data through persons. 34

In order to make good use of the different kinds of interconnected data that the DPRR RDF server makes available beyond persons, one needs to know in some detail what is there and how it is organised. This is a place where an *rdf4j* workbench mechanism available in the RDF server comes in to be useful. The workbench's “types” display shows all the types of data identified in the DPRR RDF collection, and is a useful starting point for browsing DPRR's RDF statements. Generally, one can navigate to the types display from the browser pages presented by the server via the menu of options on the left side. Here is a direct link to it: http://romanrepublic.ac.uk/rdf/repositories/dprr/types (A screen shot of what a browser shows for this is shown in the appendix as Figure 8.) 35

Some of these items that are then displayed (the ones that begin with the prefix “owl:”, “rdf:” and “rdfs:”) are types of data that are generic to RDF and are therefore perhaps less useful for a data investigation about DPRR. However, the ones that begin “vocab:” are the names for types of data that are specific to DPRR; “vocab:Source”, for instance, asserts that there is a type of data called “Source” in DPRR. Looking through the list of types specific to DPRR which are identified by the “vocab:” prefix one finds other types that are immediately identifiable: “vocab:Person”, of course, but also “vocab:SecondarySource”, “vocab:Praenomen” and perhaps “vocab:RelationshipAssertion” given what has already been said about “vocab:PostAssertion”. 36

Clicking on, say, “vocab:SecondarySource” causes the server to list all RDF statements that make reference to it. One can see quite a range of different kinds of statements about “vocab:SecondarySource”, including a comment associated with it, which tells us that “vocab:SecondarySource” is “A modern source. DPRR is primarily built by harvesting data from 19th, 20th and 21st century scholarship.” A little below this assertion is the list of Entities that are asserted to be Secondary Sources. Only their URIs are given here so one cannot immediately tell what secondary sources they represent, but all URIs in this display are clickable, so by choosing, say, http://romanrepublic.ac.uk/rdf/entity/SecondarySource/1 (A screen shot of what a browser shows for this is shown in the appendix as Figure 9.) one can see that Secondary Source 1 is *Broughton MRR 1*; later shown to be *The Magistrates of the Roman Republic, Vol. I* 37

This kind of browsing through RDF data is typical of one of the main uses of the *rdf4j* workbench displays that have been incorporated into the DPRR RDF server. They allow one to develop a feel for the meaning of the data simply by browsing through the data itself. However, not all the types of data are immediately understandable in this way, and their relationship between each other can still be difficult to grasp. Thus, DPRR data also has what is called an *ontology*: a formal description (written in OWL [OWL 2012], another RDF-related technology) of the types of data in DPRR (called “Classes” in OWL) and their relationships to one another. DPRR's ontology is described in [Bradley 2017b], web page “The DPRR Ontology” and presents all the kinds of entities in DPRR and the relationships between them. 38

Two Perspectives on Uses for RDF and LOD

We have now briefly introduced several of the mechanisms the DPRR RDF server makes available to the world (the query-oriented SPARQL mechanism will be introduced later). It is time, therefore, to step away from its specifics to think about what this approach — providing a data-oriented historical site like DPRR as Linked Open Data (in the sense that Tim Berners-Lee conceives of it) — might mean for a humanities scholarly community. 39

Most of those people in the digital humanities who are currently working on the challenges of LOD focus on work that is often described as “enriching the global graph”: making explicit the links between *different* internet-accessible data collections. We can see work of this kind in projects like *Pelagios* [Pelagios nd] and, perhaps more particularly, *SNAP-DRGN* [SNAP-DRGN nd]. DPRR/RDF, however, does not look like recent historically oriented LOD initiatives such as these. So what is its connection, in and of itself, to the LOD perspective? In spite of the different connection that DPRR has to LOD than the “enriching the global graph” initiatives have, I believe that DPRR's RDF should still be interesting to the humanist LOD community. One needs to start by thinking more about the two different kinds of engagement with LOD materials by web users which appear at different points in time in Berners-Lee's conception of Linked Data. 40

Berners-Lee's first conception of the Semantic Web was described in the early 2001 *Scientific American* article “The Semantic Web” [Berners-Lee et al 2001]. Here we see the authors proposing a data and semantically-rich extension to the already existing document-oriented web in a way that would allow ordinary folk without formal training in digital semantics to exploit this semantic richness. The authors give a number of imagined examples of agent-based software that could automatically exploit formal semantic data across different sources. One example (see page 36) tells us of a user who sends her agent software off to make an appointment with a medical specialist for her mom. To do this requires the agent to find specialists that fit with mom's prescribed treatment, then match up the appointment calendars for mom and those specialists. The software agent also needs to take into account other parameters such as distance to the appointment, and the need for physical therapists. Allowing a user's software agent to perform this kind of complex task reliably requires that the material it works with must be highly structured and have appropriate software-accessible semantics formally available so that the software agent can, on its own without human intervention, connect it together correctly and exploit it. In the ideal Semantic Web described by Berners-Lee et al in 2001 a human user would be able to safely delegate this task to their agent software and wouldn't need to worry about the details of how the agent did the job, although if she was interested she could ask the system how it went about carrying out the task and, since the computation would be based on structures that semantically mirror parts of our human understanding of the world, receive an answer that could be understood. 41

[Berners-Lee et al 2001]'s 2001 agent-oriented vision has proven to be quite ambitious. As a consequence there has been work in Computer Science to explore the somewhat simpler task of trying to make semantic web data help ordinary, non-technical users better search for things in which they are interested in the vast global internet-wide data graph. Some of this work involves trying to find ways to enrich google-like searching (which is centred primarily on very sophisticated Natural Language retrieval principles (NLP) applied to the WWW's text-oriented documents) with semantically-structured material expressed in RDF and its associated technologies. When researchers tried to build systems that could jointly exploit RDF-like structured data as well as the text in Web pages they found it to be a real challenge. One of the issues was that independent but semantically related data collections were likely to have differing internal structures and might well use different vocabulary in their formal structure for what were the same or similar concepts: a condition called “Heterogeneous Datasets” by some researchers. A good summary of some of the thinking in this area from a few years ago can be found in [Freitas et al 2012]. It is not quite clearly spelled out in this article, but an important assumption seems to be that the tools that they were interested in would ideally support querying that could be characterised as coming from what I am calling here an “intuitive user”. 42

Consider Google as an example of an existing service which is also conceived of as serving an “intuitive user”. Most Google users are not familiar with the range of material that the web possesses when they start a Google search, and they phrase their question without knowing the structure or vocabulary applied to materials on the web. In this sense, their querying is intuitive. Similarly, some of the engines that Freitas *et al* describe are meant to allow users to ask questions in a natural language without knowing much about the domains the data represents. These engines use a combination of NLP techniques combined with a sophisticated understanding of relevant RDF data with their ontologies that describe them, to provide a better query result than NLP could deliver on its own. The aim is to allow users to come with what are intuitive text-oriented questions and get richer, more trustworthy, results than they would get from the NLP approaches against text-oriented documents alone. There is a good summary of more recent thinking in this area in [Noy et al 2019]. 43

Of course, recent work by Google and others has shown that text-oriented big data strategies can achieve remarkable 44

things with only vast amounts of almost-raw text as data without needing large amounts of hand crafted formal semantic data at all. Thus, it would seem that if the 2001 Semantic Web vision is ever going to be achieved, the emergence of platforms that have rich, widely available, semantic data expressed in RDF and its associated technologies, combined with AI software of the kind envisioned here that can make use of it, are still something for the future.

45 Perhaps as the challenges of implementation of the ideas in the 2001 article became clearer, Berners-Lee began to think about the benefits of having the data without the sophisticated AI-like framework that would be needed to make the more sophisticated ideas of the 2001 article work. This is the situation we find in Berners-Lee's 2010 TED talk that I mentioned earlier. Here, the users of Berners-Lee's global data are not the kind of intuitive user with their natural language query that I have just described; a user which would need to be supported by substantial Artificial Intelligence-like methods hidden from him/her. Instead, Berners-Lee gives examples of people exploiting the power of formally structured data through “mashups” which explicitly join together bits of previously disconnected global data to gain new insights into the material. In one of Berners-Lee's illustrations we see a person joining together data about what streets a new municipal water pumping station served with demographic data about those streets, and then being able to show how this town's new station was disproportionally serving the wealthier parts of the town. This kind of working with disparate data from different sources requires something quite different from its user than the intuitive engagement of the Google-like NLP+Semantic-data approach that Freitas *et al* and Noy *et al* are exploring. If someone wishes to join up data from different sources like this they cannot be an intuitive user and take an intuitive approach based on only a limited understanding of the data one is querying. Instead, to join them together they need to understand in some detail the semantic structure and significance of their data sources, and know how to formally join them correctly.

46 The important point for us here is that the Berners-Lee TED talk's researcher's discovery of the link between the new water plant and the people it served was made not with the aid of an intuitive google-like query, but by the deliberate bringing together of two sources of structured data in a way that no one else had done. To achieve this, the data analyst needed, in some way, to be the opposite of intuitive. Instead, s/he could only create new information when s/he thoroughly understood the semantics of the pieces of data s/he is working with and understood how they connected together. Furthermore, only in this way could the strength of the argument that arises from this water plant example come out of the semantic juxtaposition of the materials.

47 Freitas *et al* characterise this kind of interaction with data and the type of “structured query” that can be expressed against it as “crisp” and seems to equate “crispness” of response with “precise answers” of the kind given by database formal queries in languages like SQL [Freitas et al 2012, 29]. These interactions with data are not like queries that are conceived of as Google-like semi-natural language expressions, where one cannot actually be sure either that the result one gets matches a natural human understanding of the query or that one gets all the material that a human would consider relevant to the question asked. Instead, these crisp structured queries have a kind of processing model that, to the degree that the data being queried can be considered to be an accurate representation of its material (admittedly, an important qualification) and inasmuch as one can express what one is interested in in the formal nature of the query language, allows one to be sure of the completeness and accuracy of the result. Although Freitas *et al* don't explain what they mean by “crisp” and “precise” in their article, it is, I expect, in this area that their sense of these terms resides.

48 Does DPRR's RDF server allow for this kind of engagement with its data? Can a classical scholar engage with the formally based mechanisms of DPRR with an intention that is similar to Berners-Lee's water plant mashup example? Certainly formal “crisp” queries of the form and spirit that Berners-Lee's 2010 examples require are available through the DPRR RDF Server's support of the SPARQL [SPARQL 2013] query language.

DPRR and SPARQL

49 What is SPARQL? Wikipedia starts its article about SPARQL by saying “SPARQL allows users to write queries against [...] data that follow the RDF specification of the W3C”. It works by allowing the SPARQL query creator to specify a pattern to look for in the RDF graph, and to display parts of the selected bits that match the pattern as results. This certainly is not the place to provide a tutorial on SPARQL, but here is an example of a query in it:

```
PREFIX vocab: <http://romanrepublic.ac.uk/rdf/ontology#>
select ?name ?officeName
where {
   ?person a vocab:Person;
      vocab:hasName ?name;
      vocab:isSex <http://romanrepublic.ac.uk/rdf/entity/Sex/Female>.
   ?assertion a vocab:PostAssertion;
      vocab:isAboutPerson ?person;
      vocab:hasOffice ?office.
   ?office vocab:hasName ?officeName
}
```

Execute

(A screen shot of what a browser shows when this is submitted is shown in the appendix as Figure 10.) 50

The query looks for graph patterns in the DPRR RDF data that show women who are also recorded has holding offices, and displays the woman's name and the name of the office. It is expressed in the SPARQL language, and the reader can doubtless see that it is not a trivial matter to learn to create queries of this kind, particularly for those without knowledge of related query languages such as XQuery for XML [XQUERY 2018], or SQL for relational databases [SQL 2018]. However, once it has been learned, it provides a powerful way to explore a complex set of RDF data, such as that found in DPRR. 51

The SPARQL query presented here in this article is given in the context of an HTML form that allows one to directly send the query to the RDF Server and receive the result. To do so, push the “Execute” button. Soon thereafter you should receive a response from the Server showing, in a table, the names and offices of all women recorded as holding offices in the DPRR dataset (or, click here to see a screen image in the appendix of the beginning of the server's response to this query). You might recall that our first view of material from the DPRR RDF Server was of its SPARQL Query screen. And indeed, the query text shown here could be copied and pasted into that screen and run from there, and would have produced essentially the same result as what one gets from the above form. 52

The form above causes the RDF Server to return its result embedded in a light wrapping of HTML that makes it more suitable for human browsing. However, the query can also be run so that it returns results in a structured form more suitable for further processing. Here is the same query set up in a form that causes the result to be returned in JSON — a format suitable for further processing by platforms such as Python or Java (if you are curious about JSON, a good starting point is Wikipedia's definition). Results can also be returned in CSV format which can be opened as a spreadsheet, although this is not shown in this example. 53

```
PREFIX vocab: <http://romanrepublic.ac.uk/rdf/ontology#>
select ?name ?officeName
where {
   ?person a vocab:Person;
      vocab:hasName ?name;
      vocab:isSex <http://romanrepublic.ac.uk/rdf/entity/Sex/Female>.
   ?assertion a vocab:PostAssertion;
      vocab:isAboutPerson ?person;
      vocab:hasOffice ?office.
   ?office vocab:hasName ?officeName
}
```

Execute

(A screen capture of the beginning of the display generated by the query is shown in the appendix as Figure 11. How your browser displays JSON data may be different from how Firefox showed it to me.) 54

Can SPARQL Queries Further Study of the Roman Republic?

Having now briefly seen SPARQL as a querying mechanism against the DPRR dataset, perhaps the reader will still not find it obvious how such a thing could be relevant to the furtherance of the study of the Roman Republic. I can see three possible concerns: 55

- Both Berners-Lee's mashup builder and DPRR's SPARQL query engine require a complex set of technical skills that one would think does not match well with the normal skill-set profile of someone interested in DPRR.
- Whereas Berners-Lees examples draw data from disparate sources and joins them together to make their point, DPRR is, by itself, a single source. Berners-Lees is making the point that the strength of LOD as a new way to look at data arises precisely from the way that it allows sources that have not been brought together before to be joined. What do LOD approaches have to offer for a single source like DPRR?
- Finally, whereas Berners-Lee's examples use the connecting together of data to make political points, there are likely to be few, if any, political arguments of the kind that Berners-Lee is interested in that could come out of a study of DPRR.

Point 1: Technical skills

In order to interact with the DPRR RDF server and get the benefits that it holds one needs to understand 56

- first, formal data modelling principles,
- then understand RDF,
- then how to query RDF datasets with RDF's query language SPARQL,
- and finally how to assemble data selected from the server for further processing, perhaps to turn the data into, say, useful displays with a spreadsheet, or with, say, Python and something like Google Graphs.

This is a tall technical order. 57

Of course, the complex technological requirements needed to interact with RDF data, plus the assumption that DPRR's users are unlikely to have the technical skills needed to interact with the data directly, is *exactly* the reason why DPRR (like CCH/DDH's other digital resources) have as its main public point of access a web-oriented user-friendly front end to its complex, formally structured, relational dataset. Why, then, is the fuller functionality similar to what direct interaction with RDF enables through the DPRR RDF Server not made available from DPRR or PoMS's more “user friendly” web front-ends? There are two reasons. 58

- One is User Experience (UX) based. The assumption behind much UX work is that the user is going to be an intuitive user, and needs to get useful results from simple interactions that require little effort to understand the database and its semantic structures. The design tries to, as much as possible, follow the user interface principle “[d]on't make me think” [Krug 2014] as put forward a few years ago by Steve Krug, the UX guru. As a consequence of this UX thinking, if we expected our web users to understand that was going on with only minimal intellectual effort we had to restrict the investigation paths for our users to relatively straightforward ones. However, to get all the “semantic juice” out of DPRR, PoMS or any other relational database requires more understanding of the formal principles of the relational model and the structures of a particular database than what matches the UX understanding of a user community.
- The second reason is that the results have to also be presented in ways that suit the user and his/her browser rather than as formally-structured data that can then be readily reprocessed by software for further analysis — as a web page rather than as structured data which could be further processed — since in this day and age web pages are both accessible and, in a general sense, understandable to pretty well anyone likely to be interested in DPRR, including otherwise non-technical users.

How could a suitably trained person take advantage of the facilities the DPRR server offers? To show how DPRR's RDF server can be exploited I have created a modest “timeline” example and made it available with the server which plots the holders of the office of *consul* by their tribe. It shows how the technologies of RDF and Python can be engaged to get materials out of DPRR's data that would be difficult to do with the more “user-friendly” UX-designed web front-end. It is based on a SPARQL query which is formulated to fetch the tribe name associated with each consul holder. You can see the query that fetches the relevant data in the form below, and by pushing “Execute” you can run it for yourself. 59

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX vocab: <http://romanrepublic.ac.uk/rdf/ontology#>
select ?startDate ?endDate ?tname ?aid ?pid ?pname
where {
  ?person a vocab:Person;
     vocab:hasID ?pid;
     rdfs:label ?pname.
  ?tribeassert a vocab:TribeAssertion;
    vocab:isAboutPerson ?person;
    vocab:hasTribe ?tribe.
  ?postAssert a vocab:PostAssertion;
    vocab:hasID ?aid;
    vocab:isAboutPerson ?person;
    vocab:hasOffice <http://romanrepublic.ac.uk/rdf/entity/Office/3>;
    vocab:hasDateStart ?startDate;
    vocab:hasDateEnd ?endDate.
  ?tribe a vocab:Tribe;
    vocab:hasName ?tname.
}
```

Execute

(A screen capture of the beginning of the display generated by this query is shown in the appendix as Figure 12.) 60

Having created the SPARQL query which fetched the data needed to plot the tribe of consuls over time, the query was then embedded in a Python script which ran it directly, took the results it generated (in JSON), and used the plotting services of Google Graph to generate an HTML page that plotted the year vs tribe. The overall result can be seen here. The *timeline* materials, including the Python script, are available from http://romanrepublic.ac.uk/rdf/timeline/. I built the script in a couple of hours and, having done one, could probably do another one for a different question more quickly. 61

This is all well and good, I hear you say; but, of course, although I have the skills needed to create something like this, I am not the right person to decide whether the result that the *timeline* example provides is actually useful to the study of Roman Republic history. Only Roman Republic historians themselves can do so, since they understand whether or not 62

any connection between a person's tribe and the offices they held could be historically interesting. It is thus historians rather than someone like me that need to be directing the engagement with DPRR's RDF data. Is it possible, then, to expect historians to be able to interact in this way with the dataset: to have both the understanding of an historian and the technical skills that enable one to fetch data using SPARQL and plot it in something like Google Graph?

63 My own experiencew of teaching Python for a number of years (in a one term MA taught module) and structured data, including RDF and SPARQL (in another) in DDH's Digital Humanities MA programme has led me to believe that it is possible for students with a conventional humanities education, but with a real commitment to engage with the potential of DH methods, to learn sufficient technical foundations to be able to engage with data-oriented materials such as DPRR's RDF representation effectively. These students, with a humanities orientation in their background, were able to bring these humanities-oriented interests and curiosity to bear on their new-found technical abilities, to conceive of and perhaps construct something like DPRR's *timeline* example. Indeed, more than one student who has attended these two modules has directly told me that they believe that they came away from these modules with the beginnings of exactly this kind of ability.

64 I also have received news recently of an example of DPRR's RDF data working in exactly this way. Although the RDF server has only been available for a few months at the time this article was being written, a researcher working in an independent research project run by Professor Chris Johanson (UCLA's Department of Classics) which is entirely external to DPRR reported to me that the RDF data has made an important contribution to their work. The project team was interested in trying to generate visualizations of Rome's Rostra during funerals of important people in the Roman Republic and to show the different types of togas that would be worn by the actors playing the part of the deceased's ancestors. They found the DPRR RDF server useful because they could use their own queries to get information about how different people were related and what offices they had held (and thus, which toga would be used to represent them). To that end the team started from the results of relevant queries to the DPRR RDF server to generate a directed graph (nodes=people, edges=paternal relationships) which they could then traverse to determine the set of togas to depict. As they say in one of the emails they have sent me, they were able to use the server to fetch readily the materials they needed themselves much more directly than they could have done with either the user-friendly DPRR web front end, or, of course, if DPRR data had not been available to them at all. There is more information at [Johanson et al 2019], including specifically the visualisations at http://hvwc.etc.ucla.edu/funerals-rostra. Professor Johanson was interested enough in the RDF server to then spend some time introducing some of his students to it. He asked them to explore the data and to see what visualisations they could produce with its help. The result was the Shape of Roman History project, which contains something like 30 charts, all of which draw their data from the DPRR RDF Server.

Point 2: Single Source

65 Berners-Lee's examples of exploiting Linked Data are often classified as what are called *mashups*: the joining together of more than one data source to enable a new representation that any one data source, by itself, could not achieve. Speaking strictly, then, the *timeline* example is not a mashup because it draws all its materials from the single DPRR data server. Thus, some might argue that it does not fit well with the assumptions in the LOD movement: that it is in the bringing together of data from *multiple* sources that new insights can come. However, it is important to understand that Berners-Lee's examples in his TED talk require multiple different sources of data because each online source is relatively small and structurally straightforward piece of data; the kind of thing that can be comfortably represented in, say, a spreadsheet. This is not the situation, however, with DPRR. DPRR's original relational model already supports quite a rich kind of interaction between many different kinds of objects, and DPRR's set of RDF — which is, after all, simply an expression of DPRR's complex database — has in fact 39 different entity types, related together by 30 types of relationship and 53 kinds of data properties spread across those 39 entity types. Thus, the DPRR RDF graph is already by itself a complex interconnected graph of data of which only a handful of all the implied relationships between these objects has ever been explored. Thus it is reasonable to expect that queries which can, relatively straightforwardly, draw on the complex interconnections in DPRR's RDF alone can expose connections that have never before been considered. Many more visualisations like the DPRR *timeline* demonstration are possible without needing to go outside of DPRR's internal web of data, and some of these might well make new insights possible about the Roman Republic.

Point 3: Political Points

66 Finally, Berners-Lee's TED talk shows that, as in his water plant example, it is possible to see that much of the work enabled by LOD that draws on contemporary data could have contemporary *political* significance. Is something as potentially significant possible in the humanities? It is true that there are unlikely to be contemporary political issues that could be usefully explored by looking at the Roman Republic and DPRR. However, there *is* some evidence around that suggests that significant original ideas can be explored though data like DPRR's that have, hitherto, been unavailable or difficult to work with, and that perhaps some of these might represent truly original research that presents ideas that no one else has noted before. As [Guetzkow et al 2004] describe in their article entitled *What is Originality in the Humanities and the Social Sciences*, new ideas -- including even radically new ones that might be read as political within the humanities itself — are often valued by the humanities community. They write:

> In interviews, we found that panellists described originality, for example, in terms of the novelty of the overall approach used by the researcher (who is 'bringing a fresh perspective') in terms of the data being used (she is 'drawing on new sources of information'), and in terms of the topic chosen (he is 'going outside canonized authors'). These statements point toward a much broader definition of originality than that posited by the available literature on originality. [Guetzkow et al 2004, 192]

67 The representation of historical materials as data (rather than text) and the drawing of historical conclusions from it has been controversial within history: one recalls the debates about the use (and some would say the misuse) of statistics in the *Time on the Cross* studies [Fogel and Engerman 1974]. See Thomas Weiss's 2001 review [Wiess 2001] for a sense of the debates that arose from this work. Nonetheless, in spite of the debates it has spurred, it certainly has been, as Weiss says, “a book that has not been ignored”. Could some kind of data analysis, perhaps statistical, that is enabled by DPRR's RDF representation cause similar stimulation and consternation within the community that studies the Roman Republic? Of course, it is too early to say much about the DPRR RDF Server in this regard. However, the enthusiastic informal reports we see from users of DDH's many data-oriented historical resources such as *People of Medieval Scotland* (PoMS) and the *Prosopography of Anglo-Saxon England* (PASE) (mentioned earlier) suggests that research product presented as data rather than prose can also be useful to historians.

68 Furthermore, recent work with data in a sister data-oriented prosopography project PoMS (which is a factoid prosopography and was captured, like DPRR, in highly structured data) suggests that significant new approaches to historical materials, when available as complex structured data and explored from new perspectives, are both possible and reveal significant potential for new insights. Starting several years ago, thanks to a grant from the Leverhulme Trust, the data *behind* PoMS's public website has been used as a base for historical Social Network Analysis (SNA) experiments. This SNA analysis work on PoMS has been initially extensively reported on in [Hammond and Jackson 2017] which is an e-book of over 500 pages. As it is pointed out in the preface for this book, all the SNA analysis was enabled by PoMS's data-oriented interpretation of its sources and came from the relationships recorded directly and indirectly in the PoMS database. The data needed for the SNA work was explicitly provided by the database, and yet could not have been carried out with either PoMS's public web interface or with google-like intuitive queries that might have selected PoMS materials. To perform it effectively required access to the information behind PoMS's public interface. Since PoMS's data was organised in a relational database, the process that was used to fetch data for SNA analysis used queries expressed in the relational database's standard query language SQL [SQL 2018] that were quite different from those SQL queries used behind the scenes to drive PoMS's public interface with its particular user perspective.

69 The resulting SNA analysis showed that PoMS data could be exploited in ways that were quite different from what one could achieve through its public interface. Although the work was still in its early days when the Leverhulme grant was over, the team was even then beginning to see that this novel SNA perspective was pointing the way to possible new insights into Medieval Scottish society. As a result, work has continued exploring this SNA approach after the grant completed, and has resulted in particpation in and leading of a number of workshops and demonstrations to the growing Digital Historian SNA community. Although the fetching of data that fed the SNA analysis of PoMS was done in the formal language of SQL rather than with LOD technologies, this technical work was quite similar to what RDF, SPAQRL,

and related technologies would have enabled. If PoMS data had been available as RDF through a PoMS RDF Server this same work could have been carried out by anyone with internet access using Semantic Web technologies such as SPARQL. Furthermore, although one of the members of the SNA team, Dr Matthew Hammond, is trained as a historian, rather than as a computer scientist, he was able to master the formal language SQL well enough to, on his own, get the data in the forms that he could use for his SNA work. If PoMS data had been made available through an RDF Server like DPRR's, he could certainly have done the same work in SPARQL instead.

DPRR and Enriching the Global Graph: A Third Kind of User

As this article has shown, DPRR is a complex and interconnected collection of RDF statements which both (i) forms, within itself, a complex and disciplined graph of information and (ii) thus offers many possible routes for exploration. However, other than the standard references to RDF vocabularies such as RDFS and OWL, DPRR's RDF does not point out of itself into materials created and held elsewhere. Since the linking together of data across the entire “global” rather than DPRR's “local graph” is part of the vision of linked data, one needs to also think about what needs to be done within DPRR to bring it more into alignment with this aspect of the global graph vision. 70

As mentioned earlier, most of those people in the digital humanities who are currently working on the challenges of LOD are interested in what is often described as “enriching the global graph” — making explicit the links between *different* internet-accessible data collections. This work has sometimes been categorised as “aggregation”, and is often done by making a block of “SameAs” assertions using the `owl:sameAs` predicate or something like it. For instance, [VIAF 2010-16] describes itself as the “Virtual International Authority File”. It is a resource maintained by OCLC as a service to libraries which aims to “lower the cost and increase the utility of library authority files by matching and linking widely-used authority files and making that information available on the Web”. Thus, VIAF has the URI https://viaf.org/viaf/78769600/ for Cicero and when it is invoked one gets a web page that shows how major world libraries have identified him. Thus, this VIAF URI can be considered as VIAF's identifier for the historical person Cicero. One can then make an `owl:sameAs` assertion via an RDF triple that asserts that the person associated with DPRR's URI (http://romanrepublic.ac.uk/rdf/entity/Person/2072) for Cicero is the same person as the person VIAF identifies with their URI. This kind of work, when done with as many of DPRR's persons as VIAF has also identified, is arguably the first step in aligning DPRR's data with the larger digital world of data, at least as it exists in the context of libraries. Similar work could be done with world wide resources such as, say, WorldCat [WorldCat nd]. 71

Establishing `owl:sameAs` links between entities in different datasets to show how they connect to each other seems to be obviously a good idea that enriches the interlinking in global data, especially if one of the links is to a recognised authority, such as VIAF. Of course, DPRR is a published prosopography. The identification of people is the point of the work it represents, and hence DPRR has some reason to claim to be an authority for Roman Republic persons in its own right. Perhaps, then, in the same way as in the past many different independent researchers working on the Roman Republic often used Pauly's RE as an authority and identified people using the person identity scheme used in it, people in the future could use DPRR's URIs to identify which historical Roman person they were referring to. 72

Much of the work done in the DH that involves linking to authorities like VIAF has been carried out in the context of identifying people who appear in texts — as a reference from a spot in a digital edition of a text, say, or perhaps from a reference in a piece of research being written up as an article or a monograph — and is undertaken in the context of textual markup. This linking of a spot in a text to an authority such as VIAF (or DPRR itself) is a useful enriching process. However, the benefits are perhaps obviously greater when the links are not from a *text* (even one marked up using TEI) to RDF data such as DPRR, but between separate datasets both of which can both be queried by SPARQL, since SPAQRL's Federated Query mechanisms [Seaborne et al 2013] allows a single query to span across more than one dataset. If, for example, a dataset (let us call it “A” here) outside of DPRR had kinds of information that does not appear in DPRR about Roman persons, and if both DPRR and “A”'s dataset's associated persons could be connected through references to common VIAF URIs, it would be possible to query data that crossed both DPRR and “A”, taking advantage of the data strengths of each of them. 73

In some ways this linking work fits with the spirit of what DPRR was already doing: bringing together hitherto separate 74

Roman Republic prosopographies; although DPRR's work was based more on establishing collections between what had been separate primarily *print* prosopographies. However, although DPRR did indeed assemble materials from these various prominent, independently produced, specialist prosopographies into a single large collection they were not able to take up the further task of linking their people to a world-wide resource such as VIAF. There simply was not the time and funds available. As it turns out, this may be the place for a *third* kind of person to engage in DPRR's LoD data: someone who might be called an “aggregator”. This third kind of user arises from the fact that it is in the nature of LOD that, now that DPRR data is open and freely available through DPRR's RDF server, someone *else* with an interest in historical people from the Roman Republic that appear in VIAF or WorldCat can choose to create RDF triples independently of DPRR's research team that assert the connections between the people identified in these resources, and those identified through DPRR's person URIs and then make their collection of “sameAs” RDF triples that assert the connections available over the web. Indeed, by hosting these triples that connect DPRR entities to VIAF or WorldCat outside of DPRR itself one avoids the possible confusion by users of who did what: it will be clear that the links between DPRR and VIAF or WorldCat were done as a separate project outside of DPRR. In fact, this is one of the benefits of the conception of LOD as data distributed worldwide when it is based on the RDF technologies.

75 So far in this section we have focused on DPRR's historic persons as the centre of a linking initiative, and DPRR is, after all, a prosopography, and thus exploiting the URIs for its people through links seem like the most obvious thing to do. However, in the RDF context all of DPRR's data is open and globably available. Thus, there are URIs in DPRR that represent things other than persons, and linking these other non-person objects in DPRR to authorities elsewhere could also be useful to do. For example, DPRR has what the Romans called *provinces* as entities associated with offices. Not all the Roman provinces were geographically based, but many of them were. Thus, perhaps a sameAs link could be establised between these geographically based provices identified in DPRR and those same geographic provinces as they are identified in geographic authority sites such as *Pelagios* [Pelagios nd]. Then, if other RDF sites also used Pelagios URI identifiers in their data, these Pelagios URIs could be used as linking mechanisms to allow these two datasets to be joined together in a federated SPARQL query. If, for example, there was a set of RDF data that associated climate conditions with Pelagios places, federated queries could be used to explore if there was any evidence that climate had any effect on who got postings associated with these provinces.

A Call to Action

76 The development of the DPRR RDF server has shown that the materials developed by a data-oriented project such as DPRR can be certainly expressed as RDF, and can be served online in this way and meet the criteria proposed for Linked Open Data by Tim Berners-Lee and others. Only time will tell, of course, how useful academics who are interested in the Roman Republic will find such an expression of this kind of research, but the fact that very soon after its launch, the UCLA project interested in Roman Republic funerals found it useful is at least encouraging.

77 Now that DPRR's data has been made available directly as LOD, perhaps it is time for other data-oriented sources to be made available in this form as well. Over the years King's DDH department, in collaboration with historians and other colleagues in the arts and humanities as well as cultural heritage sector, produced a significant number of web sites that are driven by data that could readily be mapped to and delivered as RDF in the same way that DPRR's has been. So, now that DPRR's data has been made available directly as LoD, provided that appropriate resources are in place to support this work (see, for instance, this UKRI announcement), perhaps it is time for other data-oriented sources to be made available in this form as well. Indeed, at the time that this article was being prepared for publication work was just finishing up which published another of them: the People of Medieval Scotland (PoMS) data through its own RDF server in essentially the same way. You can find its RDF Server here.

78 King's Digital Lab (KDL) is now the unit at King's responsible for hosting most of the resources that were started by DDH such as DPRR and PoMS, and has kindly agreed to take up the responsibility for hosting and maintaining their RDF Servers as well. The development of RDF Servers for these projects fits well with one of KDL's current initiatives which is centered on the idea of data exposure and publication becoming a key element in the approach to a project's development: see this KDL web page and [Smithies et al 2019]. With respect to legacy projects, one of the options KDL offers to project partners is dataset deposit and the preparation of associated metadata cataloguing it. As a

consequence one of the solutions KDL has developed is a CKAN instance hosted within KDL's infrastructure (https://data.kdl.kcl.ac.uk/). Nonetheless, KDL has also seen that the DPRR RDF Server's more dynamic approach to direct data access also has the potential to fit with this part of their vision. Rather than being mediated through a web application front end, these projects' raw data might already have an important role to play to further new humanities research in their own right. As a consequence, perhaps, like DPRR, research data for other of these web resources might well also deserve to be set free for those in the humanities who are equipped to take advantage of them.

Appendix: Screen Captures

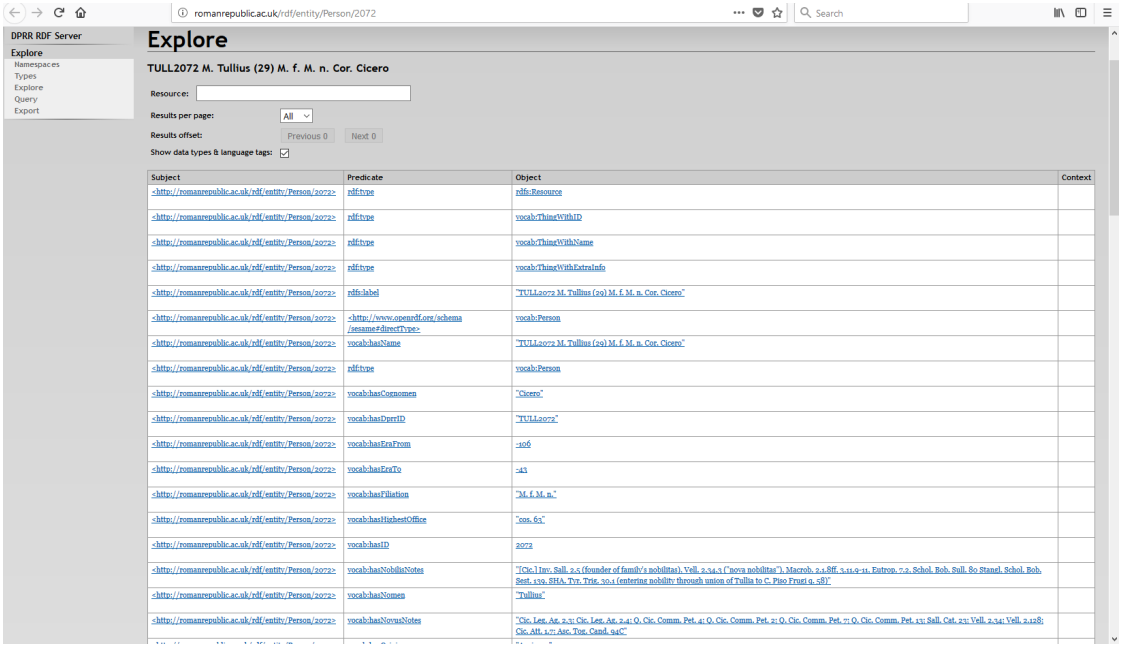


Figure 3. The DPRR RDF Server responds to http://romanrepublic.ac.uk/rdf/entity/Person/2072 (Cicero)

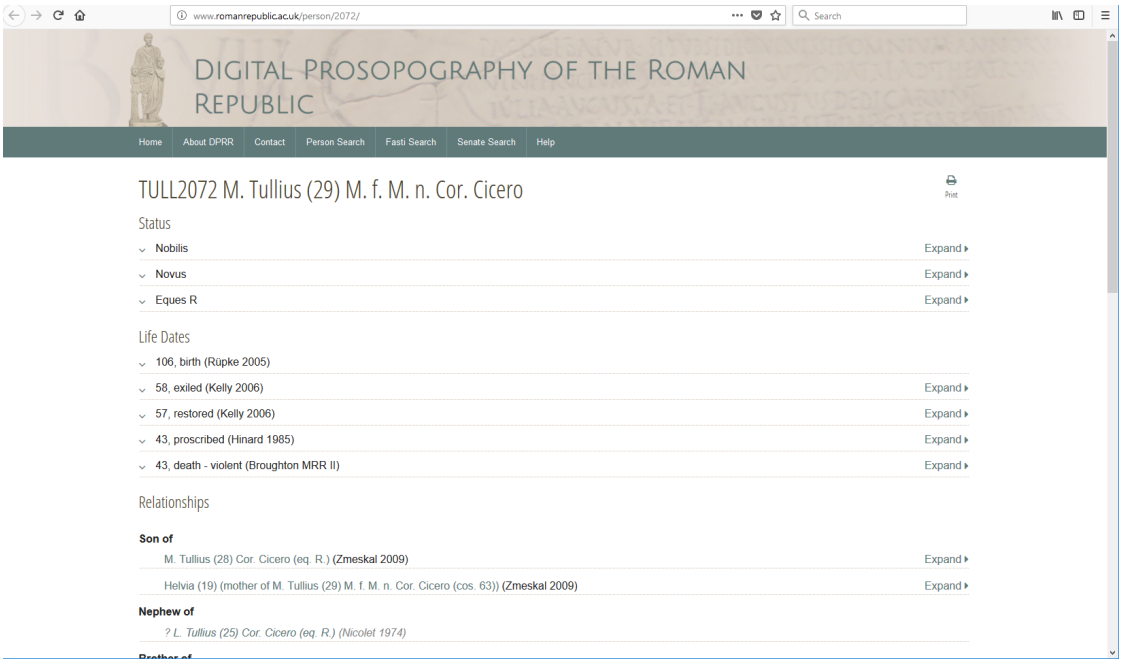
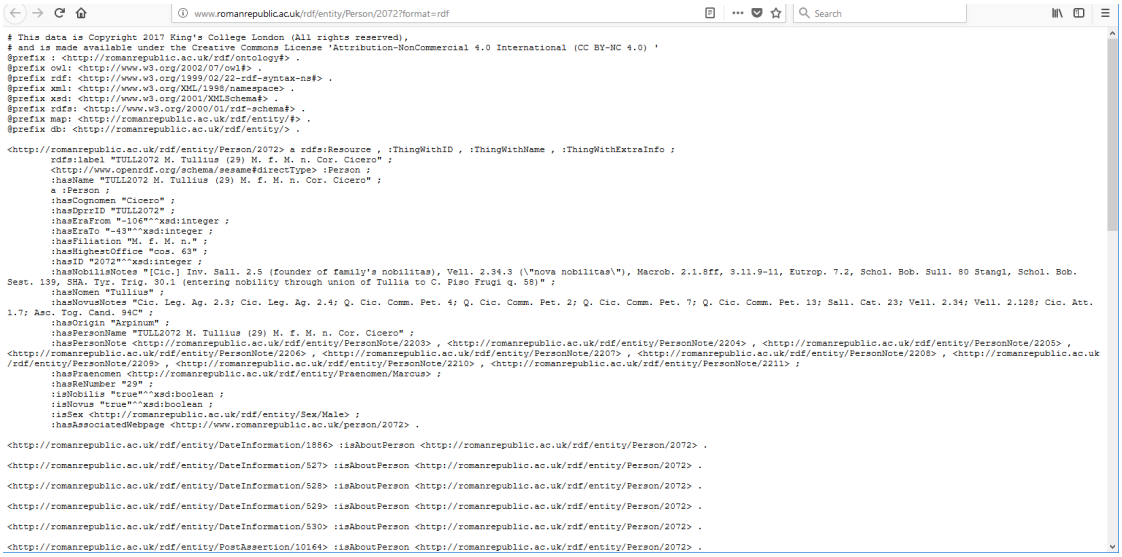
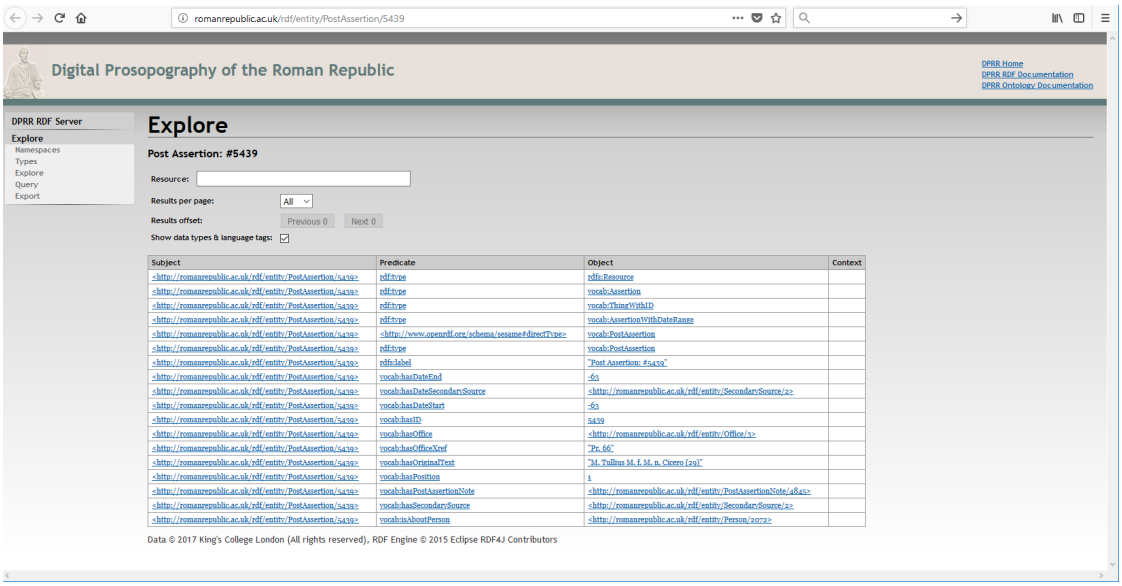


Figure 4. The DPRR browser app's response to http://www.romanrepublic.ac.uk/person/2072/



```
# This data is Copyright 2017 King's College London (All rights reserved),
# and is made available under the Creative Commons License 'Attribution-NonCommercial 4.0 International (CC BY-NC 4.0) '
@prefix : <http://romanrepublic.ac.uk/rdf/ontology#> .
@prefix owl: <http://www.w3.org/2002/07/owl#> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix xml: <http://www.w3.org/XML/1998/namespace> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix map: <http://romanrepublic.ac.uk/rdf/entity/#> .
@prefix db: <http://romanrepublic.ac.uk/rdf/entity/> .

<http://romanrepublic.ac.uk/rdf/entity/Person/2072> a rdfs:Resource , :ThingWithID , :ThingWithName , :ThingWithExtraInfo ;
        rdfs:label "TULL2072 M. Tullius (29) M. f. M. n. Cor. Cicero" ;
        <http://www.openrdf.org/schema/sesame#directType> :Person ;
        :hasName "TULL2072 M. Tullius (29) M. f. M. n. Cor. Cicero" ;
        a :Person ;
        :hasCognomen "Cicero" ;
        :hasDprrID "TULL2072" ;
        :hasEraFrom "-106"^^xsd:integer ;
        :hasEraTo "-43"^^xsd:integer ;
        :hasFiliation "M. f. M. n." ;
        :hasHighestOffice "cos. 63" ;
        :hasID "2072"^^xsd:integer ;
        :hasNobilisNotes "[Cic.] Inv. Sall. 2.5 (founder of family's nobilitas), Vell. 2.34.3 (\"nova nobilitas\"), Macrob. 2.1.8ff, 3.11.9-11, Eutrop. 7.2, Schol. Bob. Sull. 80 Stangl, Schol. Bob.
Sest. 139, SHA. Tyr. Trig. 30.1 (entering nobility through union of Tullia to C. Piso Frugi q. 58)" ;
        :hasNomen "Tullius" ;
        :hasNovusNotes "Cic. Leg. Ag. 2.3; Cic. Leg. Ag. 2.4; Q. Cic. Comm. Pet. 4; Q. Cic. Comm. Pet. 2; Q. Cic. Comm. Pet. 7; Q. Cic. Comm. Pet. 13; Sall. Cat. 23; Vell. 2.34; Vell. 2.128; Cic. Att.
1.7; Asc. Tog. Cand. 94C" ;
        :hasOrigin "Arpinum" ;
        :hasPersonName "TULL2072 M. Tullius (29) M. f. M. n. Cor. Cicero" ;
        :hasPersonNote <http://romanrepublic.ac.uk/rdf/entity/PersonNote/2203> , <http://romanrepublic.ac.uk/rdf/entity/PersonNote/2204> , <http://romanrepublic.ac.uk/rdf/entity/PersonNote/2205> ,
<http://romanrepublic.ac.uk/rdf/entity/PersonNote/2206> , <http://romanrepublic.ac.uk/rdf/entity/PersonNote/2207> , <http://romanrepublic.ac.uk/rdf/entity/PersonNote/2208> , <http://romanrepublic.ac.uk
/rdf/entity/PersonNote/2209> , <http://romanrepublic.ac.uk/rdf/entity/PersonNote/2210> , <http://romanrepublic.ac.uk/rdf/entity/PersonNote/2211> ;
        :hasPraenomen <http://romanrepublic.ac.uk/rdf/entity/Praenomen/Marcus> ;
        :hasReNumber "29" ;
        :isNobilis "true"^^xsd:boolean ;
        :isNovus "true"^^xsd:boolean ;
        :isSex <http://romanrepublic.ac.uk/rdf/entity/Sex/Male> ;
        :hasAssociatedWebpage <http://www.romanrepublic.ac.uk/person/2072> .

<http://romanrepublic.ac.uk/rdf/entity/DateInformation/1886> :isAboutPerson <http://romanrepublic.ac.uk/rdf/entity/Person/2072> .

<http://romanrepublic.ac.uk/rdf/entity/DateInformation/527> :isAboutPerson <http://romanrepublic.ac.uk/rdf/entity/Person/2072> .

<http://romanrepublic.ac.uk/rdf/entity/DateInformation/528> :isAboutPerson <http://romanrepublic.ac.uk/rdf/entity/Person/2072> .

<http://romanrepublic.ac.uk/rdf/entity/DateInformation/529> :isAboutPerson <http://romanrepublic.ac.uk/rdf/entity/Person/2072> .

<http://romanrepublic.ac.uk/rdf/entity/DateInformation/530> :isAboutPerson <http://romanrepublic.ac.uk/rdf/entity/Person/2072> .

<http://romanrepublic.ac.uk/rdf/entity/PostAssertion/10164> :isAboutPerson <http://romanrepublic.ac.uk/rdf/entity/Person/2072> .
```

Figure 5. The DPRR RDF Server responds to http://romanrepublic.ac.uk/rdf/entity/Person/2072?format=rdf (Cicero)



Digital Prosopography of the Roman Republic

DPRR Home
DPRR RDF Documentation
DPRR Ontology Documentation

DPRR RDF Server
Explore
Namespaces
Types
Explore
Query
Export

Explore

Post Assertion: #5439

Resource:

Results per page: All

Results offset: Previous 0 Next 0

Show data types & language tags:

Subject	Predicate	Object	Context
<http://romanrepublic.ac.uk/rdf/entity/PostAssertion/5439>	rdf:type	rdfs:Resource	
<http://romanrepublic.ac.uk/rdf/entity/PostAssertion/5439>	rdf:type	vocab:Assertion	
<http://romanrepublic.ac.uk/rdf/entity/PostAssertion/5439>	rdf:type	vocab:ThingWithID	
<http://romanrepublic.ac.uk/rdf/entity/PostAssertion/5439>	rdf:type	vocab:AssertionWithDateRange	
<http://romanrepublic.ac.uk/rdf/entity/PostAssertion/5439>	<http://www.openrdf.org/schema/sesame#directType>	vocab:PostAssertion	
<http://romanrepublic.ac.uk/rdf/entity/PostAssertion/5439>	rdf:type	vocab:PostAssertion	
<http://romanrepublic.ac.uk/rdf/entity/PostAssertion/5439>	rdfs:label	"Post Assertion: #5439"	
<http://romanrepublic.ac.uk/rdf/entity/PostAssertion/5439>	vocab:hasDateEnd	-63	
<http://romanrepublic.ac.uk/rdf/entity/PostAssertion/5439>	vocab:hasDateSecondarySource	<http://romanrepublic.ac.uk/rdf/entity/SecondarySource/2>	
<http://romanrepublic.ac.uk/rdf/entity/PostAssertion/5439>	vocab:hasDateStart	-63	
<http://romanrepublic.ac.uk/rdf/entity/PostAssertion/5439>	vocab:hasID	5439	
<http://romanrepublic.ac.uk/rdf/entity/PostAssertion/5439>	vocab:hasOffice	<http://romanrepublic.ac.uk/rdf/entity/Office/3>	
<http://romanrepublic.ac.uk/rdf/entity/PostAssertion/5439>	vocab:hasOfficeXref	"Pr. 66"	
<http://romanrepublic.ac.uk/rdf/entity/PostAssertion/5439>	vocab:hasOriginalText	"M. Tullius M. f. M. n. Cicero (29)"	
<http://romanrepublic.ac.uk/rdf/entity/PostAssertion/5439>	vocab:hasPosition	1	
<http://romanrepublic.ac.uk/rdf/entity/PostAssertion/5439>	vocab:hasPostAssertionNote	<http://romanrepublic.ac.uk/rdf/entity/PostAssertionNote/4845>	
<http://romanrepublic.ac.uk/rdf/entity/PostAssertion/5439>	vocab:hasSecondarySource	<http://romanrepublic.ac.uk/rdf/entity/SecondarySource/2>	
<http://romanrepublic.ac.uk/rdf/entity/PostAssertion/5439>	vocab:isAboutPerson	<http://romanrepublic.ac.uk/rdf/entity/Person/2072>	

Figure 6. The DPRR RDF Server responds to http://romanrepublic.ac.uk/rdf/entity/PostAssertion/5439

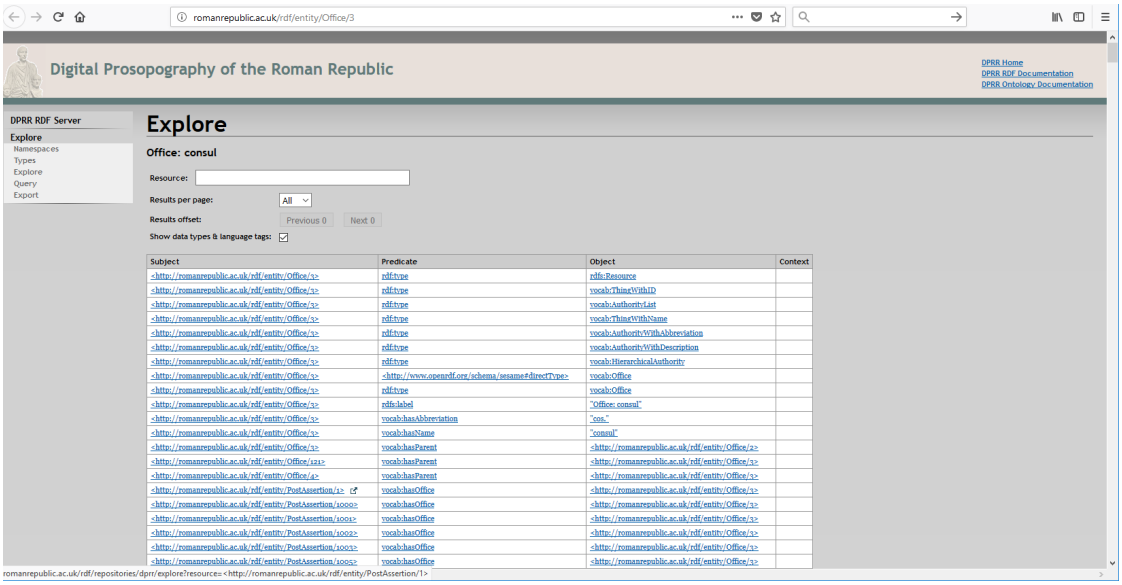


Figure 7. The DPRR RDF Server responds to http://romanrepublic.ac.uk/rdf/entity/Office/3 (Office of Consul)

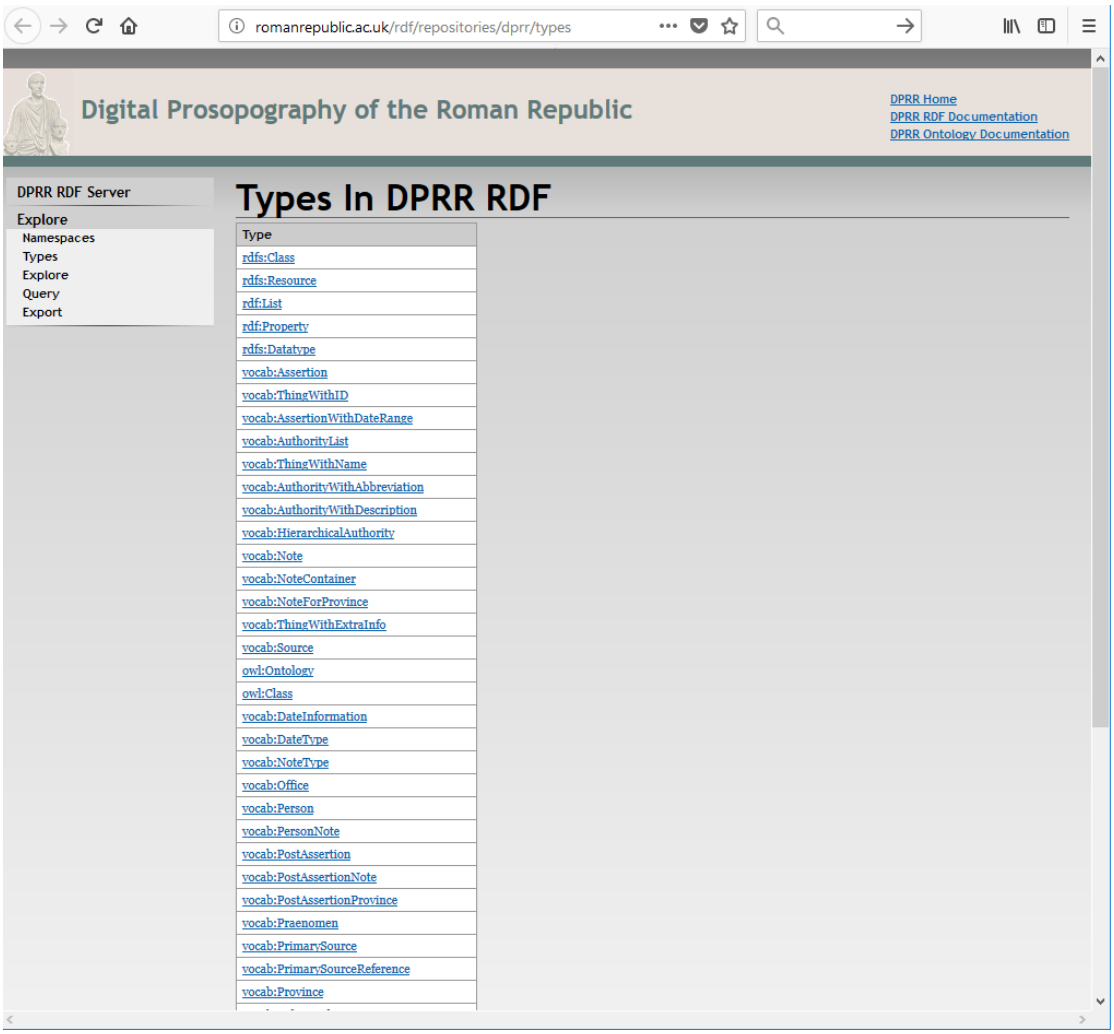


Figure 8. The DPRR RDF Server displays the list of types in the DPRR RDF dataset

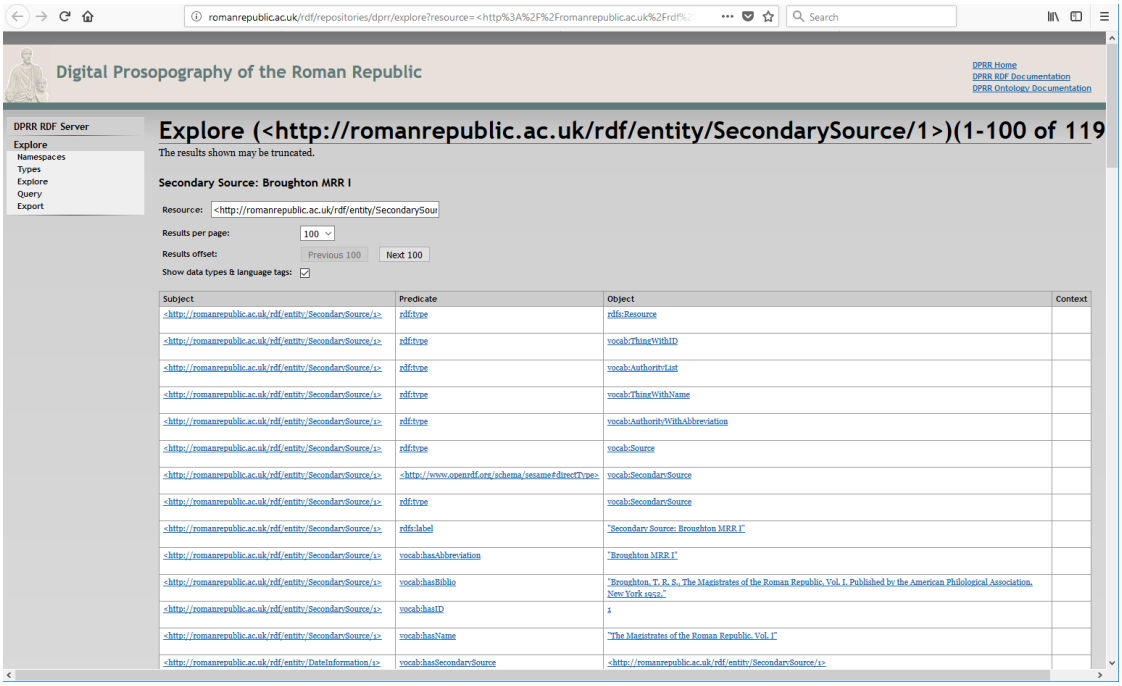


Figure 9. The DPRR RDF Server responds to http://romanrepublic.ac.uk/rdf/entity/SecondarySource/1 (Broughton Vol 1)

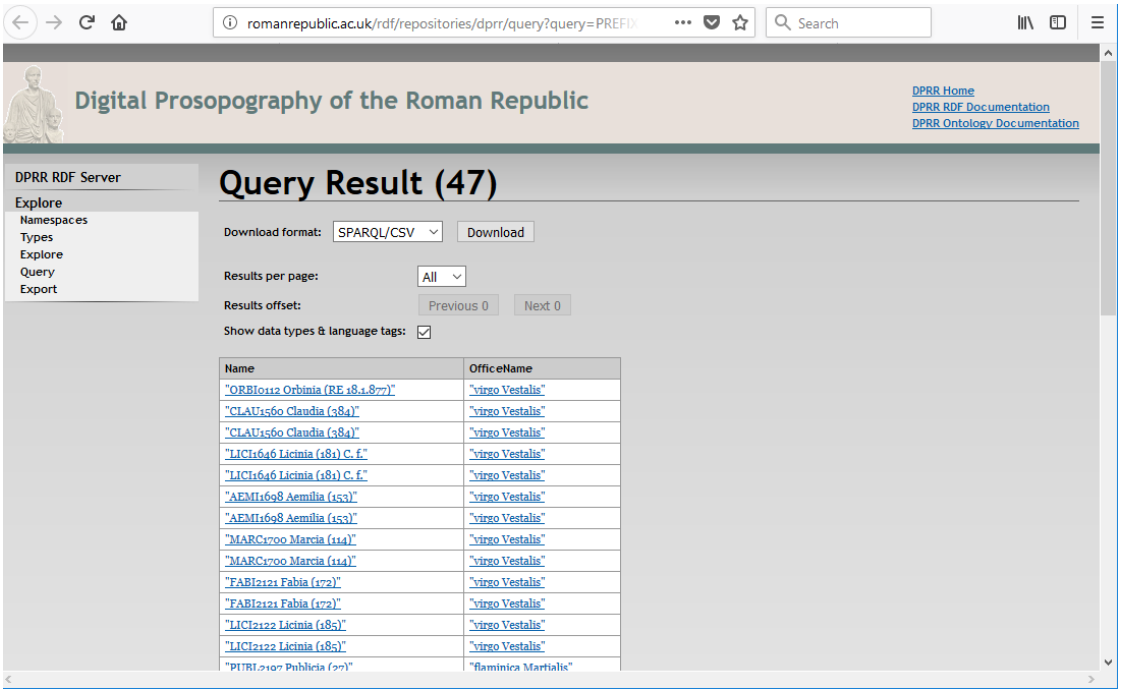


Figure 10. The DPRR RDF Server responds to a SPARQL query (list of women who held offices)

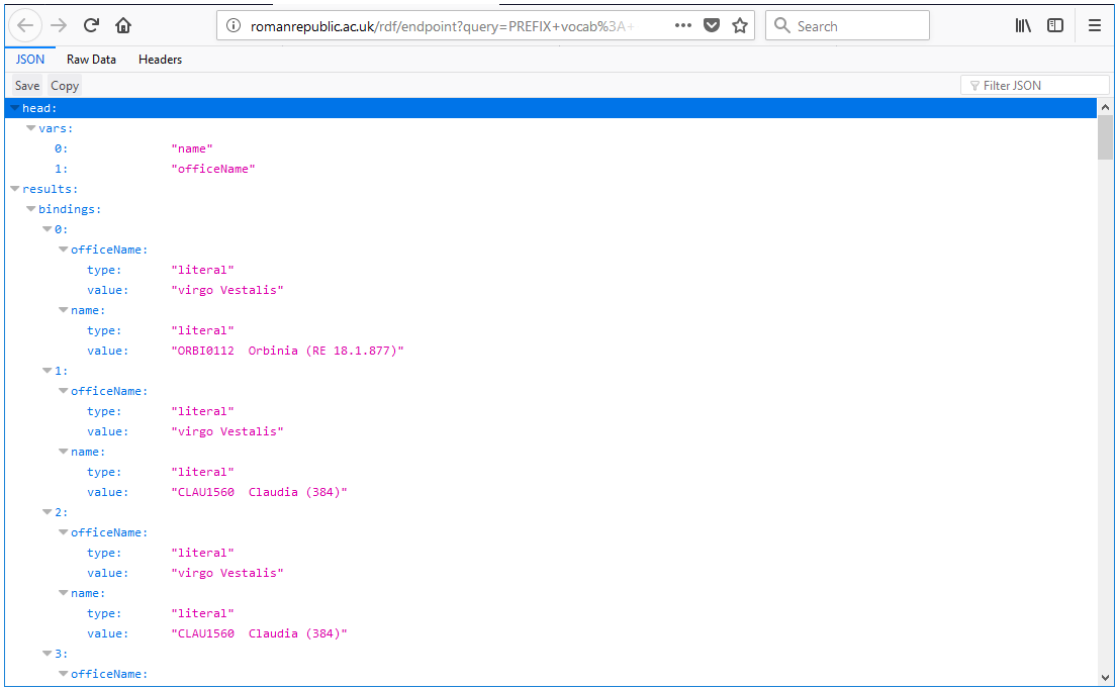


Figure 11. The DPRR RDF Server responds to a SPARQL query with JSON data

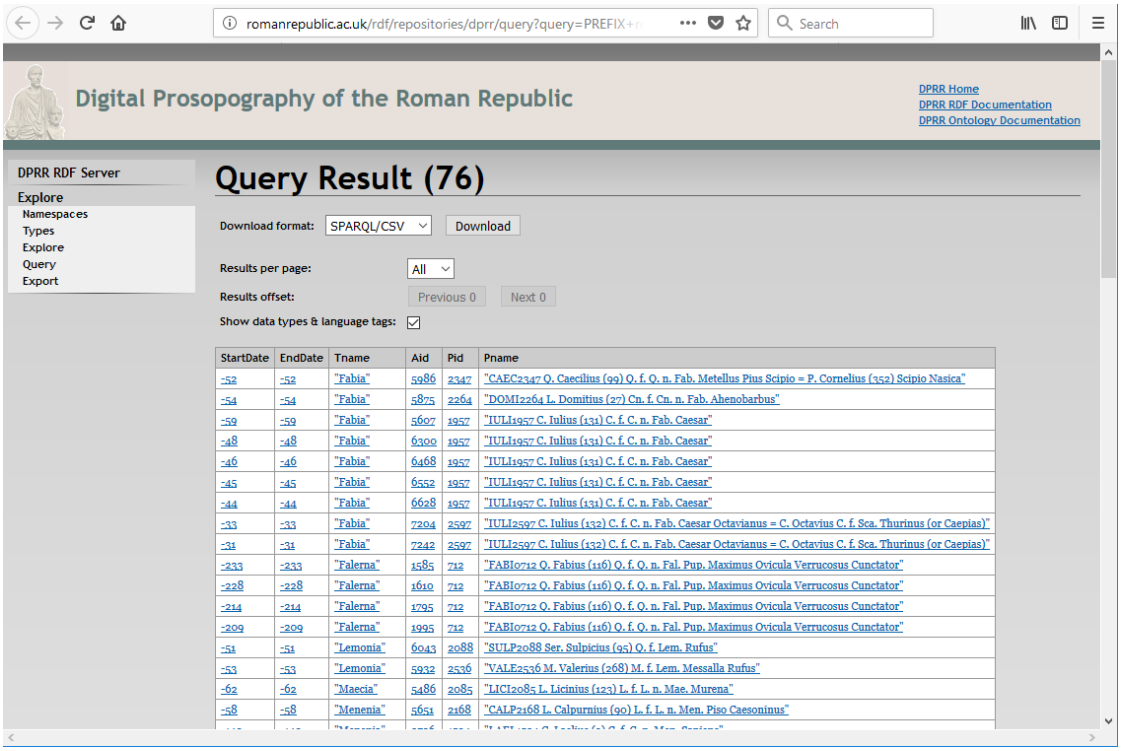


Figure 12. The DPRR RDF Server responds to a SPARQL query about tribes of consuls over time

Acknowledgements

DPRR, like all similar DH projects in which I have been involved, involves a range of organisations and people. There are, thus, several organisations and people to acknowledge here. First, one must thank the UK's Art's and Humanities Research Council, who funded the academic and technical work involved in the creation of the DPRR data and website. Second, I would like to thank the academic colleagues in the DPRR project with whom we worked for several years to create the data. Third, I must thank my colleagues at King's Digital Lab group (KDL), who have made it possible for the

RDF Server described here to be publicly hosted on their servers. And finally, I should thank colleagues at DHQ who not only provided their stellar editorial support for this article, but agreed to extend their technical infrastructure to allow this article's unusual integrated interactive components to be accommodated by their digital publishing system.

Works Cited

Beckett et al 2014 Beckett, David, Tim Berners-Lee, Eric Prud'hommeaux and Gavin Carothers, *RDF 1.1 Turtle: Terse RDF Triple Language*. W3C documentation website. https://www.w3.org/TR/turtle/

Berners-Lee 1998 Berners-Lee, Tim, *Relational Databases on the Semantic Web*. https://www.w3.org/DesignIssues/RDB-RDF.html.

Berners-Lee 2010 Berners-Lee, Tim (2010). *The year open data went worldwide*. TED Talk. https://www.ted.com/talks/tim_berners_lee_the_year_open_data_went_worldwide

Berners-Lee et al 2001 Berners-Lee, Tim, James Hendler and Ora Lassila, “The Semantic Web: A new form of Web content”. In *Scientific American* Vol 284 No 5 (May 2001), pp 35-43

Bizer 2008 Bizer, Christian, Tom Heath, Kingsley Idehen and Tim Berners-Lee. “Linked Data on the Web (LDOW2008)”. Workshop at *WWW 2008*, April 2008, Beijing, China.

Bizer et al 2009 Bizer, Christian, Tom Heath and Tim Berners-Lee. “Linked Data: the Story So Far”. In *International Journal on Semantic Web and Information Systems*. 5 (3): 1-22. doi:10.4018/jswis.2009081901. ISSN 1552-6283. http://tomheath.com/papers/bizer-heath-berners-lee-ijswis-linked-data.pdf.

Bradley 2017a Bradley, John, *Factoids: A site that introduces Factoid Prosopography*. http://factoid-dighum.kcl.ac.uk/.

Bradley 2017b Bradley, John (2017b). *DPRR RDF: Documentation* website. http://www.romanrepublic.ac.uk/rdf/doc.

Broughton 1951-2, 1986 Broughton, T. Robert S, *The Magistrates of the Roman Republic*. In series De Lacy, Phillip H. (ed) *Philological Monographs*. Atlanta: Scholars Press edition.

D2RQ nd *D2RQ: Accessing Relational Databases as Virtual RDF Graphs*. http://d2rq.org/.

Fogel and Engerman 1974 Fogel, Robert William and Stanley L. Engerman, *Time on the Cross: The Economics of American Negro Slavery*. Boston: Little, Brown and Company, 1974. xviii + 286 pp.

Freitas et al 2012 Freitas, Andre, Edward Currey, JG Oliveira and S O'Riain, “Querying Hetrogeneous Datasets on the Linked Data Web: Challenges, Approaches, and Trends”. In *IEEE Internet Computing*, Vol 16, No 1. Jan-Feb 2012. pp. 24-33.

Guetzkow et al 2004 Guetzkow, Joshua, Michele Lamond and Gregoire Mallard, “What is Originality in the Humanities and the Social Sciences?”. In *American Sociological Review*. Vol 69 No 2 (Apr. 2004). pp 190-212. Online access from JSTOR: http://www.jstor/org/stable/3593084.

Hammond and Jackson 2017 Hammond, Matthew and Jackson, Cornell, *Social Network Analysis and the People of Medieval Scotland 1093-1286 (PoMS) Database*. https://www.poms.ac.uk/information/e-books/social-network-analysis-and-the-people-of-medieval-scotland-1093-1286-poms-database/.

Johanson et al 2019 Johanson, Christopher, Marie Saldana and Benjamin Niedzielski, *RomeLab*. DOI 10.17605/OSF.IO/VGKT4. https://osf.io/vgkt4/ and http://hvwc.etc.ucla.edu/.

Krug 2014 Krug, Steve, *Don't Make Me Think, Revisited: A Common Sense Approach to Web Usability*. Amazon (3rd ed.). New Riders. ASIN 0321965515.

Mouritsen et al 2017 Mouritsen, Henrik, Dominic Rathbone, Maggie Robb, John Bradley, *Digital Prosopography of the Roman Republic*. http://www.romanrepublic.ac.uk/.

Noy et al 2019 Noy, Natasha, Yuqing Gao, Anshu Jain, Anant Narayanan, Alan Patterson and Jamie Taylor, “Industry-scale Knowledge Graphs: Lessons and Challenges”. In *Communication of the ACM*. Vol 62 No 8. pp. 36-43.

Nurmikko-Fuller 2016 Nurmikko-Fuller, Terhi, “Linked Data for Digital Humanities: Publishing, Querying and Linking on the Semantic Web”. Announced on the Oxford Summer School website. http://digital.humanities.ox.ac.uk/dhoxss/2016/workshops/LD4DH.

OWL 2012 *OWL: Web Ontology Language*. W3C documentation website. https://www.w3.org/OWL.

PASE 2016 *Prosopography of Anglo-Saxon England*. Most recently updated with results of work in the *Profile of the Doomed Elite Project* (Stephen Baxter). http://www.pase.ac.uk.

Pauly et al 1893- Pauly, August, Georg Wissowa, Wilhelm Kroll, Kurt Witte, Karl Mittelhaus, Konrat Ziegler, (eds), *Real-Encyclopaedie der classischen Altertumswissenschaft*. Stuttgart: J. B. Metzler, 1893-1980.

Pelagios nd *Pelagios Commons: Linking the places of our past*. http://commons.pelagios.org/.

PoMS 2014 *People of Medieval Scotland*. Most recently updated in 2019. http://www.poms.ac.uk.

RDF 2014 *RDF: Resource Description Framework*. W3C documentation website. https://www.w3.org/RDF/.

RDF4J 2017 @*rdf4j* website. http://rdf4j.org/.

SNAP-DRGN nd *Standards for Networking Ancient Prosopographies*. https://snapdrgn.net/.

SPARQL 2013 *SPARQL 1.1 Overview*. W3C website at https://www.w3.org/TR/sparql11-overview/.

SQL 2018 *SQL* Wikipedia article. https://en.wikipedia.org/wiki/SQL.

Seaborne et al 2013 Seaborne, Andy, Axel Polleres, Lee Feigenbaum and Gregory Todd Williamms, *SPARQL 1.1 Federated Query*. A W3C Recommendation at https://www.w3.org/TR/sparql11-federated-query/.

Smith 2017 Smith, James, Workshop description for *RDF and Linked Open Data*. http://dhsi.org/courses.php.

Smithies et al 2019 Smithies, James, Carina Westling, Anna-Maria Sichani, Pam Mellen and Arianna Ciula, “Managing 100 Digital Humanities Projects: Digital Scholarship & Archiving in Kings Digital Lab”. In *Digital Humanities Quarterly*, Vol. 13 No. 1. http://www.digitalhumanities.org/dhq/vol/13/1/000411/000411.html.

Speicher et al 2015 Speicher, Steve, John Arwe and Ashok Malhotra, *Linked Data Platform 1.0: W3C Recommendation 26 February 2015*. https://www.w3.org/TR/2015/REC-ldp-20150226/.

VIAF 2010-16 *VIAF: The Virtual International Authority File*.https://viaf.org/.

Wiess 2001 Wiess, Thomas, “Review of Time on the Cross: The Economics of American Negro Slavery”. In *Economic History Association website EH.net*. https://eh.net/book_reviews/time-on-the-cross-the-economics-of-american-negro-slavery/.

WorldCat nd *WorldCat: The World's The World's Largest Library Catalog*. http://www.worldcat.org/.

XQUERY 2018 *XQuery*. Wikipedia article. https://en.wikipedia.org/wiki/XQuery.