DHQ: Digital Humanities Quarterly
2017
Volume 11 Number 1

Fitting Personal Interpretation with the Semantic Web: lessons learned from *Pliny*

John Bradley <john_dot_bradley_at_kcl_dot_ac_dot_uk>, King's College London
Michele Pasin <michele_dot_pasin_at_gmail_dot_com>, Springer Nature

Abstract

In this paper we expand Stefan Gradmann's question at WWW2012 "Thinking in the graph: will Digital Humanists ever do so?" to consider whether humanists, more generally than just "digital" ones, might do thinking that is, at least to some useful degree, "in the graph" too. Drawing on the experience of the Pliny project, and recent work done within that project to explore how Pliny materials might connect with the semantic web, we explore ways in which structured "graph-like" thinking might be revealed in — to "peek out" from — parts of humanities research that is common to digital and non-digital humanists alike. Out of this, we propose a number of different ways that scholars might engage with the Semantic Web, and provide examples – arising from the building of a prototype extension to Pliny – of how these engagements could be dealt with. We also explore the challenge of ambiguity and incompleteness in scholarship, explain how 2D space operates in Pliny to cope, to some degree at least, with these issues, and consider the boundaries between the expressiveness of 2D space and the formal graph model of the Semantic Web. We end by proposing several possible avenues for future work that arise from our work so far.

to think is often to sort, to store and to shuffle: humble, embodied tasks [Lachance 2013]

Introduction

In his WWW2012 presentation Stefan Gradmann raised the question "Thinking in the graph: will Digital Humanists ever do so?" [Gradmann 2012, slide 2]. His aim there was to consider how the Semantic Web [Berners-Lee et al 2001] and its data oriented technologies such as RDF [Cyganiak, Wood and Lanthaler 2014] might be applied to humanities-oriented materials. Like Berners-Lee and others (see [Berners-Lee 2009] and [Berners-Lee 2010]), he claimed that digital humanists could exploit data made available through these approaches in ways that would not have been practical in the past and thereby enable the creation of entirely new research results. 1

Given that he was speaking about the data-oriented Semantic Web, Gradmann naturally took a data-oriented perspective in his paper, drawing on work done with structured text and databases. In this paper, however, we intend to start with what will appear to be a very similar question as his, but will view it from what will become a very different starting point. As a result, we believe that we will have things to say that are quite different from Gradmann. Indeed, we think that however thought-provoking Gradmann's "thinking in the graph" for *digital humanists* might be, taking our perspective allows one to explore the perhaps even more fundamental question: "Thinking in the graph: will Humanists (more generally) ever do so?". It is this more general question that we will explore here. 2

Perhaps the best bridge from Gradmann's work into our own comes through what we think of as the significant provocations contained in both the words "thinking" and "graph" in Gradmann's question. We will return to the "thinking" part a little later in this paper, but let us start with the "graph" word right away, since those readers who are not familiar with the underpinnings of the Semantic Web might be puzzled by its appearance here. The Semantic Web is an approach to dealing with globally distributed and interconnected highly structured data. For any domain to which it is applied, the material it represents is captured in an (often very large) set of simple assertions which are recorded as "triples". Each triple is made up of two digital entities representing a *subject* and *object* joined by a third digital object (called a *predicate*) that expresses a connection between them. All three components of the triple can be expressed as WWW URIs, which gives them a global scope. It turns out that a set of interrelated triples-assertions of this kind can be 3

represented in what is called a mathematical *graph* – which is a collection of nodes with connections between them. Indeed, a collection of such triples is called an *RDF Graph* in RDF's *Concepts and Abstract Syntax* document [Cyganiak, Wood and Lanthaler 2014, sections 1.1 and 3]. So, Gradmann's question is meant to focus our attention on whether digital humanists can represent things that interest them in terms of highly structured data (the assertions through triples) that can be represented satisfactorily using a mathematical graph.

4 Gradmann's example of humanities materials in his talk are rather data-oriented to start with, so one can readily see how using RDF provides an appropriate representation of their materials. Indeed, we here at DDH have had more than 20 years of experience with data-oriented collaborative project work such as that on structured prosopography (see [Bradley and Short 2004] and [Pasin and Bradley 2015], for example). We can certainly agree with Gradmann that the products of our data-oriented projects also can be thought of as both clearly useful products of humanities research that are useful to other humanities researchers, and yet are compatible with this data-oriented approach. Furthermore, by their nature as graph-oriented structured data they indeed have the potential to be suitable for further processing in, say, mashups created by independent researchers that might represent new research for the humanities in the way that Gradmann talks about. However, we have already broadened Gradmann's phrase from *digital humanities* to *the humanities*, and thus must include research work in the humanities outside of the specifically *digital* humanities. The fundamental question, then, is how, if at all, the more general processes and products of humanities scholarship (which conventionally produces research product as texts and not as data) are compatible with the Semantic Web's graph representation which is centered on a formalisation of highly structured data-like materials.

5 This distinction between data-oriented humanities projects and other humanities research is important because, as interesting as these data-oriented projects are, one must surely say that these data-oriented projects require a way of thinking about humanities materials that is quite foreign to the bulk of humanist scholars. In fact, we do not believe that structured data resources themselves, or mashup-like projects that combine data to form new expressions of existing materials from them, can be called *typical* humanities research. These structured data digital resource projects are things that most conventional humanist scholars can hopefully appreciate to be of value, and might well be things they would use as resources in their own work, but they are not things that they would think of producing as research themselves. Indeed, it is the very mathematical-like structural formality of these data-oriented products that makes them seem rather foreign. Is there, then, a place for the mathematical graph and its associated formal approaches for this majority of scholars who view their principal research sources as primary and secondary texts and do not produce resources that obviously work as graph-like datasets but instead appear in the form of books and articles?

6 Our ideas here grow out of our work on the *Pliny* project [Pliny 2009], which was started by one of the authors (Bradley) as far back as 2004 and which proposed a kind of digital framework to support scholarship. *Pliny* is meant to support humanities scholarship, and, as we shall explain in this paper, reveals a way to connect scholarship to the Semantic Web that takes quite a different tack from the data-oriented Linked Data / Mashup paradigm implied by Gradmann's 2012 paper. Instead, we believe that our approach suggests how a Semantic Web framework (while still having a data-oriented focus) could be made to fit to an interesting degree with scholarly research practice that is not apparently data-oriented at all. Thus, much of the rest of this paper aims to explore, with *Pliny* as the exemplar, how this connection between traditional scholarly practice (which is very different from the work involved in the creation of our data-oriented resources) and Semantic Web and Linked Data ideas might be usefully made.

7 We offer this article, then, as a kind of thought piece about the connection between traditional humanities scholarly thinking and formal models like mathematical graphs. Like Gradmann, we also believe that there is a need for a discussion not so much "about infrastructure but about epistemological foundations" [Gradmann 2012, slide 29]. Indeed, we believe that there is real work to be done to engage the Semantic Web with humanities scholarship, and that it will best arise out of thinking about how the epistemological underpinnings for traditional scholarship might relate to the band of technologies often called Knowledge Representation, and in particular to the formal structures the Semantic Web offers. We cannot claim that our model, as described here, provides *the definitive* answer to this issue, but it does provide at least a plausible *approach* that can lead us to ask relevant questions about possible connections between formal data and humanities scholarship, and, we believe, deserves further investigation.

8 In this paper, then, we first consider what we believe is a conventional view of what constitutes humanities scholarship, and its expression as prose text. We then look at how the *Pliny* project has provided a model that supports these activities. Then, we change the character of the article from a somewhat theoretical reflection about research to explore one way in which the work of scholarship, as *Pliny* supports it, might explicitly map to the formalisms of the Semantic Web. In this way we propose a model of how the Semantic Web could connect formally with *Pliny*'s structures, and through them, to these conventional, non-data oriented, parts of scholarship. Finally, we look at some of the kinds of future work that could flow from our approach.

Thinking and writing in humanities research

9 As we have said above, for most humanists it is probably true that the work of doing scholarly writing and its outputs may not seem to be at all compatible with the formal structures of systems like the Semantic Web. We have, in [Bradley 2014], touched upon some of the issues that explain why so much humanist research is presented in terms of, as Hayden White explains it, "a narrative prose discourse", and why language is used as the expression of this research. It is precisely *because* it is "imprecise and slippery" [Bodenhamer 2008, 224] that it is used. See, for example, [Louch 1969] and [Rüsen 1987] for a perspective on the nature of prose for humanities research from a historiographical point of view. As we considered in [Bradley 2014], this key use of text to represent the products of research fits obviously rather uneasily with the clear binary-like Cartesian structural approaches of all the conventional formal representation systems: XML, databases, and the Semantic Web.

10 Furthermore, these traditional views on what constitutes the fundamental aspect of humanities research – reading and writing – might appear to many twenty-first century *digital* humanists as representing exactly the aspects of humanities research that are outside their domain of interest. Nonetheless, we will make the case here that there is good reason to try to find potential digital aspects of them that could fit with the perspective of a digital humanities agenda. First, and perhaps most obviously, this is because even DH techniques such as, say, those associated with Franco Moretti's *distant reading* – which relies on structured, formal data as input – are only a part of an entire scholarly process and cannot be full replacements, by themselves, for traditional scholarly methods – this even for DH researchers themselves. The new data-oriented DH methods such as distant reading are often cast as borrowing of aspects of scientific methodology for the humanities. In the sciences, however, the use of scientific instruments and the collecting and processing of data from them generally is not considered to be the be-all and end-all of scientific work – one has to then publish articles that interpret this data: articles that are published either in traditional science journals, or perhaps nowadays online in major science research repositories. Indeed, scientific progress is made through the publication of ideas that arise *from* this data through the scientist's thinking about, and developing an interpretation of them. The final research product for much data-oriented scientific research is still an article in scientific prose.

11 In a similar way, formal data-driven DH approaches such as distant reading, social network analysis, or text mining cannot eliminate, on their own, the need for scholars to think about what they have found through these new techniques and then to present their understanding of the significance that has emerged from their exploration. It *is* true that, unlike traditional humanists, the DH-driven scholar may not only be reading books as sources for their research and may well be also drawing on these new DH approaches for insights and ideas. However, Moretti's work using distant reading techniques to explore the emergence of the form of the novel was not finished when the distant reading DH-part, the number crunching and graphing, was completed. He still needed to think about the meaning of his data and its analysis and indeed to write it up in that most traditional of scholarly communications: the monograph [Moretti 2005]. Furthermore, although his ideas arose, in part, out of his engagement with formal data, they also came out of his more general understanding of western literary and social culture. He put these two kinds of understanding together, added some further substantial insights of his own, and produced his book out of this effort. Moretti's practice is not unique. For the great majority of DH researchers, the work of using the DH methodologies and their tools/instruments (such as, say, those found in big data methodologies) is not the whole story of their research activities either. Significant intellectual work continues, following after the application of these tools, in that phase of research between exploration and publication. So, we intend in this paper to shift our gaze from the input/exploration part of the research to the work that follows after it and that results in an original written text as the product of the research – the external evidence of scholarship. How do we fit it with a representation of scholarship that is highly structured in the form that RDF and the other associated Semantic Web technologies are?

12 At first glance at least, as we remarked earlier, scholarly humanities writing as the expression of research output does not appear to fit well with the structured methodologies of the Semantic Web or with Gradmann's question about the place of graph-thinking in the humanities. Two things, however, begin to hint at a *place* for structured data of the kind represented by the Semantic Web (SW) in scholarly writing. If we look closely at most articles or books presenting research in the humanities, we will usually see that there *is* some formal structure at least implied in the text, both directly evident (first) in the structure represented by the flow of the argument, but also (second) in the identification and naming of themes, concepts and their connections that are presented in them. These named themes and concepts can begin, perhaps right off, to look something like things that can be expressed and shared by Semantic Web technologies, particularly ontologies. They at least hint at structure of the kind with which the Semantic Web operates.

13 However, there is an important difference between these conceptual structures and those that often are thought of as characterising the Semantic Web's ontologies and other semantic structures. The SW, through its ontology component, aims to support machine-based selection and manipulation of data by providing a formal representation of established

knowledge about the world that is both based upon a *shared* human understanding of it and that is simultaneously accessible to the technology. Indeed, it is through this formalised shared understanding of the structured data that much of the potential power of the Semantic Web comes. In the SW the expression of this shared understanding of the data is expressed in the formalisms of what are called computer *ontologies*. Thomas Gruber's [Gruber 1993] early description of an ontology as a shared common understanding of a collection of materials expressed using a formal system reveals this approach, and indeed the emphasis in the Semantic Web's ontologies for a particular domain is almost always centered on the formal definition of widely agreed terms and ideas.

14 In contrast to this, however, as Guetzkow *et al* remind us in their article "What is Originality in the Humanities and the Social Sciences?" [Guetzkow et al 2004], humanities scholars want to say something *new* about their materials, and hence will generally not want to refer simply to ideas contained in an existing mature intellectual framework that could (perhaps — and only "perhaps") be represented in a formal Semantic Web ontology. As Guetzkow *et al* state " [h]umanists and historians clearly privilege originality in approach" [Guetzkow et al 2004, 190]. In their interviews of peer review panels in the humanities they found that "panelists described originality ... in terms of the novelty of the overall approach used by the researcher (who is 'bringing a fresh perspective')" [Guetzkow et al 2004, 192]. Humanities scholars, then, are encouraged to develop their *own* voice and perspective on their material that is different from that currently established within their discipline. Their work will usually, of course, be grounded in some existing ideas shared with academic colleagues, but the best of it will also extend or perhaps even more fundamentally break with past models by introducing new themes, concepts, or connections. If it was possible – and this is a big "if" – to have a formal representation of a set of widely shared ideas for a particular humanities discipline (an ontology rich enough to capture the current understanding around the issues of, say, colonialism), then an original piece of scholarship that fits with this colonialism field might then be expressible *in part* as a kind of linked data connection to these established concepts. Even then, however, it will also almost certainly contain at least some material that is original with the researcher and that does not link to any such existing shared, formal, representation of the field. Indeed, perhaps one could claim that the aim of all scholarship in the humanities and elsewhere is to challenge the boundaries of shared knowledge and understanding. For this reason, materials that represent new scholarship would seem inevitably to go outside the natural domain of SW ontologies.

15 Furthermore, although the output that presents this original research usually eventually appears as prose text, the process to get to this presentation in prose is not captured by simply sitting down one day and letting the paper simply flow out. There is a process going on from which the ideas and how to present them emerge. We will call this process *interpretation development*, and the process of emergence that it represents is one that takes ideas for this interpretation from vagueness to relative clarity. Any structure for this interpretation does not just appear, fully formed, in the author's head at the time the article was being written. It emerges after substantial engagement with the materials with which s/he was working. Before the article writing begins, but as a part of the research process that might lead to the article, it is likely that the ideas are still only partially formed, and cannot be expressed in the rigid formalisms of, say, a Semantic Web ontology; indeed they most likely will not be initially ready for expression in the seemingly less mathematically formal medium of written prose text either. At this stage the ideas are in fact, let us say, *pre-ontological*, and bringing Semantic Web ontology technologies into the picture at this point would appear to bring formalism to bear too soon in the process of developing them. Considerable thinking needs to happen after the engagement with the materials being studied before an interpretation of these materials is mature enough to be shared with others.

16 We have just given some related reasons why the differences between humanities scholarship and Semantic Web representation through graphs and ontologies might seem to be insurmountable. Nonetheless, the rest of this article outlines the approach taken by the *Pliny* project that suggests at least a possible way to think about connections between the Semantic Web and this interpretation development activity, with its gradual development through vagueness to clarity.

17 Here, then, as promised, we return to Gradmann's phrase that we introduced at the beginning of this paper and to his use of the word "thinking" in it: for it is in the work between the gathering of materials and the writing of a personal interpretation of them that a lot of thinking has to happen. By exploring issues in the formally challenging context of more mainstream research outputs – articles, books, etc. – produced in the humanities and the kind of thinking that goes into their production, we believe that one can discover places for structure and graphs there as well. We believe that François Lachance's object-oriented, embodied, understanding of thinking (quoted at the top of this paper) suggests something important about the thinking process that goes on in traditional scholarship, and reveals an aspect of structure in that thinking that can be latched onto. Although perhaps the outward expression of this thinking-process may not at first look much like "thinking in the graph" as it is usually modelled in the Semantic Web, it seems to us that there are in fact bits and pieces of a graph approach that might well peek out, as it were, from what we can discern there and these bits of a graph model can suggest ways to consider, both positively and negatively, possible

connections between the graph, the Semantic Web, and traditional scholarly thinking in the humanities. We intend to explore some of these bits of graph model thinking that we believe are "peeking out" not so much to provide definitive answers about how they work, but to encourage further work by those familiar with the Semantic Web about the place of Semantic Web-like formal structure in traditional, non-data oriented, scholarship, and how what structure we find in it might be engaged to facilitate that aspect of scholarship in the humanities.

Pliny and the processes of Scholarly Interpretation

Work on the *Pliny* project [Pliny 2009] began in 2004 and started with building a tool (also called *Pliny*) that could support the *process* of doing humanities scholarship. Although apparently often thought of as a kind of annotation tool, *Pliny* did not turn out to be particularly about annotation, or at least not about annotation in isolation from its place in scholarship. Instead, it came to represent how a digital approach to annotation could fit with other thoughts about the representation of ideas. In this view, digital annotation became only a starting point of research, and *Pliny* became a tool that aimed to support a fuller range of humanities scholarship activities than those that an annotation tool would have done by itself. Furthermore, the work with the ideas that *Pliny* was exploring led us to come to a view about how the "thinking" parts of conventional scholarly methodology that is used not only by non-DH humanists, but by DH-oriented humanists too, could be usefully connected to the new DH instrument-oriented approaches such as text visualisation, text mining, etc. and their tools (see [Bradley 2012]). 18

Instead, then, of the *Pliny* project being focused only on the production of the *Pliny* software, the software has acted within the project as a thought-piece. It allowed us to experiment rather broadly with questions of what place there might be for technology in both fully conventional scholarship and those aspects of DH practice that, although informed by new data and approaches, continues to also follow conventional scholarly methods: observing, thinking about and interpreting what one has seen, and presenting it. It is this aspect of what the *Pliny* software was for within the *Pliny* project that connects, we think, with Gradmann's view [Gradmann 2012, slide 29], mentioned earlier, that an investigation of the epistemological foundations of scholarship was where the issues were located. The authors of this present paper are not equipped to write here about epistemology in a philosophical way; however, we hope that the thrust of this paper – indeed the work in *Pliny* from near its very beginnings in 2004 – has something to offer to this epistemological discussion that Gradmann seems to think is necessary. The intent in our paper is to draw attention to *Pliny*'s model for the process of scholarship – both conventional and DH-driven – and to explore how it might connect to some degree with the new developments of knowledge representation in the Semantic Web. 19

Some of the early thinking that led to *Pliny* appears in [Bradley 2003]. However, thoughts about what kind of software *could* be more broadly useful to humanities researchers became clearer upon Bradley's discovery of the work of Brockman *et al* in their 2001 CLIR report entitled *Scholarly Work in the Humanities* [Brockman et al 2001]. There one could see the central place of reading in scholarship, and the significance of notetaking while reading. Similar ideas about scholarship and notetaking from the perspective of an historian also appear in the work of Ann Blair in her studies on the history of notetaking [Blair 2003] [Blair 2004]. There she claims that personal notes constitute a "central but often hidden phase in the transmission of knowledge" [Blair 2004, 85]. 20

A first lesson one finds in studies of what goes on in conventional scholarship is that in almost any substantive humanities scholarly research venture one expects to start out with only an ill-defined sense of the issues one is interested in. Indeed, as far back as 1997 John Lavagnino observes this aspect of notetaking when he observes that reading was not "a mere collection of data". Instead, he claims that the role of reading in scholarly research can be found in the fact that it "generates reactions" in the reader that "subsequently" (note the use of the word) one could seek to "describe or explain" [Lavagnino 1997, 114]. Thus, the mere act of *taking* of notes and/or annotation does not, by itself, capture the central place of these notes in doing the research. Instead, it is important to see how the notes might *subsequently* assist their owner in the gradual development of new ideas that would eventually become the primary result of the research work. 21

In this light we can see a problem with much of the current work on digital annotation when it is applied to the task of supporting scholarship. On one hand, several annotation tools have been developed to support simple digital textual annotation – linking a bit of text that can act as a comment to something you are reading. Although this is useful for, say, adding public commentary to a text for teaching purposes (see discussion in this vein in [Bessette 2015] and [Dean 2015]), by itself it doesn't serve the needs of the researcher particularly well because, although it could be used as a way to record responses to the text in the way Lavagnino describes, that is all one can do with the annotations. It leaves the user there – at the beginning of a process – without good broader access to these notes that s/he can subsequently use. In contrast to this "textual" annotation approach, other work has been carried out which has been given the general name of *Semantic Annotation* (e.g. [Grassi et al 2013]. This approach has become another way to think about 22

annotation in the context of the Semantic Web and it allows one to link material in a text to elements of a formal Semantic Web structure such as an ontology. Unfortunately, as we explained above, Semantic Web annotation would seem to bring in the formal structure of the linked ontology in too soon; trying to apply an approach suitable for a predefined, formal, interpretive model to the beginning of the process before – in the case of much research in the humanities – a model is available. Although both "simple text" and "semantic" annotation are useful tools, neither engages adequately by themselves with the process of scholarly research.

23 Much of a scholar's subsequent work after annotation of a source involves struggling with emerging vague, incompletely defined, ideas, and only after a good period of time does some degree of clarity emerge, a process that we described as *pre-ontological* earlier in this paper. Once one begins to see interpretation building as this kind of a process, with perhaps a "clear-ish" conception emerging near its end, the question begins to reveal itself as being not only about formal models for the completed interpretation, but also about how to model the process to help someone *develop* it. What should a user interface and the formal structure behind it be like in order to help a researcher while they are developing their interpretation — moving from vagueness to a significant degree of clarity?

24 The work in *Pliny*, then, was not only about how to support the creation of notes in the first place, but then how these personal notes could be made available to best support the kind of intensive and extended thinking about the material that would go into the development of a new interpretation of it. Once the computer is a repository for these notes, how can it best deliver them to the user in ways that support the user's engagement with them as they struggle to work out their own understanding of the materials? *Pliny* represents an attempt to achieve a balance between conflicting needs:

- it structures the act of notetaking, annotation, and note management;
- it supports its user in the task of moving from initial partly-formed ideas through to more formally structured ones by providing formalisms when the user is ready to use them, and by not imposing them too early; and
- it provides (through its provision of two dimensional spaces) a way to cope with lack of clarity, ambiguity and vagueness.

25 An early article about *Pliny* [Bradley 2008] presents *Pliny*'s place in scholarship as supporting three phases: (a) *reading* (more recently expanded to include *exploring* by considering the place for the use of software such as a text mining or visualisation tool as a source for new materials), (b) *developing* a new interpretation, and then (c) *writing* about it. *Pliny* software is an attempt to provide a tool to support not only the first annotation and notetaking activity (phase a), but to also support the development of new ideas that might be stimulated by these personal notes in a "personal space" (phase b), and that would fit, when the ideas were mature enough, into the writing that would bring these new ideas into the public sphere in the form of a book or article (phase c). Figure 1 (similar to figure 3 in [Bradley 2012]) schematically represents materials assembled in a Pliny repository, showing how they might relate to these three activities of scholarship.

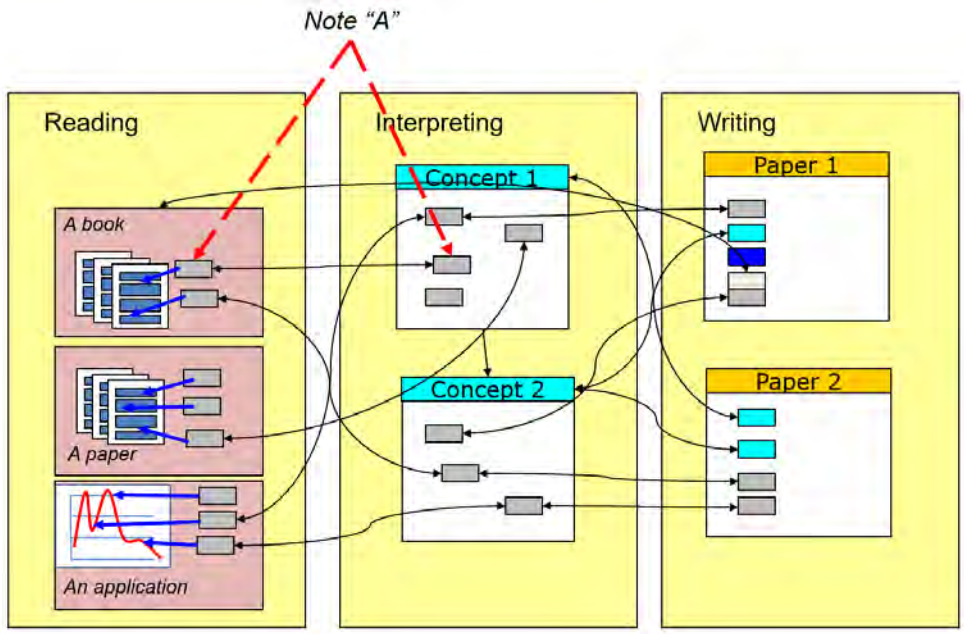

Figure 1. Schematic representation of materials in Pliny.

The left side of this figure shows the first phase: reading of both primary and secondary literature for research, and (with a bow to DH interests such as text mining or a visualisation) exploring some material using a computer application. The diagram shows small boxes with links to spots in the source texts and in a DH application display (here represented as a numerical graph). These are the annotations created by the researcher as s/he reads and explores. Initially, at least – attending to Lavagnino's comments earlier – the reader may well not be in a position to attach specific formalisms to these materials since he or she will not have developed the formalisms yet. So, instead, the notes are, most of the time, likely to be bits of personally-written text that captures the "reactions" that the reader hopes to "subsequently describe or explain" (using Lavagnino's words again) by developing a framework for them. 26

The middle area corresponds to the interpretation development phase of the research. Initially, in the earliest phases of the work there will not be any objects there. However, as the researcher thinks about the materials s/he has been exploring and works to understand them s/he will endeavour to organise their thoughts into a collection of interrelated concepts, categories, or topics. Thus, Figure 1 has in its middle area references to some of the notes that were originally created by the researcher while reading. In this phase of the research, however, they are reused by being organised under broad categories or topics (only two are shown here, and labelled generically as "concept 1" and "2" rather than with real names that a researcher would use). In addition, these concepts or topics might well be supplemented by new notes that represent new thinking by the researcher about these topics s/he has formally identified. In *Pliny* a note that started out as an annotation (such as the one labelled in Figure 1 as "Note A") can be also referenced in the different context of one of the interpretation's concept objects – it appears twice in this schematic because it then is displayed in two places by *Pliny* itself – first as an annotation attached to the text where it was created, and then a second time when it also appears as connected to the researcher's *concept 1*. Note that at this note/concept level the notes and their grouping into larger structures begin to show the kinds of formalism that is compatible with structured data. 27

Finally, when the time is right (and presumably after significantly more than 2 concepts have been recognised), the researcher draws on the ideas she has formed in the interpretation phase to put together materials (often papers) that present them. Now, the work of organising the materials changes to be one that supports the development of a text for an article that presents the ideas that were developed in the interpretation. Although *Pliny* in its publicly available form does not provide a tool to directly link the materials it holds into the text of an article, there has been some preliminary exploration in the *Pliny* project of how such a tool might operate – see [Bradley 2009]. 28

While looking at Figure 1 we are encouraged to view scholarship as process, and therefore think of this figure as, let us say, a snapshot of the data at a particular point in time. This process is somewhat organic – more like how a tree grows than how a building is built: *Pliny* is not, thus, "project oriented" in its support. In the same way that a tree starts out small with only a couple of branches but gradually becomes more complex as more branches are added, the *Pliny* model accommodates a user who needs to start small: first with notes created while reading, and then through the gradual development of ideas in the interpretation stages. As it evolves over time material is likely to develop from left to right, with (as we said earlier) much of the material at first in the "notetaking/reading" area at the left. Each of these three areas is likely, therefore, to develop at a different pace. Like a tree, where some stems become important branches, and others do not, and yet both substantial and minor branches coexist, ideas recorded in *Pliny*'s structure exist at any point in time in more than one stage of development and sophistication, with the more successful ones developing more substantially than the less successful. As with a tree, the product of the research – its fruit – is not a one-time-only affair at the end, but a continuing process during the life of the research and might occur many times out of the research structure stored in *Pliny*. Figure 1, then, does not represent the research at its end, but at some point during its life in the same way that a photograph of a tree represents the tree not as a finished product but as it is at some point in its life. 29

Although a tree is a useful metaphor here to capture the organic and ever evolving nature of research, we need take care not to press the "research tree" analogy too far. In particular, *Pliny*'s data model does not impose a hierarchical "tree like" structure. It is closer to an ever changing *graph* than a *tree*. It is perhaps what one might call the "organic" nature of this graph that is useful to remember out of this tree simile. We will have more to say about the "organic" nature of Pliny data below. 30

Annotation in Pliny

In *Pliny*'s conception of scholarly work new ideas that emerge in the mind of the researcher often emerge out of their engagement with materials in the world – often texts in books or articles, but also these days perhaps software applications such as those that support text mining, or text visualisation. The connection between these object-in-the-world and the scholar's own thoughts are often made, as Lavagnino suggests above, through annotation (or, more 31

generally, notetaking). Thus, annotation is a part of the entire *Pliny* framework, and we have done work in *Pliny* to explore annotation and more generally notetaking for a range of digital media, including both support for notetaking for non-digital sources, annotation for media-oriented digital ones such as web pages or PDF documents, and annotation within displays generated by software applications. Indeed, we believe that *Pliny*'s approach could readily be extended to include support for annotation of temporal media such as sound files or video. As a consequence, the *Pliny* approach to annotation within software provides a framework that extends ideas about annotation of media objects into thinking about how scholarly interpretation connects to the new research method of, say, text mining or visualisation. There is an extensive discussion of annotation for other media and for non-media applications and its implication within the context of the displays from software in [Bradley 2012].

Figure 2 shows an example of Pliny's approach to notetaking/annotation during reading. Here someone has been adding notes to a PDF file of Willard McCarty's 2008 article "What's going on?". 32

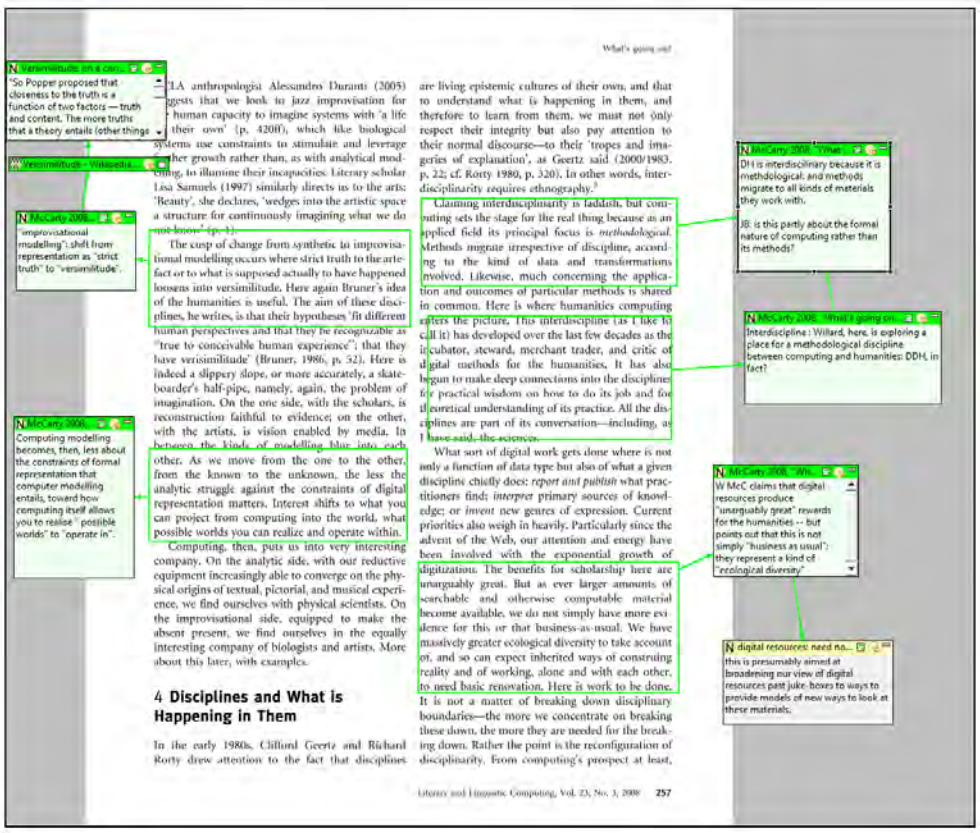

Figure 2. An annotated PDF Page in Pliny

Plinyy simulates the way annotations work on paper in two ways. First, when a researcher opens an object in Pliny with annotations on it, all the annotations are all immediately visible; nothing needs to be clicked on to see any of them. Their immediate presentation like this helps the reader remember *all* the thoughts s/he recorded on the page — something that does not happen if the reader had to open each annotation one by one. Second, annotations in *Pliny* appear to float on top of a printed text. Like annotation on paper, the *Pliny* annotator can make entirely free use of that page to hold his/her annotations and can exploit the spatial sense of the page as part of the expressive toolkit for annotation. In *Pliny*, then, an annotation, as well as linking a target to the annotation, also has itself a place in a 2D space provided by the material it annotates. 33

Supporting the development of concepts

Pliny is not only about annotating things, which was, you recall, represented only in the left "reading / exploring" area of Figure 1. How does *Pliny* support the central phase of research: the development of an interpretation? Different researchers might well use different approaches to this task, and *Pliny* is rather open ended in its support for this reason. However, we can illustrate how Pliny helps the researcher through an example which takes one particular approach. Here, the nature of research as a *process* plays a rather more prominent role. 34

35

One way to think about this process is as a gradual increase in structure: the *Pliny* user creates more structure in *Pliny* as the ideas become clearer and more structured themselves. The best way to see some of the ways that this plays out inside of *Pliny* is to examine more closely the process one could use to create a particular item about a topic called "uses of space for study" – an example first shown, albeit fully formed, in [Bradley 2008, 273]. Presenting what is actually a temporal *process* such as this in a single illustration is not practical, so the next sequence of 4 figures (Figures 3 to 7) provide a set of snapshots taken over time of the Pliny space for *uses of space for study* to show the different stages of its development.

In each figure in this sequence we have deliberately named each stage to echo some aspects of the language of *Scholarly Primitives* as presented by John Unsworth in 2000 [Unsworth 2000]. This paper has already touched on one of Unsworth's primitives: *annotating*. However, like Carole Palmer, who also makes reference to Unsworth's primitives in [Palmer et al 2009], but argues that her list is different because it comes from the somewhat different perspective; our set of primitives diverge from Unsworth's too, and for similar reasons to hers. 36

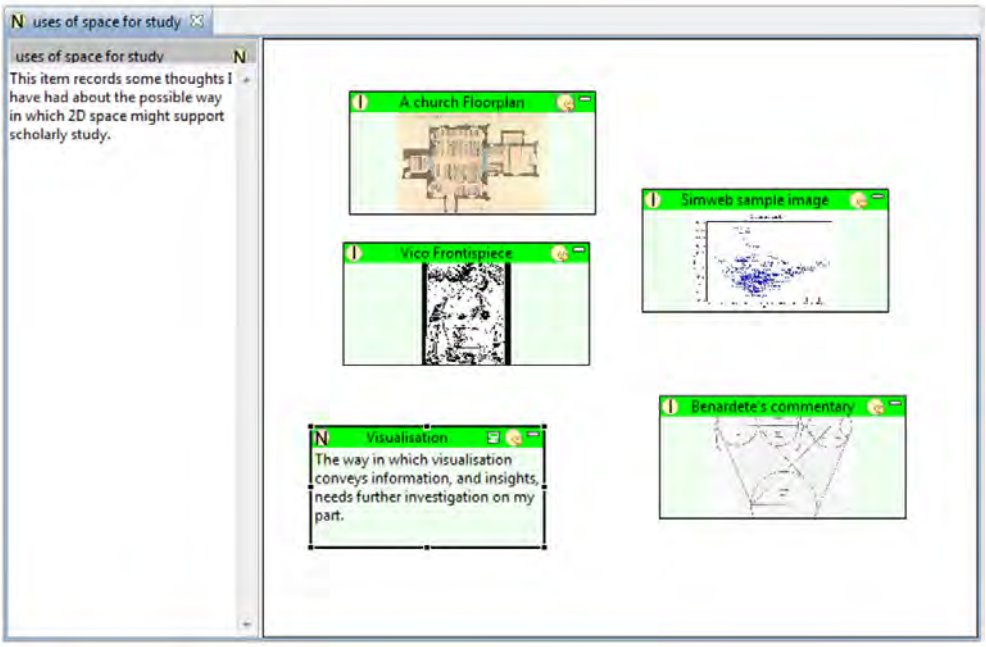

Figure 3. (a) assembling materials in Pliny

One begins by creating a holder for the topic, and naming it "uses of space for study". Then, the first step, shown in Figure 3, is *assembling* material that seems to be relevant (corresponding in part, perhaps, to Unsworth's *discovering* primitive). We see this topic space in Figure 3. The topic is named, and there is a brief description of the idea entered in the left area. However, the main place where the work is done is in the larger 2D space to the right. The *Pliny* user starts off by assembling references to things she is interested in that relate to the topic: in this case four images that show space being used in different ways. Then she adds a reference to a note on another topic she has already created called "Visualisation", which seems to relate to this one. As of yet, all she knows is that these five items feel as if they are connected to the idea of *uses of space for study*. 37

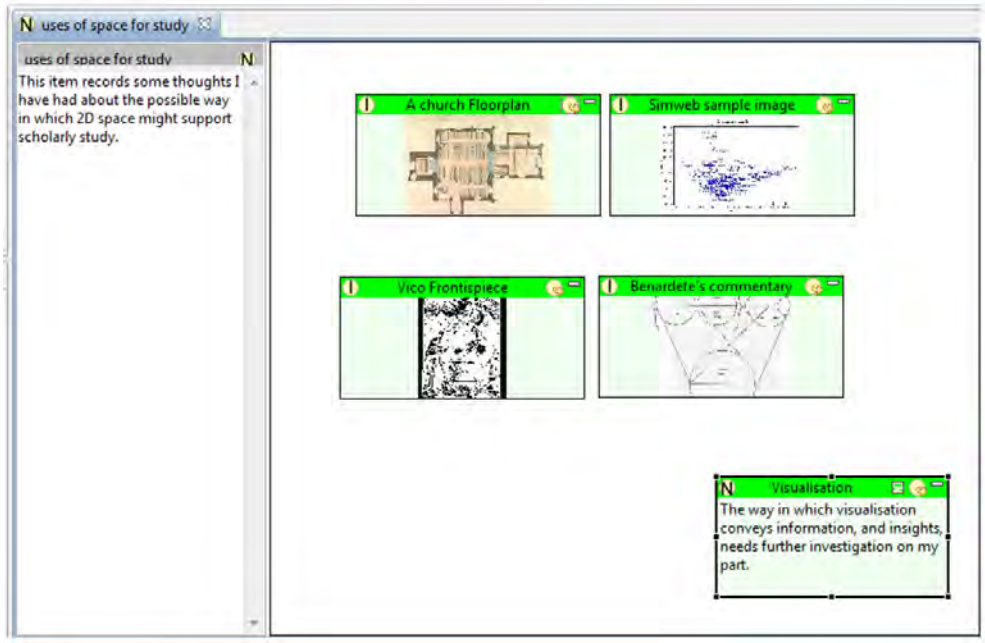

Figure 4. (b) organising materials in Pliny

38 *Figure 4 (b) organising*: Having now assembled a few items the researcher begins to notice some similarities in the use of space in two of the images, and detects what feels like a contrasting similarity in the other two. As a result, our researcher takes advantage of the possibility of proximity in the 2D space to organise them a bit, placing those which seem to be similar in some way close together. One can characterise this kind of activity as the beginning of the task of *organising* your material. Perhaps some part of this operation might broadly correspond to Unsworth's *comparison* primitive.

39 Note the importance of the 2D space for this task and the particular expressive affordances it offers. Here, no explicit links between the items have yet been asserted. Instead, there is a much more subtle and perhaps useful ambiguity available in the 2D space for suggesting relationships between items in terms of *proximity*. Putting two items close together suggests some degree of connection without requiring that it be spelled out too specifically.

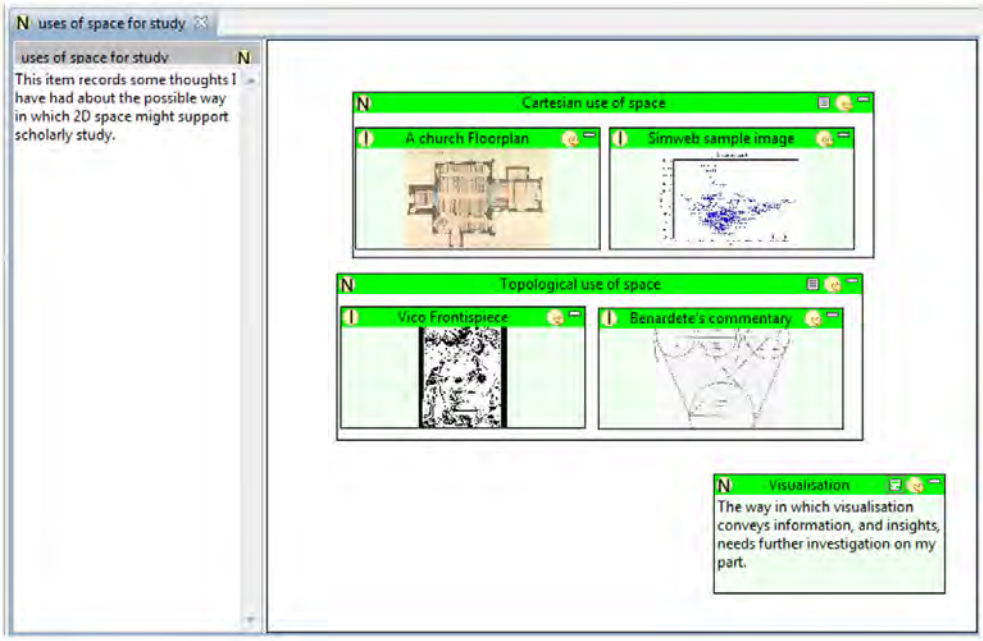

Figure 5. (c) grouping/categorising in Pliny

40

Figure 5 (c) Grouping/Categorising: Now that proximity has helped the *Pliny* user to develop a feel for possible grouping of the items that she has assembled, at some point the groupings become clear enough that she feels ready to give them names. She has already sensed that she seems to have two groups of images, and she now sees them as representing two rather different kinds of use of space. So she asks *Pliny* to put these images into two groups, which she can then name. The *grouping* and then the naming or *categorising* adds more structure to this space, and begins to express an interpretation of them.

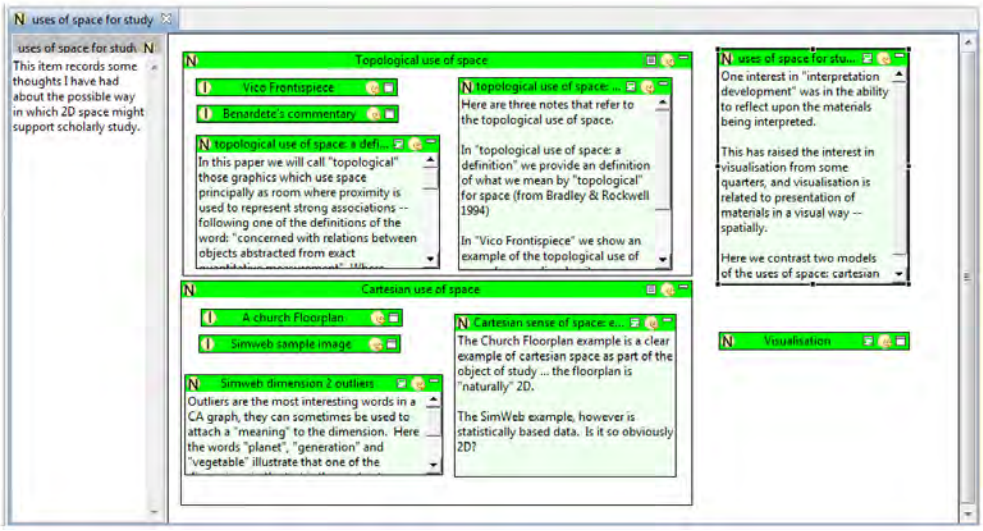

Figure 6. (d) enriching materials in Pliny

Figure 6 (d) Enriching: Having now discovered these two kinds of uses of space the user adds a few notes that record her thoughts about them, thereby *enriching* the concept space. As well as creating entirely new material for this purpose, *Pliny* also allows her to reference items that were created elsewhere in *Pliny* – introducing them in a new context, similar to what happened to the note labelled "A" in Figure 1.

41

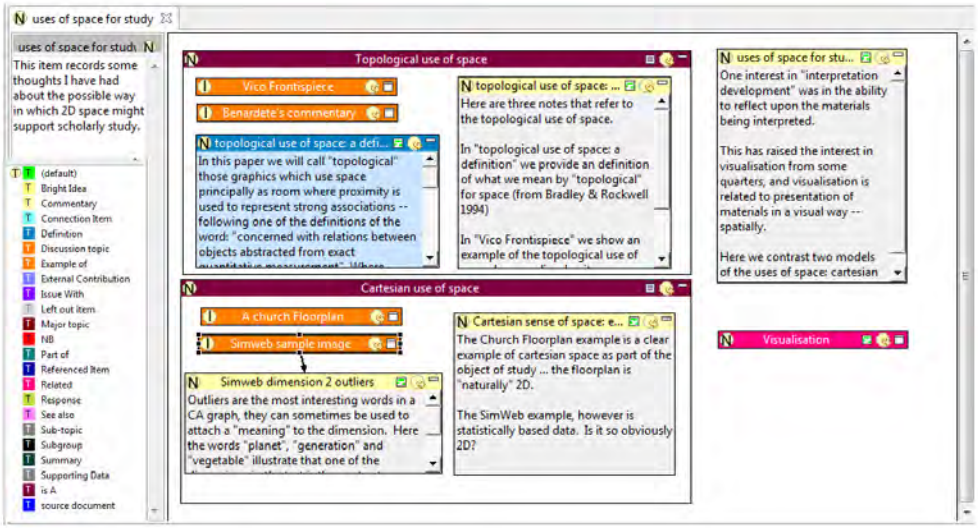

Figure 7. (e) Typing the references in Pliny

Figure 7 (e) Typing the references: Finally, now that the user has collected and organised this material, she notes that there are several kinds of connections between the topics and the things they contain: some of the objects – the images in particular – are *examples*, some of the notes are *commentary*, etc. *Pliny* allows her to assign a type to these connections to indicate that this reference is of type *example*, and that another one is of type *commentary*. *Pliny* shows these different types as different colours. Figure 7 shows the user's current set of defined *types* in the bottom left corner. She assigns these types to the different items, thereby asserting that, for example, the *Vico Frontispiece* is an *example* of a *Topological use of space*, and that the *Visualisation* topic seems to be a *related* topic to this one.

42

In summary, then, figures 2 and then 3 through 7 suggests steps in a process of developing an interpretation: 43

- *Assembling*: The user starts off by assembling materials that s/he wishes to work with.
- *Annotating*: *Pliny* provides annotation so that the user can record their responses to these materials.
- *Organising*: *Pliny* provides 2D spaces where the user can organise notes (perhaps created as annotations) and other objects to explore possible relationships between them that will hopefully eventually lead to a clear formulation of a model for these materials.
- *Grouping and Naming*: As concepts become clear the user can use *Pliny*'s grouping mechanisms in conjunction with its sense of 2D space to identify, name and organise ideas.
- *Enriching*: *Pliny*'s notes, among other mechanisms, provide a way for the user to add comments to the structures s/he has become interested in – allowing her to enrich the structure she has created,
- *Typing*: Finally, *Pliny* allows the user to attach assertions about the relationships between objects that have been associated with their concepts.

These steps in a research process move the user from preliminary reactions in the form of annotations and notes to more fully formed ideas – and within *Pliny* from less structure to more structure. Not that a researcher will necessarily be able to push all his or her ideas through to be as fully structured as *Pliny* supports. *Pliny* accommodates a mix of highly structured areas with less structured ones. 44

By examining the process involved in developing the idea of *uses of space for study* in *Pliny*, we have seen how *Pliny*'s components support this kind of work. Let us take stock for a moment now. We have thought about the activity of scholarship and how it relates to the affordances of *Pliny*. First, *Pliny* supports the *process* of scholarship, not primarily the expression of its products. *Pliny*'s approach, supporting research from annotation through concept development to write-up, provides a framework that allows its user to move from largely unstructured ideas to a more formal structure by supporting the way that this transition happens: starting off with reading and notetaking, and then through the working with these notes and the thinking about issues they raise to support the gradually emerging formalising of new ideas. *Pliny* is not project-oriented in its support. It does not require that ideas develop in step towards a sense of "completion". Instead, the *Pliny* user can combine ideas that emerge from different areas of his/her work as she pleases. *Pliny* accommodates the co-existence of certain ideas that develop more completely with those that do not. Materials in *Pliny* need not ever be finished. 45

Second, *Pliny* promotes the use of a 2D working space as a central element in its set of affordances. In the context of annotation, Pliny supports a 2D space for annotations that intentionally mimics annotation on paper, and supports the idea that annotations, as tools to assist in research, need to be fully visible whenever their resource is open. Two dimensional space is also used as a central affordance to support the assembly and organisation of materials into concepts, although actual practice of users in *Pliny* suggests that 2D-space's sense of proximity, with its degree of ambiguity, is more useful at the beginning of the effort to organise a particular topic then it is later as the concepts related to the topic become clearer. 46

Fitting Pliny to the Semantic Web

Having looked at *Pliny*'s way of modelling a part of scholarly research, it is time to return to a major theme of this paper: how does the formal structure behind *Pliny*'s support for annotation (figure 2) and concept development (figures 3 to 7) fit with the formalisms of the Semantic Web? If we can see how the two worlds connect together we have, to the extent that *Pliny*'s formal model captures a part of humanities scholarship, a way of thinking about how scholarship fits with the formal world of the Semantic Web as well. 47

The authors have explored this potential connection in a very practical way: by creating a rudimentary extension to *Pliny* in the form of a "plugin" that supported the export of *Pliny* materials into RDF and that brought the linked standard representations of the semantic web back into *Pliny*, and that thus allowed Semantic Web or Linked Data URIs to appear as *Pliny* resources. The next section of this paper, then, presents the very specific way in which this prototype plugin represented the mapping from *Pliny* to the Semantic Web and back again. Theoretical considerations may, of course, reveal more than one way to think about how this connection might be expressed and exploited, but as a consequence of the need to actually write code for this plugin, the authors here have explored one particular expression of this connection. Nonetheless, this pragmatic description of our plugin still opens up issues that connect back to the more theoretical discussion we presented earlier in this paper, and we hope readers will interpret this next, rather pragmatic and specific discussion, in that light. 48

Underlying the representation of Pliny materials in RDF is a rudimentary ontology of Pliny's objects. Its major classes are: 49

- *Resource*: *Pliny* structures a user's collection of materials as a set of Resources. Resources are sub-classed to represent types of content objects such as Web pages, PDF files or Image files. Pliny is designed to be extensible to new data types, and usually any new data type becomes a subclass of Resource. Importantly, a Pliny *note* is also a Resource.

- *Note*: As the name suggests, *Pliny*'s notes serve the purpose of being containers for a short bit of text (for example, textual annotations), but they also provide a 2D space for storing references to other objects. Hence what are described in Pliny documentation as "containment objects" (which usually become the holders for concept items such as our *uses of space for study* example) are in fact also Notes. From a Semantic Web perspective, since a note and a concept are clearly semantically different things, this conflating of Notes and Concepts is an expressive limitation that could be addressed in a future version of Pliny. Perhaps, in some future version *Pliny* would allow users to define a set of related classes for kinds of containment objects: categories such as "concepts", or "persons", etc.
- *Anchor*: *Pliny* provides anchors (currently only rectangles, although this could be extended to other shapes) to identify areas on 2D resources such as images or PDF file pages. In an annotation interpretation of Pliny data, a Pliny Anchor can be an annotation's target.
- *Reference*: Any resource (note/conceptual object, image, PDF file, etc.) can be referenced from within another resource. In the *uses of space for study* example above, for example, the inclusion of the *Vico Frontispiece* as an example of one of the uses of space was done by means of a reference to the resource that held the frontispiece image.
- *Reference type*: References can be typed (we saw the typing being applied in Figure 7). As we've seen, reference typing enriches the semantic expressiveness substantially, and a *Pliny* user can define any reference type s/he wants, defining, for instance, types such as "is an example of", or "is a commentary on". Thus, reference types could become the basis for a user-based ontology separate from Pliny's entity ontology we are describing here since, as we shall see, they can be viewed as a set of predicates in exported RDF triples. In addition, when used in the context of annotation, the reference type can also be usefully interpreted as a "kind" of annotation.

Now that we have done some thinking about *Pliny*'s data as RDF data, two quite obvious questions arise:

- *Reaching out to the Semantic Web*: How can the data that we have shown as accumulating inside of *Pliny* be transformed into RDF, the language of the Semantic Web, and
- *Drawing in from the Semantic Web*: how can the linked data in the Semantic Web be most usefully connected with the model of scholarship that *Pliny* presents?

We first explore the *Reaching out* part; mapping *Pliny*'s representation of an interpretation to the Semantic Web world. To do this we look first at the part of *Pliny* that supports annotation of digital objects (as shown in Figure 2) to see how these annotations might map to the Semantic Web, and then we explore how the concept development part of *Pliny* (shown in Figures 3 to 7) could map to there as well. 50

About annotations: in order to explore the annotation functions like those in Figure 2 we made our prototype RDF export mechanism use the OA annotation ontology ([OA 2013] and [Sanderson et al 2013]), extending and updating some previous work ([Jackson 2010]) that also explored how to map Pliny data to (the older) OAC ontology. The task was not entirely trivial because *Pliny*'s annotation tools (for, say, images and PDF files) were not designed to model simple annotations in terms of objects such as notes and targets. Instead, *Pliny*'s underlying model operates in terms of target areas, notes, and links between them all placed in the 2D space. This approach supports the sense of annotation that mirrors annotation on paper, as described earlier, but also enables a richer set of relationships between a collection of notes, targets and connections between them than just notes attached to targets, which is the sense of annotation given in the OA. Our RDF exporter, then, had to take data described in these more general terms and express them as annotations that were conformant with the OA. 51

Because of space constraints on this paper, it is not possible to explore in detail how this mapping was done, but an example will suffice. Figure 8 shows a slightly simplified version of the RDF (turtle notation) generated by our exporter for the annotation called "claiming interdisciplinary" (showing in the top right of Figure 2). This data could be published through a triplestore for public use, although this has not been done at this point in time. As we hope the prefix expansion makes clear in the figure, the *jb* prefix identifies data that belongs to a particular *Pliny* user, and would be a different prefix for a user with a different name. 52

Figure 8. RDF and Pliny annotation

The fragment labelled "A" describes the PDF file that has been annotated. This item, identified with a URI jb:resource-99, is identified as a *PDF object* (elsewhere defined as a subclass of a Pliny resource), and is then assigned attributes that are largely self-explanatory. Fragment "B" contains the RDF for the bit of text that has been applied as an annotation. This resource is identified as a *Note*, with a name and contents as shown. Fragment C defines the target of the annotation, which is here an area on a page in the PDF file. We use the "fragment selector" model given in the Open Annotation specification ([Sanderson et al 2013, section 3.2]). Finally fragment D defines the annotation by connecting the target area and the Pliny note. The annotation entity is identified as jb:lo-119, and is typed as a `RelatesTo`. `RelatesTo` is one of the *reference types* that the user has defined, and is identified as a subclass of an `oa:Annotation` elsewhere in the generated RDF. 53

Figure 8 perhaps gives a sense of how the annotation parts of Pliny have data that can map to the Semantic Web via the Open Annotation ontology. However, as Figures 3-7 show, Pliny is not only about annotation. The construction of this personal interpretation for *uses of space for study* is not an annotation exercise – if it was, what is being annotated? Thus, we cannot use the Open Annotation ontology to model what is going on in Figures 3 to 7. How is this material handled in any translation to RDF? 54

First, as [Bradley 2008, 274–6] points out, the data structure *Pliny* uses behind these displays can be thought of, to some extent at least, as a graph with typed nodes and links. As mentioned earlier in this paper, this same kind of mathematical graph is the foundation model for RDF [Cyganiak, Wood and Lanthaler 2014, section 3], and it turns out that Pliny's own graph model maps quite comfortably into RDF's "subject predicate object" representation. As a consequence, the RDF exporter we built used a different translation strategy for data from these 2D "concept spaces" than for "annotation spaces". The key was to focus on the relationship between the note and its holder in the 2D space. The *Visualisation* concept shown in Figure 7 is referenced by the *uses of space for study* concept, and its reference there is typed as a *related* item. The *Vico Frontispiece* image is shown as *an example of* the *Topological use of space*, and the *Topological* item, itself, *is A* use of space for study. 55

The mapping from the objects in the Pliny space to RDF triples, is actually rather straightforward. A fragment of the generated RDF is shown in Figure 9. 56

```
jb:resource-80 a jb:objectType_Note ;
       pliny:hasName "uses of space for study-base" ;
       pliny:hasContent "This item records some thoughts [...]"
[...]

jb:resource-85 a jb:objectType_Note ;
       pliny:hasName "topological use of space" ;
[...]

jb:resource-72 a jb:objectType_Image ;
       pliny:hasName "Vico Frontispiece" ;
[...]

jb:resource-87 a jb:objectType_Note ;
       pliny:hasName "Visualisation" ;
       pliny:hasContent "The way in which visualisation [...]"
[...]

jb:resource-85 jb:refType_is_A jb:resource-80 .
jb:resource-87 pliny:refType_RelatesTo jb:resource-80 .
jb:resource-72 jb:refType_Example_of jb:resource-85 .
```

Figure 9. RDF representation of a Pliny Topic

We have now discussed reaching out from *Pliny to* the Semantic Web. What happens when we think about what was called earlier *drawing in*: bringing aspects of the Linked Data/Semantic Web world into the user's scholarly space by making them available in *Pliny*'s workspace? One can see two rather different kinds of linking activities. The first is very similar to Semantic Annotation which is touched on very briefly earlier in this paper, but the second was based on the idea of reaching out *into* the graph-like linked data world to consider and annotate parts of that web as it currently exists. This second type of connection makes part of the Semantic Web itself an object for study in its own right. 57

First we can see in Figure 10 our *Pliny* RDF plugin's support for annotation with Semantic Web objects ("semantic annotation") in action. Here it is used to link some parts of the illustration to concepts from *DBPedia*. This figure, similar to figure 2 in [Bradley 2008, 268], is of the frontispiece from Vico's *New Science,* and, as it did in the earlier illustration, uses annotation to identify several of the philosophical concepts represented in the image. 58

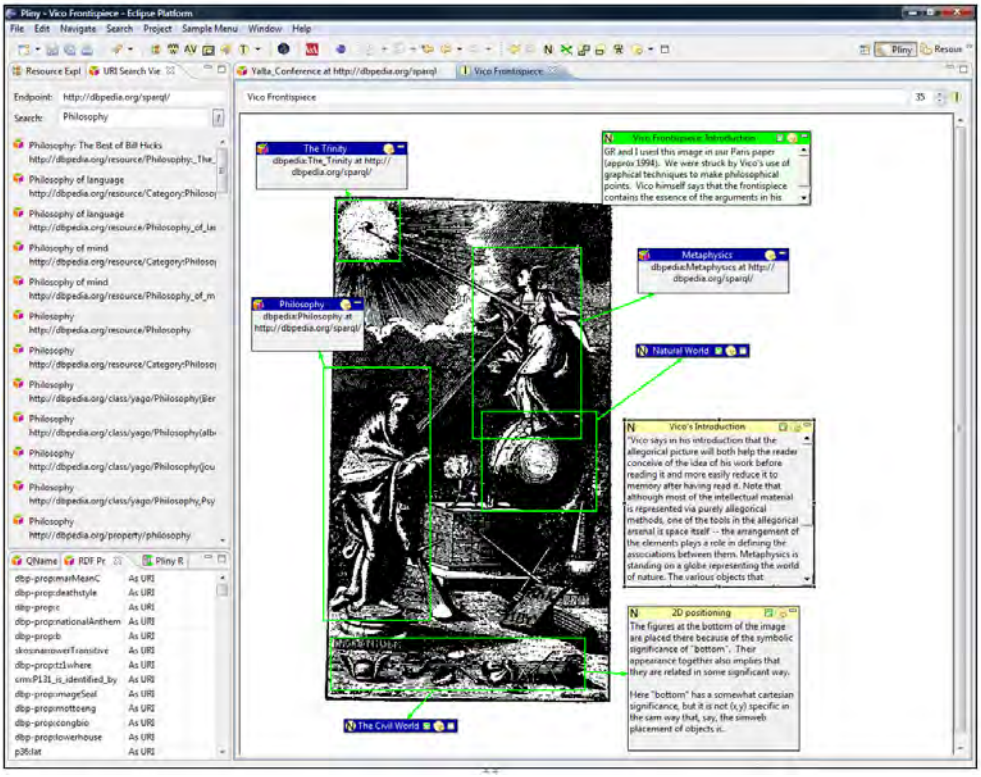

Figure 10. Semantic annotation in Pliny

In the 2008 figure, the concepts such as *Metaphysics*, or *The Trinity* were all labelled and identified as concepts with internal Pliny objects. Our RDF extension, however, allowed us to refer instead to URIs that represented the concepts in the global Semantic Web: here we can see several of these concept objects (*The Trinity*, *Philosophy* and *Metaphysics*) are actually references to their corresponding URIs within DBPedia. These links/annotations to Semantic Web URIs that identify these concepts co-exist with other kinds of objects: here we also see commentary in the form of notes, as well as links to other concepts such as the *Natural World* and the *Civil World* which the user has not chosen to connect to the Semantic Web as URIs. To make it possible for a user to locate a global URI for concepts like *the Trinity*, we developed a rudimentary query mechanism (shown on the top left side of Figure 10) as a prototype tool that allowed the user to query any SPARQL endpoint (here, to *DBPedia*) to find URIs within it that had `rdfs:labels` containing a given word – here "Philosophy". Then, having found a suitable URI in the results of the query, the user could simply drag it from the list onto the *Vico Frontispiece* image to generate a reference to it. This kind of reference to concepts by referencing a URI on the Semantic Web is, of course, exactly Semantic Annotation of the kind we touched on very briefly earlier in this paper. 59

The second kind of engagement of the Semantic Web world – which we described as making a part of the Semantic Web as "an object of study" involves the use of the structure of the Semantic Web itself as a resource for one's scholarship. This can, again, be thought of as annotation, but not in terms of annotating text or an image with Semantic Web URI's (as Semantic Annotation does), but the other way round: annotating the graph of the Semantic Web with personal materials, references to personal research concepts and with the user's own thoughts as notes, etc. Instead of (as Semantic Annotation does) annotating documents with references to the Semantic Web, we here annotate the Semantic Web personally with references to elements of a personal collection of documents and notes. 60

We built a prototype tool for this that suggests how this somewhat more radical engagement with linked data and the Semantic Web might work: see Figure 11. The display is a little crude still since the software that implements it is still at the prototype stage – but it is suggestive of what one could mean by making the Semantic Web an object of study in the *Pliny* sense. 61

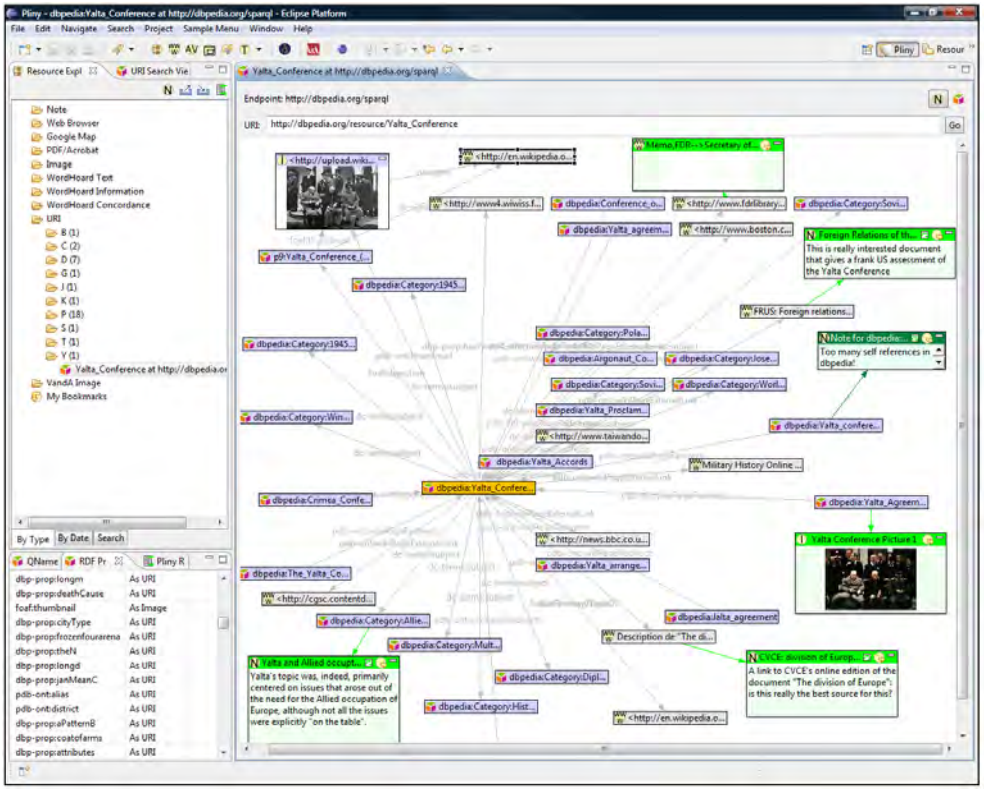

Figure 11. Annotating the Semantic Web in Pliny

Our prototype can be directed to any RDF repository that supports SPARQL, and here it shows a part of *DBPedia*'s web of linked data as a graph – here centred around *DBPedia*'s URI for the Second World War's Yalta conference (shown on the screen in orange). Most of the displayed graph is a representation of the RDF triples that *DBPedia* holds and that connect to the central *Yalta Conference* URI, and these closely related URIs appear simply as little boxes that identify related *DBPedia* objects. However, a few of DBPedia's URIs are actually URLs to web pages, and one is a URL to a photograph taken during the conference. We built a small component (shown bottom left) that allows the user to tell our Pliny display about this. Although the layout here is still crude, we hope that the graphical presentation suggests how one might present this web of objects. 62

What is interesting here are the objects shown (on the screen) in green. These, although mixed in here with the mainly RDF data from *DBPedia*, are instead Pliny objects created by the Pliny user as a personal commentary on this part of the Semantic Web, and displayed here as intermingled with it. Some are notes that comment on *DBPedia* materials (such as the note that observes that the FRUS item referenced in *DBPedia* is a "really interesting document" (shown near the top right). Two of them are links to a web pages (top right) and to an image (bottom right) that is not referenced in the *DBPedia* materials. You can think of them as personal annotations added to this small corner of the Semantic Web. 63

In the same way that *Pliny* allows users to personally annotate a web page, an image, or a PDF file with responses at the moment they have them and then later use these notes in personal deliberations, one can here annotate the Semantic Web with personal responses to parts of it, and can then fit these reactions into later thinking. The intermingling of Semantic Web URIs with Pliny objects in this prototype display allows for the creation of a personal space between, but connecting, objects being annotated and the public Semantic Web. 64

RDF and 2D Space

The reader can probably see by now that 2D space plays a significant role in *Pliny*. First, 2D space has a relatively obvious role in *Pliny*'s annotation support since an annotation *target* for an image or page of a PDF file is defined by an area on a 2D image or PDF file 2D page. As shown in the example from Figure 8, their position is expressed in RDF through the mechanism that the OA provides for this: the `oa:fragmentSelector` construct (area C). However, there are other objects in Figure 2 to which Pliny also gives a spatial position: the annotation notes themselves. The placement on the page of these annotation notes is not exported at present by our code because its `oa:Body` definition – used to represent the contents of the annotation notes – provides no mechanism for expressing it. Therefore, although 65

of course the spatial layout for these notes could still have been exported as RDF, they could not be presented purely in terms of the OA ontology.

66 Having discussed the place of 2D space in the annotation parts of Pliny, we need also to think about its use elsewhere. The most obvious other use – as a tool to support an emerging interpretation (shown in Figure 3 to 7) – makes 2D space into a central mechanism by which fuzziness and ambiguity is accommodated. It provides an open, flexible, structure for organising materials, and offers some particular affordances, most obviously *spatial proximity* (but others too: see the discussion in [Bradley 2008], as a means of expressing unclear relationships between objects. Figure 4 shows the use of spatial proximity in this way most clearly. By putting objects close together the user tentatively suggests some sort of connection between them without needing to pin down too much yet what this connection might be. Although the structuring affordances shown in Figures 5, 6 and 7 add further richness to the data recorded in this space, and reduce the need for 2D's particular expressive characteristics such as spatial proximity, even in the final image (Figure 7) of this sequence the expressive elements provided by 2D space have not been fully eliminated. The layout of the objects there provides a kind of visual enhancement of the presentation of the idea that, say, merely presenting the same associations hierarchically, would not possess.

67 Think again about the 2D space shown in Figure 4 where the user organised the four images into two groups to reflect a perceived, but yet unclear, relationship between them. If this area was exported into RDF, one would see the references to the four images appearing but there would be no evidence of any possible connection between them, even though the user is consciously using spatial proximity to represent possible connections. The actual placement of the four objects, in terms of their X-Y co-ordinates can certainly be expressed in the RDF. However, even if we added code that would record their placement, the possible connection the user is intending to present by her arrangement is still only at best implicit in the RDF. The tentative connections are presumably evident to the Pliny user (she is deliberatively using the 2D space's sense of proximity to suggest these possible connections, after all), but this possible connection is not yet formalised in the data structure. In contrast, as soon as containment is brought into the picture to more formally express the relationships (see Figure 5), the relationship between the two sets of objects is also made explicit in the exported RDF.

68 Does this matter? Does it matter that the user is exploiting the 2D space to express something about the materials without, as it were, telling *Pliny* formally about it? Indeed, does this present a serious challenge to the task we have proposed in this paper: to find a place for "graph thinking" in scholarly interpretation development? We think the issue is not as bad as it might at first appear to be. This is because although, in *Pliny*, spatial proximity, with its advantage of vagueness of expression, plays a role in assisting its user during their early thinking about a topic, it seems clear that the process of interpretation around a topic involves a gradual replacing of vague understandings with clearer, more precise, formulations of ideas. Thus, although as we have seen one could argue that the whole idea of spatial proximity (rather than spatial positioning) is to some extent incompatible with standard RDF expressiveness, it is also likely that in fact 2D proximity becomes a less prominent element of expressiveness as the process of developing this concept is carried out in Figures 3-7. Although our RDF exporter does not export spatial position data for concept objects, as the ideas develop further (in the later figures in our *uses of space for study* example) they become expressed by other means, and these other representations *are* already preserved by our RDF exporter. As a concept becomes more mature, it would seem that the spatial information becomes less important as a carrier of information, and the graph representation by itself begins to carry more of the substance of the material.

Possible Future Work

69 Although we believe Pliny's approach opens up the subject of how one might fit formal Semantic Web modelling with a broad range of fundamental humanities research processes, we cannot claim here that we have a mature and complete approach ready to go. One way to continue research is to explore how the issue has been considered in other fields, such as in mathematics or the sciences. Here, our article's reviewers have made several suggestions of work that might bring relevant insights.

70 First, work has been done in the area of teaching of mathematics to explore how students develop a richer understanding of mathematical principles. See, for example, [Gravemeijer 1999] and [Lange 2008]. Second, interesting work has been done under the category of "Personal Knowledge Models" in [Völkel 2011], and this PhD thesis draws together a rich bibliography of related work primarily from a Computer Science perspective. Third, one reviewer noted that the topic of differences between the expressive possibilities of different media raises an interesting perspective on our issues here, and suggested [Elleström 2014] and [Eide 2015] for further reading. Finally, one issue raised in this paper is the tension between humanities research and formal models; often mathematical models of the kind that

underpin many concepts in science. Here, [Nersessian 2008] has been suggested by one reviewer as a good starting point.

Returning to our work with Pliny, obviously, further work needs to be carried out in a broad range of different areas, including the place of spatial proximity in the formal model, or the issues around exploring the semantic web itself as an object for study. 71

We provide here four further examples of areas for development and/or research: 72

- adding more RDFS semantics into *Pliny*,
- extending *Pliny* to include a tool to support argumentation and writing,
- resolving issues of personal and public in *Pliny* data, and
- dealing with materials that change over time.

We discuss each of these examples below as possible future work that, although not yet done, is suggestive of some research potential in this area.

Adding more RDFS semantics to *Pliny*

If one uses *Pliny* to develop an emerging conception such as that suggested by our *uses* example (illustrated by Figures 3 through 7), it would seem that one could reach a level for formalism that begins to connect semantically to conceptions in RDFS [Brickley and Guha 2014] and even the Owl Ontology language [OWL 2012]. However, RDFS allows for the expression of more kinds for formal relationships than *Pliny* currently supports, so perhaps one interesting way forward would be to add more of RDFS concepts into *Pliny*. Users can already define new types in *Pliny* as a way to express kinds of relationships between objects, and you can see this in operation in Figure 7. Indeed, in our prototype RDF export we interpreted Pliny's "relationship types" as RDF Properties, and it would be easy to extend Pliny's rather rudimentary typing model into a hierarchical organisation organised under RDFS's `subPredicate` concept. Similarly, it would be relatively straightforward to extend Pliny's note/container model to make them operate as RDFS Classes, including support for sub-Classing. If domain/range specification mechanisms were also added, *Pliny* would provide an environment that would allow for a user to progress from very general note taking to working with almost all the formalism of RDFS to express an ontology for their material in a single, integrated, environment. Although we have not tried to extend *Pliny* in this way, it seems to us that the process involved could be quite straightforward. 73

Extending *Pliny* to include a tool to support argumentation and writing

The reader has probably noticed that *Pliny* does not actually get to the final phase of research: the actual writing up of the article. *Pliny* suggests a formal model for a scholar/user's concepts etc., and proposes a third phase (shown in Figure 1) that uses *Pliny*'s existing tools to help users assemble materials from their *Pliny* data into objects that would represent materials for a written article. Furthermore, materials from *Pliny* can already be exported in a form that a word processor can use, so a transfer from concepts and resources in *Pliny* into a writing process is already possible to some limited extent. However, once this is done the formally expressed links from the resultant writing to the concepts and resources *Pliny* holds are, of course, left behind. That is as far as the current *Pliny* model gets, and our own experience of writing with *Pliny*'s support suggests that there is a large amount of work that the writer needs to do after the materials have been assembled in *Pliny*'s article-specific areas to turn *Pliny*'s representation of concepts into prose text. Can *Pliny* be extended to provide better support for this part of the work by presenting mechanisms that better preserve the connections between the representation of concepts that it already holds with the article that in the end presents them? We have not done much work in this area (although see [Bradley 2014] for some early thinking). 74

Much of this work on *Pliny* will involve a better understanding of what a scholar is actually doing when s/he turns the concepts in *Pliny* into written text. Earlier in this paper we made a brief reference to the structure of an argument, and there are hints of what this process might consist of in [Brockman et al 2001]. Thus, providing integrated tools to allow the user to interactively develop an argument would appear to be a sensible way to proceed. Are there ways in which the argument structure can be usefully represented as structured data? One possible approach is to build tools that work with existing formal argument modelling models. 75

One such model is CIDOC-CRM's CRMInf ([CRMInf 2015] and [Doerr, Kritsotaki and Boutsika 2011]) ontology. CRMInf's website says that it provides an ontology to model argumentation and inference and claims that in this way it provides a formal language for "making the inference structure (such as a mathematical proof) and the belief in this structure implicit to the argumentation event" [CRMInf 2015]. And, indeed, one of the important tasks of a written paper is surely to present a position on some subject in terms of an argument, with inference as one of the modes of expression of that argument. 76

77 One possible problem with the use of CRMInf to support writing in the humanities is that argumentation, or indeed scholarly writing more generally in the humanities, does not necessarily always progress by means of inference and deduction. Martin Mueller's recent comment in the discussion about CRMInf in Humanist (starting in Humanist 28.812) asking if such an approach "would ever help a literary critic make sense … of Shakespeare's Sonnet 116" (in [Mueller 2015]) we think raises, albeit obliquely, the same issue. We suspect that Mueller does not believe that CRMInf's model of formal inference is, in fact, the principal way by which classic humanities scholarship is carried out or expressed. Indeed, theorists like Joris van Zundert [Van Zundert 2014] have argued that there are humanists who "tend to reason abductively" rather than deductively or by inference. If so, a tool to assist these individuals with the preparation of text that was driven by arguments based on formal deduction might be of limited value.

78 As far as *Pliny* is concerned, we have begun, as very preliminary thinking in this area, to conceive of a new tool for *Pliny* that would be a bit like a word processor, but would also present some visualisation of an argument structure and would also allow a user to preserve links into his/her *Pliny* structure from the article's text. We are attracted at present to the rather loose conception of how an article's ideas become organised that was briefly presented in Douglas Engelbart's famous 1962 *Augmentation* Report [Engelbart 1962, 81–88]. There Engelbart expresses a looser sense of dependence rather than rigorous inference, by observing that any particular idea in an article is often dependent upon one or more ideas that had been presented earlier. Engelbart calls these dependency relationships *antecedents*, and his approach does resonate with our own experience of writing the article: that part of the task is getting the flow of the text right in ordering things so that for any idea presented its antecedents are also present and presented in advance. Perhaps we might start further work here by exploring Engelbart's softer-edged "antecedent" approach in addition to CRMInf's more formal and induction-oriented model.

Publishing Pliny structure: personal and private materials

79 A third major area of further research that emerges from the ideas in this paper arises when one thinks about whether the publishing of a *Pliny* structure representing aspects of its user's interpretation can be, in and of itself, considered as a research product. As we mentioned earlier, the idea of publishing, or more informally sharing in some way, some *Pliny* materials is a *little* like the idea Gradmann proposes when he talks of publishing "personal digital curation workflows". However, although the characterising of "curation workflows" as the objects being published works well for Gradmann's Linked-Data and Mashup model, it does not work as well for *Pliny*'s personal conceptual space context. Our example presented above, the *uses of space for study* concept, and others one might create in *Pliny*, simply *are not* a "curation workflows". However, the question still remains about whether materials that are created in *Pliny* are themselves publishable: sharable as data with others in a form that is adequately representative of the ideas they present. The work described here does take us one step in this direction in that, as we have seen, we have done some basic thinking about this when we defined how *Pliny*'s data could be exported as RDF. However, what issues around making *Pliny* personal data structures public still remain?

80 Part of the challenge here relates to what we called earlier the *organic* nature of *Pliny*'s data, and the fact that a *Pliny* repository will hold a mix of both unfinished-personal material which is not intended for sharing and mature material that could be made public and sharable with others. Here, the "fruit of the tree" analogy for scholarly research products mentioned earlier is somewhat misleading. Nature packages up mechanisms within the tree's fruit that, once it is dropped to the ground, allows it to produce a new tree entirely independently of its tree-parent which created it. An apple by itself is sufficient (with soil, sunshine and water) to produce a new apple tree. Thus, although, as the parallel suggests, our "organic" graph might indeed contain areas that could be published as "fruit" for the use of other researchers, the graph model does not impose an independence of these publishable components from other objects in a user's *Pliny* repository, and in this sense is quite different from the kind of self-contained packaging that the apple represents. Instead, one of Pliny's concept objects, such as our *uses of space for study* object, does not stand on its own – it is in fact almost entirely created out of references to *other* parts of its creator's space. Furthermore, as we have indicated above, any user's *Pliny* repository will contain a mixture of mature material and less polished material. How does one assist the *Pliny* user in managing this issue: making public enough related material so that the exported material can be made sense on its own, and yet excluding personal material that is not ready for public viewing.

81 The issue of private vs public material has been worked on in connection with annotation, and contains some thinking that is relevant to the larger picture, beyond annotation alone, of *Pliny*'s data structures too. We are thinking here of the work by Catherine Marshall back in the early 2000s that is still cited today. See, in particular, her suggestive study of the nature of digital annotation practices for personal and for shared use in "Exploring the relationship between personal and public annotations" [Marshall and Bush 2004]. More recent related work can be seen in [Congleton et al 2009], where the researchers asked students to mark up maps that would be shared with different kinds of groups, ranging from small groups the students knew to larger groups that they did not. Marshall and Bush's observation, included in the

[Congleton et al 2009] work, was that when people took what had been a set of personal annotations and decided to make them public they first eliminated most of them, and for the ones to be made public they carried out significant additional work to polish them for publication.

Not all Pliny's materials are annotations, but nonetheless, we assume here that the process a user might go through in turning personal materials into public will be similar to that shown in the annotation studies. Hence, for *Pliny* the issue arises from how the system might help the *Pliny* user manage his/her collection of materials with regard to what should be made public and what should not, especially given the organic nature of the data Pliny users create. RDF 1.1's concept of *named graphs* (see [Carroll et al 2004]) might fit with this issue. 82

Dealing with materials that change over time

One curious tension in RDF is that in a world of continued changing materials, RDF's representation seems to be surprisingly static. There is no sense of timestamping of triples, for example, and no way to attribute triples to individuals who asserted them. Indeed, triples act as second-class objects within RDF with no way to attach attributes of any kind to them. RDF's "reification" process (see [Nguyen, Bodenreider and Sheth 2014] for a discussion) is meant to deal with this issue to some extent, but is, at least as of yet, poorly semantically defined and thus poorly supported by software that supports RDF. This lack of a sense of time in Semantic Web technologies is what provokes a static sense of data in RDF representations. Considering that one would think that public data would be constantly evolving and changing over time and could contain the conflicting views of many different individuals from around the globe, this is surprising. In all fairness, however, perhaps it is true that RDF was simply not meant to deal with time-related issues when it was designed, and it was then expected when RDF was created that more work in this area, as in others, was still necessary. 83

Of course, although *Pliny*, unlike the Semantic Web, is focused more on the individual than a global world of data, it has to deal with time too, since a scholar user's ideas represented in *Pliny* also continue to evolve, meaning that, like the semantic web more generally, the graph component of the *Pliny* dataset also changes constantly. 84

Changes in the Semantic Web over time is still a research topic. [Khurana 2012], although an MA thesis, seems to us to provide a good overview of the issue and possible approaches to deal with it. We are particularly interested in the connection between communities of graph contributors and temporal graphs talked about there. [Davis 2009], although now rather dated, also presents some approaches too. [Halpin and Cheney 2014] explore a provenance model as a basis for modelling the evolution of a data domain on the Semantic Web. Van de Sompel *et al* explores evolving data in the context of versioning of documents, borrowing ideas from software versioning systems and content management systems such as VCS [Van de Sompel et al 2010, 3]. Although this work does not address all of the issues that *Pliny* data raises it may provide suggestive ideas. Also, work arising out of *Pliny*, if properly framed, could well constitute further research of interest to computer science on time-evolving data. 85

Conclusions

With this paper we hope to start some discussions about aspects of how the Semantic Web might fit with a broad range of research activities as they are practiced in humanities scholarship. Semantic annotation – an approach to information handling that offers some appeal as one way to connect reading-as-scholarship with the structures of the Semantic Web – provides only one perspective on how this scholarship-Semantic Web connection might work. As it turns out, although not originally conceived of in Semantic Web terms, *Pliny* provides a model for formalising several aspects of traditional scholarship that centre on note-taking and concept development, and which are still largely compatible with RDF data structures: in particular RDF's graph representation. This "graph part" of *Pliny*, by providing a conceptual link with the Semantic Web, allows us to think in a relatively rich way about the possible interactions between actual scholarly practice and the Semantic Web than more Semantic Web driven approaches such as semantic annotation does. Furthermore, we have also explored a little how the process of the emergence of ideas in scholarship is not entirely well served by graph-oriented Semantic Web constructs, and how a different model – two-dimensional space – provides at least one alternative way to think about how to support the process of emergence of ideas. 86

Pliny's model, by focusing on supporting the *process* of scholarship rather than the representation of its end results, also provides a framework which allows us to engage with the question of how interpretation building, as it might actually be done by scholars, can be better fit with the potential of the Semantic Web. This fitting together must be an important thing to keep in mind if we wish to crack into the mainstream world of scholarship with the Semantic Web – to see how and to what extent scholarship, as it is carried out by most scholars, can be part of "thinking in the graph" as Gradmann put it. Any approach to capture scholarly practice primarily via the highly formalised parts of the Semantic Web through approaches such as semantic annotation does not readily engage with the key work of humanities 87

scholarship: the development of a new personal perspective on a body of material, and – if the idea is persuasive to others – its gradual adoption into the body of shared thinking in a particular humanities discipline.

88 In fact, we think that *Pliny*'s model of scholarship support, while revealing aspects of scholarly work that are graph-like and compatible with the Semantic Web, also reveals limitations that are inherent in a graph-based data model. In particular, incompleteness and ambiguity – central facts of life in most scholarship – are handled in Pliny's user interface by the provision of a true 2D working space, and we have suggested in this paper how proximity in 2D space can be used to deal with at least some aspects of incompleteness and ambiguity in a formal digital representation. However, we have also shown that the essence of 2D spatial proximity, as an element of expression by a scholar, does not fit entirely comfortably with RDF and the Semantic Web. Position data can be readily expressed in RDF, but once there it cannot be used by itself to exploit in standard Semantic Web ways the significance of, say, spatial proximity as a relationship between two items. Interestingly, however, *Pliny*'s approach *also* suggests that as the ideas become mature and enriched the 2D aspect of their representation becomes less significant. Thus, for those mature items apparently not much information is lost when spatial information is ignored. However, for less fully formed items – where 2D proximity is still playing a significant semantic role – this might be more of an issue.

89 *Pliny*'s approach, with its particular way to assist a scholar through a process of developing his/her understanding, provides a data model which combines a graph representation (which fits well with current trends in formal structured data in the Semantic Web) with the use of 2D space as an exploratory tool (which provides a mechanism to deal with ambiguity and lack of clarity that is an inevitable result of the process of developing an interpretation of a body of materials, but which fits significantly less well with the Semantic Web). *Pliny*, with its attempt to model the scholarly process in the way it does, encourages one to think about how intellectual work in the humanities might better fit with the broad world of open, linked data. In this way, it allows us to think about, as Gradmann suggested [Gradmann 2012, slide 22], a "scholarly use of semantic technology" that moves "beyond emulation of Annotation". It also offers us one way to think about how to move the discussion of representing humanities research in terms of the Semantic Web to one that is based upon the epistemological issues that Gradmann claims is "where the issues are located".

Acknowledgements

90 This paper started out as a more formal and expanded presentation of first an informal presentation at a NeDiMah meeting in 2012, then as a paper at DH2013 with the same title by the same authors. Its text, however, has been redrafted several times, and has benefited greatly from the comments and suggestions of its reviewers, for which the authors offer many thanks. Some previous work on *Pliny* was funded by the Mellon Foundation thanks to the award of their MATC award program.

Works Cited

Berners-Lee 2009 Berners-Lee, T., "The next web". TED 2009 presentation. Online at https://www.ted.com/talks/tim_berners_lee_on_the_next_web

Berners-Lee 2010 Berners-Lee, T., "The year open data went worldwide". TED 2010 presentation. Online at http://www.ted.com/talks/tim_berners_lee_the_year_open_data_went_worldwide

Berners-Lee et al 2001 Berners-Lee, T., Hendler, J. and Lassila, O. "The Semantic Web", *Scientific American*. 284.5. (2001): 34-43.

Bessette 2015 Bessette, L.S. "How do you Annotate in Your Class?" *The Chronicle of higher Education ProfHacker* blog. May 11, 2015. http://chronicle.com/blogs/profhacker/how-do-you-annotate-in-your-class/60101

Blair 2003 Blair, A. "Reading Strategies for Coping with Information Overload ca. 1550-1700", *Journal of the History of Ideas*. 64.1 (2003): 11-28.

Blair 2004 Blair, A. "Note Taking as an Art of Transmission", *Critical Inquiry* 31.1 (2004). University of Chicago: 85-107.

Bodenhamer 2008 Bodenhamer, D. J. "History and GIS: Implications for the Discipline". In Knowles, A. K. (ed.). *Placing History: How Maps, Spatial Data, and the GIS Are Changing Historical Scholarship*. Redlands, CA: ESRI Press. pp. 219-234 (2008).

Bradley 2003 Bradley, J. "Finding a Middle Ground between 'Determinism' and 'Aesthetic Indeterminacy': a Model for Text Analysis Tools", *Literary and Linguistic Computing*. 18.1 (2003): 185-207.

Bradley 2008 Bradley, J. "Thinking about Interpretation: *Pliny* and Scholarship in the Humanities", *Literary and Linguistic Computing* 23.3 (2008): 263-79. doi: 10.1093/llc/fqn021.

Bradley 2009 Bradley, J. "Can Pliny be one of the muses? How Pliny could support scholarly writing". Peer reviewed poster given in the Digital Humanities 2009 conference, University of Maryland, USA, 23 June 2009. Poster online at http://pliny.cch.kcl.ac.uk/docs/A0-poster-pliny-3.pdf (accessed 13 November 2013).

Bradley 2012 Bradley, J., "Towards a richer sense of Digital Annotation: Moving beyond a 'media' orientation of the annotation of digital objects". In Digital Humanities Quarterly. 2012.6.2. Online at http://www.digitalhumanities.org/dhq/vol/6/2/000121/000121.html (accessed 13 November 2013).

Bradley 2014 Bradley, J. "Silk Purses and Sow's Ears: Can Structured Data deal with Historical Sources?" *International Journal of Humanities and Arts Computing.* Vol. 8 No. 1, April 2014: 13-27. doi: 10.3366/ijhac.2014.0117.

Bradley and Short 2004 Bradley, J. and Short, H. "Texts into databases: the Evolving Field of New-style Prosopography" in *Literary and Linguistic Computing* 20.Suppl. 1 (2004): 3-24.

Brickley and Guha 2014 Brickley, D. and Guha, R.V. *RDF Schema 1.1: W3C Recommendation 25 February 2014.* W3C publication. Online at http://www.w3.org/TR/2014/REC-rdf-schema-20140225/

Brockman et al 2001 Brockman , W. S., Neumann, L., Palmer, C.L. and Tonyia J. Tidline *Scholarly Work in the Humanities and the Evolving Information Environment*. Council on Library and Information Resources (2001).

CRMInf 2015 *CRMinf: the Argumentation Model*. Website at http://www.ics.forth.gr/isl/index_main.php?l=e&c=713" (accessed 17 April, 2015)

Carroll et al 2004 Carroll, J.J, Bizer, C., Hayes, P. and Stickler P. "Named Graphs, Provenance and Trust". Bristol: HP Laboratories. Online at ttp://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.1.2197

Congleton et al 2009 Congelton, B., Cerretani, J., Newman, M.W. and Ackerman M.S. "Sharing Map Annotations in Small Groups: X Marks the Spot". *Human-Computer interaction: INTERACT 20090: 12th IFIP TC conference*: Springer Publishers

Cyganiak, Wood and Lanthaler 2014 Cyganiak, R., Wood D., and Lanthaler, M. *RDF 1.1 Concepts and Abstract Syntax*. World Wide Web Consortium (W3C) (2014). Online at http://www.w3.org/TR/rdf11-concepts/ (accessed 2 April, 2015)

Davis 2009 Davis, I. "Representing Time in RDF: Part 1". Blog *Internet Alchemy.* Entry Aug 10, 2009, Online at http://blog.iandavis.com/2009/08/representing-time-in-rdf-part-1/

Dean 2015 Dean, J. "Back to School with Annotation: 10 Ways to Annotate with Students." *Hypothes.is* blog. 25 August 2015. https://hypothes.is/blog/back-to-school-with-annotation-10-ways-to-annotate-with-students/

Doerr, Kritsotaki and Boutsika 2011 Doerr, M., Kritsotaki, A. and Boutsika, K. "Factual argumentation — a core model for assertions making". *Journal on Computing and Cultural Heritage*, 3.3 (March 2011). doi: 10.1145/1921614.1921615. Online at http://dl.acm.org/citation.cfm?id=1921615 (accessed 20 April, 2015)

Eide 2015 Eide, Øyvind. *Media Boundaries and Conceptual Modelling : Between Texts and Maps*. Houndmills, Basingstoke, Hampshire: Palgrave Macmillan, 2015.

Elleström 2014 Elleström, Lars. *Media Transformation: The Transfer of Media Characteristics among Media*. Houndmills, Basingstoke, Hampshire: Palgrave Macmillan, 2014.

Engelbart 1962 Engelbart, D.C., *Augmenting Human Intellect: A Conceptual Framework*. Stanford Research Institute, Report AFOSR-3223 1962. Online at http://www.dougengelbart.org/pubs/papers/scanned/Doug_Engelbart-AugmentingHumanIntellect.pdf

Gradmann 2012 Gradmann, S. "The Web & Digital Humanities: What about Semantics?" A paper given at WWW 2012, Lyon France 20 April 2012. Online at http://www.slideshare.net/DM2E/what-about-semantics-stefan-gradmann-www2012-lyon-france (accessed 13 November 2013).

Grassi et al 2013 Grassi, M., Morbidoni, C., Nucci, M., Fonda, S. and Piazza, F. "Pundit: augmenting web contents with semantics", *Literary and Linguistic Computing.* Vol 28 No 4 (2013): 640-659.

Gravemeijer 1999 Gravemeijer, K. "How Emergent Models May Foster the Constitution of Formal Mathematics", *Mathematical Thinking and Learning* 1.2, pp. 155–177.

Gruber 1993 Gruber, Thomas R. "Toward Principles for the Design of Ontologies Used for Knowledge Sharing", *International Journal Human-Computer Studies.* Vol 43, pp. 907-928

Guetzkow et al 2004 Guetzkow, J., Lamont, M., and Mallard, G. "What is Originality in the Humanities and the Social Sciences", *American Sociological Review,* 69.2 (Apr 2004): 190-212. Online at http://www.jstor.org/stable/3593084 (accessed 13 November 2013).

Halpin and Cheney 2014 Halpin, H. and Cheney J. "Dynamic Provenance for SPARQL Updates", *The Semantic Web – ISWC 2014*. Conference proceedings October 19-23, 2014. Springer International Publishing, Doi: 10.1007/978-3-319-11964-9_27

Jackson 2010 Jackson, L. S. *RDF-encoding Pliny annotations in the Open Annotation Collaboration project*. Technical report #ISRN UIUCLIS--2010/2+OAC. Graduate School of Library and Information Science, University of Illinois at Urbana-Champaign. Online at https://www.ideals.illinois.edu/handle/2142/15418" (accessed 13 November 2013).

Khurana 2012 Khurana, U. *An Introduction to Temporal Graph Data Management.* Computer Science Department, University of Maryland MA thesis. Online at https://www.cs.umd.edu/sites/default/files/scholarly_papers/Khurana_SchPaper_1.pdf

Lachance 2013 Lachance, F. (2013). *Humanist* Vol. 27 No 284: 13 Aug 2013.

Lange 2008 Lange, C. "SWiM – A Semantic Wiki for Mathematical Knowledge Management", in S. Bechhofer, M. Hauswirth, J. Hoffmann & M. Koubarakis, eds, "The Semantic Web: Research and Applications", Vol. 5021 of Lecture Notes in Computer Science, Springer Berlin Heidelberg, pp. 832–837.

Lavagnino 1997 Lavagnino, J. (1997). Reading, Scholarship, and Hypertext Editions. *The Journal of Electronic Publishing*, 3.1 (Sept 1997). Online version available at http://cds.library.brown.edu/resources/stg/monographs/rshe.html

Louch 1969 Louch, A. R. "History as Narrative." *History and Theory* 8,1 (1969): 54-70.

Marshall and Bush 2004 Marshall, C. and Bush A.J.B. "Exploring the relationship between personal and public annotations". Proceedings of JCDL '04 conference. New York: ACM. Doi:10.1145/996350.996432

Moretti 2005 Moretti. *Graphs, Maps, Trees: Abstract Models for a Literary Theory*. New York: Verso (2005).

Mueller 2015 Mueller, M. *Re: 28.824 an argu[m]ent & belief system*. In Humanist Vol 28 No 843.

Nersessian 2008 Nersessian, Nancy J. *Creating Scientific Concepts*. Cambridge, Mass.: MIT Press.

Nguyen, Bodenreider and Sheth 2014 Nguyen, V., Bodenreider, O., Sheth A. "Don't like RDF Reification? Making Statements about Statements Using Singleton Property". *Proceedings of the WWW '14 conference*. April 7-11 2014 Seoul, Korea. New York: ACM: Doi: 10.1145/2566486.2567973

OA 2013 *Open Annotation Data Model: Community Draft, 08 February 2013*. W3C online publication at website http://www.openannotation.org/spec/core/ (accessed 12 August 2015).

OWL 2012 W3C OWL Working Group, *OWL 2 Web Ontology Language: Document Overview (Second Edition)*. W3C publication. Online at http://www.w3.org/TR/owl2-overview/.

Palmer et al 2009 Palmer, C., Teffeau, L.C. and Pirmann, C.M. *Scholarly information practices in the online environment: themes from the literature and implications for library service development*. Technical report, OCLC Research (2009). Online at http://www.oclc.org/research/publications/library/2009/2009-02.pdf (accessed 13 November 2013).

Pasin and Bradley 2015 Pasin, M. and Bradley, J. "Factoid-based prosopography and computer ontologies: Towards an integrated approach". *Literary and Linguistic Computing*. Vol 30 No. 1 (2015), pp. 86-97. Published online June 29, 2013 doi:10.1093/llc/fqt037.

Pliny 2009 *Pliny: A Note Manager*. Website at http://pliny.cch.kcl.ac.uk/ (accessed 13 November 2013).

Rüsen 1987 Rüsen, J. "Historical Narration: Foundation, Types, Reason." *History and Theory* 26,4 (1987): 87-97

Sanderson et al 2013 Sanderson, R., Ciccarese, P. and Van de Sompel, H. *Open Annotation Data Model: Community Draft*. World Wide Web Consortium (W3C) (2013). Online at http://www.openannotation.org/spec/core/ (accessed 13 November 2013).

Unsworth 2000 Unsworth, J. "Scholarly Primitives: what methods do humanities researchers have in common, and how might our tools reflect this?" In symposium on "Humanities Computing: formal methods, experimental practice" sponsored by King's College, London, May 13, 2000. Online at http://people.lis.illinois.edu/~unsworth/Kings.5-00/primitives.html (accessed 13 November 2013).

Van Zundert 2014 Van Zundert, J. "Big Data no boondoggle". *Humanist* 28.452. October 29, 2014

Van de Sompel et al 2010 Van de Sompel, H., Sanderson R., Nelson, M.L., Balakireva, L. L, Shankar H., Ainsworth, S. "An HTTP-Based Versioning Mechanism for Linked Data". Presented at *LDOW 2010* conference April 27, 2010, Raleigh, USA. Online at http://events.linkeddata.org/ldow2010/papers/ldow2010_paper13.pdf

Völkel 2011 Völkel, M., *Personal Knowledge Models with Semantic Technologies*, Books on Demand, Norderstedt.