

Fitting Personal Interpretations with the Semantic Web

John Bradley

(Department of Digital Humanities, King's College London)

john.bradley@kcl.ac.uk

Michele Pasin

(Nature Publishing Group)

michele.pasin@gmail.com

Abstract

The emergence of the semantic web has led to work to see how it could be made to fit with humanities scholarship. One of the approaches has been to explore semantic annotation which connects parts of text to formal semantic representation of a knowledge domain. In this paper we explore why semantic annotation, although a most useful activity, does not capture the true heart of humanities scholarship, and we propose a model, emerging from work done in Pliny, for more richly capturing scholarly work in terms that are compatible with the semantic web. 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 with these issues, and consider the boundaries between the expressiveness of 2D space and the formal graph model of the semantic web.

Fitting Personal Interpretations with the Semantic Web

John Bradley, Michele Pasin

to think is often to sort, to store and to shuffle: humble, embodied tasks

Humanist Vol. 27 No 284: 13 Aug 2013, by
François Lachance

The Semantic Web was originally developed by Tim Berners-Lee in the late 1990s (see Berners-Lee et al 2001), but many of its ideas arise out of one of the sub-domains of computer science called *Knowledge Representation*. As Wikipedia tells us "**Knowledge representation** (KR) is an area of artificial intelligence research aimed at representing knowledge in symbols to facilitate inferencing from those knowledge elements". John Unsworth recognised quite early on that knowledge representation had important things to say to the humanities: "For humanities computing, knowledge representation is a compelling, revelatory, and productive way of doing humanities research -- and in many ways, it is what humanities computing has been doing, implicitly, for years." (Unsworth 2001, para 1)

Knowledge Representation implies highly structured data, and within King's Department of Digital Humanities we have had much experience of the power of knowledge representation in that form. The authors have worked on more than 20 substantial projects from fields such as history, classics, musicology, theatre studies and art history where structured data was front and centre. In all of these projects we found that important aspects of humanities scholarship *were* being represented by the structures built for them. In almost every case our discipline partners could see key ideas that they were interested in made evident in new ways they had not originally expected and this facilitated unexpected kinds of exploration.

As compelling and revelatory as this data representation has been for these partners, it is clear that the KR way of working is substantially foreign to most humanists. The primary product of scholarly research is almost always books, chapters, articles; narratives of various forms. Recently there have been claims within the DH community that this focus on articles and books as a kind of old-guard position: those old guys are protecting their turf – and probably some of this is in fact true. However, many scholars (Rüsen 1987 p. 87, Louch 1969, p 540, Bodenhamer 2008, p 224) have recognised that textual narrative, with its potential subtlety, offers the appropriate medium for presenting humanities research. Furthermore, even someone as influential as Franco Moretti in his book *Graphs, Maps, Trees* (Moretti 2005), with his idea of "distant reading", resorts to narrative (the book!) as the way to present his research even though the techniques he used to get them are significantly non-traditional.

At first glance, this kind of research output, the book or article, which perhaps represents the preferred output for, say, 95% of humanists, does not seem to be compatible with knowledge representation or the semantic web, and the structuring of material that is implicit in it. For Semantic Web technologies such as RDF (Klyne and Carroll 2004), the formal representation used is what mathematicians call a *graph*. The question, then, is as Stefan Gradmann stated it in his presentation at WWW2012: “Thinking in the graph: will Digital Humanists ever do so?” (Gradmann 2012, slide 2). Indeed, there is an even more important question: “Thinking in the graph: will Humanists (more generally) ever do so?” If one wants to find a place for the Semantic Web in the humanities, how does one square the circle of, on one hand, most humanities scholarship being presented as largely formally unstructured prose narrative, and on the other, the high degree of formal structuring of the semantic web which is strongly non-narrative in nature?

One of the efforts to bring structure into the humanities has been to explore *semantic annotation* (See definition Ontotext 2013). We will first take a look at current practices in semantic annotation and suggest why, by itself, it misses out on much of what humanities scholarship is really all about. We will then explore a different approach to introducing formal structuring representations into traditional humanities, and consider how this alternative model could provide a richer way to connect scholarship to the semantic web.

Semantic Annotation

Unlike conventional annotation, which is usually thought of as connecting a section of a text to a small text, semantic annotation links a section of text into a formal structure that captures the semantics of some domain. The idea of semantic annotation can be best given by an illustration – see Fig. I.

In the top part of Fig. I one sees a bit of text about the Yalta conference that followed World War II. In the bottom part of the figure one sees a graph representing some of the information about that conference represented as structured data and coded according to the conventions of the CIDOC-CRM. We see semantic annotation between the text excerpt at the top represented as connections – perhaps expressed by markup– between the names of Churchill, Stalin, Roosevelt and Yalta and the objects in the structure at the bottom that represents them. The graph shown in this figure is, perhaps, misleadingly simple: imagine how its complexity would grow if, say, one added other information to it about the main actors, or about Yalta, or more generally about World War II.

In February, 1945, Joseph Stalin, Winston Churchill and Franklin D. Roosevelt met again. This time the conference was held in Yalta in the Crimea. With Soviet troops in most of Eastern Europe, Stalin was in a strong negotiating position. Roosevelt and Churchill tried hard to restrict post-war influence in this area but the only concession they could obtain was a promise that free elections would be held in these countries.

From "Spartacus Educational": <http://www.spartacus.schoolnet.co.uk/2WWyalta.htm>

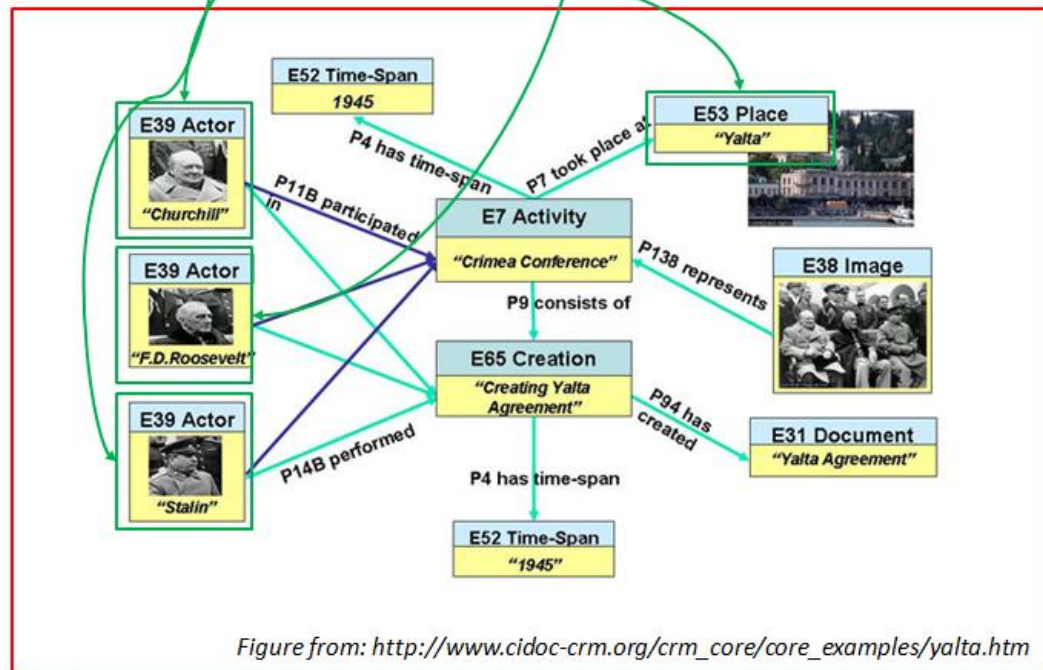


Fig. I. Semantic Annotation

Clearly, semantic annotation material like this that connects text to a formal representation of ideas is predicated upon the idea that there exists a formal representation, an *ontology*, of a body of relevant knowledge (here, people, places, documents, etc.) to link to. Although not much of the humanities has been mapped into this kind of formal representation, there has been for some time substantial work of this kind in other fields, and this has made semantic annotation a prominent activity there. As Wikipedia notes in their article "Ontology Engineering":

Life sciences is flourishing with ontologies that biologists use to make sense of their experiment. For inferring correct conclusions from experiments, ontologies have to be structured optimally against the knowledge base they represent. The structure of an ontology needs to be changed continuously so that it is an accurate representation of the underlying domain.

The Alzheimer research forum, for example, supports the use of the SWAN annotation tool that, it claims "... uses Semantic Web technology, to tie together statements made in scientific publications or on the Web to scientific evidence, biological terminologies, and knowledge bases" (ARF 1996-2013, web page <http://www.alzforum.org/res/adh/swan/default.asp>).

When a user connects a bit of text in a research paper to an entity in an ontology representing knowledge about the field, that link enriches the formal structure captured in the ontology. This is clearly an example of Linked Data at work. However, note as well the parallel research work represented by the continual updating of these ontologies by the life sciences community to reflect current understanding (Sansone *et al* 2012).

Although not as developed as in the Life Sciences, there has been some work on semantic annotation in the Digital Humanities. We at DDH have used Jamie Norrish's Entity Authority Tool Set (EATS, see EATS 2013 and Norrish and Stevenson 2008) to store data about entities such as people or places and link them to texts in several of our project, although EATS does not structure its entities as rigorously as Semantic Web technologies can. Furthermore, there are two environments for semantic linking that seem to be aimed at humanists and that operate much like SWAN. *Pundit* (Pundit 2013, see Grassi *et al* 2013) provides a web browser-based environment that is aimed at "augmenting web pages with semantically structured annotations". It claims to allow one to "easily turn web documents into a semantic knowledge network by pulling from and enriching the Web of Data". In one of their online examples the viewer is shown how to tag a reference in Wittgenstein's *Philosophical Investigations* to the idea of the "language game" and link it to the Wikipedia/DBPedia's URI for the "language game" concept. A similar ability to link text to structured data is provided by the software *SWickyNotes* (Morbidoni and Nucci 2012), which is billed by its developers as supporting "semantically annotating web pages and digital libraries". Both *Pundit*'s and *SWickyNotes*'s interface for Semantic annotation are excellent examples of semantic annotation tools in a humanities context.

The linking of text to a predefined structure of an existing formal ontology as *SWAN*, *Pundit* and *SWickyNotes* do is doubtless useful work and enriches texts in exactly the way envisioned by the Semantic Web. However, it is important to recognise that by enabling the linking of text to predefined ontologies created by others one limits the kind of things that one can say about the text. Indeed, one could characterise this work as a kind of, let us say, junior research activity: "junior" in the sense that in a large textual project it would be given to research assistants to do under the direction of a more senior researcher. In contrast, almost universally, senior researchers write papers and books as products of their research and encourage their early career colleagues to do the same. It would seem that semantic annotation does not capture much of this work that humanist scholars do.

Glimpses of Formal Structure in Scholarly Thinking

Scholarly Writing may not seem to be at all compatible with Knowledge Representation. Two things, however, suggest a place for KR in the picture: First, if we look closely at most articles presenting research in the humanities, we can probably see that there *is* formal structure at least implied in the text, both directly evident in the structure represented by the flow of the argument, but also in the identification of themes, concepts and their connections that are presented in them. However, as Guetzkow *et al* remind us (2004), scholars want to say something new about their materials, and hence will not simply want to refer to ideas contained in an existing mature intellectual framework that could perhaps be represented in a formal ontology. These scholars are trying to develop their *own* voice and perspective on their material that is different from that currently established within

their discipline: perhaps drawing on some existing ideas, but also extending or perhaps even more fundamentally breaking with them by introducing new themes, concepts, or connections.

A second place for KR-like structure is perhaps more surprising. We claim here that the pre-writing work that resulted in the development of ideas that went into the scholarly article presents an opportunity for structure that can be thought of, at least in part, compatible with KR-, indeed graph-like, thinking. As we suggest later in this article, perhaps the concepts, themes and so on that exist in the scholar's head and that underlie an article can be viewed as forming this compatible structure, and the process of developing this structure before the article describing it is written – which we will be calling here *interpretation development* – is the place where this underlying structure can be most clearly visible. This structure for the interpretation does not emerge, fully formed, in the author's head at the time the article was being written. It emerges after substantial engagement with the materials s/he was working with. Before that it is likely that the ideas were still only partially formed. There is a *process* around the emergence of ideas going on here.

But does this process engage with materials in ways that are at all compatible with the Semantic Web? Up to now most work in the SW has focused on the representation of already highly structured fields of knowledge such as the Life Sciences ontologies. RDF and the rest of the Semantic Web toolkit do not appear to provide a way of representing new concepts and ideas *before* they are formally clear. Indeed, the RDF technologies would appear to bring formalism to bear too soon in the process of developing ideas in the humanities. The rest of this article outlines the approach taken by Pliny that provides at least a possible way to think about connections between interpretation development and the Semantic Web.

Pliny and 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* is not particularly about annotation, or at least not about annotation in isolation from its place in scholarship. Instead, it was meant to combine a digital approach to annotation with other thoughts about the representation of ideas, of which annotation is only a starting point, into a tool that supports a fuller range of essentially traditional humanities scholarship than annotation does by itself.

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

A first lesson one finds, then, is that in almost any substantive humanities scholarly project, one expects to start out with only a vague sense of the issues one is interested in. John Lavagnino (1997) observes this aspect of notetaking when he observes that reading was not "a mere collection of data" (like the semantic annotation processes we have seen earlier are). Instead, he claimed that the place of reading in scholarly research is 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" (p 114). The mere

act of *taking* of notes and/or annotation did not, by itself, capture the central place of these notes in doing the research. It was also 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.

In this light we can see a problem with much of the work at present on 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 to something you are reading. Although this is useful for, say, adding public commentary for teaching, 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. On the other hand, semantic annotation – of the kind done by the tools talked about earlier in the article – brings the formal structure of the linked ontology in too soon by 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.

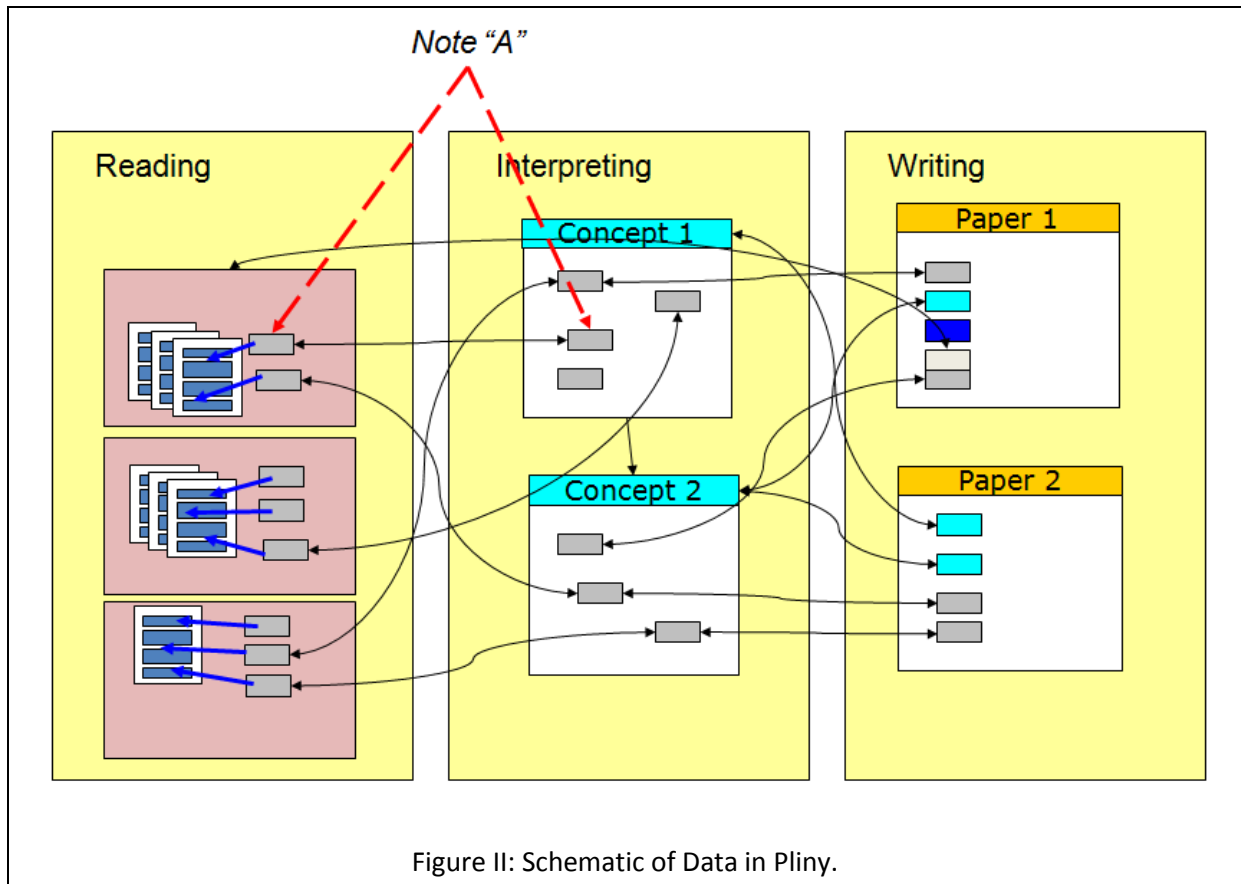
Much of a scholar's subsequent work involves struggling with vague, incompletely defined ideas, and only after a good period of time does some degree of clarity emerge, a process that much of the time should be thought of as *pre-ontological*. Once one begins to think of interpretation building as this kind of a process, with perhaps a clear conception emerging much later, 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 that helps a researcher while they are developing their interpretation?

The work in *Pliny* then was not only about how to support the creation of notes in the first place (through, say annotation), 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 was a repository for these notes, how can it best deliver them to the user to support the user's engagement with them as they 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 2 dimensional spaces – a way to cope with lack of clarity, ambiguity and vagueness.

An early article about *Pliny* (Bradley 2008) presents *Pliny*'s place in scholarship as supporting three phases: (a) reading, (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 brought these new ideas into the public sphere in the form of a

book or article (phase c). Fig. II (similar to Fig. 3 in Bradley 2013) represents schematically materials assembled in a Pliny repository, showing how they might relate to these three activities of scholarship.



The left side of this figure shows the first phase: reading of both primary and secondary literature for research. The diagram shows small boxes with links to spots in the source texts. These are the annotations created by the researcher as s/he reads. Initially, at least – attending to Lavagnino's comments earlier – the reader may well not be in a position to attach specific formalisms to the text – she hasn't developed the formalisms yet. So, instead, the notes are 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) by developing a framework for them.

The middle area corresponds to the development of interpretation phase of the research. Initially, there will not be many objects there. However, as the researcher thinks about the materials s/he has been reading and works to understand them s/he will endeavour to organise his/her thoughts into concepts, categories, or topics. Thus, Fig. II shows in its middle area the notes that were originally created by the researcher 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 a researcher would use), and then these topics being supplemented by new notes that represent new thinking by the researcher about the topics s/he has formally identified within *Pliny*. In *Pliny* a note that started out as an annotation (such as the one labelled "Note A") can be also referenced in the different context of one of the interpretation 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*.

Finally, when the time is right (and probably after more than 2 concepts have been recognised), the researcher draws on the ideas she has formed in the interpretation phase to put together papers that present them. Now, the work of organising the materials changes to be one that supports the development of a narrative for the 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 narrative of an article, there has been some exploration in the *Pliny* project of how such a tool might operate – see Bradley 2009.

While looking at Fig. II we must think of scholarship as process, and think of it as a snapshot of this at a particular stage. 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 with more branches added, the *Pliny* model accommodates starting small with notes created while reading through the gradual development of ideas in the interpretation stages. 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 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 that grows off of the research structure stored in *Pliny* many times during the life of the research. Fig. II, 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.

Although a tree is a useful metaphor here to capture the organic 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*. As it changes over time material is likely to develop from left to right, with first much of the material in the "notetaking/reading" area at the left, but then a shift as more concepts begin to be recorded in the middle area. Each of these areas is likely, in fact, to develop at a different pace.

Annotation in Pliny

Looking for a moment at *Pliny*'s approach to notetaking during reading – often categorised as annotation – we see an example of *Pliny*'s interface for annotation of documents in Fig. III. Someone has been adding notes to an article – here a PDF file of McCarty's 2008 article "What's going on?".

Pliny simulates the way annotations work on paper. This is in contrast to pretty well all the digital annotation interfaces produced recently. These interfaces, when first opened, show the object that has been annotated sprinkled with markers that indicate where annotations have been added. To see any annotation itself, the user has to request it – perhaps by, say, clicking on the annotation symbol that is currently displayed, and they may only be able to see one annotation at a time. In conventional, paper based annotation, and in *Pliny*, the user's experience of annotation is quite different. First of all, when researcher turns to a printed page that with annotations, they are all immediately visible when the page is looked at – nothing needs to be clicked on to see an

annotation. 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. In the same way that the annotator can make entirely free use of a page of paper to hold his/her annotations and indeed has the option of exploiting the spatial sense of the page as part of the expressive toolkit for annotation, in *Pliny* one can use the 2-D space of the screen page here to hold the annotations. In *Pliny*, then, an annotation is not only a link to a bit of another resource, but has a place in a 2D space provided by the material it annotates.

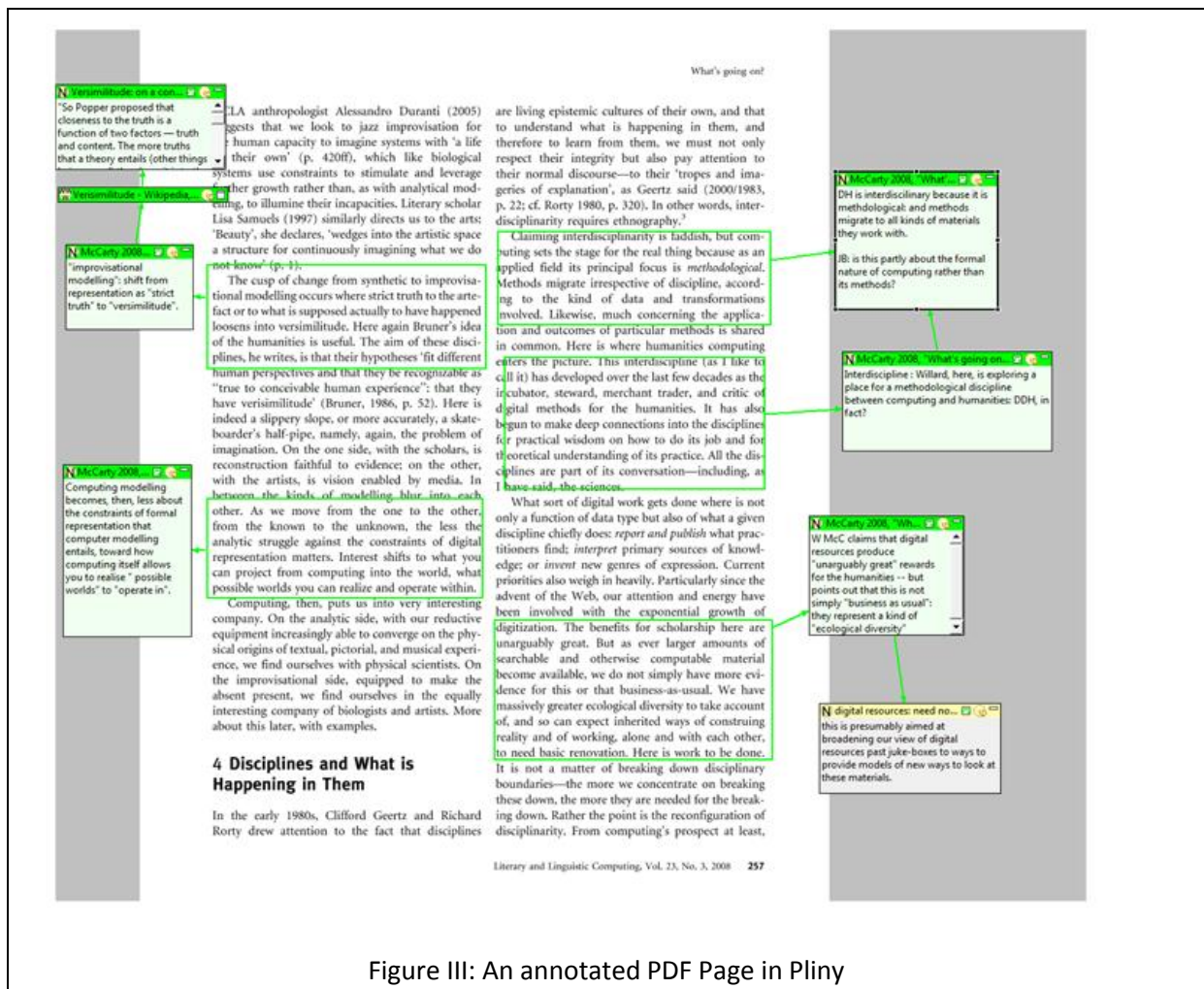


Figure III: An annotated PDF Page in Pliny

Supporting the development of concepts

We have now had a brief overview of *Pliny*'s affordances as they relate to annotation. However, *Pliny* is not only about annotating things, which was, you recall, represented only in the left "reading" area of Fig. II. How does *Pliny* support the central phase of research: the development of an interpretation? Here, the nature of the *process* of research plays a rather more prominent role.

One way to think about this process is as a gradual increase in formalism: 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 (p. 273). Presenting a temporal process such as this in an illustration is difficult, so Fig. IV shows a set of snapshots over time of the Pliny space for *uses of space for study* at different stages during its development, identified as (a), (b), (c), (d) and (e).

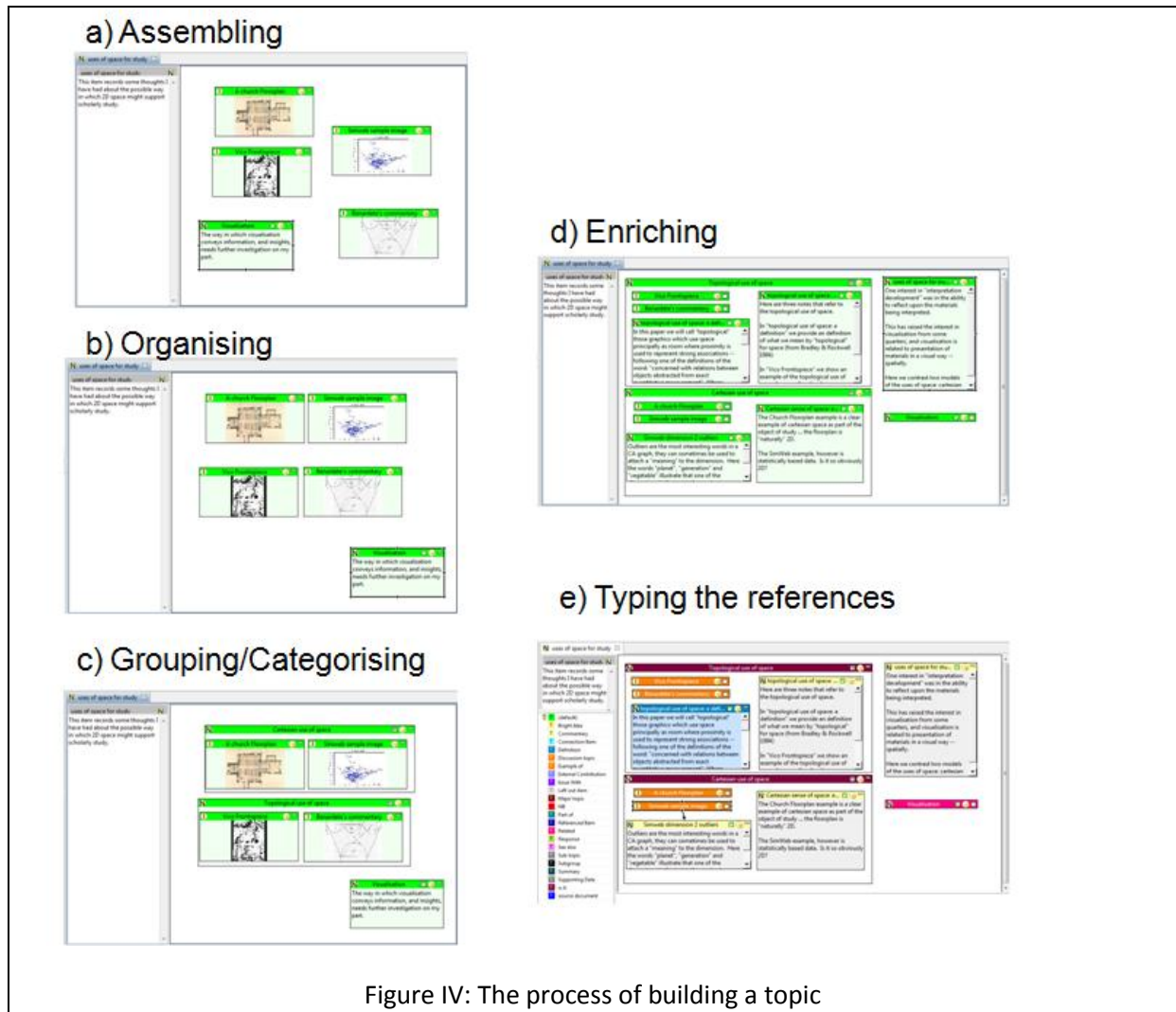


Figure IV: The process of building a topic

In Fig. IV I 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 Carol 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 of an information scientist; my set of primitives diverge from Unsworth's too.

The first step, shown in part (a) of Fig. IV, is *assembling* (corresponding in part, perhaps, to Unsworth's *discovering* primitive). One begins by creating a holder for the topic, and naming it "uses of space for study". We see it here in part (a) of Fig. IV, with a brief description of the idea also entered on the left area. The main place where the work is done is in the larger right 2D space that *Pliny* provides. The *Pliny* user starts off by assembling references to things that relate to the topic she are interested in: here four images that show space being used in different ways, and in addition a reference to a note on another topic the *Pliny* user 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*.

Fig. IV b: Having now assembled a few items the researcher begin to notice some similarities in the use of space in two of the images, and what feels like a contrasting similarity in the other two. As a result, the user takes advantage of the possibility of proximity in the 2D space to organise them a bit, placing those with a similarity of interest close together. One can characterise this kind of activity as the beginning of the task of *organising* of your material. Perhaps some part of this operation might broadly correspond to Unsworth's *comparison* primitive.

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 subtlety and perhaps usefully ambiguity available in the 2D space for established 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.

Fig. IV c: Now that proximity has helped the *Pliny* user to develop a feel for possible grouping of the items that she has assembled in this space, at some point the groupings might become clear enough that she will feel ready to give them names. Among the four images she sees two groups, representing two rather different kinds of use of space – so she asks *Pliny* to put these images into two groups, and to name them. The *grouping* and then the naming or *categorising* adds more structure to this space, and begins to express an interpretation of them.

Fig. IV d: Having now discovered these two kinds of uses of space the user adds a few notes that record his/her thoughts about them, thereby *enriching* the concept space. As well as creating entirely new material for this purpose, *Pliny* allows her to also reference items that were created elsewhere in *Pliny* – introducing them in a new context, similar to the Note A in Fig. II.

Fig. IV e: Finally, now that the user has collected and organised this material, she notes that there are several kinds of connection 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 one is of type *commentary*. *Pliny* shows these different types as different colours (not visible in the print version of this article, of course). Fig. IV e 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.

In summary, then, Fig. IV suggests steps in a process of developing an interpretation:

- *Assembling:* One starts off by assembling materials that one wishes to work with
- *Annotating:* *Pliny* provides annotation so that the user can record your responses to these materials.
- *Organising:* *Pliny* provides 2D spaces where one can organise notes perhaps created during annotation and other objects to discover relationships between them that will hopefully eventually lead to a clear formulation of a model for your materials.
- *Grouping and Naming:* As concepts become clear one 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 users to add comments to the structures they have become interested in – allowing them to enrich the structure they have created,
- *Typing*: Finally, *Pliny* allows the user to attach assertions about the relationships between objects that have been captured within 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* allows. To recognise this, *Pliny* accommodates a mix of highly structured areas with less structured ones.

By examining the process of 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*. We have focused on two aspects of *Pliny* that also characterise scholarship. First, that *Pliny* supports the *process* of scholarship, not only its products. *Pliny*'s approach from annotation through concept development provides a framework that allows its user to move from largely unstructured ideas to more formal structure by supporting the way that this transition happens: starting off with reading and notetaking, and then through the using of these notes 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 don't. Materials in *Pliny* need not ever be finished.

Second, *Pliny* promotes the use of a 2D working space as a central element in its set of affordances. In the context of annotations the idea that there is a 2D space to hold annotations on top of the annotated resource 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 the actual practice of users suggests that 2D-spaces sense of proximity, with its degree of ambiguity, is more useful at the beginning of the effort to organise a particular topic than it is later as the concepts related to the topic become clearer.

Fitting Pliny to the Semantic Web

Having looked at *Pliny*'s way of modelling the process 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 (Fig. III) and concept development (Fig. IV) 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. The authors have explored these issues by creating a rudimentary extension to *Pliny* in the form of a "plugin" that supported the export of *Pliny* materials into RDF, and that allowed Semantic Web or Linked Data URIs to appear as *Pliny* resources.

Underlying the representation of Pliny materials in RDF is a Pliny ontology. Its major classes are:

- *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, usually any new data types become 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) are in fact Notes too. For our current Semantic Web perspective, this conflating of Notes and Concepts is an expressive limitation that should be addressed – perhaps by allowing 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) to define 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. During the *uses of space for study* development above, the inclusion of the *Vico frontispiece* as an example of one of the kinds of space usage was done by means of a reference object.
- *Reference type*: References can be typed (we saw the typing being applied in Fig. IV (e)), and this supports the *typing* scholarly primitive that was mentioned earlier in this paper. As we've seen, reference typing enriches the semantic expressiveness substantially, and the user can define any reference type s/he wants, defining, for instance, type such as "is an example of", or "is a commentary on". Thus, reference types become the basis for a user-based ontology separate from *Pliny*'s ontology 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 too, two quite obvious questions arise:

- (a) *Reaching out to the Semantic Web*: How can the data that we have shown as accumulated inside of *Pliny* be transformed into RDF: the language of the Semantic web
- (b) *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?

Thinking first of the *Reaching out* part; connecting *Pliny*'s representation of an interpretation in terms of the Semantic Web world; we look first at the part of *Pliny* that supports annotation of digital objects, and how the annotations done there might map to the Semantic Web. To explore this question we made our prototype RDF export mechanism use the OAC's annotation ontology (OAC 2009), extending the work that had been done previously described in Jackson 2010. The task was not entirely trivial because *Pliny*'s annotation tools (for images and PDF files) were not designed

to model simple annotations in terms of objects and targets. Instead, the underlying model is in terms of *Pliny* objects: target areas, notes, and links between them all placed in the 2D space. This enables the sense of annotation that mirrors annotation on paper, as described earlier, but also enables a much richer set of relationships between a collection of notes, targets and connections between them. Our RDF exporter had to take data described in these terms and express them as annotations that were conformant with the OAC.

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. Fig. V 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 Fig. III). The *jb* prefix identifies data that belongs to a particular Pliny user.

```
@prefix pliny: <http://pliny.cch.kcl.ac.uk/base/> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix oac: <http://www.w3.org/ns/oa#> .
@prefix dcterms: <http://purl.org/dc/terms/> .
@prefix jb: <http://www.kcl.ac.uk/ddh/jb/plinydata/> .

jb:resource-99 a pliny:objectType_PDF-Acrobat ;
  A   pliny:name "McCarty 2008, \"What's going on?\"" ;
      pliny:url <http://llc.oxfordjournals.org/content/23/3/253.full.pdf+html> .

jb:resource-92 a pliny:objectType_Note ;
  B   pliny:name "McCarty 2008, \"What's going on?\": Note 36" ;
      pliny:contents "DH is interdisciplinary because [...]" .

Jb:target-99 a oac:SpecificResource ;
  C   oac:hasSource jb:resource-99
      oac:hasSelector jb:selector-99

jb:selector-99 a oac:fragmentSelector ;
      dcterms:conformsTo <http://tools.ietf.org/rfc/rfc3778>
      rdf:value "#page=5&viewrect=279,158,214,74"

jb:lo-119 a jb:refType_RelatesTo ;
  D   oac:hasBody jb:resource-92 ;
      oac:hasTarget jb:target-99 .
```

Figure V: RDF and Pliny annotation

The fragment labelled "A" describes the PDF file that has been annotated. This item, resource-99 is identified as a *PDF object* (elsewhere defined as a subclass of a Pliny resource), and 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, so it is described in those terms. 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 between 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. As mentioned above in our discussion of them, *jb:refType_RelatesTo* is identified as a subclass of an *oac:Annotation* elsewhere in the generated RDF.

So, that is the annotation part of *Pliny* dealt with.

That is not the whole story, of course, since, as we claim here, *Pliny* can also be used to help a user develop a personal interpretation of the materials s/he is working with. How is this material handled in any translation to RDF?

First, as Bradley 2008, pp.274-6 points out, the data structure *Pliny* uses behind these displays can be thought of as a graph with typed nodes and links. This kind of mathematical graph is the foundation model for RDF (Klyne and Carroll 2004, section 3.1), and *Pliny*'s graph model maps quite comfortably into RDF's "subject predicate object" representation. Indeed, there is no need here for the OAC framework since the structures of notes and other *Pliny* objects is not usefully thought of in terms of annotation. As a consequence, the exporter we built used a different translation strategy for data from these 2D interpretation spaces. The key is to focus on the relationship between the note and its holder in the 2D space. The *Visualisation* concept is referenced by the *uses of space for study* concept, and is identified 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*. A fragment of the generated RDF is shown in Fig. VI.

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

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

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

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

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

Figure VI: 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 that was described 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.

First we can see *Pliny*'s support for annotation of images being used as a link to RDF representation of concepts from *DBPedia* in Fig. VII. This figure, similar to Fig. 2 in Bradley 2008 (p. 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.

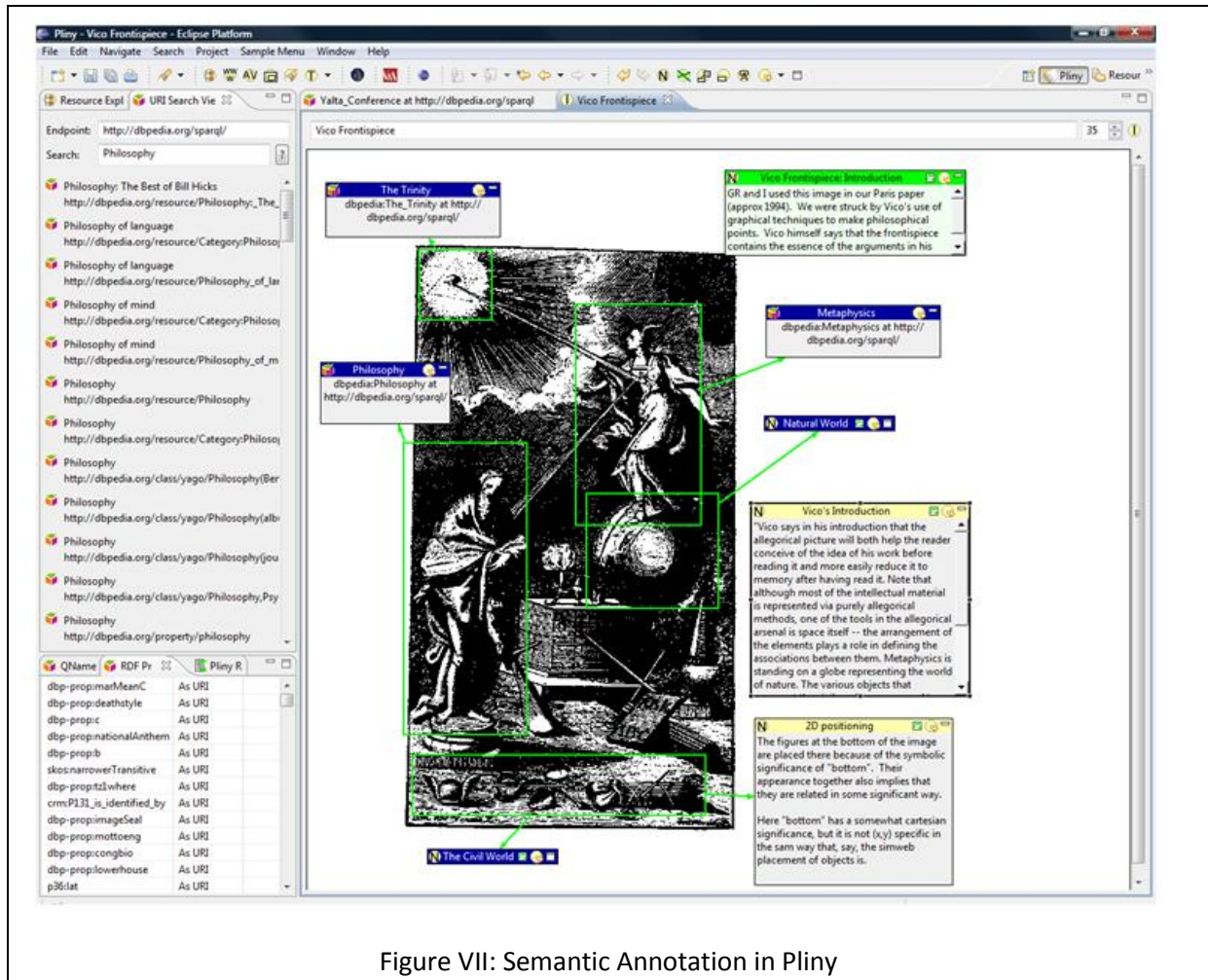


Figure VII: Semantic Annotation in Pliny

In the 2008 figure, the concepts such as *Metaphysics*, or *The Trinity* were labelled and identified with internal Pliny objects as concepts. Our RDF extension, however, allowed us to instead refer 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 connected to the Semantic Web as URIs. To allow a user to locate a URI for a concepts like the Trinity, we developed a prototype query mechanism (shown on the left side) that allowed the user to query any SPARQL endpoint (here, to *DBPedia*) to find URIs within it that had *rdfs:labels* containing the word

"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.

The second kind of engagement of the Semantic Web world involves the use of the structure itself of the Semantic Web 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 (semantic annotation), but the other way round: annotating the graph of the semantic web with personal materials, references to your own research concepts, with your 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 with references to our own collection of documents and notes. 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 Fig. VIII. 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.

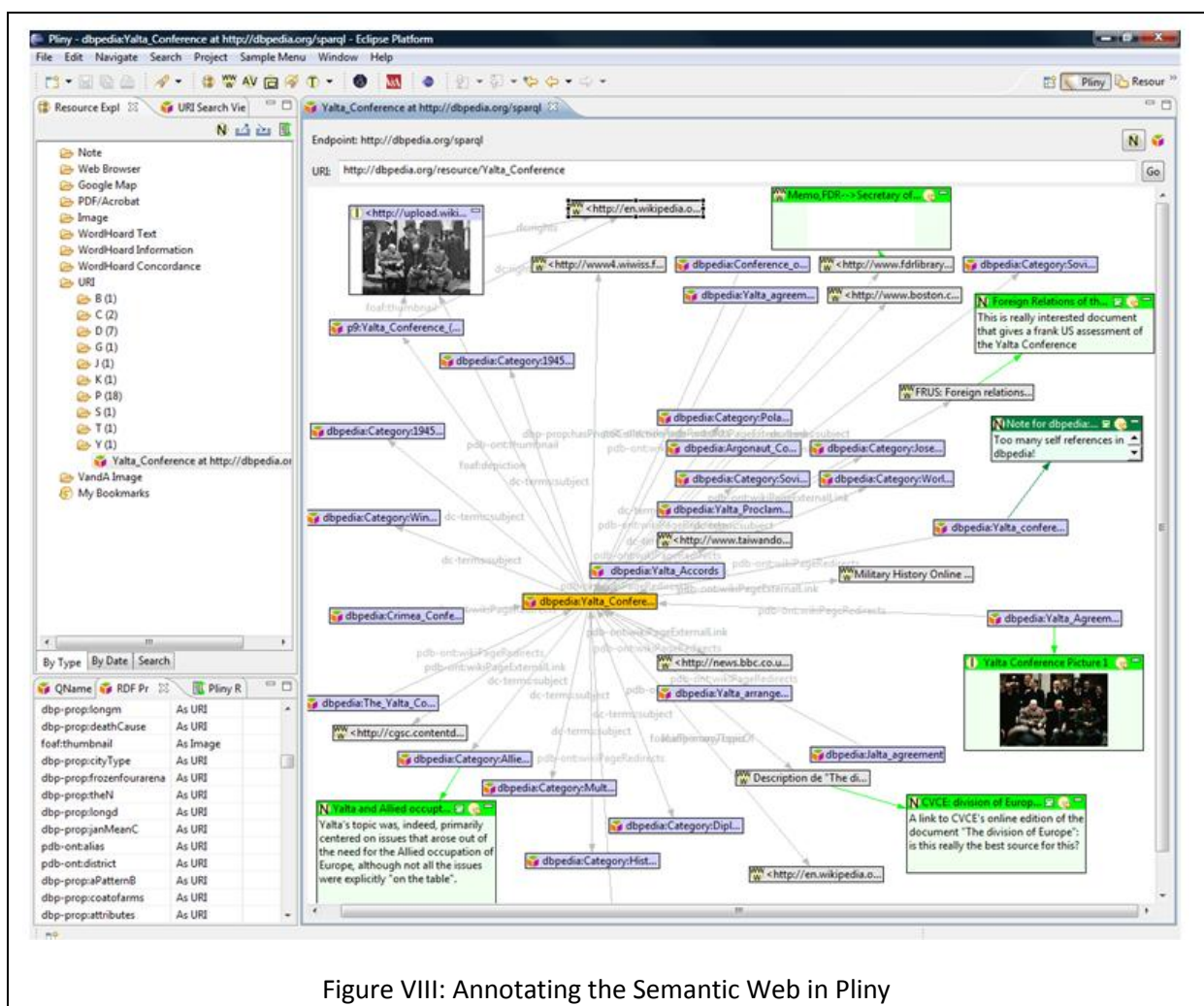


Figure VIII: Annotating the Semantic Web in Pliny

The display can be directed to any RDF repository that supports SPARQL, and here it shows a part of *DBPedia*'s web of linked data – here centred around *DBPedia*'s URI for the 2nd World War's Yalta conference (shown on the screen in orange) – as a graph. Most of the graph is a representation of the RDF triples that *DBPedia* holds and connects to the central Yalta Conference

URI, and 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 file of a small photograph taken during the conference. We hope that the graphical presentation suggests how one might present this web of objects.

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 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 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 a kind of commentary that the *Pliny* user has added on the linked data provided by *DBPedia*; as personal annotations added to this corner of the semantic web.

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

RDF, 2D Space and "scholarship as process"

Pliny provides 2D space as an affordance that is meant to accommodate fuzziness and ambiguity within the scholarship process. Space provides an open, flexible, structure for exploring the organising of materials through the use of proximity between objects, and the informal use of spatial areas into which to place things. Fig. IV (b) shows this kind of use of space most clearly – by putting objects close together the user tentatively suggests some sort of connection without needing to pin down too much yet what this connection might be. Although the structuring affordances shown in IV (c), (d) and (e) add further richness to the data recorded in this space, and reduce the need for 2D's particular expressive characteristics, even in the final image, the expressive elements that 2D space is providing has not been fully eliminated: the layout of the objects facilitates a kind of visual expressiveness to the presentation that, say, presenting the same associations hierarchically, would not possess.

However, 2D space, as an expression of the expressiveness of relations between ideas, has almost entirely disappeared out of the exported RDF materials we described earlier. Of course there remains 2D co-ordinate data in the definition of the targets of annotations for digital images or PDF pages through the OAC's *fragmentSelector* construct, because the target is, in fact, given in terms of its position on the area where the underlying object being annotated is presented. However, other 2D information, either that manages the placement of the annotation notes on the page, or in the placement of materials in *Pliny* objects such as *uses of space...* simply is not exported.

It could have been exported as RDF, of course, but its significance as RDF data would be rather limited: there are no mechanisms in, say, the SPARQL standard for finding objects that are – in a true 2D sense – close to other objects. Indeed, one could argue that the whole idea of *proximity* in the sense it is used in 2D is to some extent incompatible with standard RDF expressiveness. Thus, using the RDF exporter to export materials from Fig. IV (b) means that important information about

the relationships between the items is simply lost. However, we have already noted that 2D proximity becomes a less prominent element of expressiveness in IV (c), (d) and (e), so more of the "information" in the figure is preserved by our RDF exporter the further one travels along the process presented in Fig. IV. Indeed, it seems that the materials in Fig. IV (e) are just about as well represented as an RDF-expressed hierarchical structure which, is far less than full 2D.

Conclusions

So, what conclusions can be drawn from all this?

First, semantic annotation, as a way to connect reading-as-scholarship with the structures of the semantic web, provides only one perspective on how the Semantic Web might connect with humanities scholarship. Although not conceived of in Semantic Web terms originally, *Pliny* provides a model for formalising a part of traditional scholarship that is centred on note-taking and concept development. The formal model in which *Pliny* materials are stored is largely compatible with RDF data structures: in particular RDF's graph representation. This Graph part of *Pliny*, by providing a link to the semantic web, allows us to think in a richer way about the possible interaction between scholarship and the Semantic Web than semantic annotation does.

Pliny's model, by focusing on supporting the *process* of scholarship rather than the representation of its end results, provides a framework which allows us to engage with the question 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 real world of scholarship with the Semantic Web – to see how scholarship can be part of "thinking in the graph" as Gradmann put it. Semantic annotation, with its assumptions about links to predefined formal systems, doesn't capture the key work of humanities scholarship: the processes of 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.

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 shown 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, one can see that the very nature of 2D proximity does not fit entirely comfortably with RDF and the Semantic Web. Position data can be readily exported into RDF, but once there it cannot be used to capture the proximity nature of a relationship between two items. Interestingly, however, *Pliny's* approach is one in which as the ideas represented mature and become enriched, the 2D aspect of the representation becomes less prominent. Thus, for items which have been fully structured using *Pliny's* facilities, apparently not much information is lost in the export. However, for less fully formed items – where 2D proximity is still playing a significant semantic role – this might be more of an issue.

Pliny's approach, with its particular way to assist a scholar through a process of developing 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.

Admittedly, the fact that overwhelmingly scholarship is published as narrative text means that what happens in the process of taking a structure like that held in Pliny and turning it into prose text is missing entirely from this picture, and there are lots of reasons for thinking that this step is an important one for the evolution and emergence of scholarly meaning. Nonetheless, one can imagine the crafting of public Pliny objects around a set of ideas that comes at least close to capturing the key ideas that the scholarly article – as an endpoint for the expression of these ideas – represents. There may well be better ways to deal with the process of building and expressing an interpretation than *Pliny's* approach with combined graphs and 2D space. However, if we are going to find them, we need to do some serious work in this area.

References

- ARF** (1996-2013). *Alzheimer Research Forum*. Website at <http://www.alzforum.org/> (accessed 13 November 2013).
- Berners-Lee, T., Hendler, J. and Lassila, O.** (2001). The Semantic Web. *Scientific American*. 284.5. pp. 34-43.
- Blair, A.** (2003). Reading Strategies for Coping with Information Overload ca. 1550-1700. *Journal of the History of Ideas*. 64.1 (2003) pp 11-28.
- Blair, A.** (2004). Note Taking as an Art of Transmission. *Critical Inquiry* 31.1. University of Chicago. pp. 85-107.
- Bodenhamer, D. J.** (2008), History and GIS: Implications for the Discipline. In Knowles, A. K. (ed.) (2008). *Placing History: How Maps, Spatial Data, and the GIS Are Changing Historical Scholarship*. Redlands, CA: ESRI Press. pp. 219-234.
- Bradley, J.**(2003). Finding a Middle Ground between 'Determinism' and 'Aesthetic Indeterminacy': a Model for Text Analysis Tools. *Literary and Linguistic Computing*. 18.1. pp. 185-207.
- Bradley, J.** (2008). Thinking about Interpretation: *Pliny* and Scholarship in the Humanities. *Literary and Linguistic Computing* 2008 23.3 pp. 263-79. doi: 10.1093/lc/fqn021.
- Bradley, J.** (2009). Can Pliny be one of the muses? How Pliny could support scholarly writing. Peer reviewed paper 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, J.** (2012), Towards a richer sense of Digital Annotation: Moving beyond a 'media' orientation of the annotation of digital objects. *Digital Humanities Quarterly*. 2012.6.2. Online at <http://www.digitalhumanities.org/dhq/vol/6/2/000121/000121.html> (accessed 13 November 2013).

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

EATS (2013). *Read the Docs: EATS*. Website at <https://readthedocs.org/projects/eats/>; last accessed 13 November, 2013.

Gradmann, S. (2012). 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, M., Morbidoni, C., Nucci, M., Fonda, S. and Piazza, F. (2013). Pundit: augmenting web contents with semantics. *Literary and Linguistic Computing*. Vol 28 No 4. pp. 640-659.

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

Jackson, L. S. (2010). *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).

Klyne, G and Carroll, J. (2004). Resource Description Framework (RDF): Concepts and Abstract Syntax. World Wide Web Consortium (W3C). Online at <http://www.w3.org/TR/rdf-concepts/> (accessed 13 November 2013).

Lavignino, J. (1997). Reading, Scholarship, and Hypertext Editions. *The Journal of Electronic Publishing*, Sept 1997 Vol 3 No 1.

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

Morbidoni, C. and Nucci, M. (2012). *SWickyNotes: Sticky Web Notes with Semantics*. Website at <http://dbin.org/swickynotes/> (accessed 13 November 2013).

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

Norrish, J. and Stevenson, A. (2008). Topic Maps and Entity Authority Records: an Effective Cyber Infrastructure for Digital Humanities. *DH2008 Book of Abstracts*. Oulu, Finland: University of Oulu. pp. 166-168.

OAC (2009). *Open Annotation Collaboration*. Website at <http://www.openannotation.org/> (accessed 13 November 2013).

Ontotext (2013). *Semantic Annotation*. Web page at <http://www.ontotext.com/kim/semantic-annotation> (accessed 13 November 2013).

Palmer, C., Tefteau, L.C. and Pirmann, C.M. (2009). *Scholarly information practices in the online environment: themes from the literature and implications for library service development*. Technical

report, OCLC Research. See <http://www.oclc.org/research/publications/library/2009/2009-02.pdf> (accessed 13 November 2013).

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

Pundit (2013). *Pundit Website*. Online at <http://www.thepund.it/> (accessed 13 November 2013).

Rüsen, J. (1987). Historical Narration: Foundation, Types, Reason. *History and Theory* Vol 26 No 4. pp. 87-97.

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

Sansone, S.-A., Rocca-Serra, P., Field, D., Maguire, E., Taylor, C., Hofmann, O., Fang, H. et al. (2012). Toward interoperable bioscience data. *Nature Genetics*. 44,2 (February 2012). pp. 121-126. Online at <http://www.nature.com/ng/journal/v44/n2/full/ng.1054.html> (accessed 13 November 2013).

Unsworth, J. (2000). 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).

Unsworth, J. (2001). "Knowledge Representation in Humanities Computing". Online at <http://people.lis.illinois.edu/~unsworth/KR/KRinHC.html> (accessed 13 November 2013).