

Thanks to Jarom McDonald (session chair) Attention of DH: the Semantic Web and Linked data



•The Semantic Web is one of the subdomains of the computer science field of **Knowledge Representation** 

•We have had **enthusiastic statements from no less than John Unsworth** about the relevance of KR to the humanities, through what was called back then Humanities Computing

•Clearly some potential recognised then



•Furthermore, in the context of my department, we have had much positive experience with Knowledge Representation, in the form of database Structured Data

•I am "mister structured data" at DDH at King's, and have been responsible for many collaborative projects where structured data has been a key element. Here is a selection of them: ...

•For all these projects, and for many more we have done at DDH, **important insights and understanding** of our **colleagues from history, from classics, from music, and art history** were successfully and usefully captured in highly structured terms.

•Clearly, then, at least in these cases **important aspects of humanities scholarship were being represented** by the structures built for these projects. In almost every case, it has been evident that our discipline partners could see **key ideas that they were interested in in this data made evident in new ways** they had not originally expected, and available for new kinds of exploration.



... some potential recognised then!

•Although we have had, then, success with various humanities-based KR projects, this **approach is not seen as fitting with the approaches of most humanist scholars.** 

•KR technologies impose a highly formal representation of the material it presents. For the semantic web, the formal representation of knowledge is what mathematicians call a "graph".

•Here is a small graph – taken from **the CIDOC-CRM examples** – is shown here and charts the relationship between the players, the documents, and the event of the Yalta conference at the end of World War II.

•The question, then, is as Stefan Gradmann stated it in his presentation at WWW2012 in Lyon France: "Thinking in the graph: will Digital Humanists ever do so?"

•Indeed, we think an even more important question must be: "Thinking in the graph: will Humanists (more generally) ever do so?"

•One approach that, we believe, is trying to fit the Semantic Web with humanities scholarship is being taken under the name "**semantic** annotation".

•While semantic annotation might suit certain kinds of humanities research work, we think **we can do better** than this, and in this talk will **present a different approach** that suggests a richer kind of interaction between humanities scholarship and the semantic web.



For this talk I will be following this plan:

•Introduce semantic annotation as one way to link humanities scholarship to the semantic web

•Suggest **why semantic annotation unfortunately misses out** on much of what humanities scholarship is really all about

•Show a different approach to introducing formal structure into traditional humanities research

•Explore how this formal structure might provide a richer way to connect scholarship to the semantic web.



•So that is semantic annotation?

•Unlike conventional annotation, which is usually thought of as connecting a small text to a spot in a larger text, semantic annotation links a section of text into some sort of formal structure that captures the semantics of the text. Here, in this image – borrowed from OntoText – we see a bit of text linked to a structure that represents some places referenced in the text, and identifies "XYZ" as a company.

•Semantic Annotation activities are predicated upon the idea that there exists a formal representation of a body of relevant knowledge (here, places, companies) to link to.



•I was first aware of substantial work on semantic annotation in the Life Sciences. One of the influential pieces of software for them is the SWAN annotation tool, shown here in operation.

•The user, while reading an article on Alzeimer's Disease in the left area spots a reference to a particular gene. She can use **the area on the right** to locate the digital entity for that gene in one of the **life science ontologies** that emerged to formally model some part of recent research, and establish a link from the text in the form of an annotation to it.

•Since the annotation is to an entity in a **formal structure representing knowledge about, say, genes**, we can characterise this kind of annotation as "semantic". The link **enriches the formal structure** captured in the ontology **by connecting it to scholarly texts**.

•This is an example of Linked Data at work!

## Wikipedia: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."

•This semantic annotation activity – linking some text to a formal model of understanding of the materials the text is talking about – is **possible in the Life Sciences because the field already has a large number of formal ontologies** that can be linked to representing a broad range of related fields of research. 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.

8



•So, if semantic annotation is flourishing in the life sciences, **is there any hope for it in the humanities?** 

•We at DDH have carried out semantic annotation with a simpler but similar environment built around **Jamie Norrish's Entity Authority Tool Set** – EATS, which allows us to formally identify **entities** (people, places, etc) that turn up in our projects and then link them to TEI marked-up text. Here we see EATS at work in our **Schenker** project – the famous 20<sup>th</sup> century music theorist –being used to facilitate a reference to the composer Beethoven in Schenker's notes to the entity representing Beethoven in the project's EATS entity repository.



•Although the EATS entity management environment does not structure its entities as rigorously as could be done with the use of a Semantic Web Ontology in the way that SWAN does for the Life Sciences, we know of two environments that seem to be aimed at humanists and that provide support for exactly this: evidently bringing humanists even closer to the kind of semantic annotation of the kind that is active in the Life Sciences already.

•Pundit provides a browser-based environment that is aimed at "augmenting web pages with semantically structured annotations". It places itself in the "Linked Data" world by providing an environment which, it claims, allows one to "easily turn web documents into a semantic knowledge network by pulling from and enriching the Web of Data".

Pun	dit in o	pe	ration	
Saving annotation 9d776 Being annotation 9d776160 Creating a new annotation	1150 (New 1882)(1994)(1711)		Next 3. C Provious	O Help My Kenie ★ Avoutations ★ Christian Municipani O ▼ 4           Destination
16) Es wird die Dezim eingeführt. Der Lernend Beleh: in der Rehee wehterzählen bis 'Hat andere Zahlwöhrt h Drag, Glockenschla Drag is Gruppen zu Volkaschulen winklich get 7 In (7) bringt B Gegenstein etwa nach eingen Week Gegenstand bringt, ohne @ er habe sich nun die Wolfchweiten, ahnlich die Bezahung erses Behefts Wir können allso unter von dem Fall, in dem es i	Comment and tags Comment Linguage on Accomment here - 3 Tags C bitstrik Linguage game C bitstrik	• • • • • •	- Dividi province) PREVIEW Language Game Missing description Types : lane More Info	In a difference does not lis, however, in the act of pointing and pronouncing the word or at maxing 7 accompanying 8, but in the reise which the demonstration (pointing & sigs in the whole training and in the use which in ansate of an interpractice of by means of this language. <sup>1</sup> Ore might hink that the difference occub be described in the different cases we point to different kinds of objects. But suppose I point with the different cases we point to different kinds of objects. But suppose I point with an object and the start we made normality different in the two cases. <sup>(1)</sup> Add "meaning" the different cases we point to different kinds of objects. But suppose I point with an object and the start we made normality different in the two cases. <sup>(2)</sup> Add "meaning" the different cases we point to different particularly tempts us to this and or object as different that his answer is correct. If on the other hand, we took for correct methed cases an ensing the ecolurity of the starts, we are not correct the start and in the mass to concentration on a startand on the colour part leads normality and can be submitted on the submitted on the colour point parts of the submitted is the new that can be the submitted on the colour point parts of the submitted is the new that concents on a startand nor the colour part and the submitted the interpret of the lang order, or lise in the act of put atter in the place indicated in the web difference on the interpret of the sing order, the part atter in the place indicated in the web that the difference of the sing order, we that an applicate base of the word "there", but shade the submitted the sing order on the colour parts and the base of the word "there", but shade that shade the submitted the sing of the sing order is and pronounce the word "there". But should have parts that the high equite a place base that methed the the pointing gesture in this case is part of the practice of atom methed.

Pundit supports conventional textual annotation, but here we see it supporting a semantic one. The text being annotated at the time this screen was captured is Wittenstein's *Philosophical Investigations*— we can see previous annotations to the text represented by the little three-dot symbols that have been scattered through the text ... and we can see here the panel that turns up when one wants to add a link to a linked data-like formal representation of Wittenstein's idea of the "language game" -- apparently taken from Wikipedia/DBPedia's large collection of URIs.



Another piece of work that we are quite impressed with supports semantic annotation and goes by the odd name of "SWickyNotes": for "Sticky Web Notes with Semantics".



•Here we see SWickyNotes in operation. In it the user is identifying a fragment of text – the folktale Hansel and Gretel – as an example of one of the concept categories defined in one of the ontologies available for it: here "pathos". We have expanded SWickyNote's "New Note" screen and shows where a user records that the selected bit of text is an example of the Rhetorical devince of "pathos". You can see the available subjects showing in the bottom left area: including the Rhetorical device of "pathos".

•We think both Pundit's and SWickyNotes's interface for Semantic annotation are excellent examples of semantic annotation tools for a humanities context.



•One of the important things one can observe, however, from the kinds of semantic annotation shown in all the systems I have briefly shown you here – SWAN, Pundit and SWickyNotes – is that the kind of activity that they support feels like a kind of, let us say, "junior" research activity. By linking text to predefined ontologies created by others, one is limited to the kind of things that one can say about the text. Doing this is doubtless useful work and enriches texts in ways that can be exploited by the digital environment – exactly in the way envisioned by the Semantic Web. However, it is "junior" in the sense that one can imagine getting this kind of semantic annotation done in a large textual project by giving it to research assistants to do under the direction of a more senior researcher.

•Most of the time, in fact, semantic annotation does not represent the kind of work that humanist scholars do. OK, so what do they do instead?



As we all know, the primary product of scholarly research is almost always this kind of thing – books, chapters, articles – narratives of various forms, here represented relatively arbitrarily by a scholarly article written by Joan Holmer which appeared in the *Shakespeare Quarterly*. Holmer is publishing a new view on aspects of Shakespeare's *Romeo and Juliet* that show evidence of influence from Vincentio Saviolo's fencing manual.

•At first glance, this kind of research output – the book or article – which, I would think, represents the preferred output for, let us say, 90% of existing humanists – does not seem to be compatible with either textual annotation or semantic annotation approaches. If we read Holmes's article, we can see **that there** *is* **structure here**, 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 the text. But it certainly is not presented in the highly structured forms of knowledge representation.



Hence, the characteristic of humanities research as being "about writing books". For almost all the humanities community, the product of their research is text in the forms of books, chapters, articles, etc etc. The quotation we see here from an article by Jörn Rüsen apparently made by Hayden White – the prominent historiographer – about history: that is manifestly presented in the form of textual narrative, would be a view echoed by many in the other humanities disciplines. This view is supported by a number of other historiographers.



•Now, we have all have heard the claims that the product of humanities research is articles and books presented by the DH community as a kind of old-guard position: those old guys are protecting their turf – and probably some of this is in fact true. However, this ready dismissal doesn't represent the whole story. We see here a quote from an article by the prominent American historian David Bodenhamer who has, as it turns out, embraced data-technologies in the form of GIS systems to support his research – explaining, even so, why the character of narrative particular well suits the needs of historians: it is exactly the character of imprecision of words that, in the hands of a skilled writer, can capture the complexity and ambiguity of doing history. Surely a similar statement would be made by the broader community of humanist scholars as well.

•With the strong preference, for most scholarship, of presenting it in narrative as we see here, 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 expressed in terms of largely formally unstructured prose narrative text with its ability to deal with ambiguity and contradiction, with on the other hand the semantic web with its high degree of formal structuring of its material is strongly non-narrative in nature?



•First off, since the humanities researcher wants to say something new about the sources he or she is working with, they want to write about things that do not already have a formal structure existing for them. Holmer, like all scholars, is trying to develop her *own* representation of their material that is different from that currently established within her discipline: perhaps drawing on some existing ideas, but also extending or perhaps even more fundamentally breaking with it.

•Secondly, we must remember that this article represents a **product of the research**, rather than the research itself. It is the process that *got* to the article that is the research. What is this research process like?

•The first thing to think about here is that perhaps by the time the article is ready to be published Holmer has developed her ideas sufficiently that they could be represented clearly in text, and perhaps many of them could even be formulated in the formal language of the semantic web. However, the ideas that the articles contain probably didn't appear, fully formed, in Holmer's head as she started the research that resulted in this article. They emerged after substantial engagement with the materials she was working with. Before that it is likely that her ideas were still only partially formed. There is, thus, a process here, and I will be calling this process "interpretation building".

•But does this process result in materials that are at all compatible with the Semantic Web? Up to now most work in the SW has focused on the representation of highly structured fields of knowledge. We have not seen RDF and the rest of the Semantic Web toolkit providing adequate mechanisms to support the *development* of new concepts and ideas *before* they are formally clear. Indeed, they would appear to bring formalism to bear too soon.



•Our work on the Pliny project began in 2004 and soon came to be a project that worked on how to build a tool that could support the process of humanities scholarship. 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 plus other thoughts about the representation of ideas, of which annotation is only a starting point, into the process of building a tool that supported a full range of humanities scholarship.

•Pliny has represented an attempt to achieve a balance between conflicting needs:

•It structures the act of notetaking, annotation, and note management

•It provides – through its use of 2 dimensional space – with a way to cope with ambiguity and vagueness

•It supports its user in the task of moving from initial partly-formed ideas through to more formally structured ideas by providing formalisms when the user is ready to use them, but not imposing them too early.



•Pliny's design was influenced by thinking about a place for computing in scholarly research that goes back to one of our responses to the development of the TACT text analysis system by one of us in the late 80's and 90's, and the sense that the text analysis orientation of TACT didn't really address the needs of very many humanists.

•The ideas about what could be useful to humanities researchers were better crystallised when one of us discovered the work of Brockman, Neumann Palmer and Tidline in their 2001 CLIR report entitled "Scholarly Work in the Humanities". Here one could see the key element of reading in scholarship, with the significance of notetaking from the reading as a element in the process of doing humanities scholarship. The mere *taking* of notes and/or annotation in the first place was not, by itself, the central place of these notes in doing the research. Instead, the notes became key tools to assist in the gradual development of new ideas that would become the primary result of the research work.



It was here that we could see interpretation building, as research, as a process. Once we began to think of interpretation is a process, with perhaps a clear conception emerging at the end, but only at the end, the question begins to reveal itself as being not only about formal models for the interpretation when it is done, but also about how to model the process so that technology could help someone *develop* it: what should the user interface and the formal structure behind it be like that helps a researcher develop their interpretation? For the researcher, even if in the end there will be a structure emerge that captures important aspects of their ideas, during its development much of this work is vague, incompletely defined, and "pre-ontological" indeed, it is important to note that in almost any substantive humanities research, one expects to start out with only a vague sense of the issues one is interested in, and only after a long period of time will some degree of clarity emerge. Indeed, even when the ideas are publishable in the form that Holmer presents them in her article, they are likely not to be fully formally expressible using the formalisms of, say, knowledge representation.



•The sense of interpretation development as a process that produces its results over time is caught by this quote from John Lavagnino when he reflects on the place of reading in the interpretation process. Here he notes that reading was not merely a data collection exercise (like the semantic annotation processes we have seen earlier are). Instead, he claimed that it triggered reactions in the reader that subsequently (note the use of the word) one could seek to identify or explain. We see here Lavagnino placing reading and notetaking as only the beginning of a larger process that, if it is useful, it has to support.

•In this light we can see a problem with much of the work on annotation at present, when applied to much scholarship in the humanities. Simple digital textual annotation, by itself, 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 here, that is all one can do. It leaves the user there – at the beginning of a process. On the other hand, semantic annotation – of the kind done by the tools we talked about earlier – brings the formal structure in too soon – trying to apply an approach suitable for a predefined, formal, interpretive model, to the beginning of the process before a model is available.



Early in Pliny's development – in 2005 – we developed this schematic to graphically represent the three phases of humanities research: reading, developing a new interpretation, and then writing about it. Pliny software was an attempt to provide a tool to support not only the annotation and notetaking activity shown here on the left, but to also support the development of new ideas that might be stimulated by these personal notes in the "personal space" shown here in the middle, 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.



•The place of the notes, then, as a tool for further thought, was made rather prominent in Pliny; recognising (one hoped) a central role for them in what Ann Blair here calls the "central but often hidden phase in the transmission of knowledge".

•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 could it best deliver them to the user to support the user's engagement with them in this way?



•Thinking of the three phase diagram I showed you a moment ago, using the ideas around notes, note taking and note management to give a structural perspective transforms that figure into this structurally suggestive representation of the place for notes in scholarly interpretation development.

•To the left is still the area showing the reading of materials related to the research – both primary and secondary literature is likely to contribute to the work. The diagram shows small boxes – the annotations created by the researcher – linked to portions of texts in the documents. Initially, at least – attending to John L'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 hand-written text that captures the "reactions" (using L's word) that one hopes to subsequently describe or explain by developing a framework for them.

•The middle area corresponds to the development of interpretation phase of the research. By themselves, the notes added to the texts in the left area represent only the beginnings of research, but they allow the researcher to gradually become aware of groupings of these notes into relevant topics or concepts that interest him/her. Here we see, in the middle area the notes being organised under broad categories or topics (only two shown here).

•Finally, when the time is right (and probably after more than 2 concepts have been recognised), the researcher draws on the ideas he has formed in the interpretation phase to put together papers that present them.



•Here, then, we see Pliny's interface for annotation of documents in the form of PDF files. Someone has been adding notes to an article – here McCarty's 2008 article "What's going on?"

•Note the way that the annotations are displayed. The intent was to focus on traditional annotation. Hence, annotations in Pliny appear on top of a printed text. In the same way that all the hand-written annotations on a printed page are immediately visible when the page is looked at – nothing needs to be clicked on to see their contents. In the same way that the annotator can make entirely free use of the 2-D page to hold his/her annotations, one can use the 2-D space of the page here to hold your annotations as well. In the same way that one can use colour to differentiate your annotations, you can use colour in Pliny to differentiate them here.



•But Pliny is not only about annotating things, which was, you recall, represented only the left "reading" side of our diagram. How does Pliny support the central phase of research: the development of an interpretation?

•Here we think of this development is a kind of gradual increase in formalism: the Pliny user adds structure as the ideas become clearer and more structured themselves. Let's take a moment to see what I mean by this by examining the process one could use in Pliny to create a particular item about a topic called "uses of space for study". I have deliberatively named the stages one goes through in working with this idea to echo the language of "Scholarly Primitives" as presented by John Unsworth at King's College London 2001.

•The first step is assembling. One begins by creating a holder for the information about the topic. We see it here named as "uses of space for study", and with a brief description of the idea entered on the left area.

•The main place where the work is done in in the right area: a 2D space that Pliny provides to allow us to organise materials. We start off e by assembling references to things that relate to the topic we are interested in -4 images that show space being used in different ways, and a reference to a note on another topic we have already created called "Visualisation", which seems to be related to this one. As of yet, all we know is that these items feel as if they are connected to the idea of "uses of space for study".



•Having now assembled a few items we begin to notice some similarities in the use of 2D space in the floorplan and the Simweb sample, and a what feels like a contrasting similarity between the Vico and the Benardete's image. At first, we take advantage of the possibility of proximity in the 2D space to organise them in this way, placing those with a similarity of interest to us close together. The "visualising" object still seems relevant, but separate from the ideas we are developing about the images themselves, so it has been moved to the bottom right.

•Note the importance of the 2D space for this task and the particular expressive affordances it offers. We often see graphs presented as laid out on a flat surface and it is easy to confuse a graph as a 2D object. It isn't really – it is a kind of, let us say, 1.5 dimension object where the structure between items is really represented by the links between them. Here, no explicit links are actually present, and there is a much more subtle and perhaps usefully ambiguous mechanism available here in 2D space for established relationships between items in terms of proximity. In fact, there is good evidence in the description of the working practices of scholars in dealing with their notes that this characteristic of 2D space is usefully a part of their work at some stage. For example, we find references to researchers taking their note cards and trying to find relationships in them by building small stacks of cards and shuffling them about on a large flat surface to help them explore possible relationships between them.



Proximity has helped us to explore possible relationships between items of interest. At some point the relationships become clear enough that we are ready to give them a name. They represent two rather different kinds of use of space – so we ask Pliny to put these images into two groups, and we name the groups accordingly – with the images that seem to exemplify the two categories contained in their respective groups. The naming and the grouping adds more structure to this space.



Having now discovered these two kinds of uses of space we add a few notes that record our thoughts about them. The "Simweb dimension 2 outliers" note is actually a note that was created earlier. But its comment suggests to us a particular aspect of the significance of the 2D space here – so we show this note in this space too.



Now that we have collected and organised some materials we note that there are several kinds of connection between the topics and the things they contain. Pliny allows us to assign a type to these connections which shows up as different colours. You see my current set of "types" in the bottom left corner. We assign 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 these past few slides we have seen one way to use Pliny's 2D space metaphor to help us develop new ideas, and develop and preserve a richer understanding of a particular topic to . However, the 2D space can be used in other ways too. Here is a different approach, where the interface has being used to allow a user to create a concept map representation of some ideas. Each of the items in the diagram can hold references to relevant materials, and/or notes about the ideas – exactly as our "uses" space ended up doing.



•In summary, then, we see the process of developing an interpretation in Pliny.

•One starts off by assembling materials that one wishes to work with

•Pliny provides annotation so that you can record your responses to these materials.

•Pliny provides 2D spaces where one can organise your notes and other objects to discover relationships between them that will eventually lead to a clear formulation of a model for your materials.

•As concepts become clear you can use Pliny's grouping mechanisms in conjunction with its sense of 2D space to identify, name and organise your ideas.

•Pliny's notes, among other mechanisms, provides a way for you to add comments to the structures you have become interested in – allowing you to enrich the structure you have stored in Pliny

•Finally, Pliny allows you to attach assertions about the relationships between objects that you have capture in your concepts.

•As I hope you can see by this point, these steps in a research process move the user from preliminary reactions to texts in the form of annotations and notes to more fully formed ones – and within Pliny from less structure to more structure. Not that a researcher will necessary be able to push all his or her ideas through to be fully structured. Pliny accommodates a mix of highly structured areas with less structured ones, to recognise this.



•Returning now to thinking about how scholarship can fit with the semantic web, we look at Pliny's way of modelling the process of scholarly research. We have noted that Pliny encourages a process that introduces structure for representing the ideas the researcher develops through his/her work. Direct Semantic Annotation – discussed earlier, doesn't accommodate the complexity of actual original scholarship because it forces structure to represent the concepts right away.

•So, if we try to connect Pliny to the formal structured world of the semantic web, how does it do? If we can see how the two worlds connect together we have, at least from the perspective of Pliny, a model for the formal linking of one model for humanities scholarship to the semantic web.

• There are two quite obvious questions:

(a) Linking out: How can the data that we have shown as accumulated inside of Pliny be transformed into RDF: the language of the Semantic web

(b) Linking in: 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 "Linking 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. Here we see an annotation added to a page of the article "What is Originality in the Humanities". The annotator has an issue with the claim made in the text about the "third contribution".

•Pliny's annotation mechanism – perhaps not surprisingly – maps quite well onto the notation of the formal annotation ontology that I mentioned earlier: here we see it with the OAC dialect. Let me take a moment to talk you though selected bits of the corresponding RDF "turtle" representation shown here at the top of the screen.

- 1. First, the "jb:" prefix refers to me. I have created a Pliny type of "IssueWith" which the first bit of RDF shows is a Pliny "type", and also a kind of Annotation from the perspective of the OAC.
- The second block of triples shows me using the IssueWith object for an annotation on the text, and I use the OAC's predicates "hasBody" and "hasTarget" to connect the material I have collected together as an issue to the spot in this particular page.
- 3. The third block establishes the annotation target area as an OAC Target, and establishes it as a part of the PDF document
- 4. The fourth block defines the PDF file that contains Guetzkow's article, provides a title for this file, and provides a URL that points to the location of the article in the WWW.
- 5. The fifth block defines the Pliny note object that contains the materials contained in the annotation (not shown here in this little RDF snippet), with its title.

•We don't have time to discuss this representation in more detail, but suffice it to say that the OAC representation of annotations works quite well – with a few extensions – with annotations as they operate in Pliny.



•So, that is the annotation part of Pliny dealt with – how about those parts of Pliny that support the development of an interpretation?

•Well, first of all, it turns out that the data structure Pliny uses behind these displays can be thought of – to a large extent at least – as a graph of nodes with typed links. Here is our "uses of space for study" screen again, with Pliny's (admittedly rather crude) display of the graph structure that it implies and that it, in turn, links to. There is no time to explain here in much detail, but note that the graph with its linked types maps quite comfortably into RDF's "subject predicate object" representation.



•.. and indeed, here we see a partial representation of this data as a set of RDF triples, as Pliny would export them. 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. However, the connections between the objects that are presented can, in the most part, be captured by RDF and, thereby, exported to the linked data world.

•One thing to note: although we have made much of the importance of 2D space in Pliny – the actual 2D information that Pliny maintains about placement of object is not represented in these RDF triples. For this item, which has 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 playing a significant semantic role – this might be more of a problem.



•So much, then for the linking out from Pliny to the semantic web. What happens when we think about the issues around linking in: bringing aspects of the linked data/semantic web world into Pliny's workspace. We have explored this idea by creating a rudimentary extension to Pliny in the form of a "plugin" that allowed Semantic Web or Linked Data URIs to appear as Pliny resources.

•When we took up Pliny's way of thinking about interpretation we could see two rather different kinds of linking activities. One would be very similar to the model of semantic annotation that we saw earlier in this talk, but the other would be based on the idea of annotating parts of the semantic 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 understanding of annotation of images being used as a link to RDF representation of concepts from DBPedia. The image, the frontispiece from Vico's *New Science*, has been annotated with references to concepts that Vico will refer to in the text of his book. The annotated objects identifying 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 in Pliny 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 identified inside of Pliny, but not connected to the broader outside world of the Semantic Web as URIs.

•The kind of annotation shown by these three links is, in fact, very similar to the kind of semantic annotation as conceived of in the semantic annotation tools we spoke about earlier in this talk. Within Pliny they can co-exist with other kinds of annotations – as is shown here.



•The second use of the Pliny plugin is, we think, a more radical engagement with linked data and the semantic web. Pliny's display is a little crude still since the software that implements it is still at the prototype stage – but we hope it is suggestive of what we could mean by making the semantic web an object of study in the Pliny sense. The prototype is still, we're afraid, a kind of work in progress.

•The display shows a part of DBPedia's web of linked data – here centered around DBPedia's URI for the 2nd World War's Yalta conference – as a graph. Most of the objects on the graph are representations of the RDF triples that DBPedia holds surrounding the Yalta conference. Most of them in the little boxes are simply URIs that identify related DBPedia objects, but a few are URLs to web pages, and one is a URL to a small photograph taken during the conference. They are all from the DBPedia triple store and are here laid out in a graphical way that shows the links between them. We hope that the graphical presentation makes it a little easier to visualise this web of objects – although, to be frank, there is still work to be done to make this visualisation work better, I think.

•What is interesting here, from a Pliny point of view, is the objects shown 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. We see here several notes, a link to a web page, and a link to an image of the conference 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: see here, for example, the observation made by the Pliny user that the FRUS document referenced by DBPedia is an interesting, and frank, assessment of the Yalta conference made by the US government.

•Note the significance of this display. In the same way that Pliny allows the user to personally annotate a web page, an image, or a PDF file with responses, at the moment he has them, of his study of these objects, and then use these notes later in his deliberations – one can here annotate the semantic web with personal responses to parts of it, and fit these reactions into later thinking.

•The intermingling SW URIs with Pliny objects allows for the creation of a personal space between objects being annotated and the public Semantic Web.

## Conclusions

- The highly structured semantic web has the potential to link richly with traditional scholarship.
- Semantic Annotation provides only a partial answer to how humanities scholarship might connect to the semantic web.
- Pliny provides a model for thinking about formalisation of traditional scholarship (in the same way that WP formalises a part of writing).
- Pliny's data model combines graph structure with true 2D space: balancing a need of formality with ambiguity.
- Graph part of Pliny, at least, links clearly to the Semantic Web and enables a richer thinking about scholarship and connections to SW than annotation, either textual or semantic, does.
- In this way, perhaps Pliny proposes a richer model of interaction between scholarship and the SW, with a place for individual scholar's emerging understanding to be accommodated.

41

## •So, what conclusions can be drawn from all this?

•First, in our view, Semantic Annotation provides only one perspective on the place of the semantic web with humanities scholarship.

•Second, focusing now on Pliny itself: Pliny provides a model for formalising a part of traditional scholarship that shows ways in which traditional scholarship could benefit from computing.

•In Pliny's case, its data model 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. There may well be better ways to deal with the process of building an interpretation than Pliny's combined graphs and 2D space, and someone interested in connecting the semantic web to humanities scholarship could, perhaps, usefully engage themselves in thinking about what they might be.

•The 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 "direct annotation" does.

•In this way, Pliny provides a framework to allow us to engage with the question how interpretation building, as it is actually 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. 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 about the humanities is what really is going on.

•Pliny with its attempt to model the scholarly process provides one way to think about how intellectual work in the humanities might better fit with the broad world of open, linked data. There may be better ways still than what Pliny does – but if we are going to find them, we still need to do some serious work in this area.

•Thank you!