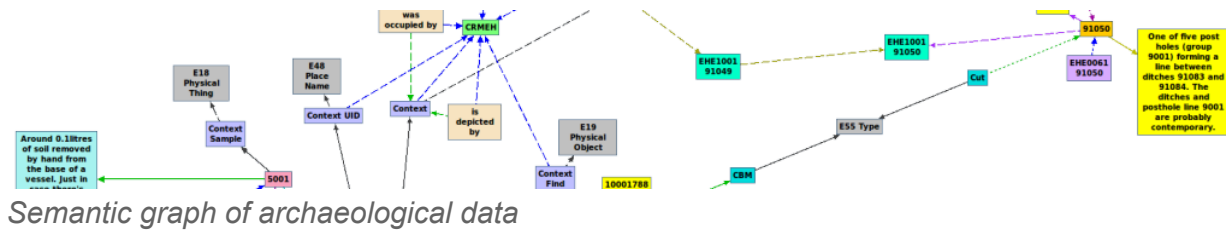


A DAY IN THE LIFE OF A POSTGRADUATE RESEARCH STUDENT

July 11, 2014 Paul Day of Archaeology 2014, Digital Archaeology Geographic Information Systems, Geosemantic Resources, relational database, research software, Spatial data infrastructure, Web Feature Services



Another year, another Day of Archaeology! Having moved on from the world of commercial archaeology, the source for my posts in 2011 ([here](#) and [here](#)) and 2012 ([here](#)), I'm now into the second year of my computer science PhD, investigating Geosemantic Resources for Archaeological Research (GSTAR) based in the [Hypermedia Research Unit](#) and [Geographic Information Systems Research Unit](#) at the University of South Wales. Today is a busy day (as usual!) wrapping up a pilot study for my PhD, writing up my Transfer Report (to move from MPhil to PhD proper) then later on, doing a bit of paid consultancy work to keep the wolf from the door. Whilst I am lucky to have a fully funded place, taking on a PhD later in life, as I have have, when you have a mortgage and two kids to support is challenging to say the least. But thoroughly rewarding nonetheless; I'm hopeful that my research can really make an impact on how we use digital heritage data within the historic environment sector and further afield but I guess we'll just have to wait and see...

THE FUSION OF SEMANTIC TECHNOLOGIES AND GISCIENCE

My literature review identified two strands of discourse within two distinct domains, each looking at how to deal with geospatial data within semantically enabled frameworks. To give you an idea of this, see the figure below which shows a Positive Stratigraphic Unit (ie some kind of layer, deposit or structure) modelled using the CRM-EH ontology and including spatial relationships. There's some more detailed description of this modelling [here](#) including various other classes used in archaeological excavations.

Firstly, within the web science arena, researchers are trying to integrate geospatial data directly within semantic resources. Geospatial data is structured using ontologies and held within [triplestores](#) alongside all other data, with geometries stored as Well Known Text (WKT) or Geographic Markup Language (GML). Both these formats are means of representing geometries using plain old text which can be embedded within semantic structures. This geosemantic data can then be accessed via 'endpoints' or APIs (ie web services) using an extension to the SPARQL language called [GeoSPARQL](#) which handles spatial objects and operators in addition to semantics. Such an approach also facilitates integration of other Linked Geospatial Data resources, such as those provided by the [Ordnance Survey](#).

Secondly, within the GIScience arena, researchers are looking at leveraging the existing capabilities of Geographic Information Systems (GIS) and Spatial Data Infrastructures (SDI) alongside semantic technologies such as triple-stores, SPARQL endpoints and the like. This way, the established highly tuned, highly efficient spatial services such as Web Feature Services (WFS) can do their thing, optimised for handling large amounts of complex geospatial data, whilst the newer semantic systems can contribute, providing the necessary semantic support.

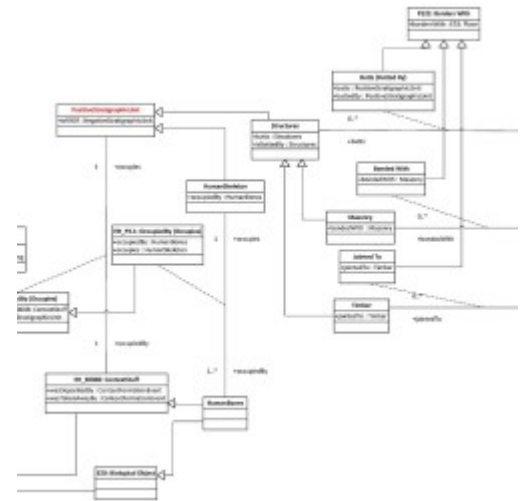
My pilot study has been looking at these approaches with a view to their application within archaeological resources and has resulted in a system I have used to investigate the pros and cons of different methods. It's a [Java](#) application, built using the [Eclipse IDE](#) and using [Maven](#) to handle the various external libraries I need to work with spatial and semantic data. These include the [Jena](#) framework (for semantic support), [GeoTools](#) (the Swiss army knife of geospatial Java programming), [GeoServer](#) (a lovely GIS server), some supporting libraries for handling CSV data. There are also some other bits and bobs such as the [Jetty](#) and [WebLogic](#) web servers to serve applications and provide access via [http](#). The data store I'm using for both geospatial and semantic data is [Oracle 12c](#), the very latest incarnation of this powerful application, which has good support for geospatial data, RDF, SPARQL and GeoSPARQL via it's updated *Spatial and Graph* component.

The data for this study comes from the Archaeology Data Service ([ADS](#)) and is one of the Channel Tunnel Rail Link ([CTRL](#)) projects which has useable spatial data available for download and was also one of the resources used in the [Stellar project](#) (published as [Linked Data at the ADS](#)), so the data is already available in RDF format compliant with the CRM-EH extension to the CIDOC CRM ontology. The availability of the semantic data allowed me to steam on ahead without having to spend time preparing semantic data from a typical relational structure, although this is fairly easy to do with the [Stellar Toolkit](#) (as I did on the [Colonisation of Britain project](#)).

The write up from this is forming a big chunk of my Transfer Report, currently in draft to be wrapped up and submitted imminently. A fuller write report will form a chapter or two in my final thesis. There's a bit more technical info over on my [GSTAR blog](#).

A BIT OF WORK ON THE SIDE...

A condition of my funding is a limitation on the amount of paid work I can do on the side. To be fair, this is entirely understandable; a PhD requires concentrated attention and distractions can be highly counter-productive. But, I do need to feed my family and pay the bills so I do the full amount (6hrs/week) of consultancy for various clients through my own business, [Archaeogeomancy](#), and through the digital



Modelling Positive Stratigraphic Units

heritage specialists [Archaeovision](#). Having such a restricted programme means I have to be picky about which projects I take on but the flipside of this is that I only take on projects I can reasonably undertake and which interest me in some way. A bit of a change from working for a commercial company and having lots of management tasks to contend with and having to support whichever projects were sent my way. Although I do now have to do all my own accountancy and admin... 😞

I'm currently working on some systems development for a client assessing police and fire stations across England, a lovely piece of work which is generating a gazetteer of sites including all manner of interesting data. The system I'm developing is a fairly typical relational database backend with an associated GIS for a spatial perspective and producing cartographic outputs for reporting. Whereas my PhD largely uses Open Source software for everything, as that is where the development focus is with academic research software, this project is back to my old toolkit of Microsoft Access, ArcGIS and some bespoke tools built around this. The bulk of this system is now complete, a draft version has already been handed over, and today I'll be loading a revised dataset and adding additional tools to deal with specific tasks the users need to accomplish.



The Old Fire Station, Church Lane, Lincoln by Lincolnian

It's always good to branch out too and another ongoing project I need to do some work on over the weekend is a non-archaeological spatial information system. It's a system for assessing potential ecological impacts and strikingly similar to the kinds of approaches we archaeologists use to undertake Environmental Impact Assessments and to complete Heritage Statements and the like. So whilst there isn't a heritage angle, my skills and experience can be brought to bear on GIS work in another environmental discipline; something current students ought to bear in mind when they come to look for jobs!