Semantic Technologies Enhancing Links and Linked Data for Archaeological Resources

Keith May
English Heritage, UK. *Keith.May@english-heritage.org.uk*

Ceri Binding
University of Glamorgan, UK. *cbinding@glam.ac.uk*

Doug Tudhope
University of Glamorgan, UK. *dstudhope@glam.ac.uk*

Stuart Jeffrey
University of York, UK. *stuart.jeffrey@york.ac.uk*

Abstract:
Outcomes from the STELLAR (Semantic Technologies Enhancing Links and Linked data for Archaeological Resources) Project are presented. The basis for this research is the need to widen access to archaeological datasets, many of which rarely reach publication. The tools and methodologies presented will allow archaeologists, other related domains, or non-specialist third parties to cross search different datasets using querying tools to ask new research questions of the previously un-connected data. Because the data has been 'semantically enabled' by mapping it to a Conceptual Reference Model (CRM), such research questions could be of greater complexity and at a broader perspective than previously possible. The conceptual reference modelling also opens possibilities to investigate the basis for more implicit relationships and interpretations not previously searchable in the underlying data. The semantic technologies employed are based on standard representations of domain vocabularies and the underlying core ontology, an archaeological extension (CRM-EH) of the CIDOC CRM. Methods for mapping data to the CRM and extracting semantic RDF representations from the datasets are described. STELLAR templates and online applications are presented. The need for controlled terminologies using SKOS W3C standards for Thesauri and vocabularies is emphasised and further work on developing a SKOS template and the potential of user-defined templates for further purposes is discussed.

Key Words: *Linked Data, Semantic Web, CIDOC CRM, Ontology, Semantic Interoperability, Digital Archives, Cross-searching*

Introduction and Background

The STELLAR project (http://hypermedia.research.glam.ac.uk/kos/stellar/) was a collaboration between Glamorgan University, English Heritage and the Archaeology Data Service. The aims of the STELLAR project were to develop methods, tools and associated guidance documentation (http://hypermedia.research.glam.ac.uk/resources/STELLAR-applications/) with tutorials and reports, to enable non-specialist archaeological users to carry out mappings of their archaeological data sets to CIDOC Conceptual Reference Model (CRM) based ontological models. The online tools facilitate the easier conversion of

archaeological data to Resource Description Framework (RDF) formats by non-specialist users and those less familiar with the CIDOC CRM ontology (Doerr et al. 2011). A final outcome of the project will be for various archaeological (but non-CIDOC CRM specialist) users to convert their own data sets to Semantic Web and Linked Open Data (LOD) formats (RDF/XML), thereby testing and proving that the tools and methodologies work effectively. As a final output the project has published a group of Linked Data sets in the ADS triple store that was created for this project (data.archaeologydataservice.ac.uk/). These Linked Data have been created using STELLAR tools to show that a range of different original datasets can be mapped to the CRM-EH (the English Heritage archaeological extensions of the CIDOC CRM; http://hypermedia.research.glam.ac.uk/kos/CRM/) and thereby made available as interoperable archaeological data.

The STELLAR project developed from many of the ideas and experiences of the previous three-year long Semantic Technologies for Archaeological Resources (STAR) project (http://hypermedia.research.glam.ac.uk/kos/star/). The STAR project developed a semi-automatic tool for extracting the appropriate sets of RDF triples that express the data mappings to the CIDOC CRM and CRM-EH ontologies. However, that tool required certain expert knowledge from STAR team members to operate. The aim of STELLAR is to generalise and significantly enhance the data extraction tool and the methods developed by STAR so that, based on their archaeological knowledge of the data, third party archaeological data providers will be able to use it in combination with the mapping guidelines produced. The extracted data can be represented in standard formats that allow the datasets to be cross-searched and linked by a variety of Semantic Web tools, following Linked Data methodologies. The extraction tools enable semi-automated use by non-specialist users to map their data sets to CRM-EH and extract archaeological datasets converted into RDF/XML representation conforming to CIDOC CRM. The project team have also produced best practice guidelines and tools for generating Linked Data corresponding to extracted datasets and based upon experiences from publishing Linked Data from various data providers. These guidelines and the tools are available from the STELLAR website at http://hypermedia.research.glam.ac.uk/kos/stellar/.

Semantic Technologies and Linked Data Uses for Archaeology

There has been a considerable growth of interest in the Semantic Web since it was first publicised by Tim Berners-Lee (Berners-Lee et al. 2001). The idea of the Semantic Web - as perhaps would be distinguished say from Frege's use of a concept (Frege 1893, 17) of the Semantic Web - can mean a lot of different things to different people and even Berners-Lee's notion has changed somewhat and been refined in the course of the last ten years. According to the W3C (2011), "The Semantic Web provides a common framework that allows data to be shared and reused across application, enterprise, and community boundaries."

The Semantic Web extends the functionality of the World Wide Web (WWW), which is based on web browsers using HTML (Hyper-Text Markup Language) first pioneered by Berners-Lee. The Semantic Web in addition to HTML browsers, uses technologies such as RDF (Resource Description Framework), XML (Extensible Markup Language) and OWL (Web Ontology Language) for referencing (using RDF/XML) data objects as represented by persistent URIs (Uniform Resource Identifiers). The semantic part of this is commonly handled by the use of ontologies (such as CIDOC CRM) for representing not just the data in machine readable formats, but also represents more complex relationships between the different data items using such machine (and sometimes human) readable scripting languages such as

OWL or SPARQL for semantic querying over data.

The aim is to describe in a computer format the common structures and meanings in data sets in a way that enables computers to carry out processing that is akin to how humans deduce or infer new knowledge or understanding when reasoning over data sets. The intention is for computers to be able to automatically obtain more meaningful results from searches - rather than simply relying on the returns from simplistic keyword matches – and thereby to enable researchers, either humans using computers or 'robot' style search engines, to perform automated information gathering and new research outputs.

Interest in the general principles of how to create and use online technologies to weave a 'web of data' rather than just HTML documents began to filter through to archaeological computing circles in the early years of the century. Interest was initially confined to a few who were familiar with existing work on ontologies in particular, such that the first papers mentioning the Semantic Web appeared in sessions where the CIDOC CRM began to be presented (Doerr et al. 2010) and practical applications were beginning to be investigated in 2003-04 (Cripps and May 2010).

The importance of these technologies for archaeology has been previously presented at CAA (Binding et al. 2010; May et al. 2009) but it is worth reiterating some key foundations for the research work again here.

- The first and perhaps most practically significant point, is that there was considerable appeal to archaeologists in an approach that simply asked that existing data be mapped to a more conceptual model for it to be usable.

- The CRM modelling approach is based on *mapping* and formalizing the existing knowledge of the domain experts, so archaeologists can use it *without* changing their existing and underlying, or separately held, database systems.

- The CRM's conceptual framework enables the modelling of complex and often only implicit conceptual processes for analysing archaeological data which could not be easily represented by conventional data modelling techniques. For example, representing the complex spatio-temporal relationships between concepts such as Phasing and Grouping.

- The event based modelling of the CRM suited many core archaeological activities. Archaeology deals primarily with *Events in the present* (e.g. archaeological investigations) that record data and interpretations about other *Events in the past* (as revealed and documented by archaeological enquiry).

- The extensibility of the CRM could allow local extensions of the modelling for Archaeological processes and data (sub-classes), while maintaining compatibility with the core entities (classes) of the CIDOC CRM ontology.

- This has demonstrated the ability to hold data in ways that can relate archaeological data to other closely related disciplines such as environmental, geological, or biological domains, and thereby demonstrate principles of inter-disciplinary research and interoperability.

- Using the CRM for modelling provided the advantages of OO modelling without pre-determining an OO or relational implementation.

- Using an existing ontology such as CRM should provide greater standardisation and interoperability with other data sets.

Further work on the STAR project explored the potential and demonstrated the feasibility of using semantic technologies, particularly

through application of a domain ontology, for cross-searching both free text, in the form of archaeological grey literature, and associated data items in an RDF triple store (Tudhope et al. 2011).

Implementation Requirements and Nature of Linked Data Outputs

In the STAR project, all the data that was mapped to the CIDOC CRM and CRM-EH was incorporated in the STAR online demonstrator application. Because of this the data in RDF could all be served from a single server and there was no requirement that the data should be interoperable with any data outside the STAR applications. For STELLAR the overall aim is to create RDF data that can be added to the general body of Linked Data in the Linked Open Data 'cloud' (Armbrust et al. 2009). Because of this the output data needs to conform to the W3C guidelines for Linked Data (http://www.w3.org/standards/semanticweb/data). These include the use of persistent URIs - discussed in more detail in section 7. While STAR worked with internal project URIs, STELLAR will make the linked data available with URI references that will remain stable on the web over time so that other people can consistently use them and link to them with their own data, thus adding value, and hopefully more meaning, to the inter-connected data.

STELLAR employs a more recent version of the CIDOC CRM than available for STAR, which has made available a set of URIs for the various ontological entities (http://erlangen-crm.org/current/).

Triple statements and RDF outputs

The main output from the archaeological STELLAR template conversion is an RDF/XML file containing a series of RDF statements based upon the triple statements depicted in the CRM-EH model and using the CIDOC CRM relationships between the parent classes of the CRM-EH extensions. By running a dataset through the conversion program the user is effectively mapping their data to the CRM according to the series of statements represented by the CRM-EH and CIDOC CRM ontologies.

In the RDF representation of the model there is a triple statement (two entities connected by one relationship) representing the relationship between a finds object and the material that it consists of.

An example of a single instance of this RDF triple for one data item (a finds object with ID 6001) converted by the STELLAR_Finds template would look like the following when expressed in RDF:

```
<rdf:Description rdf:about="http://
STELLAR/crmeh/EHE0009_6001">
<ecrm:P45_consists_of>
<crmeh:EHE0030_ContextFindMaterial><rd
f:value>Copper alloy</rdf:value>
```

In this example the finds object is identified by it's URI «http://STELLAR/crmeh/EHE0009_6001» - which incorporates the unique finds number 6001- and the finds object has been stated to consist of the material Copper alloy.

In this way a series of RDF statements are produced by each STELLAR template corresponding to the relationships that the CRM model identifies to hold between the archaeological data entities.

The Use Case for Stellar Approaches and Templates

The STAR and STELLAR projects, held a number of workshops to review user needs and requirements for cross-search and interoperability between project datasets from different organisational systems. We

identified four key concepts: *Contexts; Groups, Finds; and Samples* involved in archaeological activity (May et al. 2009) and have developed a data extraction template with related data for each concept. In addition, for the STELLAR purposes of making multiple different projects available as Linked Data, we needed a further template to distinguish the data originating from one project from that of another project. Following discussion around the various issues of how to identify and name different forms of archaeological 'site', this template has been called Investigation_Projects (Fig. 1).

STELLAR focuses on the broader shared archaeological concepts that enable searching between different sites (or investigation projects) that have used the practice of 'single context recording' to record individual units of archaeological significance and the stratigraphic relationships that hold between those *Contexts*. The STELLAR templates also cover the 'grouping' of contexts into larger *Groups* of either shared structural or morphological significance for synthesis and analytical reporting or phasing purposes. STELLAR also includes the general processes of identification, typology and dating of particular *Finds* objects, along with the taking and recording of different types of *Samples* and various notes associated with the different concepts. Attributes include materials, measurements, time periods, location, IDs, Notes, Types of element, etc. Attributes that are not present in particular data sets can be omitted, so if there is no data matching a certain field in the template it will be ignored by the extraction program and no RDF created for that attribute.

Development of CRM-EH for Interoperability

The templates correspond to a core set of elements within the CRM-EH, a subset of the full model for STELLAR purposes. The CRM-EH was originally designed (Cripps et al. 2004) to encompass a range of archaeological activities, including excavation processes, such as stratigraphic relationships, but also covering finds recording, analysis and conservation; sampling; and environmental processing, etc. Because some aspects of the model, such as stratigraphic relationships between individual contexts, are most meaningful for cross-searching or inference over collections of site-specific data (i.e. data from projects in a closely related geo-spatial location), some parts of the CRM-EH model are therefore more appropriate for a system enabling intra-site analysis than for STELLAR's immediate purposes of interoperability between distributed projects.

STELLAR Template Methodology

A number of standardised templates have been produced that enable the semi-automated extraction of data from a database and the creation of the appropriate RDF relationships for the data according to the CRM-EH ontological model.

If an archaeological user is content that the CRM-EH adequately represents the basic relationships in their data set, they can use the templates to produce RDF versions of their own data by mapping the relevant data fields in their input data set to the corresponding concepts and fields in the template and thereby generating RDF through the STELLAR tools which can subsequently be published as Linked Data.

A template has been produced for each of the main areas of archaeological data that have been identified by STELLAR as useful for interoperable cross-searching and querying. Figure 1 shows the main archaeological templates and the related fields for Contexts, Groups, Finds, Samples and Investigations related data. In addition, a further template covers more detailed sample measurement data and this illustrates that more of the CRM-EH modelling could be incorporated as templates but we have chosen in this project to focus

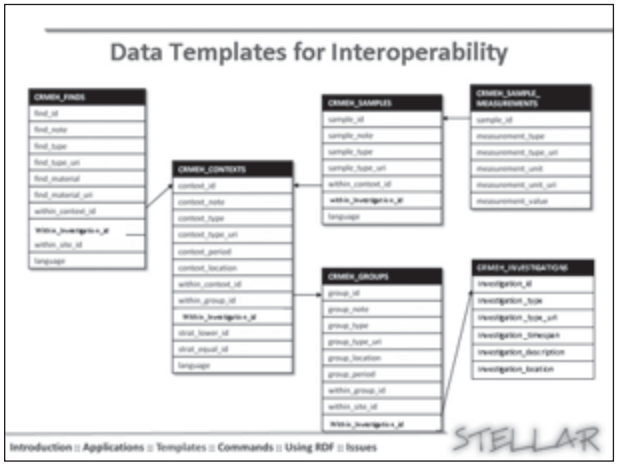

Figure 1. STELLAR Archaeological Templates for Interoperability.

more specifically on some key data entities that we believe, from workshop feedback and other testing, will generally prove most useful initially for interoperability and cross-searching.

The current set of CRM-EH templates are 'internal' to the tools and any modification of the internal templates requires rebuilding the STELLAR application. In response to workshop feedback, a facility for user defined 'external' templates has been additionally provided for converting data to any user-defined textual form. These templates operate in conjunction with the STELLAR.Console application (see below) and a tutorial on the user defined templates (http://reswin1.isd.glam.ac.uk/stellar/tutorials/tutorial2.html) is available on the project website that explains their creation and use. This allows users to tailor existing templates or to define completely new templates for their purposes. For example, a template for the CLAROS classical art project (http://www.clarosnet.org/) format has been provided.

STELLAR Applications

Two applications, STELLAR.Console and STELLAR.Web, have been produced to enable the production of RDF. Users (data providers) select a suitable template and provide it with the appropriate data input. The data can

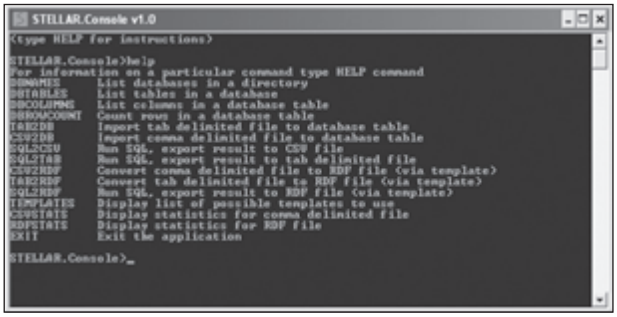

Figure 2. list of commands displayed in STELLAR. Console.

either be an SQL file (for the console tool) or a previously prepared CSV file (for the web tool). The template applications identify a particular column name in the input data and then process each row of data in turn, using the values in each column of the matching data field.

Choosing a template corresponds to making a mapping to the CRM and CRM-EH entities associated with the template. The user provides the input required for the chosen template, choosing which of the optional elements to supply.

STELLAR Console

STELLAR.Console is a command line utility application to perform a variety of data manipulation and conversion tasks related to the aims of the STELLAR project. Files of delimited tabular data (TAB, CSV) can be imported and consolidated to a local database then queried using SQL. A series of templates can then convert the query results to RDF in a robust, consistent and repeatable way. When converting to RDF, the user specifies which template to apply. The user also supplies a file with the SQL commands that will generate the required input for the given template from the internal database. Batch processing is possible with STELLAR.Console, which has a wide choice of methods for expressing the inputs to the templates (Fig. 2). Further details about the use of these methods along with associated

guidance are available for download from the STELLAR website.

STELLAR.Console commands may be either entered interactively or supplied to the program via a text file (for use in sequential batch processing).

STELLAR Web

STELLAR.Web (Fig. 3) is a simpler browser based application to perform the CSV2RDF and RDFStats functionality using the same templates as used in STELLAR.Console. Often there will be other means available for producing the initial tabular delimited data (CSV files) so this application is designed to allow users to work on their own pre-processed data files to produce the RDF output.

To use the application, the user chooses the relevant delimited data file. The source data file to be uploaded must be in comma delimited file

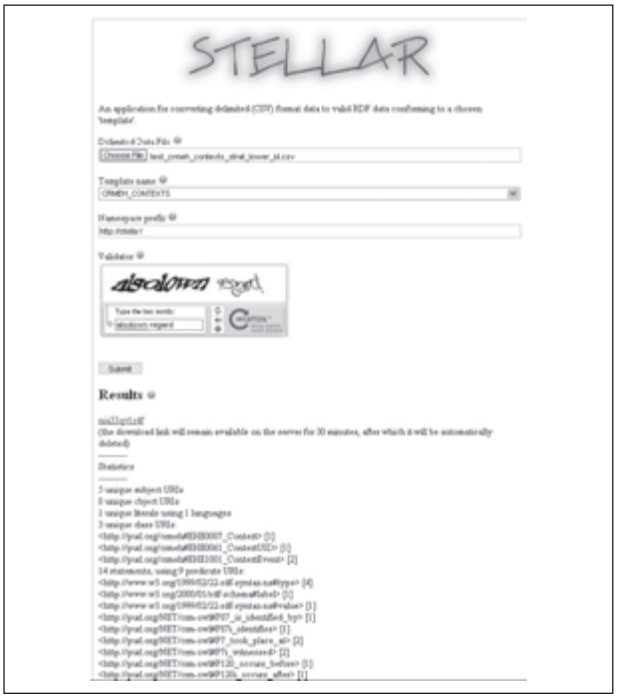

Figure 3. STELLAR.Web interface showing Context template processing.

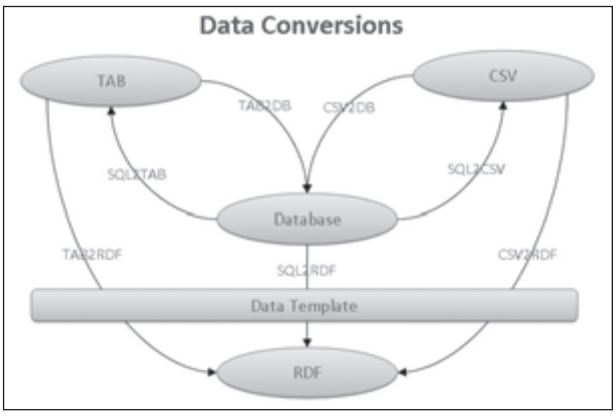

Figure 4. Principle data conversion programs and data flows.

format (CSV) with the first row containing the column names to be recognised by the chosen template. The user then selects the applicable template from the drop-down list of Template Names. The user will need to input a default namespace prefix. The namespace is a URI that will prefix all entity instance identifiers in the resultant RDF file. If the resulting data is intended for publication as Linked Data then the choice of this should be made with careful consideration of the various issues pertaining to how to make URIs persistent (Berners-Lee 1998). The validator control checks the word matching to confirm that it is being used by a human (this is present to prevent possible abuse/misuse of the system by online web-bots).

After pressing the submit button, the resultant RDF file will be created and is available for download via a hyperlink under “Results”, along with some statistics relating to the RDF file which are also displayed. These statistics can help in assessing whether the data created matches what was expected from the conversion process.

Data Conversion Approaches

The templates enable conversion of data to RDF from both CSV and TAB delimited formats. Figure 4 illustrates which of the

Figure 5. extract of resultant RDF, showing context 1015 stratigraphically below context 1016.

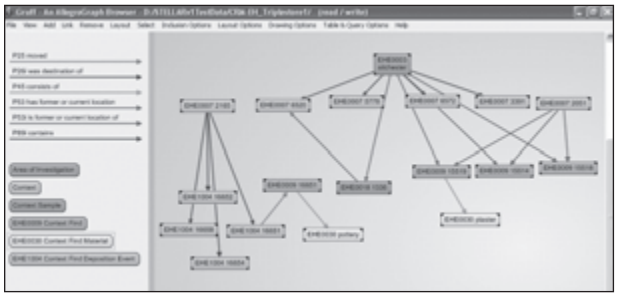

Figure 6. Allegrograph Gruff RDF Browser showing CRM-EH concepts and instances of Silchester RDF data.

various STELLAR programs can convert data from existing databases to either CSV or TAB delimited formats and then subsequently in to RDF.

Another advantage of using the template approach is that it enables consistent treatment of the construction of the URIs that are produced in the resulting RDF. This is very important if the output is intended for linked data uses. For consistency in the CRM-EH templates the conversion process creates URIs for entities based on the following pattern: {data namespace}{entity type}_{identifier/value}.

The data namespace will be a combination of an "organisational_name+unique_project_code+identifier/value". So for example, given an example[1] data namespace of "http://stellar/silchester/", a context with a context_id of "1016" would be given a URI of "http://stellar/silchester/EHE0007_1016". Examples of the generated URI's can be seen in the RDF extract shown in figure 5. In practice more individualistic examples of project codes are needed. Including the individual organisations name/identifier along with site codes helps generate unique URIs.

The CSV2RDF template, which is used by the STELLAR Web application, converts the data from a comma delimited (CSV) file into RDF. If the RDF file name is not supplied, it is generated based on the name of the delimited file (e.g. "myData.csv" generates output of "myData.csv.rdf"). The namespace (/ns) parameter is a URI which will be prefixed to all entities in the RDF output, if not supplied a default temporary namespace URI ("http://stellar/") will be used which can be manually replaced later. The optional "/noheader" flag indicates that the comma delimited file contains no "header" row of column names and in such cases the default column names will be automatically generated.

Outputs and Uses of the Data

The RDF outputs from the STELLAR templates can be used in a number of ways. If loaded into an RDF triple store database then the data can be queried using SPARQL queries. So a group of different data sets converted to RDF could be cross-searched in similar ways to the querying carried out by the STAR demonstrator interface (Tudhope et al. 2011). The RDF can also be investigated using RDF browser tools, such as Allegrograph Gruff which enables visualisation of the data as RDF graphs showing the entities

1 This namespace is only as an example. For linked data, a more persistent URI would be needed. The example data is taken from the Roman Town Insula IX database, hosted by ADS and made available as CSV files via http://dx.doi.org/10.5284/1000259

and relationships between them and also in tabular form.

Significantly, by using the persistent domain namespace and generating the data with persistent URIs, the RDF data could be incorporated into the linked data cloud and therefore becomes searchable alongside any other datasets published in this way.

The inclusion of the stratigraphic relationships as part of the data generated by the contexts template (Fig. 1) means that as well as cross-search between different project data sets the CRM-EH relationships could also enable more specific querying of the relationships within a data set, so a query such as "find me examples of floors with coins stratigraphically below them" would be possible.

As a final outcome of STELLAR a number of archaeological datasets have been selected by ADS for publication as linked data and this work is currently ongoing. In the course of the STELLAR project, ADS used and tested the STELLAR template tools and associated guidance documentation and provided feedback where helpful revisions could be made and where further guidance was necessary.

Issues of Controlled Vocabularies

Search and retrieval of data can be enhanced by using controlled vocabularies and this is further aided if those controlled vocabularies can also be represented in standard RDF formats (Binding and Tudhope 2010). It was therefore also important that another STELLAR output should be a template for creating SKOS format RDF (http://www.w3.org/2004/02/skos) of both concepts and concept schemes (Fig. 7). For the initial STELLAR research work the vocabularies were merely derived from lists of terms used in the glossaries of databases that were converted by STELLAR tools. However, the STELLAR templates have been designed for use of URI referencing of online persistent

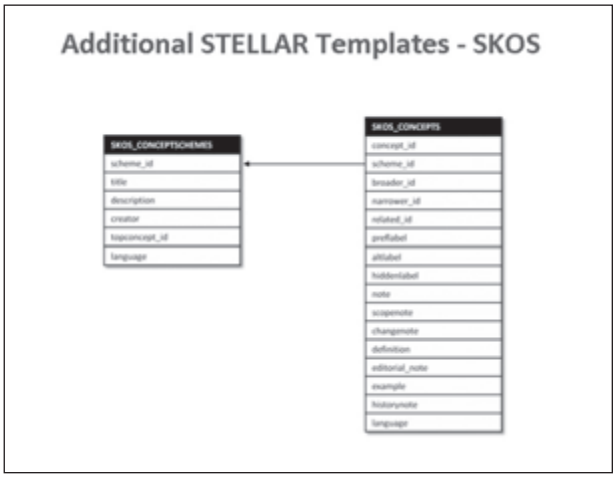

Figure 7. STELLAR templates for conversion of controlled vocabulary to RDF.

URIs for vocabularies and future work is planned to incorporate standardised controlled vocabularies and national thesauri in SKOS format (e.g. the English Heritage National Monuments Thesauri – NMT http://thesaurus.english-heritage.org.uk/thesaurus.asp?thes_no=1).

The resulting SKOS format vocabularies, thesauri or glossaries can be linked back to any data that uses them for controlled terminology by the presence of a "uri_type" field (e.g. find_type_uri) which would then reference the persistent URI for the vocabulary term rather than just the string given by a database instance in a look-up table. This approach raises further questions about how to manage such usage of a number of similar but separately developing archaeological glossaries. These issues are discussed below.

Discussion and Conclusions

STAR and STELLAR have shown that the CIDOC CRM, and the CRM-EH archaeological extension, can be used as the semantic mediation language that enable archaeologists to map similar conceptual records together in previously unconnected archaeological datasets. These conceptual mappings from different archaeological databases then facilitate

semantic cross-searching and querying of those otherwise previously unintegrated data sets.

Linked Data potential uses

At present there is limited archaeological data available as linked open data, but this is very likely to change as more people become familiar with the technologies and the tools for publication and interrogation become more widespread. Although more archaeological data has become available as digital data online in the last 10-15 years, it is still relatively difficult (in fact currently not feasible) to cross-search and interrogate all these different resources at a level of detail to answer detailed research questions (Richards and Hardman 2008, 106).

Being able to search for and ask questions about the relationships between all examples of a particular type of archaeological feature, context, find or sample is therefore seen as one more step towards better understanding of the broader research questions that can be considered through Research Frameworks (Olivier 1996) at local, regional, national and broader thematic levels. Since the majority of archaeological data in England (and elsewhere) is recorded in the course of development control activity, a wide range of organisations are currently producing data, in a wide range of outputs, across the country. The kind of interoperability and cross-search afforded by the STELLAR tools may therefore offer a glimpse of a future, where such distributed data sets might be studied at a broader level, rather than just as project specific datasets.

The potential for Linked Open Data is therefore largely untapped at present. If archaeological data is made available with persistent URIs, then there is the opportunity for new paradigms of cross-search and semantic enquiry involving both free-text reports and associated data. Perhaps most significantly, there is the ability to track the dynamic processes associated with interpretations and reinterpretations made about data recorded during excavation, passed through post-excavation and scientific analysis, let alone what is finally published. As well as the opportunities for interoperability and knowledge generation from datasets that have previously never been interrogated together, there is the opportunity to build intra-site analysis and interpretive query mechanisms which allow much more dynamic re-engineering or 'front-loading' of the archaeological analysis processes, while the excavation is still ongoing, as originally envisaged by the Revelation project that led to the development of the CRM-EH (May et al. 2004). This could lead to significant 'smart' systems for identifying and targeting key aspects of excavation data while still on site.

Controlled terminology and semantic issues

In order to make these kinds of cross-search and interoperability truly feasible at the levels required by Research Frameworks, there is a need to be able to better integrate the terminologies used in the disparate datasets. At present most such archaeological vocabularies, where they are controlled or co-ordinated in any way, are only standardised for use at an organisational level. If many (or all organisations) use different variations of context, finds and samples terminology then trying to cross-reference these may face serious practical problems that prohibit interoperability. This is one area where linked data may make a major contribution to the development of archaeological information on the semantic web if it incorporates the development of online terminology resources with definitive URI identifiers that the broader heritage sector can reference. Future developments might see a combination of unified glossaries, where that is feasible, combined with (machine traversable) mapping between different organisational vocabularies, via definitive URI identifiers on the linked data web.

Issues of 'published' data

One issue that emerged during the STAR and STELLAR projects was the question of how to deal with data from different stages in the archaeological process. At least three different stages of data production can be recognised in the life-cycle of archaeological data. Firstly, there is data deriving immediately from the excavation/recording stage. Secondly, there is data that derives from further analysis of excavations or recording projects. Thirdly, there is data that is disseminated as part of a final publication. While the CRM-EH would enable exploration and tracking of the evolution of interpretive ideas associated with data through these different stages of the archaeological process, and the STAR project managed to produce interoperable data from all three, it is less certain how best to reflect the levels of confidence that might be associated with interpretations (e.g. about dating or material composition of objects) made directly on an excavation, compared to a final publication following detailed analysis. For STELLAR linked data, so far we have opted to present linked data only for data sets that we know have already been published. This is partly because the previous experience suggests that in the first instance it makes most sense to compare like with like, but also partly to avoid the issues raised about how to signify 'interim' results, in an online environment. It is likely that identifying that a dataset is an 'interim' dataset will need to be something recorded in the metadata of the linked data datasets (for example, as potentially described by the Vocabulary of Interlinked Datasets, http://rdfs.org/ns/void-guide). There is still considerable further work to be done in this area by organisations such as the Open Knowledge Foundation (http://okfn.org/) involving provenance, attribution, copyright and sustainable methodologies for making such metadata for different aspects of the process searchable and persistent.

Dataset publication and ownership

Another issue raised by linked data publication relates to ownership of datasets. While it is fairly commonplace for archaeologists to make their data available to others, prior to formal publication, for the purposes of ongoing research, it is another matter if the outputs of that research would result in linked data publication. Ideally this issue would be resolved by all organizations publishing their own datasets under appropriate unique and persistent domain names, but for the present such an ideal world seems some way off.

At present the resource logistics and limited visibility of linked data may mean that it will still be a number of years before it begins to be more widely adopted by the archaeological, or even broader historic environment sector, although perhaps interoperability of shared vocabularies might come sooner. However it is likely that this adoption of linked data technology may be accelerated if new approaches and systems for citation of data (e.g. http://www.datacite.org/) become more widely used in the sector.

One area being explored further by the STELLAR team in future research will be to consider publication of Linked Data for resource discovery in association with fuller online digital archives. In such an approach users could use semantically enabled search tools to query across a suitably selected level of resource discovery Linked Data, and then URI references might enable linkages through to the deeper archive data for download and re-use, once the results of queries confirm the presence of suitably rich data archives.

Bibliography

Armbrust, M., Fox, A., Griffith, R., Joseph, A. D., Katz, R., Konwinski, A., Lee, G., Patterson, D., Rabkin, A., Stoica, I., and Zaharia, M. 2009. "Above the Clouds: A Berkeley View of Cloud Computing." *Communications of the ACM* 53 (4).

Berners-Lee T., 1998. "Cool URIs don't change." http://www.w3.org/Provider/Style/URI.html.en

Berners-Lee, T., Hendler, J., and Lassila, O. 2001. "The Semantic Web." *Scientific American* 284 (5):34.

Binding, C., and Tudhope, D. 2010. "Terminology web services." *Knowledge Organization* 37 (4):287–98.

Binding, C., May, K., Souza, R., Tudhope, D., Vlachidis, A. 2010. "Semantic Technologies for Archaeology Resources: Results from the STAR Project." In *Computer Applications and Quantitative Methods in Archaeology (CAA2010), Granada*.

Cripps, P., and May, K. 2010. "To OO or not to OO? Revelations from Ontological Modelling of an Archaeological Information System." *Beyond the artefact – Digital Interpretation of the Past. Proceedings of CAA2004 – Prato. 13-17 April 2004.*

Cripps, P., Greenhalgh, A., Fellows, D., May, K., and Robinson, D. 2004. "Ontological Modelling of the work of the Centre for Archaeology." http://cidoc.ics.forth.gr/technical_papers.html

Doerr, M., Gill, T., Stead, S., and Stiff, M. (eds). 2011. "Definition of the CIDOC Conceptual Reference Model, Official Version of the CIDOC CRM version 5.0.4. December 2011." ICOM/CIDOC CRM Special Interest Group. www.cidoc-crm.org/docs/cidoc_crm_version_5.0.4.pdf

Doerr, M., Schaller, K., and Theodoridou, M. 2010. "Integration of complementary archaeological sources." In: *Beyond the artefact – Digital Interpretation of the Past. Proceedings of CAA2004 – Prato. 13-17 April 2004.*

Frege, G. 1893. *Grundgesetze der Arithmetik*. Jena: Pohle. Partially translated in: Firth, M. 1964. *The Basic Laws of Arithmetic: Exposition of the System*. Berkeley and Los Angeles: University of California Press.

May, S., Attewell, B., Cripps, P., Cromwell, T., Crosby, V., Graham, K., Heathcote, Jones , C., J. L., Lyons , E., May , K., Payne , A W., Reilly , S., Robinson , D., Schuster , J., Stonell-Walker , K., Walkden , M. 2004. Revelation Assessment Report Report Number: 78/2004. English Heritage.

May, K., Binding, C., and Tudhope, D. 2009. "Following a STAR? Shedding more light on Semantic Technologies for Archaeological Resources." *Proceedings Computer Applications and Quantitative Methods in Archaeology (CAA2009), Williamsburg*.

Olivier, A. 1996. *Frameworks for our Past: A review of research frameworks, strategies and perceptions*. English Heritage.

Richards, J., and Hardman, C. 2008, "Stepping back from the trench edge." In *The Virtual Representation of the Past*, edited by M. Greengrass, and L. Hughes, 167-168. Ashgate.

Tudhope, D., May, K., Binding, C., and Vlachidis, A. 2011. "Connecting Archaeological Data and Grey Literature via Semantic Cross Search." *Internet Archaeology* 30. http://intarch.ac.uk/journal/issue30/5/toc.html

W3C. 2011. „W3C Semantic Web Activity". World Wide Web Consortium. Accessed December 12, 2011. http://www.w3.org/2001/sw/