**Presentation proposal for 9th European Networked Knowledge Organisation Systems (NKOS) Workshop at the 14th ECDL Conference, Glasgow, Scotland**

# KOS-based tools for archaeological dataset interoperability

**Ceri Binding, Douglas Tudhope**
**Hypermedia Research Unit, University of Glamorgan**

## 1. Aims

The presentation will discuss the role of different kinds of KOS in providing semantic cross search of archaeological datasets and grey literature, drawing on work for the AHRC funded STAR project. It will go on to discuss plans for the next stage of work in the ongoing STELLAR Project - *Semantic Technologies Enhancing Links and Linked data for Archaeological Resources.*

STELLAR is in collaboration with the Archaeology Data Service (ADS) at York University Archaeology Department and English Heritage. The ADS has a mandate to provide a digital repository for outputs from research funded by various bodies and holds a wide range of datasets from archaeological excavations. However datasets and applications are currently fragmented and isolated.  Different terminology and data organisation hinders search and comparison across datasets. Currently, archaeological data is rarely reused and re-examined in light of evolving research questions and interpretations.

## 2. Methods

The STAR project addressed these concerns by developing semantic and natural language processing techniques to link digital archive databases and the associated grey literature, via an overarching core ontology framework, the CIDOC Conceptual Reference Model (CRM), extended for archaeological purposes by English Heritage.

The work has required methods to be developed in mapping datasets to the core ontology, extracting semantic web representations in RDF and developing semantic search techniques that operate over RDF generated from various datasets (and also grey literature via Natural Language Processing techniques).

Terminology web services have been developed based upon SKOS thesaurus representations. The service is based on a subset of the SWAD Europe SKOS API, with extensions for semantic concept expansion. These terminology services allow access to the SKOS thesauri and glossaries in a variety of (browser neutral) user interface widgets. They can be employed in a wide variety of applications for both data entry and display purposes, where access to controlled terminology, browsing of concept structures or query expansion is required. SKOS concepts are linked to CRM entities.

SPARQL techniques have been developed to cross search the RDF representations of the five archaeological datasets employed and the grey literature. A query builder dynamically generates SPARQL statements required for cross search.

## 3. Outcomes
The presentation will describe the STAR mapping tool and semantic search Demonstrator.

### 3.1 Mapping tool
A domain expert generated spreadsheets showing the key mappings from the various datasets to the CRM-EH. These selections from the different databases were extracted via SQL queries, and stored as separate RDF files (simplifying the process). This intellectual work was significantly assisted by a mapping and data extraction tool, which automatically generates the RDF statements after the interactive mappings are set up.

### 3.2 Semantic Search Demonstrator
The existing STAR Demonstrator will be presented, with a short demonstration of the cross search and browsing of the instance data and the conceptual model. A common RDF data store holds the CRM-EH ontology, associated thesauri and the extracted instance data, both from the datasets and the grey literature. The demonstrator focuses upon semantic search (via SPARQL, the Semantic Web RDF query language) afforded by a user interface which seeks to hide the complexity of the underlying ontology.

### 3.3 STELLAR ongoing project
STELLAR aims to generalise and extend the data extraction tools produced by STAR to facilitate their adoption by third party data providers. The extracted data will be represented in standard formats that allow the datasets to be cross searched and linked by a variety of Semantic Web tools, following a Linked Data approach.

The presentation will discuss the mapping/extraction tool, plans for extending it for third party use and guidelines that will be used to assist third party users map datasets to the ontology. Linked data design issues will be discussed, including domain name issues, URI generation, linked data expression of (SKOS) vocabularies, mapping relationships between local dataset glossaries and standard cultural heritage glossaries and thesauri.

**References**
http://hypermedia.research.glam.ac.uk/publications#kos
http://hypermedia.research.glam.ac.uk/kos/star
http://hypermedia.research.glam.ac.uk/kos/stellar

**Contact**
Douglas Tudhope,
Faculty of Advanced Technology,
University of Glamorgan,
Pontypridd, Wales, CF37 1DL, UK
Email: dstudhope@glam.ac.uk