**Keith May – English Heritage Research & Standards Group**
**Mobile Tel: 07990970976**
**Keith.May@english-heritage.org.uk**

**CRM SIG – extended abstract**

# Integrating cultural and scientific heritage: archaeological ontological modelling for the field and the lab.

The initial work of the English Heritage archaeological teams was prompted by a need to model the archaeological processes and concepts in use by the Centre for Archaeology (CfA), to inform future systems design and to aid in the integration of archaeological information in interoperable web based research initiatives.

This work has already successfully produced a model and diagram representing the broader archaeological process in considerable detail and complexity (Cripps, et al). Early on in the original modelling project decisions were made to attempt to limit the degree of minute detail that the modelling would attempt to show. For example, the concept of a context record sheet was only modelled as a CRM Information Object (E73) but the model does not show all the particular 20-30 fields of data contained on that pro-forma context record sheet. Hence the degree of what I term ‘granularity’ for the purposes of this paper, were fixed in the early stages at a fairly high to medium level.

Further work, as preparation for a forth-coming project called STAR (Semantic Tools for Archaeological Resources), aiming to incorporate the use of the CRM modelling with FACET tools (Tudhope, et al), has lead to working on more detailed mapping of a specific area of the archaeological science based, palaeoenvironmental recording section of the model. This has successfully resulted in a number of enhancements to the existing model but has also highlighted issues of how balanced the modelling will remain, and to what extent we will need to model other areas to the same degree of granularity, in order for it to be self-consistent.

This paper will show how the Archaeological ontology diagram has been further enhanced to show the general archaeological activities represented by the different conceptual entities. It will highlight in particular the Archaeological Science entities to show which elements already have been modelled and where additional modelling of a particular archaeological science terminology now fit in.

The main aims of the work on enhancing the existing model were:

- Presentation of elements of the original CfA model needed to be enhanced
- Integration of the Archaeological Science Thesaurus terms
- Address issues of ambiguities between Finds and Environmental archaeologists over usage of different terminologies based on Ecofact/Artefact distinctions for objects.
- Provide the basis for better integration of archaeological science information with other archaeological data already modelled.

The recent peer-group revision in England of the Archaeological Science Thesaurus made mapping to CIDOC CRM quite timely. This paper will briefly introduce key elements of the Thesaurus of Archaeological Science Terms and Definitions which has been developed by a working party of archaeological scientists headed by some of my colleagues at English Heritage. Their aim in developing this thesaurus is to make the results of archaeological science research more consistent, and thereby readily searchable, for researchers, and non-specialists. There is a lot of archaeological Grey Literature resulting from the scientific analyses carried out on material from archaeological investigations, which is not readily available to other archaeological scientists.

The archaeological science thesaurus is hierarchical with six main categories, each using a controlled vocabulary terminology, that could be used to structure database fields. The six main categories of the thesaurus are:

**Object type:** Type of remains (item) worked on e.g. animal remains,
**Material type:** The material of which the item is composed e.g. bone, teeth, skin
**Modification state:** The physical condition of an item of interest, particularly documenting its state of preservation or changes subsequent to its use.
**Aspect:** A feature of the remains, divided into natural aspects e.g pathology, and those features resulting from modification by humans e.g evidence of working decoration etc.
**Investigative technique:** The scientific or statistical technique used to investigate the item.
**Recovery Method:** The technique used to gather physical material for further analysis.

This paper will give an overview of the approach taken to mapping these categories and their typologies, and will present the latest results of the mapping of the archaeological science thesaurus to the CIDOC-CRM. It will consider several problems encountered, such as, how, and even whether, to distinguish in the modelling between the different treatment of of Artefacts from Ecofacts during archaeological post-excavation processes by Finds archaeologists and Palaeoenvironmentalists? The paper will consider in some more detail the category of ‘Object Type’ as an example to illustrate some of the questions and problems encountered in these differences of definitions between the archaeology and science based terminologies.

The paper will also introduce some further issues encountered in defining the different degrees of granularity in the modelling. From the thesaurus hierarchy, Object Type has three levels of narrow term. Do they all need to be mapped? What is the best way represent these differences of granularity in the modelling diagrams, alongside concepts such as ContextRecord (EH_E0048), which has a multitude of other data fields, that so far we have chosen not to model in such detail?

Other categories in the Archaeological Science Thesaurus (e.g. Material, Investigative Technique) only have two, or less, narrow terms and therefore, if mapped in entirety, would be mapped to different degrees of granularity.

A related issue has also been identified with the original modelling of the concept of the archaeological context and has raised issues about how to reflect these differing degrees of granularity using 2D ontological modelling diagrams, or modelling software such as Protégé.

More pragmatic issues of how best to update the diagrammatic representations of the model, were also addressed when developing a representation of the original CfA model using Protégé software. This enabled the addition of the new archaeological science mappings more readily without a major overhaul of the existing diagram. This work has largely been carried out in partnership with Doug Tudhope and Thanos Zafiriu at Glamorgan University as part of a DELOS project. Work still continues on finalising this modelling.

The Protégé modelling of the archaeological science data has only partly been completed to date in the time available, although further work is planned as part of an upcoming three year AHRC funded STAR project. Although in the early stages, so far some issues have been encountered with the Protégé tool, such as how best to output results to less ontology aware 'domain experts' who need to verify the modelling of their particular activities.

**Bibliography**

Paul Cripps, Anne Greenhalgh, Dave Fellows, Keith May, David Robinson
Ontological Modelling of the work of the Centre for Archaeology
CIDOC CRM Technical papers, September 2004.

D Tudhope, C Binding, D Blocks, D Cunliffe
FACET: thesaurus retrieval with semantic term expansion
Proceedings of the second ACM/IEEE-CS joint conference on Digital Libraries 2002 - portal.acm.org