Paper presented at CIDOC2012 - Enriching Cultural Heritage, Helsinki, Finland, June 10–14, 2012.

The documentation of conservation and scientific analysis at the British Museum: Lessons for Enriching Cultural Heritage

Jonathan Whitson Cloud

British Museum

Jonathan WHITSON CLOUD
Technical Support Officer - Documentation
Collections Services
The British Museum
Great Russell Street
London WC1B 3DG

Tel: +44 (0)20 7323 8880
jwhitsoncloud@thebritishmuseum.ac.uk

LONG VERSION FOR PRINT

Abstract

The British museum has been recording museum object conservation and scientific research effort and outcomes since the appointment of Dr Alexander Scott in 1918. As with other areas of documentation the emergence of digital technologies has provided new opportunities for recording, retrieving and sharing this information particularly alongside collection catalogue records. The British Museum has taken some steps along this path moving, in some areas rapidly, from long standing paper records through systematic digitisation to the production of linked data. This paper will look at the history and data types involved, noting the needs of Conservation and Science documentation as they differ and coincide with the documentation of museum objects and what lessons can be learnt for enriching collections knowledge generally.

Introduction

One, perhaps the most common, method of Enriching Cultural Heritage entails placing diverse but associated data sets one beside another, linked if not integrated, added if not

absorbed. I intend this morning to take the British Museum's experience of recording and linking one such set of associated knowledge, Conservation and Scientific Research (CSR), picking out some issues that are both distinctive to that knowledge set and to the wider task of enrichment. These issues will be drawn firstly from the British Museum's experience and history and then from a summary analysis of the outputs of Conservation and Scientific Research documentation.

History and experience

A timeline of the recording and sharing of CSR data at the museum might include:

- 1753 British Museum founded based on the collection of a scientist (Hans Sloane)
- 1918 Scientist Dr Alexander Scott appointed, initially to deal with effect of war time storage
- 1976 First computerised collection catalogue system at BM, CSR data not include
- 1990 Conservation recorded in 3rd catalogue system (but not linked to catalogue records)
- 1995 Science recorded in unix SAS database (not linked to conservation or catalogue)
- 2002 Integrated Conservation Module in 4th (current) catalogue system
- 2007 Catalogue data displayed directly to the web (but not conservation data)
- 2009 Scientific output recorded in MS Access system (remains unlinked to conservation or catalogue)
- 2010 Both Conservation and Science data added to Collections-on-Line via an experimental RDF publication stream
- 2012 Integrated Science module added to current catalogue system linking with conservation
- 2012 Catalogue and CSR data expressed in CIDOC-CRM RDF available at BM endpoint

Science in museums

Science at the British Museum goes back to its founding roots, coming as it does from the Enlightenment thirst for empirical knowledge, an analytical and systematic approach to the world, its peoples and their products.

Perhaps then the first thing to notice is the refinement of what we mean by science in a museum context from the broad categorization of the world to the detailed analysis of the composition of objects. Where systematic classification was the driving impulse of the earliest museums, including the BM, interpretation has now become the central focus (alongside preservation) and enrichment can be seen as a continuation of that change including, as it might, strictly non empirical data.

Rate and extent of change

This timeline is obviously skewed by my choice and awareness of significant events in CSR documentation at the BM. I have included none of the radical improvements in card indexing or object numbering that undoubtedly occurred between 1753 and 1976.

It is worth thinking about how records were produced, retrieved and shared prior to the advent of computers. Paper reports were written and filed. The effectiveness of retrieval would, then as now, depend on the knowledge management skills applied. How well organised was that filing cabinet, how consistently was that object referenced? The same techniques we promote today: consistency of knowledge structure and entity referencing. Presentation would be by way of knowledge emerging in lectures, catalogues, display labels and publications. In other words CSR knowledge was used to enrich the understanding of the collection by presenting that knowledge alongside other catalogue knowledge in the places knowledge was shared. So perhaps the changes are not so great.

That is not to deny the impact of computerization. The structure given to data and the links that can be established between information nodes and sets far exceeds anything even the best run filing cabinet could manage. It is the ability to link ideas and entities, and to access data

remotely that marks computer records from paper ones. CSR files can now, thanks to computerisation, be seen linked to the objects they refer to from anywhere with an internet connection.

Fragmentation and union

One phenomenon that can be seen at several points in the timeline seems to be common with new technologies, that of fragmenting once coherent data sets before (sometimes) bringing them together again; one system not integrating with another, the digitised hard to compare with the non digitised and the rdf hard to compare with the not rdf etc. The impact of not having digital CSR records at various times integrated with the catalogue system has had a fragmenting effect on the BM's collection management and our experiment with rdf has prevented the CSR data being searchable with the catalogue data. In the case of rdf we expect the phenomenon to be short lived. The walls between internal systems are also becoming, we expect, less solid as a result of the renewed focus on coherent knowledge management provoked by Linked Data. Systems are being gathered into environments.

Archives as digital assets

The decision to scan paper CSR records rather than capture the content in a granular, categorised, database format was a resource restricted one but does have some positive enrichment consequences. Firstly the development from hand written to typed to electronically generated reports tells a story of its own and these items are of themselves cultural heritage objects. Secondly the quantity of material scanned vastly outweighs what would have been possible to capture as data, with the resources available. And having got the scans into the public domain something like crowd sourcing might be possible to get the

granular data out. A great deal of other enriching material will be archival/non digital and scanning, for all its data dead-ended-ness, will have an important role to play.

Proximity to catalogue

Perhaps the most difficult conversations held on the way to adding our CSR output to our collections web site was how it should be presented. On a separate but linked web site/page? Un-separated and entirely contiguous with catalogue data? We eventually settled on tabs, a position half way between these two extremes. The broader question is when does enrichment become diversion, dilution and confusion?

Authorship and authority

Another long running conversation before presenting CSR data on the web was whether conservator's and scientist's names should appear. In the case of conservators some effort went into removing names but for the scientists their names on their reports were an indication of scholarly ownership and authority.

Both scholarly and community generated content will generally come with an author's name or some sort of identifier and, the scholars at least will argue, some ascription as to whether that author is a scholar or a community member may also be welcome.

Sensitivity

Some material may not be suitable for public consumption. For the most part the BM takes a relaxed view on this; as a rule that we release all material unless there is a very good reason not to. In the case of CSR material we are still experimenting with exactly where this line falls. Reports that relate to non BM objects have been withheld as have those that are incomplete. The correspondences in many of the Project reports and references to other

conservators and museum staff have been removed. We have added a link to every record pointing to a page describing our release policy. It is worth remembering that none but the most recent of this material could have been imagined to be likely to be published as written. Moderation is required and this is time/effort consuming.

Data Types

So is Conservation and Scientific Research knowledge difficult to capture and share?

A table of CSR activities, what they act on and their outcomes:

Activity	**Acts on**	**Produces/Recorded as**
Conservation Report/Survey	Single Object, Groups of Objects, Non Objects	Text, Measurements, Categorisation, Date/timelines, Images, Annotated Images
Preventative Conservation/Science	Groups of Objects, Non Objects	Text, Measurements, Categorisation, Date/timelines, Images, Annotated Images, Technique specific File types
Conservation Intervention	Single and Groups of Objects	Text, Measurements, Categorisation, Date/timelines, Images, Annotated Images,
Scientific Analysis	Single Objects, Groups of Objects, Conceptual Groups (Object Types)	Text, Measurements, Categorisation, Date/timelines, Images, Annotated Images, Technique specific File types

To break that down into unlabelled data types we have:

- Associations/links
- Categorisations
- Measurements
- Dates
- Text
- Images
- Technique Specific File Types

The top four of these are standard database data types so storage at least isn't an issue.

Taking each in turn:

Associations/links need entities to link to. For single objects this should be simple enough with catalogue records providing the r reference-able id. For multiple objects some sort of collection or group record may be useful but is not essential. Technically the BM has struggled to link CSR records to large numbers of single object records but that is a limitation in the implementation of our current system, fixable but non-trivial. Conceptual groups such as material, technique or culture, can be linked to via their thesaurus term. Another good reason to have strong thesauri.

Categorisations are the simplest and most effective of database tools. Some thesauri, look-up lists, etc will be shared with the object catalogue. CSR specific thesauri the BM has identified include analysis techniques, conservation techniques and conservation materials. Other thesauri may be required for other enriching data sets but the rule as ever is don't create when you can re-use. Our CSR data currently shares the following thesauri: Object Type, Object Material, Bibliography and Location.

Measurements are measurements; however the scope and quantity can be exceptional in comparison to catalogue object records. Just starting with the elements gives 98/112/114/118 things to measure. One thing computers do well however is deal with large data sets so this shouldn't be a problem.

Dates are dates but it is worth noticing that nearly all CSR activity is Event based. Whilst many current collections knowledge systems struggle with this, including our own, the CRM manages it very well. Another narrative aspect of Conservation work in particular is that, as a result of the activity, objects are actually physically changed. This is I think unique in all museum activity as an intended outcome. Again in our experience the CRM looks to offer a better model for recording this than our current system.

Text in the form of the narrative report is the central outcome/output of CSR activity. In database terms however blocks of text are not optimal. One solution is to store it in a dedicated content management type system. This isn't something the BM has looked at very deeply but recent discussions on community generated content may encourage development, as with the fragmentation point made above, probably as part of a coherent knowledge environment.

Images and **Technique Specific File Types** are often described as digital assets and can be considered together or apart depending on the power of the digital asset system. The issue here is storage capacity and, in the case of the Technique Specific File Types, sharing the files away from their native environment i.e. the machine that generated them. The BM has not resolved these issues

So, in short, CSR knowledge is no harder to capture or give structure to than any other knowledge set.

In conclusion Conservation and Scientific Research has a long tradition of enriching our knowledge and understanding of Museum objects. The British Museum's experience of presenting CSR knowledge alongside catalogue records has provided us with many valuable lessons for enriching cultural heritage generally with further data sets still. At its simplest the rule seems to be that structured data and reference-able entities technically allow stress free sharing alongside other structured data with reference-able entities. The issues that remain are human, political and behavioural. Not a conclusion that should surprise anyone in documentation.