

# **Joined up thinking: linking and sharing Science Museum collections and content**

## **Daniel Evans & Ailsa Jenkins**

By way of context the Science Museum in London has its origins in the South Kensington Museum that opened in 1857. The collections management system contains records for more than 200,000 objects and 100,000 people and organisations and there are a growing number of exhibitions being documented.

We are going to look at the steps that the museum took to link and share its collections and content. Firstly, to bring its diverse electronic and digitised content together. Secondly to get more value from the content and thirdly to create a platform for content that can be added easily in the future.

So what have we been doing? The first step in the process was to bring diverse content together as Stories, Objects & People. The aim was to make core museum content available on the main Science Museum website and to link it together.

At the moment we have a veritable smorgasbord of content spread across various places. Currently on the Science Museum website there are pages that originate from the Web Content Management System (Sitecore) featuring people and objects that are in no way connected to the CMS.

'Brought to Life' (exploring the history of Medicine) features stories and 4,000 objects. It was built into the museum's website with a technical solution that relied heavily on the integration of data. This is now a static site as to add more object records would be costly and labour intensive.

There are also relational databases; the Science Museum Group's Collections Online is the web front end for our collections database from the Collections Management System (Mimsy XG). We also have the Digital Asset Management System (iBase) and the Archives Management System (AdLib).

On completely separate web sites there is 'Making the Modern World' and 'Ingenious' that needed a one off import from legacy documents as XML.

There is also the Science & Society Picture Library, again, completely separate. There is no one stop shop where you can find all things Alexander Fleming (or Charles Babbage, or James Watt...) that we have in the collections. A lot of time and money has been invested in our existing content and we are looking to get much more value from it.

We were looking for a flexible solution that would provide a more effective platform for adding future content. As data and content is added and improvements and data cleaning is carried out in the source systems new and revised text it will appear online.

As part of the process other organisations were consulted (BLib, BM, Archives Hub).

Ontologies were considered; Dublin Core (DC), Dublin Core Terms (DCT), Friend of a friend (FOAF), Simple Knowledge Organization System (SKOS), CIDOC Conceptual Reference Manual (CIDOC CRM), Europeana Data Model (EDM) and schema.org.

As much information as possible is to be published in linked data format (RDF) or resource description format. Data was extracted from all these various systems and consolidated into a triple store (more of that later).

So, what will it look like?

The layout was refined following user-centred testing and we will shortly be launching an object page for every object. So that a visitor can dig deeper if they want to the content will be displayed as:

- Context and discovery, including an image
- Connected themes
- Connected people and organisations
- Objects
- Other [such as in-house object names]

There is also a Fact file that includes information about each object, such as Maker, Place made, Materials, Date, Object number, Source. Also included will be Location text from the CMS revealing gallery or storage information so it is clear what is on display and where. Whilst sited near the bottom of the web page this cataloguing information is foundational and crucial to the workings of the linked data.

Have we achieved this outcome, to create a critical mass of content; compelling, coherent overall offer? We think that we have. Content will be more than a catalogue as context will link the object to the world outside.

Technically this has been achieved. We have built a new infrastructure with linked open data at its heart. Data was extracted from the variety of systems and consolidated into a triple store. We have adopted a pragmatic ontological approach to maintain the integrity of our data sources, while leaving open the possibility of translating this for a greater variety of ontologies in the future. We expect that the data model will evolve continuously.

So on to step 2.

The aim was to open our content up to wider use and wider engagement. Information will be released to the public in a variety of reusable formats (including microdata, APIs and feeds).

From the analytics we know that rich engagement with our online content takes place across the internet, on blogs, on twitter, but this engagement isn't reflected on our own web presence. The party is over there in the kitchen, but we are over here stuck in the garden shed watching them have all the fun.

The wire frames for the new web site were user-tested and a user-centred design developed. We learned that users had relatively little interest in the object data itself but it was the richness of the linkages driven by our data that greatly increased the content's value.

An example of engagement is the *Victorian "rather sinister" artificial hand and arm*. This object was blogged on Boing Boing on 31 August 2010 and drove 2,662 visits directly, but 45,843 visits started on that page over following 2 months through retweets, blogs, other social bookmarking sites.

< <http://boingboing.net/2010/08/30/victorian-rather-sin.html> >

To encourage reuse we are licensing most of our content under a Creative Commons license (CC BY NC SA 3.0) whilst our metadata is to be placed in the Public Domain through (CC0 1.0). Our Museum content is at the heart of our engagement with our audiences. We want to make it as easy as possible for people to use our content

wherever possible. We also want as much of this activity as possible to be reflected back on the website.

There is a twitter feed <@sciencediscoveries> and a Comments page for each object. To encourage sharing we have made it easy to tweet, and to share on Facebook and Google plus.

We are actively encouraging re-use of content so there is a citation and guidance on copyright and licensing so that images can be download or embedded. We want them to credit us, and if they are making money from our assets then we should share it. So on to step 3. Linked data + Open data = Linked Open Data.

We can link our data with other data, and re-use it in different ways and we want others to do that too. This a key idea of the Semantic Web introduced by Tim Berners-Lee in 2001.

Selecting info from a relational database you have to know what the content is to be able to query it.

It is different using triples. A centralised triple store has been created that contains the RDF representation of content. The data was extracted and consolidated from the Museum's disparate sources. RDF triples (Subject, Predicate, Object) used in linked open data are not dependent upon their containing database needing to be defined in advance. What was crucial was that each piece of data had a URI (uniform resource indicator).

There is also a SPARQL endpoint (an open language to query info) that permits federated searches.

Here's the example of the data.gov.uk site releasing data sets in a number of formats, allowing it to be combined, for example, as an overlay to an open source map of London to show something that in isolation you could not have seen.

A graphical example is the London Cycle Hire Journeys and Pollution map from UCL where the thicker the line indicates more journeys and the redder lines show more pollution. This combines data from 3.2 million cycle journeys from Transport for London (TfL) with pollution data and routing, buildings data on a map.

<<http://mappinglondon.co.uk/2012/london-cycle-hire-and-pollution/>>

Another example is the Archives Hub working on a Linked Data project called Linking Lives. < <http://archiveshub.ac.uk/blog/2011/08/locah-linking-lives-an-introduction/>>

Standard ontologies were used, but with an almost 'pick and mix' approach. The ontology contains:

- Object
- Agents (People and Organizations) (from FOAF)
- Assets
- Rendition
- Document
- Exhibition
- Place
- Concept (from SKOS)

Where possible linking will take place from Places to Geonames, People & Organisations to VIAF (Virtual International Authority File) and generally to Wikipedia/DBpedia, BNB (British National Bibliography).

So we have achieved our aims; of linking our content together better, and linking it to our Science Museum brand; opening our content up to wider use and wider engagement; experimenting with what we (and others) can achieve with Linked Open Data. We have created a flexible platform with long term value that is futureproof. Time will tell and Stories, Objects & People will be coming to the Science Museum web site < <http://www.sciencemuseum.org.uk/> > in the near future.