CIDOC 2014

Title: Using crowdsourcing to enhance collections data and the online user experience
Author: Alex Bromley
Institution: Museum of London
Contact: abromley@museumoflondon.org.uk

Keywords: crowdsourcing; museums; London

The Museum of London collections cover a wide variety of disciplines and material from prehistoric flint tools to the 2012 Olympic cauldron. Although the exact number of objects is unknown, it is estimated at approximately 2 million individual items, not including bulk finds from our archaeological archive. Over the years various inventories, audits and digitisation projects have attempted to catalogue the material but there is a limit to what can be achieved given our finite resources. One way of solving this is through public crowdsourcing of documentation.

While core cataloguing for inventory purposes will always be done by the Museum staff or supervised volunteers, aspects which can be crowd-sourced include the identification of places depicted as well as granular indexing of printed material where no digital version exists. Another type of information which the museum requires is data which enables better resource discovery of our online collections.

With these issues in mind, the Museum of London has been piloting different methods of crowdsourcing over the past year. These projects have been part of a wider programme known as *Opening Up to Digital Environments* funded by Arts Council England.

Using three different models, the Museum has created online applications to present and gather collections information:

- Model 1 – a closed system for specialists
- Model 2 – an open system for all users
- Model 3 – a closed system for school children

This paper explores the reasons behind the Museum's decision to experiment with crowdsourcing, how they identified the participants, the platforms and techniques used for building the data gathering applications, as well as some lessons learned about the exercise and how the data might be used.

From the outset a decision was made to utilise the Museum's collections online infrastructure for delivering the projects - the Collections Information Integration Module (CIIM). This piece of middleware is used for extracting information from our collections management system (MIMSY XG), transforming data where necessary, then creating a SOLR index on which to build web applications. It also comes with the ability to 'augment' resources, adding data which is not required in our collections management system but is necessary for online delivery. Three web applications were built to capture the crowd-sourced data which in turn was stored in a new module of the CIIM, the User Generated Content (UGC) module. This system keeps the crowd-sourced data separate from collections information, has an API as well as an admin interface where contributions can be verified or removed.

Model 1- Surrey Canal photographs
The museum houses an extensive archive of material relating to the London Docklands and inland waterways. One such collection is a group of photographs of the Surrey Canal, much of which is now covered up by post-war development. After discussion with the curator, the project manager identified approximately 100 images for inclusion in the closed project. While many of the images had a brief title, the exact location where the picture was taken is often unknown. Users were asked to identify the premises, street, neighbourhood and date if known. As this was to be a 'closed' system, a suitable group had to be identified. After consultation with the curator, a local history group in south-east London offered to assist. Volunteers were issued with user names and were invited to go through each image in turn. If they could not offer any information, they could 'skip' and do the next one.

Model 2 – London Street Views 1840
Between 1838 and 1840 London publisher John Tallis created a series of 88 pamphlets titled a *Tallis's London Street Views*. These guides illustrated London's main commercial streets (or parts thereof) and included a street history, advertisements, a local map and, most useful, an elevation street view and business directory. Eighty-seven of these guides are in the Museum of London's collection. These are highly valuable resources for researchers of Victorian London but were only available to visitors to the Museum Library. Even simple online publication would not have made them wholly accessible and it was going to need a more granular type of digitisation and some crowdsourcing to make them useful. Thirty five of the pamphlets were chosen for the project and each double page elevation, together with the index, was digitised in-house and the resulting media records 'augmented' in the CIIM with GIS information (Longitude and Latitude) and other data to aid navigation. Out of the 35 pamphlets, 20 ‘thoroughfares’ and squares were identified which roughly corresponded to the modern street layout of central London.

This was the most ambitious of the three projects and a web application was developed which interfaced with Google Maps API to present a map-based entry point to the digitised resources as well as a modern-day Street View of the elevation. A 3-D ‘walk through’ of the elevations completed the offering.

Users are given the opportunity to register and then navigate through the resources, highlighting and identifying buildings in each elevation, then transcribing the information from the printed indexes to digital form. All the data is sent to the UGC module and it requires 5 people to agree on the premises description before it appears in the digitised index.

The site has proved popular since launch in March 2014 and now virtually all 20 thoroughfares or squares have been catalogued by users.

Model 3 – Tag London
One of the limitations of the Museum's collections online system is its reliance on the traditional Museum system of object naming as the sole method of object classification. This can throw up some unhelpful search results – for example a search for 'painting' won't return records with an object name of 'oil on canvas'. Therefore our evaluation identified a need for an 'object type' function which would help researchers go straight to the material they want. In addition, there was a need to provide an index of objects according to 'period', something which is only an optional field in our core data standard.

A recent strategic objective to 'Engage every school child in London' led the team to

consider whether local schools could be recruited to do this 'object type' tagging and contact was made with teachers to gauge interest. Four schools agreed to take part in a pilot – with children ranging from 8 to 14 invited to 'tag' approximately 1000 objects over a three week period in planned lessons.

Incentives were added to the project with children awarded 'medals' for completing 10, 25 and 50 records. Quality was to be maintained by arriving at a consensus based on the number of pupils who entered the same data. The following represents a summary of the verification rules:

- Each different object is seen by at least 10 different users
- Each question requires 70% of users to agree the same answer before it is 'verified'
- If less than 70% of users agree on the same answer, that data is considered 'not verified'
- Once 10 different users have answered all the questions on a given object it is not offered to any more users and is considered 'complete'

Lessons learned and next steps

The learning from the crowdsourcing projects has been significant, and with the help of the evaluation that is now being carried out, the museum stands to benefit hugely from the experience and knowledge that it has developed from the projects.

Each project had different objectives, audiences and methodologies and analysis of consensus data has not yet been completed. However the following interim observations can be made:

Model 1 – a closed system for specialists

- Contributions to the closed system have so far been limited with only five registered users.
- The task of finding a target audience and engaging them should not be underestimated. A relationship needs to be developed and very often the digital confidence of specialist history groups is limited and they may prefer to speak in person to a curator.
- The data which has been gathered may be imported into MIMSY XG to enhance object records.

Model 2 – an open system for all users

- Publicity is important. Use social media channels wisely to attract the right type of audience.
- Think about the level of consensus which is achievable bearing in mind the size of your user base. If you set the level too high you may not have enough visitors to verify all the data.
- The user interface has been a barrier to some users (30% found it 'difficult' to use). A lesson here is to think about what you want to achieve by crowdsourcing and don't clutter the screen with additional information and functionality peripheral to the main task.
- There are some very enthusiastic crowd-sourcers out there who will do most of the work. Many people will visit just to browse.
- Quality of the data has so far been good with only 31 data items 'reported' by users out of a total of 730. Out of these, 22 were removed and marked 'incorrect' by the system administrator.

- Based on this success, the Museum may add more digitised maps over the coming year.

Model 3 – a closed system for school children

- There is a tension between providing terminology lists that are understandable by children and yet are useful for documentation purposes. A mapping table was necessary to cross-walk terms the Museum might use to describe objects and those a child can relate to.
- Thought needs to be given to whether the classes relate to the teaching of 'History' or 'Computing' and where the most value will come from for the pupils and teachers.
- Incentives and 'gamification' are a good way of promoting participation although this should not encourage completion rates above quality.
- Ensure that objects presented to the audience are a cross section of your collection to encourage engagement and learning.
- The project may be scaled up in the future depending on the results of consensus data analysis.
- The UGC module has an API which may be integrated with the collections online index to aid resource discovery.

Links

London Street Views 1840: http://crowd.museumoflondon.org.uk/lsv1840