

Title: The collection catalogue as the core of a modern Museum's purpose and activities

Author: Tanya Szrajber

Position: Head of Documentation, The British Museum

Contact: Tszrajber@britishmuseum.org

The original title of this paper has been modified slightly, since the expression 'collection catalogue' has replaced 'collection database'. The main reason is that many institutions around the world do not have the resources to acquire and maintain a database, relying on other means instead. A collection catalogue can take many forms. Ideally, it is an online database, but it can also be a spreadsheet, a set of manual or typed records, or for intangible cultural heritage, sound and video recordings. The eventual aim is to digitise the information and make it publicly available. As long as the information is presented in a structured format, this goal is achievable thanks to the development of standards such as SPECTRUM and CIDOC-CRM. The second reason is that the revised title reinforces the main point of this paper. Recently, a great deal of attention has quite rightly been paid to technological and other developments surrounding the use and re-use of collection data. It is my intention to highlight the process which generates the data in the first place, the complexities of which are not always fully appreciated, ie. cataloguing.

Museums and other repositories of collections share a basic set of objectives, and meet them to a greater or lesser degree. Generally speaking, these are: documenting the collection; its care and preservation; providing physical and virtual access to it; developing, researching and interpreting the collection; developing and sharing expertise; engaging communities and the general public. The internationally recognised collections management standard SPECTRUM details twenty one procedures relevant to museum collections. These are: Pre-entry; Object entry; Loans in; Acquisition; Inventory control; Location and movement Control; Transport; Cataloguing; Object condition checking and technical assessment; Conservation and collections care; Risk management; Insurance and indemnity management; Valuation control; Audit; Rights management; Use of collections; Object exit; Loans out; Loss and damage; Deaccession and disposal; Retrospective documentation. Many institutions, public and private, may not require or have the resources for implementing the full range of procedures. However, cataloguing is identified in SPECTRUM as one of eight primary procedures, and is defined as 'The compilation and maintenance of key information, formally identifying and describing objects. It may include information concerning the provenance of objects and also collections management documentation e.g. details of acquisition, conservation, exhibition and loan history, and location history. It need not bring together in one location everything known about an object, but should provide cross-references to any other relevant information sources known to the organisation.'

In my opinion, cataloguing is the most fundamental of the SPECTRUM procedures, since without appropriate information about the objects, the others could simply not take place. Producing a collection catalogue is a collaborative effort between various specialists, principally curators (and other academics), documentation specialists and IT specialists. These three categories of expertise are not mutually exclusive and in some cases the roles are conflated for beleaguered individuals who have to cope with all three. Documentation specialists facilitate the transfer of curatorial knowledge about the objects to the system

available to that institution. They do so by applying existing standards or developing others more specific to a collection, in collaboration with curators, and the same applies to terminologies.

The main point, however, is that for all the specialists engaged in the process, the aim should be to produce high quality data, recording scholarly information about the objects and their context. There are certain fundamental considerations which affect the attainment of this goal, and the following are based on my experience of working at the British Museum. The most challenging requirement is to establish a balance between quality and quantity, and this is mainly determined by the size and nature of a collection. If it comprises a few thousand relatively homogeneous objects, the luxury of devoting more time to cataloguing each can perhaps be afforded, but this is less realistic when dealing with millions of objects of varying typologies and spanning broad cultural and historical contexts. In addition, some objects by their very nature have more complex information to record, as they may include inscriptions, detailed provenance and production, iconography, etc. Funding and other resources, such as staff and equipment, are clearly further significant factors and these, together with the nature of the collection, should be considered when planning and setting targets for digitisation projects. Establishing clear guidelines and standards at the outset saves a great deal of time and frustration. Their authorship varies, depending on their content. Documentation specialists are responsible for general data input and system guidelines, including where to record the various categories of information, the format of some data such as dates and location codes, and features such as repeat fields, etc. Others necessitate a collaboration between documentation and curatorial staff, and reflect to a great extent the individual institution's established conventions. They can be generic, such as how to record the names of people and places, inscriptions, or bibliographic references, as well as discipline-specific, such as what information to record about clocks, coins, prints, archaeological, ethnographic or other types of collection.

The process of cataloguing is significantly assisted by the availability of relevant terminologies. Here the decision is whether to import existing vocabularies or create in-house varieties. In the case of the British Museum digitisation project, as with many others, the latter solution was chosen. The main reason was simply that the terms which formed their infrastructure were extracted from existing indexes relating to the collection, and were thus entirely relevant but also familiar to curators. In addition, the sheer breath of the collection in terms of historical, geographical and cultural origins meant that no existing terminology resources were sufficient to cover the British Museum's requirements. Those developed internally, which are continually worked upon and improved, include many local and foreign terms since they often have no direct English equivalents. The wide range of British Museum terminologies include relatively stable drop-down lists, polyhierarchical thesauri, and sophisticated authorities, most especially the Biographical Authority for recording the names of people and institutions. Candidate terms can be created by the users, and are then vetted to ascertain whether and how they should be incorporated. The successful application of these terminologies in the British Museum database, as well as the numerous requests by other institutions to use them, indicate that this was the right path to follow.

My experience of training colleagues to catalogue all varieties of objects in the collection has demonstrated that despite all the rules, guidelines, standards and terminologies, cataloguing remains unpredictable. Every object is unique in some manner, even those produced in matrices or moulds or in series. It is therefore essential to engage curators directly in the process, not just by providing information to documentation specialists, but by creating and editing the records themselves. This achieves the best results and also a sense of shared ownership and pride in the project. However, in the current climate of economic austerity, the need for museums to engage in fundraising has led to increasing demands on curators' time, and exhibitions and other public events make up a significant proportion of their work. Cataloguing is not a high-profile activity in museum terms, since the results are not immediately evident unless expressed in a publication, and is often relegated down a list of more pressing tasks. The choice of system, especially of database, is therefore of the utmost importance and should allow the recording of complex information in an intuitive and logical framework. There should ideally be a balance between structured and unstructured data, e.g. terminology-controlled fields and free-text. Training should be tailored to the needs of the user, and apart from general system instructions, creating records should focus on the type of object to be catalogued.

The use and re-use of cataloguing data is most prominently demonstrated in online collection databases, one of the most exciting and groundbreaking developments of the last few years. The ability to search for object information outside the confines of the holding institution has revolutionised research and the very notion of public access to collections. But it presents many challenges, and there are various degrees of success in its implementation. Online access to collection data is a paper in its own right and so is not discussed here in any detail. However, there are also fundamental issues to consider, which in a sense also apply to the physical access to museums. These considerations include the audience to which the information is intended, how to deal with public comment and whether to charge a fee. In the case of the British Museum, the collection data is taken directly from the internal database, i.e. it is not modified in any manner for the online version, although some fields are withheld if their publication might present a security issue. Since the database was originally created for internal use it is academic in nature. It was decided that a compromise in standards to satisfy different users was not necessary, and indeed re-writing over two million records was not feasible. As it turned out, the online database is accessed by a varied and astonishingly large number of users, with around one million visits each month, mainly for research purposes. The overall quality of the data and the profusion of high quality images are the main reasons for its success, as well as a clear and inviting interface. In addition, the terminologies are also available at the point of search and on the results pages too. Public comment should be part of the process of publishing collection databases online, and in the case of the British Museum, has contributed to improvements in the data, the underlying system and the interface. The issue of whether to charge or not, especially with regard to images, is complex and subject to debate. However, similar issues arise with regard to charging for museum entry, and it has been successfully argued that the public has a right to free access to publicly funded collections, whichever form they may take. This is certainly the case for The British Museum, which was founded during the Enlightenment and

maintains the principle of free access to all visitors, to promote learning and understanding through its worldwide collection.

It is apparent to those of us working in the field of documentation how much progress has been made in the last thirty years. Museum documentation is now a recognised discipline which underpins most if not all museum activities. Thanks to organisations like Collections Trust and ICOM, there is a much greater availability of standards to choose from, including those which act as main reference points for the community, such as SPECTRUM and CIDOC-CRM. The technical developments which have taken place significantly facilitate the creation and mapping of data, and include a great variety of databases, languages, tools, and 3D technology. Significant advances have also been made in readily available technology, such as desktop and mobile computing. But perhaps it is the universal access to information which has had the greatest impact on the perceived importance of collection databases, most obviously through the Web, a relatively recent phenomenon, and now including social media such as Facebook. The possibilities are endless and are rapidly increasing. Another critical factor is the awareness of the huge potential which readily available collection data offers for research and collaboration. Apart from individual websites, aggregations and portals, the research into and development of linked data opens up new horizons as manifested in digital research environments such as WissKi, ResearchSpace and Getty Scholars' Workspace™.

However, such groundbreaking and rapid developments in the field of collection documentation and technology generally can lead to some confusion and even a sense of exclusion. Regarding the choice of terminologies and authority fields, the choice includes Getty vocabularies, museum-specific vocabularies, discipline-specific terminologies, multilingual thesauri. It is also difficult for many people to grasp the complex nature of current research and development, with concepts referred to in a variety of ways, and often as acronyms, such as 'ontology', 'reference model', 'framework', 'system', 'schema', e.g. CIDOC-CRM, GIS, SKOS, BIBI, FOAF, PROV. Finally there is the debate concerning the best method to provide public access to multiple sources of collection data, and the various options have their protagonists. Thus there are networks, portals, aggregators, SPARQL endpoints (e.g. Europeana, Ariadne, NMS, Michael Culture Association, Culture.fr, Joconde, MusIS, DAPHNE, CulturaItalia, CultureCloud.

We have to recognise that nobody can know everything, that each discipline has its own specialism even though there are some polymaths who manage to cover several. So while some experts specialise in researching and developing the best methodologies and environments for harmonising, accessing and re-using collection data, others specialise in generating these data. It is essential to maintain strong collaboration and effective communication between these two channels in order to attain the common goal of public access to multiple sources of high quality collection data. Despite the challenges ahead, a great deal of progress has already been achieved and we are in an exciting new era of collection documentation.