Cultural Heritage Metadata: A View from the Ground

Abbi Roberts, Reader in Information Management, Robert Gordon University

Introduction

My poster presentation at the CILIP CIG conference was an overview of a research proposal which I created as part of my postgraduate Information Management course. It examined how metadata can be used in cultural heritage organisations to denote the historical and cultural context of heritage objects and documents. It brought in strands of thought from archive and museum studies alongside issues surrounding specific technological implementations for metadata management. The literature revealed the lack of standardisation in this particular area of data curation and the difficulties of creating 'one size fits all' approaches in an area where every collection and archive has its own idiosyncrasies and where the information needs of users can be very different. This is a vast field of study, and the proposal gave a very broad overview of ideas and methods which have been used in the past and suggested potential research questions for practitioners about what they valued most in contextual metadata.

Since then, I have been involved as a volunteer with an archive project helping to arrange and catalogue a small collection at Guildford Cathedral, with the ultimate aim of providing access for the public via an online catalogue. Given my background reading on this topic, it was the perfect opportunity to get involved in the creation and use of heritage metadata in a real-world setting. In contrast to the broad, theoretical view of metadata management, I discovered that the emphasis within projects such as this comes back to the question of accessibility for the user. The often limited time, resources and funding available mean that metadata solutions must be efficient but above all fast and relatively cheap to implement. Furthermore, it often falls to volunteers to sustain the collection once the initial setup project has been completed, therefore solutions must also be user-friendly and easily maintained by non-expert staff. This article aims to re-examine some of the themes of my original research proposal in terms of these realities, in order to give a new perspective on metadata creation and curation within cultural heritage collections as experienced by users, archivists and cataloguers 'on the ground'.

Digital Surrogates and Web Access

Integration of collections with some kind of online presence is very often one of the key aims of heritage cataloguing projects. Indeed, in any modern archive, an important part of collection management is the creation of digital copies of documents which can be used for reference and preservation, or for inclusion in promotional materials and Online Public Access Catalogues (OPACs). This requires metadata to be created for three tiers of documents:

- the original objects, be they photographs, books or artefacts
- a high-quality digital surrogate for reference, preservation (in case of accidental loss or damage) or to enable fragile or degraded items to be examined without need for further handling
- one or more lower-quality digital images intended for use on websites or in literature to provide information or promote the work of the archive.

These discrete documents all require different metadata which reflects their separate uses but which is also interlinked to show their relationships with each other. For example, a document's OPAC entry may contain a low-quality digital surrogate whose metadata may point to both its location in the file plan and the location of its high-quality partner. The high-resolution image is normally kept for internal use, or in some cases for use by researchers specifically requesting access.

CC BY

Physical locations of artefacts can also be related to the virtual location of their digital surrogates via internal metadata. This arrangement can be seen as a kind of merging of collections metadata (e.g. the OPAC) and records management metadata (e.g. the internal digital file location). The two structures may be similar or different depending on the particular nature of the collection or the storage needs of the organisation. The important point is that the line between organisational records and collection metadata is becoming more blurred in modern collections, which may have potential implications for metadata management in the future.

Automatic Metadata Translation and Extension

Within online OPACs, technological enhancements can be used to automatically augment metadata to aid both cataloguers and users. One example which has been used in the Guildford Cathedral archive concerns the recording of dates. Researchers coming to an archive or collection catalogue may have an exact date for a particular document, or they may have nothing more accurate than a decade or even a century. A catalogue or resource can be made more accessible for these users if the implementation of the metadata takes this into account, particularly as an archive will tend to have more accurate information on a document than an average user. This can be achieved via the use of automatic metadata translation, which can be implemented very easily in date fields. Using this method, a document catalogued under the date "1957", for example, can also be automatically listed as "1950s" (and by extension "1950-1959"), or even "20th Century" or "1900-1999" if appropriate, with no extra cataloguing work needed. The process also works in reverse, so that a user searching for an exact date will be able to find a document catalogued as "1950-1959" by searching for "1957". Cataloguing under terms such as "circa 1957" will automatically add metadata to the document for dates three years either side. This programmable metadata translation could even be tailored to the needs of a particular archive. For example, if there is a significant watershed date around which documents fall, the system can take this into account, for example making "pre-1957" and "post-1957" a part of the metadata translation. This feature can be implemented in many OPACs and can save time and resources for cataloguers, as well as increasing potential routes of access to documents.

User-generated Metadata

In some ways, selection and implementation of metadata is no longer the sole domain of archivists and cataloguers. Many collections will allow users some limited access to metadata creation in the form of keyword tagging. However, users are less likely to follow controlled vocabularies, which can lead to problems in terms of the accuracy, completeness and consistency of the metadata they create. Take, for example, the many ways in which "World War Two", "Second World War", "WWII" etc. can be conflated. For these reasons, user-generated metadata is often implemented as a 'bolt-on' aspect of the final web OPAC and is unlikely to become an intrinsic part of the back-end archival description. Yet, it can be a crucial part of the public engagement aspects of a collection, allowing users to make connections or observations about the documents in their own words. In this way, allowing this kind of sandbox area where 'bad metadata' can exist may be a legitimate way of providing avenues of access which may be otherwise closed to some users. In some instances, it may even act as a two-way process of communication if, for example, a user has personal knowledge relating to an archive document. Carefully managed, this can be a crucial way of drawing in users and augmenting a collection with knowledge which would otherwise be lost.

Hierarchies for Context

The most important function of metadata and archival description – context – must not be forgotten. In small collections and in time- and resource-limited projects, it is often more effective to denote context purely via a hierarchical structure. This is used efficiently in the Guildford Cathedral archive to delineate the separate strands of documentary provenance. For example, the photograph collection has a separate hierarchy to the diocesan records though there may be instances where the two collections overlap. Metadata standards such as ISAD(G) can then be used to provide further granularity and cross-referencing by ensuring that specific information is included in each record.

The popularity of the ISAD(G) standard reflects its ease of use and the relevance of its elements, yet in some ways it is the hierarchical structure which provides the most helpful context. It can give an immediate overview of a collection, showing the different aspects at one glance. It also provides a ready-made method of describing the physical location of documents, as items can be kept together depending on where they sit in the hierarchy. This aspect can too often be neglected in the rush to digitise archives, and it is also useful in any ongoing project where documents are constantly being taken in and out of the archive. In this way, the hierarchy structure itself becomes a valuable metadata tool which carries a lot of information very efficiently, even before drilling down to the item-level metadata.

Conclusion

This overview of some of the metadata methods and concerns within small heritage projects brings home the message that the user is now very often the focal point to the organisational decisions made for the preservation of collections. Ease of access, new ways of encouraging engagement and the use of digital images make for a constantly changing environment in archives and collections. Underlying this, however, are still established methods and approaches adopted by information professionals about how metadata should be used, as well as technological considerations about how to ensure access and save time and resources. In terms of real-world projects, it is clear that this will always mean striking a balance between creating the best possible information structure for a given collection, and working efficiently in short timescales. Whilst it is important to examine metadata standards and how they aid description, too much preoccupation with finding the 'perfect' standard is perhaps unhelpful. In this sense, my experiences within a real archive project corroborate many of my findings in the literature – namely that heritage organisations are constantly innovating and looking for a combination of solutions which allows them to best serve the needs of their collections. If quality and provenance of the archive information can be best preserved through using more than one metadata management method, then these needs are being served.

I would like to thank the People's Cathedral Project team at Guildford Cathedral for giving me the opportunity to experience a cataloguing project first hand. More information can be found at http://www.guildford-cathedral.org/about/the-peoples-cathedral.

Further Reading

DEAN, K., 2014. Digitising the modern archive. *Archives and Manuscripts*, 42(2), pp.171-174.

ROSS, S., 2012. Digital Preservation, Archival Science and Methodological Foundations for Digital Libraries. *New Review of Information Networking*, 17(1), pp.43-68.

STERNFELD, J., 2011. Archival Theory and Digital Historiography: Selection, Search, and Metadata as Archival Processes for Assessing Historical Contextualization. *The American Archivist*, 74(2), pp.544-575.