Outsourcing Authority Control at UCL

Thomas Meehan, Head of Current Cataloguing, University College London

Introduction

Cataloguers at UCL (University College London) Library Services[1] perform the work of identifying authors and subjects relating to bibliographic records, but do not create authority records and generally do not edit them. This has been outsourced since 2006 to an external supplier, Library Technologies, Inc. (LTI)[2] based in the U.S. This article provides an overview of how this came about. UCL Library is a large academic library on a number of sites, mostly based in central London. We have been using the Aleph library management software since 1999.

Background

Before 2006, UCL used no authority control at all. That is to say, we did not have any authority records; we simply had the author and subject browse indexes in Aleph. We had begun to make more serious efforts to align our headings on LC forms, by checking new headings in particular against the Library of Congress Authorities website[3]. Although some cataloguers had experience of editing and using authorities in other work places, there was no established body of expertise at UCL, especially in how they worked on Aleph.

As part of a wider report into the practices of major university libraries in the U.S. in 2002[4], a colleague investigated manual vs outsourced authority control and recommended either loading the complete LC Authority Files into Aleph or outsourcing the work. A further internal report in 2004[5] compared in-house and outsourced solutions for authority control. It found that the costs of performing the work in-house were prohibitive in terms of staff time and money; it recommended outsourcing the work to a vendor and signing up to a vendor's update service to maintain the authority file.

Choosing a Vendor

We investigated authority vendors and undertook to investigate two suppliers. We submitted the same large test file of bibliographic records to both vendors and analysed how well they performed on a number of criteria. Both vendors performed very well and could easily have been chosen in the absence of the other.

Both matched headings very well, often where UCL's entry of the name required some interpretation of incorrect coding, punctuation, and - in the case of the "Königlich[e] Akademie der Wissenschaften zu Berlin" and the "Polyglott [sic] Bible"- incorrect spelling. Some headings were easily identified by one and not the other, and vice versa. One supplier was especially good at matching the names of film actors whereas the other excelled at names of kings, popes, and corporate bodies: we have a sizeable number of both.

LTI - the successful vendor - were particularly effective at tidying up many of the small authority errors that had crept into our catalogue over time. For instance, they managed the complex rearrangement of subject strings where UCL had retained the wrong order of subdivisions or used headings with places which may not be subdivided geographically, e.g. converting "Art, Modern$y21st century$zSpain$zValencia$vExhibitions" to "Art, Spanish$zSpain$zValencia$y21st century$vExhibitions". They were able to intelligently add or change qualifying information to conference headings, for example changing "$cLondon" to "$cLondon, England" and "$cManchester, Eng" to "$cManchester, England".

LTI were also able to deal with changing jurisdictions, something that was particularly significant with UCL's important collections at the School of Slavonic and East European Studies (SSEES). For example, a heading

which UCL had as "World War, 1939-1945$zYugoslavia$zSerbia" was converted to the soon to be out of date but correct "World War, 1939-1945$zSerbia and Montenegro$zSerbia". LTI made similar correct changes with respect to Czechoslovakia and the Russian Federation. Similar changes were made with obsolete language names, such as changing "Icelandic, Modern" to "Icelandic" in a complex Biblical uniform title.

In addition, the introduction of authority control presented an opportunity to tidy up other aspects of our data, such as punctuation and MARC21 indicators. Both vendors under consideration offered this service to some degree, either using specifications provided by us or using their own routines. LTI offered the latter and performed a number of very useful corrections with little need on our part to specify what was required. Examples included removal of some incorrect square brackets from uniform titles, even if no match with an authority record was made, rearrangement of conference subfields and accurate re-punctuation, and the correction of first indicators in name headings. LTI also correct the second indicator of 245 fields according to their language and the first word of subfield $a.

Initial Implementation

Before proceeding with LTI, we had to set up Aleph properly too. Aleph is fully capable of dealing with authority control but needed some configuration as we had not used it for authorities before. Generally this was a case of setting up separate databases for Library of Congress and Medical Subject Headings (MeSH) authority records - some libraries keep them together - and making sure that these were indexed and linked properly to the bibliographic indexes so that users and cataloguers benefit from the see references and additional information available. An Ex Libris consultant aided the systems team to perform much of this work, and helped to set up the export and import routines that would be needed to submit and retrieve any files.

We submitted the entire file of over a million bibliographic records to LTI via ftp in late 2006. They processed the file, amending headings to match the LC Authority file, their own additional records, or MeSH as appropriate. We retrieved this file of amended bibliographic records as well as the authority records that went with them, and a number of reports giving statistics and details of linked and unlinked headings. The bibliographic and authority records were imported and indexed on Aleph. The whole process took about six weeks. While this was happening, we could not amend any of the records we had sent as all changes would be overwritten when the new versions were returned. This was obviously very disruptive although we could download and add new records.

The initial run was generally very successful. There were some ongoing configuration problems in some of the obscurer corners of authority control, which took a lot of work to iron out. For example, duplicate see references, such as the non-preferred "ABC" which can refer to both the Australian Broadcasting Corporation or the American Bibliographical Center; the differentiation in display of LC and MeSH terms where terms match - such as Cancer being a see references in MeSH but a preferred term in LC; or, the clashing of subfields v and t in Library of Congress Subject Headings (LCSH) for terms like "Correspondence".

Ongoing Authority Work

It was obviously important that we maintain our authority file. We still do not create any authority records ourselves. We check our now enhanced indexes where authority records are now visible. If we do not have an authority record for a heading, we check the LC Authorities or LCSH through Classification Web[6] or the LC Authorities site, and MeSH through the online MeSH browser[7] We do not need to download or

create any authority records at this stage as this is taken care of by ongoing updates by LTI.

We signed up for LTI's Authority Express (AEX)[8] service from the beginning. For this, we send off all new and amended bibliographic records once a month. The process and results are much the same as for the initial authority work - we send off the bibliographic records which are returned to us amended and with matching authority records and reports - but the turnaround time is a matter of a few hours. In practice, this means we stop amending existing records at four o'clock one afternoon and resume normal cataloguing the following morning.

The AEX service does not deal with headings or authority records that are amended, so changes in headings over time, such as the addition of death dates to authors, or changing terminology in LCSH are not catered for. LTI's Authority Update Processing (AUP)[9] service takes care of this. We have not yet subscribed to this service but hope to do so soon. We have manually taken care of some pressing problems where, for instance, the addition of a death date means that an old name heading without it becomes out of step with a newly acquired name-authority heading with it. However, this is not scalable in the long run.

There are in addition two bulk updates in particular that we would like to take advantage of, and that LTI should be able to help us with. The first of these is non-Roman script see references, which were recently added to many authority records. We have significant Hebrew and Cyrillic script collections, and we have made efforts to enhance the original script elements of some of these records, especially for Hebrew script. These see references would greatly help the discovery and streamline the cataloguing of non-Roman scripts. The other bulk update is of course RDA, which changes many authority headings in various ways, such as "William, of Ockham, ca. 1285-ca. 1349" to "William, of Ockham, approximately 1285-approximately 1349" as well as adding much information to the record itself, such as gender (MARC21 375 field) and field of endeavour (MARC21 372 field). Clearly it will not be possible to edit these manually. LTI should be able to perform the latter as part of a special AUP run[10].

Conclusion

Outsourcing authority work was for us the only realistic way to implement authority control at UCL. There was a lot of work in setting up, testing, and choosing a vendor, but our systems team were excellent and the resulting day-to-day workflow is much the same as before but more rigorous. Cataloguer identification of current headings and corrections to our indexes is still vital, if not more so, especially on new records created from scratch, but we have only had to venture briefly into editing authority records themselves. The greatest obstacle in some respects is the large amount of down-time needed to process and index a large file of bibliographic records like ours. Six weeks is a long time for a cataloguing backlog to build up and to be unable to push material through for users to consult, although there are of course creative ways round this. However, once done, it should not have to be repeated. LTI themselves have been hard to fault: capable, reliable, and helpful.

Outsourcing authority control has suited our situation and enabled us to implement it effectively from scratch very quickly. Although there is some loss of freedom, especially as we decided to only use authority records provided by LTI, relying only on outsourced records means we know that we are closely in step with orthodox LC and MeSH headings. This keeps the file easier to maintain and will be a boon when the correct standard forms of names and subjects effectively form the basis of links in a world of shared and linked data.

References

[1] UCL Library Services. http://www.ucl.ac.uk/library/

[2] Library Technologies, Inc. http://www.authoritycontrol.com/

[3] Library of Congress. *Library of Congress Authorities.* http://authorities.loc.gov/

[4] UCL Library Services. *A report into comparative cataloguing practices in "Ivy League" universities in the United States of America and UCL*. [Internal report], 2002.

[5] UCL Library Services. *Improving authority control at UCL: looking at in-house and outsourced solutions*. [Internal report], 2004.

[6] Library of Congress. *Classification Web*. http://classificationweb.net/

[7] National Library of Medicine. *MeSH browser*. http://www.nlm.nih.gov/mesh/MBrowser.html

[8] Library Technologies, Inc. *Authority Express*. http://www.authoritycontrol.com/ax-doc

[9] Library Technologies, Inc. *Authority Update Processing*. http://www.authoritycontrol.com/aupdetails

[10] Library Technologies, Inc. *Toward implementation.* http://www.authoritycontrol.com/rdatransition