

Helen K. R. Williams

Authority control at LSE: the continuing story

**Article (Accepted version)
(Unrefereed)**

Original citation:

Williams, Helen K. R. (2013) Authority control at LSE: the continuing story. [Catalogue & index](#), 172. pp. 9-13. ISSN 0008-7629

© 2013 Chartered Institute of Library and Information Professionals (CILIP) Cataloguing and Indexing Group

This version available at: <http://eprints.lse.ac.uk/53480/>

Available in LSE Research Online: October 2013

LSE has developed LSE Research Online so that users may access research output of the School. Copyright © and Moral Rights for the papers on this site are retained by the individual authors and/or other copyright owners. Users may download and/or print one copy of any article(s) in LSE Research Online to facilitate their private study or for non-commercial research. You may not engage in further distribution of the material or use it for any profit-making activities or any commercial gain. You may freely distribute the URL (<http://eprints.lse.ac.uk>) of the LSE Research Online website.

Authority Control at LSE: the continuing story

Helen Williams, Assistant Librarian, Bibliographic Services, LSE Library Services

In 2010 I wrote for both Update and C&I about the authority control project we carried out at the London School of Economics, concluding that, 'Although the project [had] been time-consuming, it [was] worthwhile. Our catalogue is now more consistent and has fewer errors, making retrieval more straightforward for users. In a library this size the catalogue is the primary way in which users identify our holdings. Our library catalogue continues to get a high score on the student satisfaction survey which shows that the hours put into this project have been fruitful.'¹ The importance of authority control is well recognised in the bibliographic community. As Michael Gorman says, 'We cannot have real library service without a bibliographic architecture, and we cannot have that bibliographic architecture without authority control'², so with that in mind our efforts on authority control work at LSE did not cease with the end of our initial project.

With an edition of C&I dedicated to the topic of authority control it seemed timely to write about how we have incorporated the results of our initial data clean into our day to day workflows, and additionally how we are dealing with the 58,000 bibliographic records, and many more associated authority headings contained therein, imported into our existing library catalogue as a result of The Women's Library collections moving to LSE.³

Back in 2007, at the point of our initial data clean, Marcive (the company to whom we outsourced our authority control project) kept a copy of our entire catalogue and all associated authority records, which enables them to provide us with 2 on-going services. On the second Monday of every month we export a file to Marcive containing all the new bibliographic records which have been added in the preceding month. Marcive run their automated processes on this file to correct any unauthorized headings, and to provide us with any authority records we do not already have for name, subject or series headings in those records. This is called the 'Overnight Authorities' service, and our export is dealt with promptly by Marcive so that when we arrive at work on Tuesday morning the cleaned files of records have been returned to us. It is important for us that this process is so efficient because while the records are with Marcive we cannot make any edits to them or they would get overwritten when the cleaned records are loaded back into our catalogue. We have used a recurring meeting reminder, which all our cataloguers have in their Outlook calendars, to remind them when the export is occurring. As well as the overnight service for newly added records we wanted to ensure that authority file maintenance continued on the existing catalogue, and this is achieved through the 'Notification Service'. This tells us whether any of the authority records already in our catalogue have been updated or deleted by the Library of Congress (LC). These authority records are loaded into Voyager (our Library Management System) by our IT team, and then using the Global Headings Change Queue (GHCQ) in Voyager we are able to manually approve each of these changed headings which in turn automates the updating of all affected bibliographic records. We choose to look at each heading manually not only so that we can be sure that each heading provided to overwrite one of our existing authority records is correct, but also so that we can see if there are any instances of this heading combined with subfields (Voyager cannot automatically update these so we deal with them manually or force them into the GHCQ ourselves) or any records where there is a conflict which will require manual resolution.

¹ Williams, Helen, K. R., *Cleaning up the Catalogue*, in *Library and Information Update*, January/February 2010, p 46-48

² Gorman, Michael, *Authority Control in the Context of Bibliographic Control in the Electronic Environment*, in *Authority Control in Organizing and Accessing Information: Definition and International Experience*, Eds., Arlene Taylor, and Barbara B Tillett, The Haworth Press, 2004, p.21

³ The Women's Library @ LSE <http://www2.lse.ac.uk/library/newsandinformation/newsArchive/2012/Womens-Library.aspx>

It is not only the Notification service which generates in-house work. When the Overnight Authorities files are sent to us we also receive a report of unrecognized headings, which automated processing could not match with an authority record, and a 'multi-matches' report, where the heading could conceivably be linked with more than one LC authority record, and automated processing could not resolve this. Both of these reports, as well as the GHCQ work are dealt with by one of our Senior Library Assistants as part of their regular work.

Our authority control work hasn't been without challenges. We discovered, having established what we thought were successful procedures, that Voyager was not dealing properly with the delete reports that are sent to us by Marcive. The system automatically rejected any files containing a 'd' for delete, assuming we would not want a deleted record loaded into Voyager, when it should instead have used those files to delete the corresponding authority record already stored in Voyager. This means we now have to manually go through the delete report dealing with records ourselves and updating affected bibliographic records.

Earlier this year the extensive changes to the LC/NACO authority file as a result of RDA meant that we received an overwhelming number of new authority records into our GHCQ. The report was so large that for several days we were unable to open the GHCQ while we found a work-around by extending the 'time-out' function in our Voyager set-up files which allowed longer for the report to process and open before the system gave us error messages and crashed.

Our most recent challenge has been the authority control related issues surrounding the merging of two separate catalogues. On 1 January 2013, custodianship of The Women's Library collection transferred from London Metropolitan University (LMU) to LSE, and throughout 2013 LSE Library staff have been working with LMU staff to ensure the successful move of the collections to LSE, ready for the opening of The Women's Library @ LSE in August 2013. This included the migration of 58,000 catalogue records for The Women's Library print collections into Voyager by 1 September 2013. This has been an extensive project and much of the work is outside the scope of this article, but authority control has been just one of the areas we have worked on as part of this data migration. LSE uses only LC authority headings in bibliographic records, but The Women's Library collection has a varied history and consequently different vocabularies have been applied to the records over time. Knowing that the 58,000 records would sit in Voyager, alongside our existing data, we felt that it was important to be using consistent authority headings across all records because bringing in variant or uncontrolled data would affect the search and retrieval functions and the quality of the catalogue as a whole.

We decided that the most efficient way to achieve this unity would be to outsource an authority control data clean of the incoming records to our existing provider, Marcive. Importantly this also meant that the newly received Women's Library records would join the copy of our entire catalogue held by Marcive, enabling the authority file maintenance I described above to apply to The Women's Library records as well as to existing LSE records. Without sending those 58,000 records to Marcive it would not be possible to extend authority file maintenance coverage across the entire newly combined catalogue, which would have serious consequences for on-going authority control procedures for the catalogue as a whole.

The cataloguing staff from LMU helpfully provided us with a raft of information relating to the cataloguing of The Women's Library materials. Particularly useful in the area of authority control was information about subject fields which, although they were no longer in use for current cataloguing, had been used to create records in the past. This meant a number of the records we received contained the following fields:

- 690 Precip descriptor string (Precis headings were formerly assigned to records created by the British Library between 1971 and 1990)
- 695 Local subject terms which were not part of any named thesaurus
- 696 COMPASS topical descriptors and

- 697 COMPASS geographical descriptors (COMPASS stands for COMPUter Aided Subject System and replaced Precis in 1991 at the British Library. It was used until 1996, and was no longer used in British National Bibliography records from 1997)

In order to make the subject data provided in these fields more consistent we decided to swap each 690, 695, 696 and 697 entry for an appropriate LC authority heading. The 41,471 Precis headings in 690 fields could be tackled with the help of Marcive. They offer a facility to flip 690 headings to 65X fields before they begin processing, and then match the Precis string to LC headings where possible. This was very successful, and only 4590 headings could not be changed using automated processing (which we will deal with manually in-house). It was not an option to have the 695, 696 and 697 fields dealt with in the same way, so instead we have internally generated reports of the 2104 headings falling into those fields so that we can manually amend the relevant bibliographic records, replacing the data in those fields with appropriate Library of Congress Subject Headings (LCSH).

Another reason for the variety of different subject headings in the catalogue is that The Women's Library has followed standard British Library (BL) practice, which has changed over time. The British Library applied LCSH to records created for the British National Biography (BNB) between 1971-1987 and from 1995 onwards.⁴ Similarly not all records created at The Women's Library have LCSH, so there were some uncontrolled 65X subject headings in the catalogue with a second indicator 4 (indicating the source of the subject thesaurus is unspecified) rather than 0 (indicating a LC heading). However a spot check indicated that a number of these headings were LC compliant (or could easily become LC compliant through automated processing). Something that we learnt in our last data clean was that Marcive does not attempt to clean subject data with a second indicator of 4 so this time around we were prepared for that and able to ask our IT team to run a global change on all The Women's Library records to change second indicator 4 to 0 on 6XX fields. This meant all those headings would be checked by Marcive, and matched with a correct authority heading where possible, and any which could not be cleaned through automation would be detailed for us in reports from Marcive which we could manually correct ourselves.

As well as the use of different subject thesauri over time, we were aware that not all names in 1XX fields were LC authority controlled. BL name authorities (standard before 1993 when the BL and LC agreed to establish a single source of name authorities by merging their 2 files⁵) had been consistently applied earlier in the history of The Women's Library collection, and so there was some variation between these and the LC headings that we have been familiar with using here at LSE in recent years. Additionally the specialist material being collected meant that authority records may well not have existed for vast numbers of names at the point of record creation, but many years later as we work on it now, it is more likely that LC name authority records will have been created for some of the names in the collection.

We contacted Marcive to request a one-off automated check of all name, subject and series title headings against standard LC authority reference files and correction of unauthorised headings in catalogue records, and loading of these corrected name authority records into our local name authority file. We supplied a test file of 1000 records which Marcive returned to us, having cleaned it according to our specifications. As with our last authority control project, it was important to check this carefully and make sure we were happy with all the changes which would soon be applied to all the records being dealt with in this project. With this in mind 4 of the Bibliographic Services team were involved in checking 1 in 20 records from the sample file.

Having sent the sample data using a MarcEdit file we realised that we faced an issue we hadn't dealt with as part of our previous project. Last time we had exported records directly from Voyager, which meant that

⁴British Library, *Metadata Services Standards: subject access in British Library bibliographic records* <http://www.bl.uk/bibliographic/subject.html>

⁵British Library, *Metadata Services Standards: authority control* <http://www.bl.uk/bibliographic/authority.html>

the file of cleaned data supplied by Marcive was loaded back into Voyager using the bibliographic record identification number (bib id) as a stable match point. The Marcive reports of outstanding corrections contain just the bib id and the affected heading, so we rely on that Voyager bib id to retrieve the records which need attention in the subsequent in-house work on those reports. However at this stage of the project The Women's Library records did not have a stable Voyager bib id. We were either accessing the data via a MarcEdit file (where the records had an LMU bib id) or on our Development Server (where the records had a temporary Voyager bib id which would not be carried across to the live system when the data was migrated). The data was in our Development Server while we experimented with various other issues, such as automatically generating holdings and items records out of bib record data or linking parent and child records, and the initial plan was to load the data into the live server only shortly before the opening date for The Women's Library @ LSE so that records were not visible to the public before the material was available for use at LSE. However if we sent the data to Marcive from the MarcEdit file it would have no LSE bib id at all, and if we sent it from the Development Server it would not have the same bib id as it would once it was in our live system, meaning the bib ids in Marcive reports would not correspond with records we wanted to edit on Voyager. After much discussion of different options our IT team loaded all the records to the live system, and then carried out a bulk suppression of The Women's Library records, having tested on the Development Server that we would be able to carry out a bulk 'unsuppression' later in the project.

The full file was then sent to Marcive for the data clean and towards the end of July we received 58,000 cleaned records, associated authority records and 11 files of reports. These consisted of unauthorized headings for corporate names, meeting names, personal names, series entries, subject headings, geographic headings and then a large number of headings which had multiple potential matches with LC authority records.

We carried out final checks on the data before IT loaded it into our live system. It is a busy time of year in the library with our Summer School students here, which meant it was not appropriate to take the system offline which would allow a quick load. However loading such a large amount of data onto the live system can slow it down, which again was something we wanted to avoid for our Summer School students. Instead the file was split into 4 smaller loads, and planned to be loaded overnight over a period of 4 days. We discovered during our initial authority control project in 2007 that loading the files and regenerating the keyword indexes at the same time practically ground the loading process to a halt, so the plan this time was to load the data without regenerating the indexes, with an awareness that this meant a short period in which the search facility used old indexes while bibliographic records contained new data. This would only affect staff working on the records as they were still suppressed from public view at this point. The data was then re-indexed by Ex Libris, our system vendor, as part of our Voyager upgrade in August 2013.

While the data was loading into the system we began examining the files of reports which Marcive had supplied with the data in order to decide on the priorities for this manual part of the process, and preparing instructions for the temporary member of staff who would be carrying out this work.

Aware that accessing the material through subject searches was of high importance we decided to deal with 65X fields first. These reports were also more manageable in size than the names-based reports. The unrecognised geographic headings report contained only 45 headings, and could be dealt with in a few hours, so this was a quick win. This was followed by the unrecognised subject headings report, which we are currently working on at the time of writing, and with 4590 headings we estimate will take 24 days to complete. In order to finish the subject related work we then plan to deal with outstanding 695, 696 and 697 headings before returning to the Marcive reports and beginning on the personal names report containing 9810 records. We have been able to employ a temporary cataloguer to work with us until the end of October, so we will continue to make as much progress through the subjects and personal names reports as we can, while keeping the unrecognised corporate names, meeting names, series names and multiple matches for future project work.

As I write (in August 2013) this project is still a work in progress. At the end of August all the records will be unsuppressed on our system ready for The Women's Library @ LSE catalogue to go live on 2nd September. Although there will still be some on-going manual work to do around authority control at this point, as outlined above, our work with Marcive means we are confident that the vast majority of authority records will have been cleaned through automated processing and that future records added to The Women's Library @ LSE will be authority controlled through our regular and robust authority procedures.