Ever evolving: metadata services and repository involvement at LSE

Helen Williams, Metadata Services Manager, The London School of Economics and Political Science

Back in 2012 I wrote an article for C&I entitled 'The past is a foreign country: transforming a bibliographic services team from copy cataloguers to metadata creators'.[1] In it I summarized how our traditional Bibliographic Services team cataloguers became what one kind blogger described as 'some kind of indispensable repository metadata ninja unit',[2] while still maintaining their usual cataloguing work. The article:

- outlined our background to the decision, and the preparation and practicalities involved in redefining our workflows;
- reflected on the successes and learning points of managing change and updating skillsets; and
- considered how we could remain at the heart of the Library's mission to support the University.

While 2012 is not entirely a foreign country to our 2017 workflows, there have been changes in the last 5 years, as you would expect.

Firstly, a couple of structural changes mean that instead of a combined Bibliographic Services team, who shared acquisitions and cataloguing work, we now have separate teams for Acquisitions and Metadata Services. The Metadata Services team currently consists of the Manager, 2 Assistant Librarians (one of whom is part-time), 2 Senior Library Assistants, and 3 Library Assistants. From this group of 8, half are involved in working on our LSE Research Online (LSERO) repository. One of our Senior Library Assistants oversees and manages repository work in Metadata Services, and our 3 Library Assistants combine work on LSE Research Online with cataloguing work. This is in contrast to 6 Library Assistants working on LSERO in Bibliographic Services days, but compensated for by our team no longer carrying out acquisitions work.

Secondly we have seen the workload itself changing. Before we took on repository work our number crunching suggested that the team could spend 21 hours a week creating repository metadata without seeing a dip in our current print cataloguing output (these hours were instead freed up from historic gift cataloguing). At the end of our first year we found our average weekly time spent on LSERO was close to this at 24 hours, but with peaks of 42 hours a week, and lows of 16 hours a week, flexibility had been the key to our success.

In 2013 we used Eprints to manage our REF submission and so the year was dominated by adding and checking, and double checking, metadata to support this, which we managed with dedicated support from a couple of members of our (at that stage, still larger) team who developed significant REF expertise. By 2014 we had the skill of juggling our cataloguing and LSERO workflows down to the fine art you'd hope for from a 'ninja unit', and our average weekly time on LSERO was still pretty consistent with an average of 23 hours a week being spent on LSERO metadata creation, shared across 3 Library Assistants, with the help of the Senior Library Assistant for repositories.

---

1. http://eprints.lse.ac.uk/46220/ Accessed 24th April 2016.
2. Playforth, Rachel, 'The value of cataloguing: CILIP Cataloguing and Indexing Group conference 2012', *The Toast in the Machine*, September 12, 2012. https://playforth.wordpress.com/2012/09/12/the-value-of-cataloguing-cilip-cataloguing-and-indexing-group-conference-2012/ Accessed 10th April 2016.

But as 2015 rolled into 2016 we began to feel as if we were regularly ‘up against it’ in Metadata Services, with all our time going on cataloguing or LSERO data creation, rather than any of the project-type work we were used to having on the go at the same time. The inbox system we use to manage incoming LSERO work seemed to be constantly overflowing and we were regularly asking our team to delay cataloguing for a few days so that we could have all hands on deck to clear repository backlogs. Scrutinising our statistics backed up my suspicion that while incoming cataloguing remained reasonably constant, we were seeing changes in the LSERO data. As Table 1 shows, while we were seeing a reduction in the number of citation-only entries being submitted to the repository, we were seeing a significant increase in the number of full- text entries being received. The metadata for these takes longer to create, both by virtue of dealing with the full text and due to the additional metadata requirements of recent years, such as funder and project details to comply with the RCUK policy on Open Access, date of acceptance, preferred Creative Commons attribution, etc., which meant that overall we were seeing a significant increase in the hours required to keep on top of this work.

	Average hours spent creating full text records each week	Average hours spent creating metadata only records each week	Average total time spent creating LSERO metadata each week
2014	13 hours	10 hours	23 hours
2015	21 hours	8 hours	29 hours
2016	31 hours	6 hours	37 hours

Table 1: Hours per week for the whole team

Our current expectation is that Library Assistants will spend approximately one third of their time on cataloguing, a third on LSERO metadata and final third on counter duties and team or Library wide activities. Counter duties and other Library meetings are fixed timetable slots, so those continued to happen, but LSERO and cataloguing time competed with each other and required a lot of flexibility and juggling in the team to respond to competing priorities, particularly on the weeks where the hours required were above the averages in Table 1. Furthermore, we were finding that although we didn’t have acquisitions work to do anymore, 3 bodies instead of 6 meant it was harder to juggle workloads and urgent requests. If 1 or 2 people are off in a team of 6, the work is still manageable, but when that’s out of 3 your team can be depleted very quickly and very unexpectedly, making the best laid plans hard to execute. While the LSERO work doesn’t pile up visibly in quite the same way as books on shelves, or boxes on the floor, we are aware of our role as service providers, both to the Library’s Research Support Services team who have overall responsibility for the LSERO service, and to our academics who submit their publications to the repository, so backlogs sit uncomfortably with us, and we were keen to find a more effective way of managing the process.

We knew that possible solutions could not involve reducing the incoming work, or increasing the people to do it, which doesn’t leave a lot of ‘wriggle room’, so instead we tried changing our service level agreements from a 3 working day turnaround to a 6 working day turnaround for most content, but with the promise of dealing with urgent requests in 24 hours. We also have incoming blog content to be archived each month, and have increased the turnaround time for this to one month, though we aim to have protected time each month to keep up with this as while the metadata is simpler, the volume of content is higher. Obviously this doesn’t reduce the amount of work that needs doing, but the longer turnaround time frame has given us a greater degree of flexibility, which in turn has reduced the need to regularly adjust workflows and get the whole team focused on LSERO at the expense of cataloguing. This makes for a less pressured feel to the workflows as well as making them easier to manage.

In 2016 we successfully bid for some Higher Education Innovation Funding (HEIF) to increase the visibility of research outputs by extending the archiving of PDF versions of LSE blog posts in LSERO. The metadata applied to this content when PDFs are archived in LSERO enhances their discovery on the web, and facilitates exploiting them as a research resource. It also supports the Library's strategic objective to collect and preserve the intellectual output of the School, by ensuring that outputs such as blog posts, which fall outside more traditional publishing channels, are still preserved, and made discoverable.

The Library was already archiving selected blog outputs (chosen either by blog editors who send content for addition, or individuals who request inclusion of their post) but this was not a systematic or comprehensive approach. The project began with LSE's top 10 public facing blogs, and the daunting prospect of systematically adding somewhere between 5000 and 11,000 records to LSERO, in a 3-and-a-half-month period, in order to retrospectively archive all content from those blogs. At the time of writing, towards the end of April, we are just ahead of schedule with 3700 records already added to LSERO and the hope of retrospectively archiving content from a select number of LSE's other 62 blogs as well.

The HEIF blogs project is being supported by a team of 3 temps and will finish at the beginning of July so it has been important to consider how we can sustain this increased amount of blog archiving work alongside all our usual incoming work. A secondary part of the project has been working with the School's IT team to automate the creation of PDFs, which has been successful, and to auto-populate as many of the metadata fields as possible, which we are still working on. If both of these are successful then we would anticipate a significant reduction in the staff time required to create archived versions of blog posts and to assign appropriate metadata, allowing us to embed comprehensive archiving of the top 10 public facing blogs into business as usual work.

This may sound optimistic, given the increased volume of work outlined earlier in the article, but the School, including a team in the Library, is currently implementing a Current Research Information System (CRIS), which we anticipate will significantly change existing repository workflows and reduce the amount of manual metadata entry required, thereby freeing up staff time to deal with other unique LSE content, such as the blogs. However, since the project is not yet complete it remains to be seen exactly what the impact of the CRIS will be, and whether our current repository workflows will look like a foreign country in a year's time.