Introduction

The future of metadata and cataloguing is a hot topic for me at present. I work alone at my University library as the sole metadata practitioner, and see hundreds or thousands of records pour into my catalogue all the time. I make global interventions to these; modifying a URL here, adding in “linkyMARC” there. I check and update authority data and keep a local authority file up-to-date. I even get to catalogue something once in a while (exciting things like a British Standard, or a Chinese language textbook!). What I don’t get to do is meticulously check over the records entering our LMS, and then catalogue the items one at a time, ensuring that all the full stops are correct or that all the subject headings actually reflect the content of the work. At a certain point, given the sheer amount of supplied metadata, I have to accept what I get and make it accessible to our end users. This may mean leaving records alone and not making enhancements that I otherwise would make had I catalogued the item myself. It means trusting that the data conforms to library standards, e.g. accurate Library of Congress Subject Headings (LSCH), accurate name headings, accurate Dewey classmarks, even accurate transcription of titles (in large e-book packages, assuming the title is even supplied), and accurate ISBD.

I was therefore, excited to be invited to a Jisc “metadata day” at Senate House, in London on 22nd May. I was invited as a representative of the “Mercian metadata special interest group” part of the Mercian collaboration. The Mercian collaboration “is a professional network formed from SCONUL higher education libraries based in the East and West Midlands of the UK”. My task was to represent the interests of the 23 HE organisations within that collaboration. These institutions cover a wide range of HE organisations, from the Open University to very new Universities such as Bishop Grosseteste. Many of these organisations are like Aston with one metadata specialist; others have small teams (and large research collections) and some have no metadata specialists. Yet each library still faces a set of the same or similar issues, such as what titles and resources do we have access to and do we have records for these? Are there records for titles we don’t have access to? Are the URLs still up-to-date, and what do we do if they need bulk changes? How do we deal with the legacy collections and metadata? And finally, most of the metadata specialists in the collaboration agree that vendor supplied meta data leaves much to be desired.

The meeting in London was conducted under Chatham House rules, so in the remaining part of this paper I will avoid relating the details of the very productive discussions held on the day. The day concluded however, with an understanding that there were two main issues around UK data supply, these being the standards for metadata quality, and licences for use and reuse of this data. The group also agreed that these issues were solvable and made a commitment to do so. What I propose to do in the rest of the paper is to reflect on the kinds of issues raised by NBK and UK data supply, as well as look at some of the standards which may emerge from these discussions.

2. It is worth noting that I volunteered to work on the metadata quality task and finish group.
Metadata quality

As already mentioned, most of the metadata at Aston is derived from 3rd parties and vendors. This data is in the form of shelf-ready MARC records for both print and electronic single purchases, OCLC WorldShare records for some electronic subscriptions, and vendor records for various other subscriptions, such as ProQuest’s academic complete. Only a very few items are catalogued in house. These tend to be foreign language materials, British Standards and some audio-visual materials. The quality of the data from these sources is quite variable, from excellent to poor, with some vendors being much better than others.

Whose standard? Which exchange format?

Since I started in post in November 2016, Aston has technically been an RDA library. RDA is, for me, my lingua Franca for cataloguing. I was taught to catalogue in AACR2, an education I am very grateful for, but for almost my entire professional career RDA is the standard I have used and know best (notwithstanding recent changes). The promise of RDA is something I am excited by as a practitioner and librarian. The idea we could make knowledge more available and more discoverable is the reason I am a cataloguer in the first place. So, when asked by our acquisition librarian, “Are we an RDA library?” I naturally replied; “Absolutely”, and therefore our suppliers need to provide an RDA record. But do 3xx fields an RDA record make? And where do FAST headings fit into this picture? Or publisher subject headings? Are the LCSH correct, accurate and reflective of the content? If RDA is more about relationships and identities, where are these in our vendor records? And then we enter the thorny issue of who owns the record rights. What changes can I make to them? Where and how can I share these records? What will the record look like in a reading list? A union catalogue? A discovery layer?

I raise these questions quite deliberately, as I do not have answers to them. I would love to know precisely what data I am getting, how I may use and reuse it and what it can do to enrich the information universe which I and our users inhabit. I would like to know that the users are able to find, identity, select, obtain and explore the resources they need or want, and that this is facilitated, made easier and possible by the metadata efforts I and my cataloguing colleagues are making each and every day. But I do not have this confidence and I do not know if the efforts I make on a day to day basis are adding the value they should.

This is not exclusively, or even primarily, a problem of derived data. Machine transcription of the primary elements of a metadata description strikes me as eminently sensible, particularly as this data already starts life digitally. Titles, authors and editors, page numbers, dates of publication and copyright, contents notes and descriptions of content are all necessary, but as a metadata specialist I’m interested in the relationships and the entities. And for those exploring the information universe this is where metadata specialists add real value. But MARC records are not really cutting it. It is true that we have relationship designators and links to authority files like NAF\(^3\) and VIAF\(^4\) but when we code our MARC records as RDA, what we mean is that we’ve added some elements and rebranded some others. As an exchange format MARC has had a long and illustrious career. However, it is no longer really required for exchange, and as an encoding schema for RDA it isn’t fit for purpose. What I really want is to utilise the potential power of RDA as schema in an encoding language that understands this data. I want to be able to push the data out to systems that understand RDA, and is accessible to our users wherever they encounter it on the web.

---

3. The Library of Congress Name Authority File
4. OCLC’s Virtual International Authority File
The future

At the end of the metadata day two task and finish groups were established, one to address standards and metadata quality, and another to address use and reuse of data. I have left off the discussion on reuse here but there will need to be a robust discussion around this. As we use more and more data from 3rd parties we still need to do our jobs as librarians, serving our immediate users as well as the wider community. Hopefully an open metadata licence can be created that is acceptable to both vendors and librarians. I stepped forward to help with the metadata quality issues however, as this personally interests me. What I want to try to achieve if possible is a set of standards that can be feasibly supplied by our data suppliers and which meet our needs as a library community. I do not know what this might look like at this time, but with Jisc asking the question and NBK giving some momentum to a UK standard, can we break the inertia around MARC? Can the BL, HE libraries, Jisc and Cilip come up with a UK encoding scheme for RDA that truly unlocks RDA’s potential? Can we separate out recording metadata elements from the encoding and exchange format? And can we have a UK policy around the standard of data, or a UK “application profile” for RDA?

I will be working with Jisc on the metadata quality task and finish group. The success of this will depend on community engagement. What we need to know is what we want and why. To this end I really encourage anyone who thinks these issues are important to contact me or others involved in the work, with comments, suggestions, criticisms or any ideas. As far as I am concerned no idea is too bizarre or revolutionary. We are all working everyday with metadata and we are the experts. Let’s try to effect the change we want to see, rather than having this change imposed on us by a world that has already moved on.