Background

The Future of Cataloguing – it’s a massive subject. Where do you start? What do you cover? What do you omit? Can you look back (you can never look back)? Do you cover data processing, computer programming, or focus on ‘traditional’ cataloguing activity – record creation, metadata sharing, standards, quality? Do you limit it to library cataloguing, or do you include the wider world of metadata and the semantic web? What are the metadata boundaries, and how far is its reach? Where do you start when it is suggested that CIGS (the Cataloguing and Indexing Group in Scotland) hold an event on the Future of Cataloguing?

Planning

CIG Scotland had been thinking of running a Future of Cataloguing event for some time, but a suitable format eluded us. It was agreed that the standard seminar-style event would not be appropriate: attendees would have a wide range of ideas regarding the future of cataloguing (catalogues, cataloguers, cataloguing, metadata, standards, formats, new technologies, new environments etc.), and although six or so speakers on the topic would be sure to be interesting, it wouldn’t cover the full range or depth that the delegates would want to explore. An alternative event format was suggested: the world café. This format is well established for discussion orientated meetings and was sold to the CIGS Committee as a “not-so-speed dating for cataloguers” event. The aim of the world café style is to engender open discussion between delegates across a range of topics, working in small groups. For the CIGS FoC world café, it was proposed that around six tables, covering six different topics on the future of cataloguing, would capture what we wanted to achieve. Each table would have its own facilitator to start the conversation, move it along, collate points, and ensure that all delegates would have a chance to share their thoughts. Each session would have a time limit, and there would be the chance for delegates to attend each table. However, when working out the session timings, it became clear that this would result in very short sessions with insufficient time for discussion, or a very long – and exhausting - day. Therefore, we decided to settle on four sessions across six topics and tables, with delegates able to decide which topic they wanted to join over the course of the day.

The next task was to find facilitators who would be happy to collate notes, keep the discussion flowing, ensure that one person did not dominate the table, to keep the conversation to the table topic, and to give a summary of discussions at their table at the end of the day. No small task, but appropriate inducements in the form of refreshments, as well as being a vital part of an interesting day, were enough! CIGS were extremely lucky in being able to identify a range of subject experts who were willing and able to act as FoC facilitators: the day would not have been half as successful as it was without them.

We decided that the event would benefit with a presentation at the beginning of the day; this would set the scene, motivate delegates and prepare them for the discussions ahead. CIGS were extremely fortunate in attracting the services of Alan Danskin, Metadata Standards Manager at the British Library. Alan has been key to the development of the BL’s metadata strategy, which we agreed would be a great introduction to the issues surrounding the future of cataloguing.

The World Café

Alan didn’t let us down, and provided a fascinating insight into the drivers behind the British Library’s new Collection Metadata Strategy (2019-23) Foundations for the Future, which was a perfect scene setter for the day. Alan started off by asking delegates how many of them had been to events previously that covered the future of cataloguing; not as many people as Alan had expected raised their hand. He asked what had the audience been doing, and a lone voice piped up, “Cataloguing!”. That set the tone nicely for the day.
Following Alan’s scene setter, delegates were asked to join their first session. The tables and topics on offer were:

1. The future of authority control, identifiers, linked data – facilitated by Diane Pennington, from the University of Strathclyde
2. The future of subject analysis and controlled vocabularies (LCSH, FAST, TGN, etc.) – facilitated by Will Peaden, from Aston University
3. The future of content standards (& data models) – convergence, divergence (RDA, ISBD, DCRM, ISAG(G), etc.) – facilitated by Gordon Dunsire
4. The future of encoding models and formats – convergence, divergence, emergence (MARC, BibFrame, DC, Schema.org, EAD, etc.) – facilitated by Jenny Wright, from Bibliographic Data Services
5. The future of data quality: data sources, supply, volume, management – facilitated by Alasdair Macdonald, from the University of Edinburgh
6. The future of cataloguing practice: what will the cataloger or metadata manager of the future look like? – facilitated by Alan Danskin, from the British Library

Delegates had 35 minutes at each session, whilst facilitators had 5 minutes to collate each session’s discussion, and further collate at the end of the day. The four sessions were intersected by a much needed buffet lunch.

Due to the diverse range of perspectives, opinions and backgrounds, and the complexity of the topics covered, clear answers and conclusions were not expected. However, delegates rose to the challenge, and many challenging and thoughtful insights were captured by our excellent facilitators over the course of the day. These are summarised below, which, due to the nature of the event and the multiple sessions, may appear ad hoc, overlapping and tangential, but provide an interesting overview of the issues, questions and challenges facing the cataloguing and metadata community today.

Summary of discussions

Each table had a selection of questions, which had been gathered from all facilitators and CIGS Committee members prior to the event. These were aimed at starting and continuing conversations, allowing the discussions to flow naturally.

Table 1 : the future of authority control, identifiers, linked data

Questions included:
- How do we handle ethical and legal (GDPR) considerations of authority control?
- What identifiers are proving most useful?
- Are our library specific authorities too much of a silo, or can they teach and influence the direction of authority management?
The discussions raised concerns over the different vocabularies that are available; how should these vocabularies be linked together? Which vocabulary is useful for which audience? How do you choose the appropriate vocabulary for your collection and audience, and how does it relate to other collections, other audiences? There was much curiosity regarding this topic, but also concerns on how to implement solutions, what would make sense for the cataloguer on the ground, or indeed if this was necessary.

It was questioned where linked data was actually linking to and from, especially in the MARC environment, which is still the standard environment for libraries and not easily connected to the linked data environment, highlighting the challenge – and potential - of bridging traditional metadata silos with the semantic web.

It was agreed that there is a need for linking our data more seamlessly and semantically, but questions over what sources and authorities and data could be trusted were raised. There are multiple options for managing authority control and linked data, but it is still not clear how these can be connected effectively in current systems. It was highlighted that authority standards are used in different ways in different institutions – and despite the presence of the standards, these are not always adhered to systematically or consistently. Hanging questions included: in a linked data environment would these standards still be relevant; should there be more standardisation; or were standard vocabularies and authorities becoming less important? (Which links nicely to this comic from xkcd: https://xkcd.com/927/)

Table 2: the future of subject analysis and controlled vocabularies (LCSH, FAST, TGN, etc.)

Questions included:

- Is subject analysis still a worthwhile exercise?
- Are subject indexes too Western focused, and is this a problem?
- Can library standards impact on other non-library standards on the web?
- Are FAST headings a good compromise between detailed subject analysis and heading construction, and crowd sourced subject headings?

The discussions agreed that subject analysis is a good thing, but that subjects needed to be more visible and embedded in discovery layers to realise their value and full potential. The “aboutness” that subject analysis provides is currently lost, as modern discovery layers, etc – and indeed users - focus on the use of keywords rather than utilising the value of subject analysis and subject vocabularies – has it not always been thus?

It was questioned how much subject analysis is carried out today, considering the amount of records that are created – issues of duplication, variance, incompleteness, skewed results, skewed analysis, misleading the user.

It was agreed that LCSH (Library of Congress Subject Headings), despite its global reach, had a number of challenges for the modern day cataloguer: it is still very American-centric; it is slow to change; it is difficult to influence for most institutions and cataloguers; and on the whole is not sensitive to changes in terminology, modern usage, and prone to archaic, “non-PC” terminology – one of the handicaps of such an extensive and comprehensive behemoth. Despite these challenges, it was noted that better use of LC subject headings in modern discovery layers would be of value to users, although it was queried if LCSH was the correct lens to discover content, for the reasons noted above, and whether more agile, modern vocabularies were better suited to the modern discovery environment.

A discussion regarding FAST headings (Faceted Application of Subject Terminology) revealed that some delegates considered there to be a lack of unity in the application of the headings; some delegates viewed FAST as a dumbing down of subject analysis; that FAST headings resemble a keyword approach, and don’t offer detailed or effective subject analysis. Opportunities to link FAST with other vocabularies and data were highlighted, with use of IDs and URIs in a linked data environment – one of the benefits of adopting a modern subject vocabulary such as FAST. Others questioned how more external links would work with existing systems in use, highlighting some of the linked data concerns raised above.
The ethics of subject headings was also discussed, and also the extent to which end users identify with and use subject headings, or rely on keywords for their discovery – is this a failing of subject analysis and vocabularies, their lack of visibility in current discovery systems, information literacy, or simply user preferences and the power of modern indexing and search techniques ... or indeed has it ever been thus?

Table 3: The future of content standards (& data models) – convergence, divergence (RDA, ISBD, DCRM, ISAG(G), etc.)

Questions included:

- How important is it for cataloguers to understand the data models underpinning the content standards?
- Is ISBD still relevant?
- Will RDA become a standard beyond the library world? Will RDA take hold of the whole library world?
- Are developments in the standards (such as RDA) fit for purpose? If so, how? If not, why not?
- Is there any prospect of a uniform standard applicable across the information world?

As can be imagined, this was an interesting table to attend. There was a mixed response to whether or not cataloguers need to understand the underlying data models for content standards. There were arguments to be made for both sides: on the one hand, knowing about data models can help cataloguers with trouble shooting, data manipulation and how to interpret cataloguer’s judgement. On the other hand, it was noted that many cataloguers simply require an input form, and there is no need for burdening cataloguers with technical or abstract data modelling. The discussion then moved onto the more cerebral notion of what a cataloguer might be, as the term can cover a range of skills. It was noted that a range of users and environments might well need a range of data models, or application profiles at least, such as RDA offers. Then the inverse question was asked, how important is it for data modellers to understand cataloguers? There appeared to be agreement that it was important for systems support to understand why cataloguers do what they do, and that underlying aims of cataloguing needs to be understood beyond the cataloguing community, and vice versa. Cataloguers use their skills to describe content in order for users to discover, identify and access content; it is the content that is important, not the carrier or format. It was also noted that – some, all? - cataloguers need to be more technically minded, and more adept at data processing and manipulation, semantic web possibilities, and using data tools and techniques such as MarcEdit, OpenRefine and others.

The next discussion was regarding the relevancy of ISBD. The practical point of view is that legacy data is still important – good metadata endures - and the library / cataloguing community will have to live with that legacy and subsequent hybrid data for a long time to come. It should be possible for catalogues to co-exist with different standards. However, there is a move away from descriptions to relationships between entities when describing items, and a recognition for the need for consistent authorities to support this effectively. ISBD is still useful for consistent displays of metadata.

It was acknowledged that RDA would not be adopted across the web as the one single descriptive data model, and that is not its aim. The consensus was that RDA was the most effective global standard for describing library and cultural materials, but that a lack of a common encoding model was a limitation to realising RDA’s full potential. It was noted that RDA is making progress outwith the library world, for example, in describing museum collections, and synchronising with other data models through the Library Reference Model, with RDA vocabularies supporting linked data applications. Archives appear to be in a different situation, with delegates suggesting that they are thinking more about linked data possibilities. It was suggested that linked data would be an expensive method of describing any collection in current library or archival environments, especially if this was done as a solo exercise.

The question was also raised about the link between content standards and information literacy, as well as the role of cataloguers and trusted metadata against the rise of fake information and fake metadata.
Table 4: The future of encoding models and formats – convergence, divergence, emergence (MARC, Bibframe, DC, schema.org, EAD, etc.)

Questions included:

- How do we break through the impasse of LMS suppliers only supplying systems for formats of the past?
- Does it matter that Bibframe does not use the same data model as RDA?
- Will Bibframe deliver a new content standard to replace MARC?
- Is MARC really a dead man walking? Does it really need to die?

It was noted that MARC21 is the working format for most traditional libraries. Libraries are bound by their legacy systems and the limitations of the marketplace. There might be complaints about the development of systems with other standards, but the future format beyond MARC is still not clear, with the status of Bibframe and other emerging encoding models still embryonic. How can LMS and metadata management systems be designed with an uncertain landscape for modern encoding models? The discussion moved around to leading the necessary developments from within the library community – this links back to discussions on Table 3 regarding cataloguers needing to be more technical. It was also noted that no one else would do the system design we require for the library community, so it would have to be specified by those who understood what was required. Legacy issues with using MARC in the modern environment were highlighted, which was no surprise to delegates. However, future formats and encoding models would have to enable efficient and flexible data entry, as more and more items need appropriate metadata attached to them. New models will need to reduce level of duplication, as well as support non-expert entry. It would have to enable data to join up seamlessly with other data forms, and also enable that data to be extracted and manipulated: the consensus was for clean, granular, interoperable, adaptable and neutral encoding models, able to realise the potential of modern data models, such as RDA offers.

This table also noted that there was currently a limited business and use case for linked data in the library environment. A question was asked regarding the impact of the background format on collection management – how do the two link together in a new system?

The final comment made on this table was: Free the data (out of the MARC box)... for example, using RIMMF, which is freely available and designed to exploit the full potential of RDA. (Other format neutral systems are [not] available.)

Table 5: The future of data quality – data sources, supply, volume, management

Questions included:

- What can we do with our data derived from outside sources? How can we share it, what may we do with it?
- Does the sheer volume of e-resources render traditional cataloguing obsolete?
- Are there ways to improve vendor derived metadata?
- What are the most significant problems that you face? Are they related to metadata quality, speed of delivery, accuracy of content coverage?

This was a table which was full of lively discussion, revolving around vendor-supplied data, and how that can - and should – be improved. It was noted that there are enough issues at present when working with MARC-centred data. There were discussions as to what is good enough when considering data quality. Records supplied for bulk loading were mentioned, along with the associated issues of quality, consistency, and ensuring the correct content coverage. Resources at the libraries are spread thin, so it is not possible to check every record that is entered into the catalogue – potential for bulk data processing, analysis, correction? If not addressed, this can lead to “lost” content due to the metadata not being discoverable or leading to the correct content (e.g. simple spelling mistakes, incorrect URLs, plain bad metadata, poor authority control etc.). On the plus side, the scarcity of cataloguer time has led to more collaboration with regards to resource and metadata sharing.
It was noted that legal deposit libraries face different challenges to academic libraries, for example, using BNB numbers rather than ISBNs for matching purposes. This brought the conversation around to the idea that one size does not fit all, even though that seems to be what is received from vendors. There was a demand for better quality and customisation of records that libraries are purchasing. There is an increased burden on staffing, especially in cataloguing and systems, and an ever-increasing workload. There are many uncatalogued items (such as those without ISBNs, or lack of other identifiers), which are not able to be tackled due to lack of staff and effective data processing solutions. However, there appears to be an appetite for upgrading legacy data, along with an unsurprising ethos of sharing data, given impetus in the UK on the back of recent NBK developments and community involvement. This gives some optimism for future cataloguing endeavours, leading to more effective workflows, better quality metadata, and an improved metadata ecosystem.

There was one final suggestion: if a library is contributing to a consortial cataloguing system, and paying a subscription, should that library be getting credit back for uploading records? A system similar to using solar panels and generating electricity for the National Grid could be employed. That would an additional incentive to increase cataloguing output.

Table 6: The future of cataloguing practice – what will the cataloguer or metadata manager of the future look like?

Questions included:

- What will be the future function of the library catalogue?
- How do we attract people to become cataloguers of the future?
- What opportunities do you see for a more integrated, collaborative approach to metadata creation and maintenance?
- How do you explain the purpose and potential of metadata to decision makers in your institutions?

There was much discussion around the subject of change: changes to standards (AACR2 to RDA, for example); changes to systems (slow to change, but there are changes); changes to expectations, of users, of institutions; changes in cataloguing resources (there are fewer cataloguers now, on lower salaries). However, at the same time, there is a forging of new relationships.

Another discussion posited that librarians and cataloguers understand the value of metadata, what it can do for the user, and how far it could help with discovery and other reports, but it was not clear whether the systems underneath the metadata layer are able to utilise the richness and granularity afforded by good metadata effectively. This can prevent us realising the underlying value of and investment in our bibliographic metadata and controlled vocabularies.

With the rise of communal knowledge bases, vendor supplied data, etc., it was questioned whether libraries and librarians are losing control of the metadata. There are risks associated with having central sources of metadata, such as the de-skilling of cataloguers, at a time when deeper cataloguing knowledge is required to understand the underlying models behind cataloguing standards. What replaces this skill? Will the resulting metadata deliver the expected results? There is a wide skills base at the root of any cataloguer, but it is clear that more collaboration between cataloguers is needed. There is also a need for more data manipulation and management, which is outwith the skills range of the standard cataloguer.
Q&A

Following the summing up of the day’s discussions from the facilitators, the floor was opened to questions and comments. The need for effective training was raised. Is training for cataloguers provided at Library School appropriate to the future needs? Cataloguing is normally an optional course in a post-graduate degree, and there is a lack of focused training, especially on the technical side, which is commonly learnt on the job. However, it would be difficult to plan technical training on an academic course, as it would be difficult to keep such training neutral and non-vendor focused; the software and hardware is constantly changing – unless training tools such as RIMMF were employed? On a different level, it was suggested that there needs to be in-service training for copy cataloguers, especially in public libraries. The lack of cataloguing trainers across the UK was noted.

Leadership from the community was also highlighted. It was also suggested that individual libraries need to take a leadership role in the development of solutions, metadata standards, systems development, and data quality, and take it forward. One example would be working with vendors and suppliers to improve the cataloguing modules of different Library Management Systems, something that SCURL is actively pursuing in Scotland. It was noted that the British Library are doing a good job in this area too, but that the historical system of libraries working in silos was a barrier. Regional consortia are one way of breaking this mould, such as SCURL’s SHEDL model, or the White Rose Libraries, but the feeling was that there needs to be more community and consortial involvement, with more libraries joining consortia, and cross-consortia collaboration too.

Finally, it was noted that the value and importance of metadata is recognised much more widely in the modern information age; it is prevalent everywhere and in all spheres! It is no longer the sole purview of libraries, museums and archives – it never was, but even less so now. The day’s discussions had focused inwards, on metadata in libraries, but, the future of cataloguing and management of metadata may not lie in libraries; a cataloguer’s skill set is extensive, and these skills are very transferrable from libraries to other industries. There are vast swathes of metadata work outwith libraries, and cataloguers and the cataloguing community need to consider how we can influence metadata creation, use and re-use across all aspects of modern society.

Conclusion

The Future of Cataloguing was a sell-out event, with very positive feedback from delegates. There was a lively buzz across the whole day, with some tired delegates – and facilitators – at the end of day. Due to the nature of the event, delegates provided a much more active and fruitful contribution to the day than would be expected at a standard seminar, and this was seen as a very positive outcome. As to the points that were discussed, more questions than answers were supplied. This was expected, as the aim of the event was to spark thoughts, debates and new ideas. It helped that the facilitators revealed their passion for their subject areas, and this spurred on the discussions, leading to a very lively, energetic buzz at all the sessions.

CIGS is very happy with how well the day went, and we are pleased to see that the format has inspired a series of events across the country, notably the Mercian Collaboration and CIG. With the success of the event and the positive feedback, CIGS aim to run the FoC world cafe again in the future, possibly as a regular event, and there are many other topics to include and discussions to be had. We will work with colleagues across Scotland and the UK to consider how we can take the outcomes of this event and others forward, and how the community can address the issues highlighted in the discussions on the future of cataloguing.

CIGS would like to express their deep-felt thanks to all the facilitators, as the Future of Cataloguing world café could not have been a success without their energy, enthusiasm and expertise – there were some very tired, though we hope very happy, facilitators at the end of a very successful day.