Work in Progress: The Linked Open Bibliographic Data Project

Anne Welsh, Dr Antonis Bikakis, Natalia Garea Garcia, Simon Mahony, and Dr Charles Inskip, UCL Department of Information Studies and Dr Mira Vogel, UCL E-Learning Environments.

Background

The publication of the BIBFRAME Primer (Miller et al, 2012) provided cataloguers with a glimpse of the new cataloguing model proposed by the Library of Congress to replace MARC. Prompted by the findings of the US National RDA Tests, the Bibliographic Frameworks Initiative sought to find a structure that would allow for the linked data capabilities of RDA while accommodating the legacy data created in libraries in MARC from the 1960s – 2010s.

As such, BIBFRAME is the fourth major model with which cataloguers in the Anglo-American community are required to familiarise themselves in order to understand modern cataloguing and meet the needs of the job market, and, therefore, the fourth model that educators must teach beginning cataloguers (Table 1).

Card / Dictionary Catalogue	Paper-based system; entry points and length of description limited by concerns around physical space (size of card / page and size of catalogue cabinet / shelving unit). Each record describes the item in the cataloguer's hand entirely – its intellectual and physical contents.
ISBD and AACR in MARC	ISBD, AACR and AACR2; built on principles of the card and dictionary catalogue; monolithic record structure in which information in fields is contextualised and explained by reading the entire record; inherited the card / dictionary catalogue's description level, covering both the intellectual and physical contents of the item in hand.
FRBR in RDA	Although much is inherited from ISBD, AACR and AACR2, the principle of RDA is a version of FRBR's WEMI model: Works, Expressions, Manifestations and Items are catalogued separately and linked to each other (and to the appropriate WEMI of other records); moves away from the monolithic record, favouring instead field-level links between records, aiming for the linked data environment. Currently constricted by systems based on the MARC format, which does not accommodate RDA's linked data capacities.
BIBFRAME	Linked data solution proposed by the Library of Congress; based on RDF; data model proposes Work and Instance, allowing for simplified analysis on the part of cataloguers and, significantly, the construction of RDF triples – a foundation for linked data in the wider web environment. Many triples in BIBFRAME incorporate Annotations – pieces of data that provide information about the Work or Manifestation, such as creator, publisher or ISBN.

Table 1. Cataloguing Models Anglo-American cataloguers learn in 2015

The tension between teaching cataloguing principles and practical cataloguing (Pattuelli, 2010), the difficulties of accommodating cataloguing within the curriculum (Robinson and Bawden, 2010) and the need for a longer syllabus to be taught within a fixed time-frame (Welsh, 2013) have been well-documented. Alongside these issues, an awareness of the

---
Welsh, A., Bikakis, A., Garea Garcia, N., Mahony, S., Inskip, C. and Vogel, M. (2015) The Linked Open Bibliographic Data Project, *Catalogue and Index* 178: 15-19.

scarcity of visualization tools that students can use without advanced technical skills led teaching staff and researchers at UCL Department of Information Studies to scope a project to create a teaching tool for BIBFRAME.

Project Aims

Funded by an E-Learning Development Grant from UCL E-Learning Environments, the aim of the Linked Open Bibliographic Data project (UCL Department of Information Studies, 2014-15) is to develop a BIBFRAME dataset as an Open Educational Resource, which will help students to learn the new standard in an interactive way, and at the same time become familiar with linked data concepts and principles, RDF and other state-of-the-art web technologies. An important aspect of the project is its interdisciplinary nature, combining methods and techniques from Library Studies, Digital Humanities and Computer Science.

Another key feature of the project is collaborative co-learning with students – a concept in Higher Education that is becoming embedded in teaching practice. At UCL, it is expressed through the Connected Curriculum (UCL, 2014-15), which underpins our Teaching and Learning Strategy 2010-2015 (UCL, 2010) and is in line with the UCL 2034 Strategy (UCL, 2014), which advocates student involvement in research at all levels of study.

Within the Linked Open Bibliographic Data project, there are two forms of student involvement. The first is our Student Systems Developer, Natalia Garea Garcia, who is currently working with Anne Welsh (Project Coordinator) on data selection and Antonis Bikakis (Project Lead) to develop the product. She is also advised by Simon Mahony on what is necessary to make an Open Educational Resource – a tool that can not only be used in the core and advanced Cataloguing modules at UCL, but also by other educators and learners. Later stages of the project will see this work intensify and Simon's expertise will be crucial.

Another key activity in which Natalia is involved is product testing, and in this she works with Charlie Inskip and, indirectly, with Mira Vogel. So far, Charlie has held focus groups with UCL MA LIS students from the class of 2014-15 following a session in the core Cataloguing module on RDF led by Antonis and Natalia. This has provided us with feedback that the tool is needed to assist a non-technical audience in grasping the power and significance of RDF, and, therefore, of BIBFRAME. Although the class was highly interactive, with well-thought-out activities to guide students through creating and visualizing RDF triples, there was a high cognitive load in getting to grips with software designed for use primarily by computer programmers. This result was positive, as it verified module coordinator Anne Welsh's instinct which initiated the Linked Open Bibliographic Data project in the first place: that something designed specifically for those with no programming skills was necessary to maximise learning and understanding of the BIBFRAME model.

Focus group participants are another important group of student collaborators. Students from the Advanced Cataloguing module are working with Charlie and Natalia to test both the concepts used in the tool and the tool itself.

All of these students – Natalia and the product testers – are meeting the Connected Curriculum's 'seven dimensions of connectivity' (Figure 1) and are collaborating with the project's teaching and research staff in creating the new knowledge needed to complete the project and create the Open Educational Resource.

Later in the year, we'll also be issuing calls for practitioner focus groups, and hope that members of the Cataloguing and Indexing Group will be among the first to step forward and contribute. The contribution of current cataloguers is essential to make the Open Educational Resource as relevant as possible to the community as a whole.

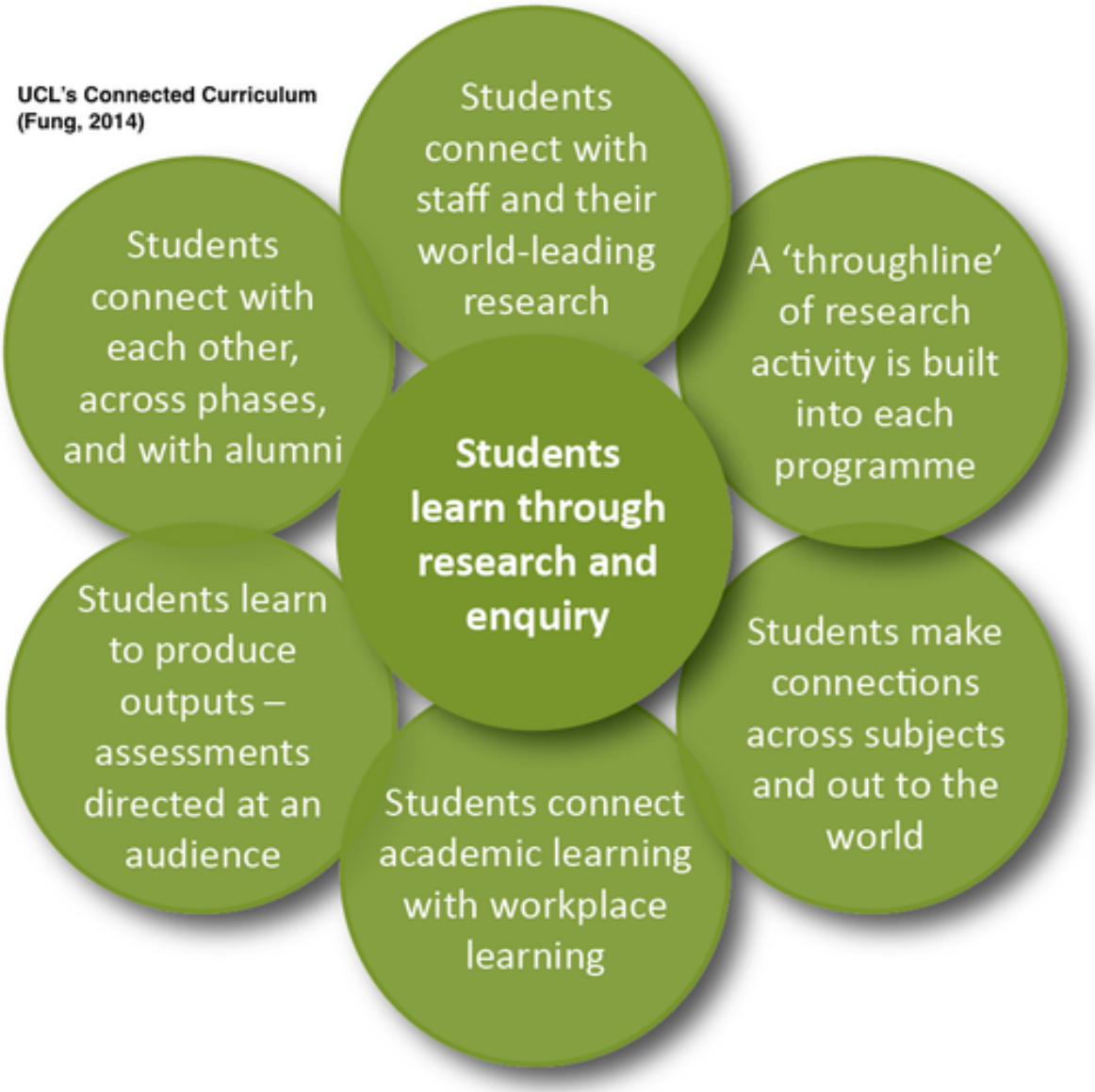

Figure 1. Seven Dimensions of Connectivity, http://www.ucl.ac.uk/teaching-learning/strategic_priorities/connected-curriculum ; full paper by Dr Dilly Fung currently in press. Reproduced with permission.

Building on Existing Knowledge

As background to their project work, and as an important part of their own individual learning in Advanced Cataloguing, Natalia and the product testers have tried RIMMF (RDA in Many Metadata Formats) (MARC of Quality, 2015) to see how it visualizes RDA. The Jane-athon data (R-balls, 2015) was released just in time for us to play around with it in class, and students also created their own sets of WEM (Work, Expression, Manifestation) data, using examples Anne pre-selected for them, including a paperback and hardback of the same

Welsh, A., Bikakis, A., Garea Garcia, N., Mahony, S., Inskip, C. and Vogel, M. (2015) The Linked Open Bibliographic Data Project, *Catalogue and Index* 178: 15-19.

book; the various editions of AACR, and a proof copy of a book and its final published version.

As well as consolidating their knowledge of RDA (building on their cataloguing policies submitted for assessment in the core Cataloguing module), and providing a neat tree structure of relationships, creating data on RIMMF highlighted issues for students in applying the Expression part of the WEMI model. The intellectual challenges of WEMI have been documented elsewhere (Parent, 2014), and we did discuss them in the core Cataloguing class practicals, but creating records on RIMMF provided students with practical examples. What exactly *is* the status of a proof copy? In theory, it's simply a different Manifestation of the same Expression as the final published copy, but does that hold true when there have been so many changes and additions between the proof and the final published version (de la Mare, 1952a and 1952b)? Similarly, there was a productive discussion around the many different releases of AACR (Joint Steering Committee for AACR, 1967; 1978; 1998; 2002), although that was easier for students to resolve. Certainly, after tackling these, and other examples, we could all see the pragmatism of BIBFRAME's simplified WI (Work, Instance) model. The binary choice of whether something is a Work or an Instance appeals.

From a pedagogical point of view, it was interesting to see which items caused issues for beginning cataloguers, and it was noticeable that the two students with workplace RDA experience had far less difficulty in reaching conclusions. However, as Deborah Lee (2014) has pointed out before, working with RDA does force cataloguer judgments at many turns because of the wide range of options built into the standard. Certainly, the more experienced students were confident in making *their* decisions, but were not dogmatic in forcing their points home.

From Natalia's point of view as Student Systems Developer, creating records on RIMMF as part of a class with other students has given her food for thought, both in terms of the structure and presentation of the tool she is building with Antonis and Simon. As Project Coordinator and Project Designer, it is important to Anne and Antonis that the project's offering be familiar enough to the cataloguing community to be useful without creating too great a cognitive load to learn how to use it.

Project Outcomes

So, you may well ask, what will our Open Educational Resource for Linked Open Bibliographic Data look like? The honest answer is that it is too early in the process to be clear on presentation: there is much development work for Natalia, Antonis and Simon to carry out, and Charlie and the product testers are still in the very early stages of their work.

By July, when Natalia and Anne are presenting a Masterclass at the CILIP Conference, we will be ready to share our experience in identifying appropriate datasets for student learning (rich in relationships that were not represented in MARC) and in converting pre-existing data from MARC to BIBFRAME using the conversion tools supplied by the Library of Congress and Zepheira (BIBFRAME.ORG, 2014). As Tom Meehan (2014) has pointed out, many of the efforts in creating BIBFRAME have, so far, focused on conversion of legacy data, and the BIBFRAME standard might have looked quite different had that not been the case. However, as students and educators within a vocational discipline, we do not see such

constraints as a problem. Rather, working with the practitioner community to solve its real-world issues, and closing the research:practice gap so often identified (Haddow and Klobas, 2004) is a core part of our research and learning. We look forward to continuing to work with the CIG community in developing teaching and learning resources and models for beginning and intermediate cataloguers.

Works Cited

BIBFRAME.ORG (2014) Tools, http://bibframe.org/tools/

de la Mare, W. (1952a) *O Lovely England. Proof Copy.* London: Faber.

de la Mare, W. (1952b) *O Lovely England*. London: Faber.

Haddow, G. and Klobas, J. E. (2004) Communication of Research to Practice in Library and Information Science: Closing the Gap. *Library & Information Science Research* 26(1): 26-43.

Joint Steering Committee for AACR (1967) *Anglo-American Cataloging Rules*. Chicago: American Library Association.

Joint Steering Committee for AACR (1978) *Anglo-American Cataloguing Rules. 2nd ed.* Chicago: American Library Association.

Joint Steering Committee for AACR (1998) *Anglo-American Cataloguing Rules*. *2nd ed. 1998 revision.* Chicago: American Library Association.

Joint Steering Committee for AACR (2002) *Anglo-American Cataloguing Rules. 2nd ed. 2002 revision.* London: Facet.

Lee, D. (2014) 'Teaching Cataloging in a Brave New World'. *Journal of Library Metadata* 14: 166-189.

MARC of Quality (2015) *RIMMF3,* http://www.marcofquality.com/wiki/rimmf/doku.php

Meehan, T. (2014) 'What's Wrong with MARC?' *Catalogue and Index* 174, http://www.cilip.org.uk/sites/default/files/Catalogue%20and%20Index%20issue%20174%2C%20March%202014.pdf

Miller, E., Ogbuji, U., Mueller, V. and MacDougall, K. (2012) *Bibliographic Framework as a Web of Data: Linked Model and Supporting Services.* Washington: Library of Congress, http://www.loc.gov/bibframe/pdf/marcld-report-11-21-2012.pdf

Parent, M. (2014) 'Implementing RDA in aTime of Change: RDA and System Migration at RMIT University'. *Cataloging and Classification Quarterly* 52(6-7): 775-796.

Pattuelli, M.C. (2010) 'Knowledge Organization Landscape: A Content Analysis of Introductory Courses'. *Journal of Information Science* 36(6): 812-822, http://jis.sagepub.com/content/36/6/812

R-balls (2015) *Jane-athon*, http://rballs.info/topics/p/jane/janeathon.html

Robinson, L. and Bawden, D. (2010) 'Information (and Library) Science at City University: 50 Years of Educational Development'. *Journal of Information Science* 36(5): 631-654.

UCL (2010) *Institutional Teaching and Learning Strategy 2010-15*, http://www.ucl.ac.uk/teaching-learning/secure-downloads/ILTS.pdf

UCL (2014) *UCL 2034: A New 20-year Strategy for UCL*, http://www.ucl.ac.uk/ucl-2034/UCL2034

UCL (2014-15) *Connected Curriculum*, http://www.ucl.ac.uk/teaching-learning/strategic_priorities/connected-curriculum

UCL Department of Information Studies (2014-15) *Linked Open Bibliographic Data*, http://www.ucl.ac.uk/dis/research/collaborativeprojects/lobd

Welsh, A. (2013) 'Teaching RDA in 2013-14: Flipping the Cat Class'. *Catalogue and Index* 173: 30-38, http://www.cilip.org.uk/sites/default/files/documents/Catalogue%20and%20Index%20issue%20173%2C%20December%202013.pdf