Catalogue and Index

Periodical of the Chartered Institute of Library and Information Professionals (CILIP) Cataloguing and Indexing Group

Editorial

September 2016, Issue184

This issue is themed around research; we wanted to showcase some of the work on cataloguing and metadata that is currently being focussed on in an academic context. We have a contribution from Anne Welsh, a lecturer at UCL, who has published widely in this field, and who is currently undertaking a PhD in cultural studies. We also have pieces from graduate students who presented at the LISDIS conference in 2015 in Huddersfield, looking at female book collectors of the 19th century, two 19th century book collectors who presented their collections to Cambridge University, and the Ladybird books from 1940-1980. We also have a review and summary of the recent CIG conference which took place in Swansea in early September; papers from the conference will feature in the next issue of C&I, but we hope this review will give you a taster.

If you are undertaking research in a topic related to cataloguing and metadata we would be interested in hearing from you, and would like to feature research articles in future issues of C&I.

Helen Garner (h.j.garner@shu.ac.uk)

Karen Pierce (PierceKF@Cardiff.ac.uk)

Contents

Introduction

This methodological communication discusses the use of MarcEdit in a recent research project and foregrounds how a tool designed for the library community to manipulate catalogue data has been repurposed within an academic methodology. As such, it discusses solutions to the research problem generated by difficulties in outputting MARC records highlighted at CIG 2014 (Welsh, 2014) and the IFLA Rare Books and Special Collections Section's Conference *A Common International Standard for Rare Materials: Why? And How?* (Welsh, 2016b) and in articles published in *Catalogue and Index* (Welsh, 2015) and *Cataloging and Classification Quarterly* (Welsh, 2016a). In doing so, it suggests ways in which metadata for a particular set of rare materials – the catalogue records for the Working Library of Walter de la Mare (Senate House Library [WdlM]) – have been incorporated in the research database and thereby moved beyond Wilson's (1968) idea of the "descriptive power" of bibliographic control to the second, greater power he defined – "exploitative power," summarized by Smiraglia (2008, 35) as "the power of a scholar to make the best possible use of recorded knowledge," and which I have previously argued is a larger purpose than those *solely* of applying international standards and creating linked data (Welsh, 2016a).

Marc data as internal library communication

When devised in the 1960s, Machine Readable Cataloging was primarily focused on the solution of contemporary workflow issues in cataloguing. As the Librarian of Congress put it at the time, "The Library of Congress early recognized that the widespread application of computer technology to libraries could come about only if bibliographic data in machine-readable form could be distributed with precision and at reasonable cost" (Mumford, 1968, [i]). Following on from conferences on a machine format for catalogue records, the MARC Pilot Project was established as "an experiment to determine the feasibility of centrally producing a standardized machine-readable record for application by local installations to serve their specific requirements" (Avram, 1968, 9). International interest in the MARC Pilot, and specifically the British National Bibliography's plan for a UK pilot, in the words of MARC architect Avram, "directed thinking toward a standard communications format suitable for interchanging bibliographic data, not only from one organization (LC) to many, but also among organizations, perhaps crossing national boundaries" (Avram, 2003, 1713). The ensuing MARC II Project had four "criteria to judge the flexibility and usefulness of the format," the first of which was "printing – bibliographic data display in a variety of forms (3x5 catalog cards, book catalogs, bibliographies, etc.)," with Information Retrieval appearing third, after "Catalog division – e.g. personal names used as author and subject may be filed together in a separate catalog" (Avram, Knapp and Rather, 1968, 3; underlining in original).

This is not to say that those involved in the initial development of library catalogue data were not interested in how researchers might use their data – merely that anticipating such use was firmly in the area of 'future-casting' for these early developers. As Avram, Knapp and Rather (1968, 4) put it, "Since so little is known about how a bibliographic record will be used in machine-readable form for retrieval, it was only possible to anticipate future applications." It is important to acknowledge that the user was at the centre of the motivation to develop machine-readable (and, therefore, machine-retrievable) data. As Byrne (1998, 3) has described, in the pre-MARC environment, "in addition to being focused on a limited number of uses, [1960s – 1990s scientific industry] commercial systems are generally designed for use only by trained personnel, not by the public. Though library systems do include many functions that are designed for use only by trained staff, a primary and essential focus of library systems is use by library patrons rather than trained personnel." She goes on to point out that "Most early library systems, designed in the early to mid-1960s [pre-MARC], were designed to serve as circulation systems. Specifically, these systems were used to maintain records of the items that were checked out by a patron and to produce overdue notices and other forms related to library circulation routines" (Byrne, 1998, 4).

CC BY

In this, we can see not only that the MARC developers were prescient to consider Information Retrieval, but that they were working in an era in which what and how we catalogued were heavily restricted by the technology available. As Whaite (2013) has argued, "an old catalogue becomes a relic of its time." Considering the date of the MARC reports cited here – 1968 – and of Wilson's seminal *Two Kinds of Power: An Essay on Bibliographical Control* – 1968 – the library catalogues that are "relics of [that] time" reflect the limited technological capabilities of the computer as opposed to the ambitions of developers and cataloguers. That famous year of revolution – 1968 – brought revolutionary changes in the world of library data as well as in society at large. We could suggest that it is important to treat with kindness the restrictions faced by our mid-20th century colleagues, since surely we too have been limited in the late 20th and early 21st centuries by our own computer systems. Surely our own ambitions – for bibliographic models focused on relationships; for data that can be reutilized easily by the world outside the library walls; for improved cataloguing workflows – are not fully reflected in the data we create as "relics of [our] time." There is a time-lag between what we want for our users and its being fully reflected in the data we create.

MARC Data as Textual Artifact

As argued elsewhere (Welsh, 2016a), the narrative arc in the story of the creation of computer catalogue data has, to this point, been focused on library workflows and on search and retrieval. As Byrne and other commentators have described, restrictions in first the hardware and software on which we relied and then the data models available to us have meant that it is comparatively recently – only since the 1990s – that it has been possible to run searches that scan the entire data relating to every item in a library system (*cf.* Byrne, 1998; Bowman, 2007; Tedd, 2007).

Certainly, by the mid-1990s, it was possible for Leeves (1995, 22) to assert, "A common view is that library housekeeping functionality is now well catered for." As library management systems moved from being standalone to networked, it was, arguably, no wonder that cataloguing theorists, seeing records liberated from the local systems of the 1970s and turnkey systems of the 1980s described by Leeves (1995) and Tedd (2007), began to consider new models for catalogue data, such as *Functional Requirements for Bibliographic Records* (FRBR), first published by IFLA in 1998. This model asserts "four generic user tasks … The tasks are defined in relation to elementary uses that are made of the data by the user" and are described as "to find entities that correspond to the user's stated search criteria …; to identify an entity …; to select an entity that is appropriate to the user's needs …; to acquire or obtain access to the entity described" (IFLA, 1999, 82, underlining in original). As highlighted elsewhere (Welsh, 2014; Welsh, 2015; Welsh, 2016a; Welsh, 2016b), it is crucial to recognise that the tasks are "generic" and the uses "elementary" – the IFLA Study Group on Functional Requirements is not saying that these are the *only* uses to which catalogue data may be put. "Search is only elementary" (Welsh, 2016a).

Indeed, catalogue data is more than solely a surrogate for the items it describes; more than simply the objects retrieved in a search. It is, as Whaite (2013) has argued, "a relic of its time"; as Smiraglia (2008) has asserted, a cultural artifact; and as Anderson (2002) has highlighted, a text in its own right. And, as text, it should be possible to interrogate it in the same ways, and using the same tools that we use on any other forms of text, from Pynchon's *V* (Tsatsoulis, 2012) to Houghton's *The Victorian Frame of Mind* (Gibbs and Cohen, 2011) and back into the world of *Early Modern Print* (Humanities Digital Workshop at Washington University in St Louis, 2013-), to name only three projects that have used methods in digital bibliography to explore texts.

MARC data has been used to good effect in the creation of several tools that are of use to bibliographers, including Copac's union catalogue of "c. 90 UK and Irish academic, national and specialist library catalogues" (Copac, 1996-) and Content Management (CCM) Tools (Copac, 2012-); Edina's SUNCAT, the Serials Union Catalogue for the UK (Edina, 2003-); various projects undertaken by OCLC, including its Linked Data Subsets and Linked Data Markup on Worldcat.org (OCLC, ©2016); and CERL's many products, such as Material Evidence in Incunabula (MEI) and the Heritage of the Printed Book (HPB) Database (Consortium of European Research Libraries, 2012-).

However, in terms of study as text in its own right, not solely as information for inclusion in databases for retrieval, attention paid to catalogue data has been scant and difficult. As Mitch Fraas (2014) put it

when describing a project he undertook to create a network analysis of former owners of the codex manuscripts at University of Pennsylvania Libraries, “I realize now that this task would have been near to impossible at most libraries where the online catalogs and back-end databases don't easily allow public users to batch download full records. Fortunately at Penn all of our catalog records are available in MARC-XML form.” Similarly, James Baker (2013) has reported a need to thoroughly cleanse data from the British Cartoon Archive before starting the quantitative analysis he wished to carry out on it.

Not all researchers possess the technical skills of Fraas and Baker. Barriers faced include practical issues such as interoperability of tools (Terras *et al*, 2016), and there can be a fundamental lack of understanding of the opportunities afforded by digital research and the datasets that are available (Mahony and Pierazzo, 2012). National libraries have striven in recent years not only to translate MARC data into linked data formats, but also to encourage scholarly engagement with library data through competitions and fellowships (Welsh, 2016a; Welsh, 2016b).

Tomm (2012) reported a methodology that bypassed the need for advanced programming and data manipulation skills:

> Analyses of catalogue records did not proceed directly from the McGill OPAC. Bibliographic data for the Klibansky Collection was exported from the OPAC, gathered in a personal bibliographic database (EndNote) for exploration and manipulation, and then exported again for further manipulation and analysis in a spreadsheet (Excel). Standard desktop software was selected for simplicity and to keep the procedure accessible to a broad group of potential users.
>
> The basic steps are:

1. Export bibliographic data from the catalogue (save .txt file)
2. Import text files of bibliographic data into EndNote
3. Manipulate data in EndNote
4. Export desired fields from EndNote in TAB Delimited [sic] format (save .txt file)
5. Open in Excel for additional manipulation and analysis (Tomm, 2012, 85)

In devising her methodology, Tomm drew on the work of earlier scholars (Gardner et al, 2010; King et al, 2011; Kwan, 2010; Schlichter and Kraemmergaard, 2010; Xu, 2011), although these are focused on the use of reference management software in literature reviews. At the time of writing, Tomm (2012, 77) agreed with Childress (2011, 144) that “Much of the available literature deals largely with citation managers and generators, more specifically reviews, comparisons, evaluations and use cases for such programs.”

However, within the Digital Humanities, we can observe the use of citation manager Zotero for more than solely reference management. The Zotero plug-in Paper Machines has been highlighted as one of the tools powering “The Digital Humanities Contribution to Topic Modelling” (Meeks and Weingart, 2012) and “allow [ing] anyone to begin exploring large collections of sources to look for trends in the data such as an increasing interest in certain subjects over time, which could point the researcher to interesting questions worth pursuing further. With Paper Machines, anyone with a collection of texts stored in Zotero can generate word clouds, phrase nets, map geo-references found in their corpus, extract structured data using DBPedia, or generate and visualize topic models. All of this can be done without having to pre-process your corpus or leave Zotero” (Crymble, 2012). Co-creator of Paper Machines, Jo Guldi has also co-authored *The History Manifesto* (Guldi and Armitage, 2014), which sets out ideas for analyzing the past in order to assist in current political decisions, focusing on the long-term rather than the short-term and therefore on methods which allow historians to conduct such research, inevitably processing Big Data through techniques in Distant Reading. The book includes an explication of the historiographical uses to which Paper Machines can be put:

> One may use the tool to generalize about a wide body of thought – for instance, things historians have said in a particular journal over the last ten years. Or one may visualise libraries against each other – say, novels about nineteenth-century London set against novels about nineteenth-century Paris. Using the tool, a multitude of patterns in text can be rendered visible through a simple graphical interface. Applying Paper Machines to text corpora allows scholars to accumulate hypotheses about *longue-durée* patterns in the influence of ideas, individuals, and professional cohorts.
>
> By measuring trends, ideas, and institutions against each other over time, scholars will be able to take on a much larger body of texts than they normally do. (Guldi and Armitage, 2014, 91).

Even before Paper Machines was developed, Zotero itself was used in projects that chose to visualize their data with other tools such as Voyeur Tools. For example, With Criminal Intent provided a plug-in that enabled users to manage their data from Old Bailey Online through Zotero (Cohen et al, 2011), while Tufts University incorporated Zotero into its Visual Understanding Environment (VUE) alongside other tools (Baepler and Murdoch, 2010, 5). The VUE Project created tutorials on how to use Zotero in 'Dynamic Content Mapping' (VUE Project, 2009b) and 'Semantic Mapping Tools' (VUE Project, 2009a).

Tutorials on Paper Machines include Emory Libraries' *Lincoln's Logarithms: Finding Meaning in Sermons* (Emory Libraries, 2013), although a recent review of *Exploring Big Historical Data: The Historian's Macroscope* (Graham, Milligan and Weingart, 2015) highlights that "the authors had to remove from the book an example using the tool *Paper Machines* because an update to the software on their computers had broken the tutorial in the time it took to write the book," and suggests that "A greater focus on core principles would have been helpful in future-proofing the text" (Crymble, 2016). Book reviewer Crymble's own project, *The Programming Historian*, currently includes three tutorials on "Distant Reading" (Froehlich, 2015; Hulden, 2014; Graham, Weingart and Milligan, 2012), and covers Zotero itself in three tutorials on "Application Programming Interfaces (APIs)" (Morton, 2013a; Morton, 2013b; Roberts, 2013).

Although Tomm did not explore the growing use of Zotero in her 2012 thesis (and some of the more impressive resources relating to it were published after she had submitted her work), its development by Digital Humanists and uptake by Digital Historians further strengthens her assertion that "As catalogues, cataloguing standards and technology outside the library continue to evolve, the ways that catalogue data can be accessed and used by researchers will continue to shift. But a barrier has been broken and catalogue data has already become useful beyond the 'silo' of the library" (Tomm, 2012, 78).

Exporting MARC data from the catalogue

In 2010, I began work on my own PhD, which is an analysis of the books in the Working Library of Walter de la Mare, housed at Senate House Library with the classmark "[WdlM]" (square brackets not indicating an insertion, but punctuation present in the original classmark). The library had recently completed cataloguing materials in [WdlM] and its companion collection of the De La Mare Family Archive of Walter de la Mare's Printed Oeuvre (Classmark [WdlM] T (again, square brackets present in the original)), and the first task I undertook was downloading records to Zotero for my own use.

As reported at CIG 2014 (Welsh, 2014), "Attempts to export to any of the reference management options did not carry the notes field through, which, given the focus of my work is largely provenance, meant that the most useful elements of the records were lost to me," and I soon discovered "that despite an impressive list of export options, there was not a single one that provided me with what I needed: a clean, tab delimited file of MARC fields that I could import into Excel. The CSV and tab delimited text files did not work correctly – even assistance from the then systems team did not result in my having a clean copy of the data" (Welsh, 2015, 5).

Although aware of developments around Zotero, my tool of choice for working with catalogue data – especially catalogue data that I need to manipulate – is, of course MarcEdit. As a librarian who qualified in the mid-1990s, who chose to work mainly in small, special libraries and small, special collections within larger

institutions, I'd been using this set of tools, described by creator Terry Reese (2004, 25) as "a free, Windows-based, metadata editing software suite that [he] develop[s] and support[s] as part of [his] contribution to the library profession" almost since its launch in 2000. Unfortunately, and possibly due to the same issues that were affecting the CSV export options, when I imported the data for my PhD to MarcEdit, I received a string of error messages, and my best attempts resulted in data that was messy beyond the limited powers of batch editing at my disposal to fix. From 2010-2014, when I presented my paper at the CIG Conference, it seemed that the only options available to me were to manually tab delimit the hundreds of records (either in notepad or in MarcEdit), or to ask the systems team to create a report just for me using their internal tools or publish the data (again, just for me). As discussed elsewhere, although some libraries are beginning to explore bespoke publishing of data for researchers, there are workflow and cost implications for any one-to-one services (Welsh, 2016a).

Moreover, PhD research should be one's own work, and it should be possible for others to recreate the steps taken to carry it out – from an ethical standpoint it might be improper to ask for bespoke systems work to be carried out on my behalf (Welsh, 2015). Just as Tomm (2012, 85) insisted that "Standard desktop software was selected for simplicity and to keep the procedure accessible to a broad group of potential users," it was important to me that any work to export bibliographic data from the catalogue and import it to my research spreadsheet, or to any other digital tools should be possible to be carried out by me using the tools I had at home – or by my examiners using the same tools, should they choose to check that part of my work.

So I reported to the systems team the issues I had faced as a library user attempting to capture and reuse the data, and I continued with my work, using 'proxy' data created manually, or hacked together as a 'next-best workaround' for a clean dataset. I continued in the belief that over the course of a part-time PhD (5-7 years), something would happen that would make it possible for me to undertake the 'real' work with the 'real' data that I desired. Within my research project, the materials most affected by my use of proxy data were those that don't fall within the areas on which I have focused for full thesis chapters. My work has been structured to consider books that may have had an influence on de la Mare's own writing, such as the poetry collections, short stories and novels he owned, but also those dealing with subjects about which he wrote, including nature, childhood and the supernatural. After an initial inspection of the books in [WdlM] in 2010-2012, my work has been focused on particular genres and subjects within the collection (*cf.* Welsh, 2013a; Welsh, 2013b). So without good-quality data, there might be a danger that materials outwith these areas and whose only appearance may be in the thesis appendices, could be neglected. Books such as Schrödinger's (1944) *What is Life?: The Physical Aspect of the Living Cell* ([WdlM] 488) and Crake's (1874) *Simple Prayers: A Manual of Instruction and Devotion for Schoolboys* ([WdlM] 489) may not take starring roles in the key chapters of the thesis, but they form part of the collection, and should be represented within the data on equal terms with items which are discussed at length.

Following CIG 2014, the idea of exporting data from the catalogue and importing it successfully to appropriate tools grew in importance. A quick assessment of the actual results of data output from a range of catalogues that I had carried out on the build-up to the conference had found similar issues in each that I tried, which had, at first, seemed incredible to me – I thought I must be doing something wrongly. However, when I asked for a show of hands from conference attendees of cataloguers who had tried all the export options available to library users, only one hand was raised – belonging to a colleague from the British Library (Welsh, 2014; Welsh, 2015). Of course, if it were straightforward to export MARC data in easily reusable states, systems teams, library management system vendors, and those publishing datasets in linked data formats would have a much easier working life – and may even find the aspects of their roles that can be described as 'MARC wrangling' would be redundant.

Importing MARC Data to MarcEdit for Mac

In 2015, my belief that "something would happen" that would allow me to export MARC data from the catalogue to MarcEdit was justified. In April 2015, Terry Reese announced that, following demand from the Mac community, he was beginning work on a version of the suite "that uses native Mac APIs" (Reese, 2015). In his own words, "MarcEdit is so fragile when being run on a Mac … [because] MarcEdit utilizes a cross

platform toolset when building the UI which works well on Linux and Windows systems, but tends to be less refined on Mac systems" (Reese, 2015). Interestingly, the reason he gave for not carrying out these developments earlier was lack of perceived demand from Mac users: "I can count on two hands the number of times I've had someone request a version of MarcEdit specifically for a Mac. And since I've been making a Mac App version of MarcEdit available – its use has been fairly low … With an active [MarcEdit] community of over 20,000, I try to put my time where it will make the most impact, and up until a week ago, better support for Mac systems didn't seem to be high on the list" (Reese, 2015). Following a community-led campaign started by Whitni Watkins and Francis Kayiwa, it became clear to Reese that there *was* demand for a stable and reliable MarcEdit for Mac: "After 8 days, it's done. In all, 40 individuals contributed to the campaign, but more importantly to me, I heard directly from around 200+ individuals that were hopeful that this project would proceed" (Reese, 2015).

The Mac Operating System has been gaining in popularity with developers in general. As this year's Stack Overflow survey of over 50,000 developers highlighted, "Last year, Mac edged ahead of the Linuxes as the number 2 operating system among developers. This year it became clear that the trend is real. If OS adoption rates hold steady, by next year's survey fewer than 50% of developers may be using Windows" (Stack Overflow, 2016). From a share of 60.4% in 2013, Windows fell to 54.5% in 2015 and 52.2% in 2016, while Linux use has grown from 19.9% in 2013 to 21.7% this year, and Mac OS X has risen steadily from 18.7% in 2013, to 20.3% in 2014, 21.5% in 2015, and 26.2% in 2016 (Stack Overflow, 2016).

It is not necessary to enter the debates between developers as to which operating system is the best to acknowledge that versions of software for different operating systems can differ in terms of features and efficiency. While MarcEdit, originally created on Windows, was "fragile" on Mac, there are other programs that are created first on Mac and then translated to Windows, and it is not uncommon to run "virtual machine applications" to provide a Mac experience within Windows, or vice versa (Pot, 2016). There are some disadvantages to running one operating system within another, including increased demands on RAM (Random Access Memory) and CPU (Central Processing Unit) use (*cf* Kissell, 2014; Joseph, 2015; Rizzo, 2013). Cataloguing programs Marc Report and RIMMF, designed by The Marc of Quality, as well as Reese's MarcEdit were developed first for Windows. When using them for research, I have relied on a Windows PC at home, while for teaching I have had to use UCL's virtual desktop on my MacBook, which utilises Citrix Receiver to create a Windows environment. The programs do seem to run more slowly on Desktop@UCL than at home on a native Windows interface, and do seem to crash more frequently.

In any case, the development of MarcEdit on native Mac UIs turned out to be the "something" for which I had been waiting. I watched Reese's (2013-) build page until it looked like work was venturing into new territory as opposed to trying to replicate features on MarcEdit for Windows (Reese, 2016), and then in March 2016 I attempted to import data from the Senate House Library catalogue to MarcEdit for Mac – *and succeeded*. Between March and April I played around with the data until I was confident importing it to Excel, firstly to create an overview spreadsheet of all the records in [WdIM], and then to create spreadsheets for specific fields that I could run through Excel and then import to Gephi to create data vizualisations.

In the style and spirit of Tomm (2012, 85) quoted above, here are the steps that I used:

1. Search Senate House Library catalogue using a "Mixed/local classmark" search for "WdlM" (983 records including [WdlM] T and [WdlM] P)
2. Save records 1-557 and export to local disk (export.txt)
3. Open in MarcEditor and check contents
4. Open MarcTools; upload export.txt and execute MARC=>MARCXML function (save as WdlM.xml)
5. Open in MarcEditor and check contents
6. Open MarcTools; upload WdlM.xml and execute MARCXML=>MARC function (save as WdlM.mrc)
7. Open WdlM.mrc in MarcJoin; select "Export Delimited tab"; set delimiter to "Tab(\t)" and check "Normalize data" box. Select desired fields. (save as .txt file)
8. Open Excel; import .txt file (save as .xlsx file).

From this point, I can save .htm files to upload to TAPoR List Words to quantify, for example, the publishers whose books constitute most of the collection. I can create appropriate edges as CSV files to upload to Gephi to create visualizations to help me think about the provenance of the books in [WdlM] – detecting differences in the level of annotation in books written by de la Mare's friends and other books he owned. The .xlsx files generated from export.txt via WdlM.mrc become the central set of data for quantitative analysis.

Being able to work with data from the catalogue itself also allows me to analyse not only what subject headings were used for the books in [WdlM] but also the consistency of their application. Data for [WdlM] and [WdlM] T were not always created 'from scratch' but derived from pre-existing records, and so it is interesting to see how the resulting subject headings were applied. As evinced by Attar's (2012) article on the cataloguing decisions about [WdlM] T, descriptive cataloguing appears to have been a greater focus than subject indexing (Attar, 2012).

Since hearing Caitlin Bailey's (2013) presentation of her MSLIS research into "the use and analysis of small collections in the study of historical thought" and her proposal of "the feasibility of such collections for the development of unique data sets" (Institute of English Studies, 2013), and reading Tomm's quantitative analysis of the subjects within the Klibansky Collection and Davies and Fichtner's (2006) breakdown of the subjects within Freud's Library, I have been keen to analyse (1) the subjects in de la Mare's Working Library (Senate House [WdlM]) and (2) their representation in the subject headings applied to the books. Being able to export data from the catalogue into MarcEdit is the first step in this analysis.

Scope and limitations of use of MarcEdit

Writing in the mid-1980s about the future of research into writers' libraries, Gribben (1986, 311) predicted that "The technology and determination that enable us to penetrate outer space will most likely also give us better means to explore the intellectual lives of our cherished authors. Word-processors, as well as other apparatuses now beyond our ken, will ultimately supplement the researcher's notecards and fileboxes, but an unquenchable curiosity about the creators and backgrounds of great literary manuscripts will continually bring forth dauntless scholars in each generation."

Certainly, reference management software would have been "beyond [Gribben's] ken," far less Williams's 21st century question, "Can we call Zotero a Scholar's Box for the digital age?" and her answer to it, "I think we can, but we need to recognize that the citations we have are still stuck in a box, in many ways. We cannot copy citations from library databases and drop them into a word processor without using a bibliographic manager like Zotero as an intermediary to capture the structured data that might be useful to my computer when I need to format a bibliography" (Williams, 2015,4).

Tomm's (2012, 85) approach to exporting data from the catalogue at McGill had the advantage of using "Standard desktop software … for simplicity and to keep the procedure accessible to a broad group of potential users," but proved difficult to replicate for the Working Library of Walter de la Mare. The tools in MarcEdit ultimately provided not only a way to extract and manipulate data, but to do so in a way that allowed for a great deal of control over the manipulation, since the suite of software developed by Reese allows not only for the execution of a range of different algorithms behind the front end, but also for files to be opened and edited directly through the MarcEditor.

However, in order to use MarcJoin, which proved to be the most useful tool for extracting specific fields, an understanding of MARC 21 fields is required. For example, to extract the dates of publication from WdlM.mrc for use in TAPoR and Gephi, it is necessary to know that MARC 260 $c is the field and subfield for publication date. The availability of the *MARC 21 for Bibliographic Records* online and free of charge means that it would not be impossible for a non-cataloguer to work out the fields they need to create the .txt file they require, but there is, clearly, an extra effort involved. Thinking of Byrne's (1998, 4) differentiation between commercial scientific industry databases and library management systems, we could not claim that MarcEdit has been designed with "a primary and essential focus … [on] use by library patrons rather than trained personnel."

That said, the ability to extract data from the catalogue for use in research by Digital Humanists – whether through the intermediary steps of a reference management system or MarcEdit – can be seen to be the beginning of the fruition of Attar's (2004, 11) prediction of "the developing function of a catalogue record as a research tool in itself, instead of a mere finding aid."

As argued elsewhere (Welsh, 2016a; Welsh, 2016b), the case for the solo researcher working on bibliographic research into an author and / or owner of a private library that has now been absorbed into an institutional library is, in some ways, smaller than the case for the international standardization of rare books cataloguing, or the publication of linked data. However, "In another way, it is much larger, leading us back round to Wilson's philosophy of the exploitative power of bibliographic control. If we can meet the needs of researchers who want to engage with our data not as a route through to 'the real' objects of their research – full-text files, books, the item for which catalog data is a surrogate – but as an integral part of their own research, then, surely, we are assisting not simply in an 'elementary' user task, but something that is fundamental to scholarship: 'the best possible use of recorded knowledge' (Smiraglia, 2008, 35)" (Welsh, 2016a).

To put it in business terms, as Williams (2015, 9) has, "Libraries must find the means by which scholars can save and sort, use and reuse the resources they find from our collections, or faculty will gravitate to for-profit research platforms that will resolve these problems but within a proprietory and private space." More motivationally, "Libraries are part of a generative process. Cards of single ideas are written, rearranged, and stacked to help build theses, which, in turn, help build books which, in turn, form bibliographies, which fill libraries. I'd like libraries to find a way back to Gessner's *Bibliotheca Universalis*, a place where the library and the scholar are connected" (Williams, 2015, 9).

Tools developed in the library domain, such as MarcEdit, provide a connection between my identity as a researcher and my identity as a librarian. I would like to hope that by working on and disseminating methodologies using such tools, the connection between Humanities researchers without a professional library background and the bibliographic data that powers their initial forays into new knowledge can be encouraged and strengthened.

Acknowledgements

Some of the research presented in this article was undertaken as part of a PhD in Cultural Studies at University College London. The author would like to thank Gladstone's Library for the award of a Revd. Dr. Murray MacGregor Scholarship in 2016, which provided a quiet space to write up some of her thesis.

Works cited

Anderson, J. 'Materiality of Works: The Bibliographic Record as Text'. *Cataloging and Classification Quarterly* 33: 39-65.

Attar, K.E. (2012) 'Modern Special Collections Cataloguing: A University of London Case Study'. *Journal of Librarianship and Information Science* 45(2): 168-176.

Attar, K.E. (2004) 'Cataloguing Early Children's Books: Demand, Supply and a Seminar'. *Catalogue and Index* 151: 8-12.

Avram, H.D. (2003) 'Machine-Readable Cataloging (MARC) Program'. In M.A. Drake (ed.) *Encyclopedia of Library and Information Science*. 2nd ed. New York: Marcel Dekker, pp. 1712-1730.

Avram, H.D. (1968) *The MARC Pilot Project: Final Report on a Project Sponsored by the Council on Library Resources, Inc.* Washington: Library of Congress.

Avram, H.D., Knapp, J.F. and Rather, L.J. (1968) *The MARC II Format: A Communications Format for Bibliographic Data*. Washington: Library of Congress.

Baepler, P. and Murdoch. C. (2010) 'Academic Analytics and Data Mining in Higher Education'. *International Journal for the Scholarship of Teaching and Learning* 4(2), http://digitalcommons.georgiasouthern.edu/ij-sotl/vol14/iss2/17

Bailey, C. (2013) 'Hidden in Plain View: The Value of the Small Collection'. Presented at: Blackburn's "Worthy Citizen": A Colloquium on the R.E. Hart Collection, London, 23 November.

Baker, J. (2013) 'On Metadata and Cartoons'. *British Library Digital Scholarship Blog*, 16 May, http://britishlibrary.typepad.co.uk/digital-scholarship/2013/05/on-metadata-and-cartoons.html

Bowman, J.H. (2007) 'OPACs: The Early Years and User Reactions'. *Library History* 23: 317-329.

Byrne, D.J. (1998) *MARC Manual: Understanding and Using MARC Records.* 2nd ed. Westport, Connecticut: Libraries Unlimited.

Childress, D. (2011) 'Citation Tools in Academic Libraries: Best Practices for Reference and Instruction'. *Reference and User Services Quarterly* 51(2): 143-152.

Cohen, D., Gibbs, F., Hitchcock, T., Rockwell, G., Sander, J., Shoemaker, R., Sinclair, S., Takats, S., Turkel, W.J., Briquet, C., McLaughlin, J., Radzikowska, M., Simpson, J. and Uszkalo, K.C. (2011) *Data Mining with Criminal Intent: Final White Paper, August 31, 2011*, http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.458.2152&rep=rep1&type=pdf

Consortium of European Research Libraries (2012-) *CERL Resources*, https://www.cerl.org/resources/main

Copac (1996-) *Explore c. 90 UK and Irish academic, national and specialist library catalogues*, http://copac.jisc.ac.uk

Copac (2012-) *Copac Collection Management Tools*, http://ccm.copac.ac.uk/

Crake, A.D. (1874) *Simple Prayers: A Manual of Instruction and Devotion for Schoolboys*. London: A.R. Mowbray.

Crymble, A. (2012) 'Review of Paper Machines, Produced by Chris Johnson-Roberson and Jo Guldi'. *Journal of Digital Humanities* 2(1), http://journalofdigitalhumanities.org/2-1/review-papermachines-by-adam-crymble/

Crymble, A. (2016) '[Review of] *Exploring Big Historical Data: The Historian's Macroscope*'. *Reviews in History*, February, http://www.history.ac.uk/reviews/review/1889

Davies, J.K. and Fichtner, G. (2006) *Freud's Library: A Comprehensive Catalogue*. London: The Freud Museum.

Edina (2003-) *SUNCAT*, http://suncat.ac.uk/search

Emory Libraries (2013) *Lincoln's Logarithms: Finding Meaning in Sermons*, https://disc.library.emory.edu/lincoln/papermachines-2/

Fraas, M. (2014) 'Charting Former Owners of Penn's Codex Manuscripts'. *Mapping Books*, 24 January 2014, http://mappingbooks.blogspot.co.uk/2014/01/charting-former-owners-of-penns-codex.html

Froelich, H. (2015) 'Corpus Analysis with Antconc'. *The Programming Historian,* http://programminghistorian.org/lessons/corpus-analysis-with-antconc

Gardner, W.L., Lowe, K.B., Moss, T.W., Mahoney, K.T. and Cogliser, C.C. (2010) 'Scholarly Leadership of the Study of Leadership: A Review of *The Leadership Quarterly*'s Second Decade, 2000-2009'. *The Leadership Quarterly* 21: 922-958.

Gibbs, F.W. and Cohen, D.J. (2011) 'A Conversation with Data: Prospecting Victorian Words and Ideas'. *Victorian Studies* 54(1): 69-77.

Graham, S., Weingart, S. and Milligan, I. (2012) 'Getting Started with Topic Modelling and MALLET'. *The Programming Historian*, http://programminghistorian.org/lessons/topic-modeling-and-mallet

Graham, S., Milligan, I., Weingart, S. (2015) *Exploring Big Historical Data: The Historian's Macroscope*. London: Imperial College.

Gribben, A. 'Private Libraries of American Authors: Dispersal, Custody and Description'. *The Journal of Library History* 21(2): 300-310.

Guldi, J. and Armitage, D. (2014) *The History Manifesto*. Cambridge: Cambridge University Press.

Hulden, V. (2014) 'Supervised Classification: The Naïve Bayesian Returns to the Old Bailey'. *The Programming Historian,* http://programminghistorian.org/lessons/naive-bayesian

Humanities Digital Workshop at Washington University in St Louis (2013-) *Early Modern Print: Text Mining Early Printed English*, http://earlyprint.wustl.edu/

IFLA Study Group on the Functional Requirements for Bibliographic Records (1998) *Functional Requirements for Bibliographic Records*. Munich: K.G. Saur.

Institute of English Studies (2013) '[Abstracts for Blackburn's "Worthy Citizen": A Colloquium on the R.E. Hart Collection, London, 23 November]', https://blackhartbooks.files.wordpress.com/2013/06/colloquium-abstracts.pdf

Joseph, C. (2015) 'Parallels, VMware, VirtualBox or Boot Camp: Best Virtualisation Tool for Mac'. *Macworld* 9 October, http://www.macworld.co.uk/feature/mac-software/best-virtualisation-app-run-windows-on-your-mac-boot-camp-vmware-parallels-3626493/

King, R., Hooper, B. and Wood, W. (2011) 'Using Bibliographic Software to Appraise and Code Data in Educational Systematic Review Research'. *Medical Teacher* 33(9): 719-723.

Kissell, J. (2014) 'Running Windows on a Mac: Why I Prefer VMware Fusion'. *Macworld,* 18 June, http://www.macworld.com/article/2364514/running-windows-on-a-mac-why-i-prefer-vmware-fusion.html

Kwan, M.-P. (2010) 'A Century of Method-Oriented Scholarship in the Annals'. *Annals of the Association of American Geographers* 100(5): 1060-1075.

Leeves, J. (1995) 'Library Systems Then and Now'. *VINE* 25(3): 19-23.

Mahony, S. and Pierazzo, E. (2012) 'Teaching Skills or Teaching Methodology'. In B.D. Hirsch (ed.) *Digital Humanities Pedagogy: Practices, Principles and Politics*. Cambridge: Open Book.

Meeks, E. and Weingart, S. (2012) 'The Digital Humanities Contribution to Topic Modelling'. *Journal of Digital Humanities* 2(1), http://journalofdigitalhumanities.org/2-1/dh-contribution-to-topic-modeling/

Morton, A. (2013a) 'Creating New Items in Zotero'. *The Programming Historian*, http://programminghistorian.org/lessons/creating-new-items-in-zotero

Morton, A. (2013b) 'Intro to the Zotero API'. *The Programming Historian*, http://programminghistorian.org/lessons/intro-to-the-zotero-api

Mumford, L.Q. (1968) 'Foreword'. In H.D. Avram. *The MARC Pilot Project: Final Report on a Project Sponsored by the Council on Library Resources, Inc.* Washington: Library of Congress.

OCLC (©2016) *Innovation Lab: Bridging the Gap between Research and Development,* http://www.oclc.org/research/innovationlab.html

Pot, J. (2016) 'Is One Operating System Not Enough? Here's the Five Best Virtual Machine Applications'. *Digital Trends*, 10 April, http://www.digitaltrends.com/computing/best-virtual-machine-apps-for-mac-linux-and-windows-pcs/#:NoOOLJxH8ABbJA

Reese, T. (2004) 'Information Professionals Stay Free in the MarcEdit Metadata Suite'. *Computers in Libraries* 24(8): 24-28.

Reese, T. (2013-) 'Current News'. *MarcEdit Development*, http://marcedit.reeset.net

Reese, T. (2015) 'Building a Better MarcEdit for Mac Users'. *Terry's Worklog* 11 April, http://blog.reeset.net/archives/1665

Reese, T. (2016) 'MarcEdit Mac: Export Tab Delimited Records'. *Terry's Worklog* 28 March, http://blog.reeset.net/archives/1917

Rizzo, J. (2013) 'How to Run Windows on Macs'. *MacWindows* 23 September, http://www.macwindows.com/winintelmac.html

Roberts, S. (2013) 'Counting Frequencies from Zotero Items'. *The Programming Historian,* http://programminghistorian.org/lessons/counting-frequencies-from-zotero-items

Schlichter, B.R. and Kraemmergaard, P. (2010) 'A Comprehensive Literature Review of the ERP Research Field over a Decade'. *Journal of Enterprise Information Management* 23(4): 486-520.

Schrödinger, E. *What is Life?: The Physical Aspect of the Living Cell.* Cambridge: Cambridge University Press.

Smiraglia, R.P. (2008) 'Rethinking What We Catalog: Documents as Cultural Artifacts'. *Cataloging and Classification Quarterly* 45: 25-37.

Stack Overflow (2016) *Developer Survey Results 2016*, http://stackoverflow.com/research/developer-survey-2016

Stalker, L. and Dooley, J.M. (1992) 'Descriptive Cataloging and Rare Books'. *Rare Books and Manuscripts Librarianship* 7(1): 7-23.

Tedd, L. (2007) 'Library Management Systems in the UK: 1960s – 1980s'. *Library History* 23: 301-316.

Terras, M., Baker, J., Hetherington, J., Beavan, D., Welsh, A., O'Neill, H., Finley, W., Duke-Williams, O. and Farquar, A. (2016) 'Enabling Complex Analysis of Large-Scale Digital Collections: Humanities Research, High Performance Computing and Transforming Access to British Library Digital Collections'. Presented at: Digital Humanities 2016, Krakow, Poland, 11-16 July.

Tomm, J. (2012) *The Imprint of the Scholar: An Analysis of the Printed Books of McGill University's Raymond Klibansky Collection: A Thesis Submitted to McGill University in Partial Fulfillment of the Requirements of the Degree of Doctor of Philosophy*. Montreal: McGill University, http://oatd.org/oatd/record?record=oai%5C%3Adigitool.library.mcgill.ca%5C%3A114196

Tsatsoulis, C.I. (2012) 'Unsupervised text mining methods for literature analysis: a case study for Thomas Pynchon's *V*'. *Orbit* 1(2), https://www.pynchon.net/owap/article/view/44

VUE Project (2009a) 'Map Based Searching and Semantic Analysis Screencast', https://www.youtube.com/watch?v=iKuBE-J7JuU

VUE Project (2009b) 'Mapping a Zotero Collection into VUE', https://www.youtube.com/watch?v=UsYDOo95ses

Welsh, A. (2013a) '"Books! –": *Pleasures and Speculations* in Walter de la Mare's Library'. Presented at: Researching the Reading Experience, Oslo, Norway, 10-12 June, http://discovery.ucl.ac.uk/1413960/

Welsh, A. (2013b) 'The Poet's Poets: Collections and Anthologies in Walter de la Mare's Working Library'. Presented at: Writers and their Libraries, London, 15-16 March, http://discovery.ucl.ac.uk/1414072/

Welsh, A. (2014) 'Metadata Output and its Impact on the Researcher'. Presented at: Cilip Cataloguing and Indexing Group Conference, Canterbury, Kent, 8-10 September 2014, http://discovery.ucl.ac.uk/1450941/

Welsh, A. (2015) 'Metadata Output and the Researcher'. *Catalogue and Index* 178: 2-8.

Welsh, A. (2016a) 'The Rare Books Catalog and the Scholarly Database'. *Cataloging and Classification Quarterly* (in press).

Welsh, A. (2016b) 'The Rare Books Catalogue as the Foundation of the Scholarly Database'. Presented at: IFLA Rare Books and Special Collections Section Conference, A Common International Standard for Rare Materials: Why? And How?, Lisbon, Portugal, 22 February 2016, http://www.ifla.org/files/assets/rare-books-and-manuscripts/Lisbon-2016-presentations/annewelshiflarbsc2016_reduced.pdf

Whaite, K.C. (2013) 'New Ways of Exploring the Catalogue: Incorporating Text and Culture'. *Information Research* 18(3), http://InformationR.net/ir/18-3/colis/paperS09.html

Williams, M. (2015) 'Library of Cards: Reconnecting the Scholar and the Library'. *Partnership: The Canadian Journal of Library and Information Practice and Research* 10(2): 1-10.

Wilson, P. (1968) *Two Kinds of Power: An Essay on Bibliographical Control*. Berkeley: University of California Press.

Xu, F. (2011) 'A Standard Procedure for Bradford Analysis and its Application to the Periodical Literature in Systems Librarianship'. *Library Hi Tech* 29(4): 751-763.

Using sales catalogues to study the female book collectors of the Nineteenth Century

Lucy Saint-Smith, Assistant Librarian, The Grey Coat Hospital School

For my MA LIS dissertation in 2014 I researched women as book collectors in the late eighteenth and early nineteenth centuries (Saint-Smith, 2014). Book collecting was highly fashionable at that time, and the male book collectors of the era (known by their contemporaries as the "bibliomaniacs") are well documented. The historical view of women as collecting for sentimental or decorative reasons, not understanding the value of the objects and therefore not really collecting in any meaningful way at all, has obscured the history of female collectors in general and book collectors in particular. Their collections are not preserved and there is a limited amount of documentary evidence or prior research into them.

My dissertation was based around a chapter of *The Book Hunter In London* by W. Roberts on women as book collectors (1895). Roberts argues that there is no such thing as a true female bibliophile, but suggests some women who may qualify, and some others who possessed interesting collections. From his list of fifteen I selected four who lived over a similar period of time and whose sale catalogues were easily available. Frances Mary Richardson Currer (1785-1861) is one of the few female book collectors who is well known. Lady Frances Vernon Harcourt (1805-1872) was a member of the Harley family who inherited the majority of her family estates towards the end of her life. Lady Sydney Morgan (1778-1859) was a famous Irish author. Miss Margaret Bothwell Drummond (1795-1862) was the descendant of a notable Scottish family, but there is next to no information about her other than the basic biographical facts. I studied their habits as book collectors using David Pearson's framework for the study of seventeenth century libraries which proposes evaluating five categories: Contents, Acquisition, Design and Storage, Motivation and Destruction and Loss (Pearson, 2012).

Each of these women's collections were sold posthumously via Sothebys auction house, so there is a catalogue that details at least part of their collections (Sothebys, 1873, 1863, 1862a, 1862b). In addition, Richardson Currer had a private catalogue made of her collection in 1833 (Stewart and Richardson Currer, 1833). The analysis of these catalogues formed the basis of my research. I elected to use Currer Richardson's private catalogue for quantitative analysis as it contained records of more of her collection than the sale catalogue and was presented as she had chosen to present it, although I used information from the sale catalogue as part of my discussion.

The main requirement from working with these catalogues was to have a data set that I could easily manipulate to draw conclusions about the collecting habits of these women. I was looking for quantitative data to both provide a profile of their collections and to interrogate specific statements e.g. that one of the things that made Currer Richardson's collections special was the large number of natural history books she inherited from her great-grandfather, physician and botanist Richard Richardson (1663-1741).

I transcribed the contents of the catalogues from copies of the originals at the British Library and, in the case of Currer Richardson, online at archive.org into an Excel spreadsheet (see Fig. 1). I used the same fields that were present in the nineteenth century catalogues, explicitly (title, author, place of publication, publisher/printer, binder, date, size and number of volumes) and inexplicitly (language). I mostly recorded the data in each field as it appeared in the original catalogue in order to save time and to make referring back to the original catalogues easier if required. The only fields that were different were place of publication and the names of publishers or printers, which needed to be uniform to be of any use and also were not always populated fields, requiring less time to check and alter. I utilised the CERL Thesaurus for authorised terms. If I had a longer time frame for the project I would have used an authority list for the authors as well.

CC BY

Figure 1.

	Title	Author	Language	Place of Publication	Publisher/Printer	Binder	Date	Size	Volumes	Subject	Subject 2	Subject 3
2	Scritti Vari ordinati	Galileo, Galilei	Italian	Florence	-		1864 8vo		23	Philosophy	Mathematical Philosophy	Astronomy
3	Journey to Katmandu	Oliphant, L	English	-	-		1852 8vo		6	History	Historical Prolegomena	Voyages and Travels, including Foreign
4	I Contemporanei Italiani, 69 Nos.	-	Italian	Turin	-		1861 8vo		1	History	Historical Prolegomena	Travels, including Foreign Topography
5	Ruins of Sacred and Historical Lands	-	English	-	-		1853 8vo		4	History	Historical Prolegomena	Voyages and Travels, including Foreign
6	British Housekeeper Book	Walsh, J. H.	English	-	-		1860 8vo		11	Arts	Economic Arts	Domestic Economy
7	Memoires	Motteville, Mme. de	French	Maastrict	-		1782 8vo		6	History	Biographical and Monumental History	Biographical History
8	Ranoltadi Lettere sulla Pittura	Bottari, M. G.	Italian	Milan	-		1822 8vo		8	Arts	Fine Arts	Architecture
9	Rivoluzioni d'Italia	Denina, C.	Italian	Turin	-		1829 8vo		4	History	Modern History	History of Italy
10	Orlando Furioso	Ariosto, L.	Italian	London	-		1822 8vo		1	Literature	Polite Literature	Poetry

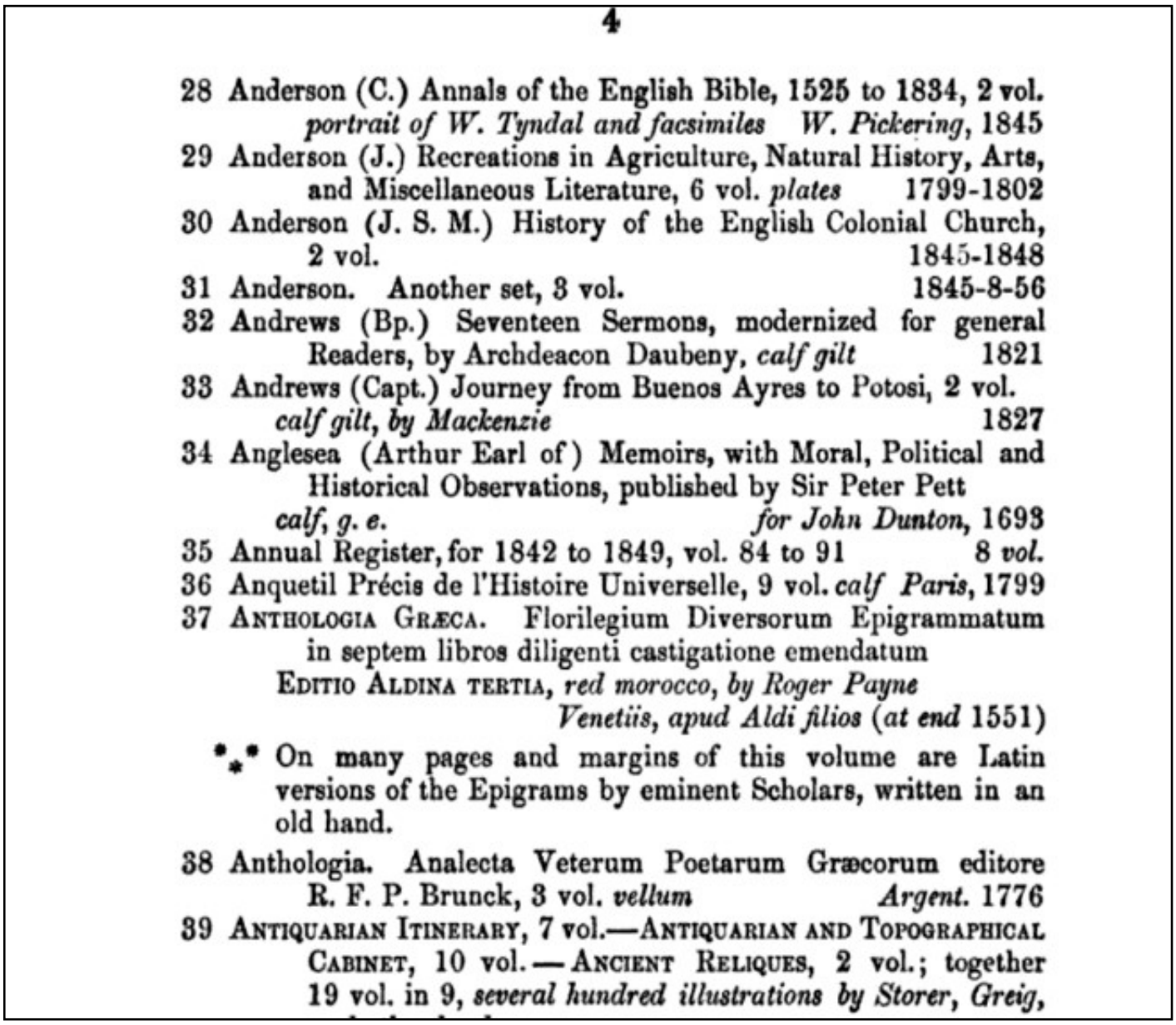
The Richardson Currer catalogue uses T. H. Horne's classification system (Horne, 1825) so I recorded this as part of the transcription and also classified the other three catalogues using this system. Horne's system is complicated with many subdivisions couched in early nineteenth century language but it had the advantages that, firstly, the largest of the four catalogues (Currer Richardson's) was already done, and, secondly, that it was a classification system of the time and so is more useful when discussing collections of the nineteenth century as it matches up with contemporary writing.

Once I had the dataset I was able to breakdown each catalogue and produce descriptive statistics for the following categories: Subject, Language, Place of Printing, Printer/Publisher, Physical Features and Date of Creation. My sample size was not large enough for inferential statistics, and it wasn't my intention to arrive at a reductive notion of a 'typical' nineteenth century woman's library. I established that each woman's library had a very different character, although there were some commonalities - each collection was mostly in English, for example, with significant numbers of books in Italian, French and Latin.

I examined several statements regarding these women's libraries using my data. This included the idea that the inherited natural history books in Currer Richardson's catalogue were particularly key to her collection, which is a 'fact' that is repeated in numerous blog posts and biographical entries about her(Angus Library and Archive, The, 2012; Gawthrop, 2002; Lee, 2004; Yorkshire Archaeological Society, 2013). My analysis showed that, although Currer Richardson had the greatest number of works classified as natural history (204 titles), proportionally Vernon Harcourt and Drummond had more than she did (Drummond 9% and Vernon Harcourt 8% to Currer Richardson's 4%). Not only does natural history not form a particularly significant portion of her library (by way of comparison she owned 1498 works on modern history) but 166 of the 204 natural history books were published after the date of her great-grandfather's death, so could not have been part of his collection. While this does not preclude the idea that she valued her great-grandfather's collection above the rest, it does not support the often repeated idea that her great-grandfather's contribution to her collection was as important as her own.

A significant limitation of my research was the fact that sales catalogues are problematic as evidence. Their primary purpose is not to be an enduring bibliographic record - it is to sell books. They have omissions, both of records and individual pieces of data. Books are often sold in lots where not all the titles are named. They are, as any catalogue, biased in favour of what the cataloguer thought was important, so the more desirable printers and publishers are recorded, but others are not. Particularly nice bindings are recorded, but others are not. This image of a page from Currer Richardson's auction catalogue (Fig. 2 from Sothebys 1862a) illustrates the variation in information recorded. It should be noted that Currer Richardson's personal catalogue is also highly selective as to what printers, provenance and places of publication it records.

Figure 2.



4

28 Anderson (C.) Annals of the English Bible, 1525 to 1834, 2 vol. *portrait of W. Tyndal and facsimiles* *W. Pickering*, 1845

29 Anderson (J.) Recreations in Agriculture, Natural History, Arts, and Miscellaneous Literature, 6 vol. *plates* 1799-1802

30 Anderson (J. S. M.) History of the English Colonial Church, 2 vol. 1845-1848

31 Anderson. Another set, 3 vol. 1845-8-56

32 Andrews (Bp.) Seventeen Sermons, modernized for general Readers, by Archdeacon Daubeny, *calf gilt* 1821

33 Andrews (Capt.) Journey from Buenos Ayres to Potosi, 2 vol. *calf gilt, by Mackenzie* 1827

34 Anglesea (Arthur Earl of) Memoirs, with Moral, Political and Historical Observations, published by Sir Peter Pett *calf, g. e.* *for John Dunton*, 1693

35 Annual Register, for 1842 to 1849, vol. 84 to 91 8 *vol.*

36 Anquetil Précis de l'Histoire Universelle, 9 vol. *calf* *Paris*, 1799

37 ANTHOLOGIA GRÆCA. Florilegium Diversorum Epigrammatum in septem libros diligenti castigatione emendatum EDITIO ALDINA TERTIA, *red morocco, by Roger Payne* *Venetiis, apud Aldi filios* (*at end* 1551)

*** On many pages and margins of this volume are Latin versions of the Epigrams by eminent Scholars, written in an old hand.

38 Anthologia. Analecta Veterum Poetarum Græcorum editore R. F. P. Brunck, 3 vol. *vellum* *Argent.* 1776

39 ANTIQUARIAN ITINERARY, 7 vol.—ANTIQUARIAN AND TOPOGRAPHICAL CABINET, 10 vol.—ANCIENT RELIQUES, 2 vol.; together 19 vol. in 9, *several hundred illustrations by Storer, Greig,*

Additionally, the sales catalogues are merely a snapshot record of the collections at a certain point in time. Such a static record fails to reflect the changing nature of a collection. There is some provenance information - again showing what sales or previous owners the cataloguers thought were important - but otherwise it is impossible to see how the collections developed. It is also impossible to say how much of the collection is being sold - Frances Vernon Harcourt's catalogue, for example, is described as being "the remainder" of her collection, but there is no other sale catalogue of her belongings. Without other information it is also difficult to say whether items were acquired for the collections by the women in question.

All this being said, sales catalogues are very important in the study of nineteenth century female book collectors for the simple reason that in many cases they are the only record of a woman's collection. Margaret Drummond, for example, despite owning a very exciting collection that included one of Carl Linnaeus' annotated copies of his work, several Persian manuscripts and numerous rare volumes of prints and drawings, is, as far as I can tell, invisible as a bibliophile outside of her auction catalogue. We therefore have to accept the limits of the sales catalogue records. We can also learn from the selective nature of the records what was considered desirable and important at the time, although not from the perspective of the women themselves.

There has been a lot of discussion recently about quantitative analysis of the library catalogue (Welsh, 2016). My dissertation has demonstrated the value of such an analysis of sales catalogues and the value of these catalogues as a research tool. This is particularly the case with collections and collectors that were not part of the accepted elite of the time and so are not well represented in the historical record. Currer Richardson is known as a bibliophile not because her collection was interesting (which it was), or because she was highly knowledgeable about her subject (which she was). She is so well known because of her relationships with the male book collectors of the time. Quantitative analysis of women's sale catalogues enables us to investigate the bibliographical lives of other women who did not have such connections, but nevertheless managed to create wonderful collections of their own.

This article is based on work done for my MA LIS dissertation at UCL, a summary of which was presented at the 2015 LISDIS Conference in Huddersfield. I would like to thank my supervisor Anne Welsh for her support.

References

Angus Library and Archive, The, 2012. A leading female book collector. The Angus Blog.

Gawthrop, H., 2002. Frances-Mary Richardson Currer and Richard Heber: Two Unwearied Bibliophiles on the Fringe of the Brontë World. Bronte Studies 27, 225–234.

Horne, T.H., 1825. Outlines for the classification of a library : respectfully submitted to the consideration of the trustees of the British Museum. Official copy for the use of the trustees. (Not published.). G. Woodfall, London.

Lee, C., 2004. Currer, Frances Mary Richardson (1785–1861), in: Oxford Dictionary of National Biography. Oxford University Press.

Pearson, D., 2012. The English Private Library in the Seventeenth Century. The Library: The Transactions of the Bibliographic Society 13, 379–399.

Roberts, W., 1895. The Book-Hunter in London: Historical and other studies of collectors and collecting. Elliot Stock, London.

Saint-Smith, L., 2014. La Femme Bibliophile: Female Book Collectors in an Age of Bibliomania (M. A.). UCL, London.

Sothebys, 1873. Catalogue of the Remainder of the Library of the late Lady Frances Vernon Harcourt. Sothebys, London

Sothebys, 1863. Catalogue of the Library of the Late Distinguished Authoress Lady Morgan. Sothebys, London.

Sothebys, 1862a. Catalogue of the Principal Portion of the Magnificent Library of the late Miss Richardson Currer, of Eshton Hall, Yorkshire. Sothebys, London.

Sothebys, 1862b. Catalogue of the Very Choice Library of the Late Miss Drummond, of Berkeley-Square Bristol. Sothebys, London.

Stewart, C.J., Richardson Currer, F.M., 1833. A catalogue of the library collected by Miss Richardson Currer at Eshton hall. London.

Welsh, A., 2016. The Rare Books Catalog and the Scholarly Database. Cataloging & Classification Quarterly 54, 317–337. doi:10.1080/01639374.2016.1188433

Yorkshire Archaeological Society, 2013. Frances Mary Richardson Currer 1785-1861 [WWW Document]. North Craven Building Preservation Trust. URL http://www.ncbpt.org.uk/folly/exhibitions/exhibitions_2013/panels/files/pdfs/panel6.pdf (accessed 8.10.14).

Compiling a complete catalogue for the Ladybird publishing Series 1940-1980: a feasibility study in children's literature and curatorship

Yvette Campbell yvette_campbell@icloud.com

Introduction

In 2010, a thesis was submitted to the School of Library and Information Studies in University College Dublin as part of my MLIS on examining the cherished Ladybird books published from 1940 to 1980. Exploring the publishing history, it questioned whether it was feasible for the series during this period to be catalogued according to rare book standards by examining the metadata used for this collections held in selected university libraries in the UK and Ireland. For a feasibility study, it was my objective to compile a descriptive catalogue of Ladybird Books from 1940-1980 in an attempt to aggregate bibliographic information regarding the early works of Ladybird Ltd that have previously been scattered across a variety of sources into one single resource.

These collections form part of the cultural heritage but they can only transmit that heritage when they are adequately accessible. In the case of Ladybird books from the mid twentieth century, they have become increasingly popular for collectors in recent years. However, for libraries and research studies, they remain curiously under-regarded, poorly documented and misunderstood. To rectify this, 5 academic library catalogues from Ireland and the UK were examined to compare and evaluate any non-conventional problems associated with standard bibliographic entries of Ladybird books from the period 1940-1980. Results showed that there was an inadequate recording of the necessary 500 Notes field for these books with some inconsistencies in the required Series field that gave incomplete and misleading information about the item being described.

In resolving any areas for concern in the library catalogues and to eradicate uncertainty regarding what has been published by Ladybird during these years, this thesis endeavoured to create a complete bibliographical tool in the form of a descriptive catalogue, the aim of which revealed a cultural and historical influence to determine whether or not this early Ladybird series is deserving of curatorship according to a rare book collection. Previous work had been done in UCD by Christine Ryan on constructing bibliographic information for the Bartlett Collection of the early Puffin Series held in the Church of Ireland College of Education (CICE) in 2004, which provided recognition and better classification for Puffin books as a learning resource.

Thus a descriptive catalogue focussing mainly on information provided in the 500 Notes field for this study may hopefully serve as a starting point in building official descriptive records of a first edition Ladybird Book collection for a public or research library in the UK or Ireland, which may function as a potential cultural learning resource.

Methodology

According to some in the publishing business, it takes quite a while to know how many books the publisher has actually sold (Lamb, 1998: 167). Therefore, the prime basis for the compilation of a catalogue for this thesis is because it has been acknowledged by the editorial director for the company, Douglas Keen, that Ladybird Books Ltd did not keep many records of the books they published at this time in the later years of the firm, nor was book production organized or researched before (Mullin. 2002). However, this loss of company records for many firms during the Second World War, which has seriously hindered book historians, can be attributed to patriotic sacrifices to waste-paper drives.

As a result of this fact, many newspaper or magazine articles concerning Ladybird Books, are ambivalent regarding an exact figure of production output during this period with regular estimations of 1000 in one to over 2500 titles in another. However, the most comprehensible and substantial work that has been researched on this collection, is the work of bookseller Robert Mullin and the team of The Wee Web (Mullin. 2002), who states that between 1940 and 1980 Wills and Hepworth/Ladybird Books, published 63 different series, collectively containing 663 books. Whether this can be taken as fact or not, one may never know as it is impossible to know how much was produced. Nevertheless, after examination, there appears to be some errors in the number of titles according to Mullin (2002). It appears that there are rather 613 titles between 1940 and 1980 as some of the titles listed by Mullin (2002) are published two or three years subsequent to 1980. Therefore 613 titles formed the body of the catalogue.

CC BY

During the production of each of these series, Wills and Hepworth produced a catalogue of titles to appear at the back of each first edition book in that individual series, while in later standard editions this was modified as just a 'title list' often with reprinted titles from later years added in sporadically, making the list un-chronological and mismatched. These were only available if one wrote to Ladybird requesting the free catalogue or list. In writing to Ladybird Books Ltd, I requested if the series catalogues were still available for perusal. Unfortunately, none of the catalogues produced during the period under review exist today.

Consequently, the assemblage of a complete descriptive and illustrated catalogue of first edition Ladybird books in contrast to a title list published during and after each series offers increased utility and identification of these books. In the recording of bibliographic information of the physical history and binding information of the book for this catalogue, the requirements differ significantly from the needs of the user for whom the catalogue is merely the key to the contents of the collection.

In establishing how best these books should be catalogued for this study, 5 web-based catalogues were investigated to discover which records of Ladybird books displayed the much needed 500 Notes field in default single record displays. The catalogues examined were academic library catalogues from the U.K. and Ireland. It should be noted that since these libraries are most likely to have an OPAC, the sample chosen for this study is not necessarily representative of catalogues in general. From surveying these records, this gave me an indication of what areas needed to be revised in the compilation of a descriptive catalogue for this book collection.

Not displaying the 500 field may hide the only field that differentiates the item from other items in the collection. Failure to display certain elements of description such as the 500 field may give users incomplete or misleading information about the item being described. Using the 300 field for Physical Description is simply not enough for older materials. It is a significant issue, with the importance given to this type of information by users highlighted by the results of a study conducted by the University of Toronto in 2002. (Carlyle & Timmons, 2002: 195). For librarians the 500 and related fields retain an acknowledged importance, especially in the cataloguing of rare books, but beyond this, these fields offer great benefit to users when searching.

To gain a professional understanding and knowledge of descriptive records for older books, I entered a search term for Ladybird books from 1939 onwards in the International League of Antiquarian Books (ILAB) database on 13 June 2010 to see if these books may be considered as antiquarian books. This search retrieved 125 hits. The earliest record retrieved was the very first book that appears in the catalogue in this study entitled *Bunnikin's Picnic Party* (1940). Unfortunately, few illustrations were available which thwarts classification due to various reprints. The latest title to appear in this database is from 2007. Although the database does not make MARC displays available to users, it was clear from the substantial binding and physical history information that the descriptions were of a standard that adequately identifies and compliments these vintage books. It was this high standard of descriptive information provided in this database that presented me with the format for the catalogue in my study.

Another reference tool used for this thesis was the Copac National, Academic, and Specialist Library Catalogue (2008). A search of Ladybird books from 1939 onwards resulted in 7115 hits entered on 15 June 2010. Although the descriptions in the Notes field was limited compared to the ILAB database, this provided my research with the correct layout design of the field titles and to correctly reference punctuation especially in the 245 Title field and the 300 Physical Details field.

The CURL Minimum Standards for Bibliographic Records (2003) was consulted for the updates in bibliographical standards especially in Section B for Post 1800 imprints on 16 June 2010. In the case of very rare books in the catalogue, the code "rbgenr" for rare book genre was adopted in the records Note field.

The field entries used in the catalogue include:

Title
Author
Publication
Physical Description
Series field
Notes field

In some instances, depending on the title, Edition Statements and Translations have been included as separate fields from the Notes field for better clarification.

Critically there is no one resource or online database that provides a complete catalogue of all 613 front cover illustrations and this has become one of the aims of this study as well as providing a complete catalogue of these books during the period under investigation.

Outcomes of Creating the Catalogue

The outcome of carefully completing a total descriptive catalogue for Ladybird publications from 1940 to 1980 has revealed a great deal about whether the vintage children's book should be catalogued descriptively in the dual disciplines of Curatorship and Children's Literature. Some books that were crucial to the successful realisation of the complete catalogue have proved frustratingly difficult to locate. Therefore, 4 out of the 613 books in the catalogue have been excluded. The contents of the catalogue should be valuable for all those interested in early children's literature regardless of whether the complete collection is available to them or not. There is no claim that this catalogue is at all without its errors or that this is the single method or approach in making this a more renowned collection of early children's books. No doubt the reader should find some errors contained within. However, from undertaking this project I have answered certain key questions relating to this feasibility study.

In terms of why these books should be given a 'rare book' status, the answer has been proven in the literature review regarding these books in the physical composition and format of the books. Also the fact that they contain material that by today's standards would be considered non-PC with illustrations depicting the style and sentimentality of a recently bygone era, constitute these books as what is now regarded as 'rare books'. For example, a good edition of the Well-Loved Tales Series 606d's 'Cinderella' (1964) accompanied by its original dust-wrapper can fetch up to £300 (Birtles 2004) on eBay and competition can be fiercely intense among collectors.

Accordingly, these books should be treated as such in an appropriate library rather than allowing this historical collection to remain undocumented. Perhaps this collection may be taught as one of the modern children's book collections in the multiple LIS disciplines of Children's Literature, Rare Books Curatorship and Book History.

References

Birtles, J. 2004, "Ladybird Books Provide Early Learning in Investment", *Independent on Sunday*, 26 Sept., [Online] available at http://www.theweeweb.co.uk/ladybird/independent_interview.php (accessed 28/5/2010).

Carlyle, Allyson and Timmons, Traci E. 2002, "Default Record Displays in Web-Based Catalogs", *The Library Quarterly*, Vol. 72, No. 2, Apr, pp. 197-204.

Lamb, Wendy. 1998, "Strange Business: The Publishing Point of View", *Signal*, No. 87, Sept, pp. 167-173.

Mullin, R. 2002, "A Ladybird History", The Wee Web [Online] available at http://www.theweeweb.co.uk/ladybird/ladybird_history.php (accessed 20/5/2010).

Mullin, R. 2002, “Interview With Douglas Keen”, The Wee Web [Online] available at http://www.theweeweb.co.uk/ladybird/douglas_keen.php (accessed 28/5/2010).S

Stevenson, Iain. (2010) *Book Makers: British Publishing in the Twentieth Century.* London, The British Library.

An exploration of the book collections of John Couch Adams and Samuel Sandars

Dr Sophie Defrance, Rare Books department, Cambridge University library

A slightly different version of this poster was created for the LIS DIS conference in 2015, and was based on my dissertation for the MA LIS at UCL, written under Anne Welsh's supervision in 2013. It aimed to present two book collectors, John Couch Adams (1819 –1892) and Samuel Sandars (1837-1894), who bequeathed their collections to Cambridge University Library. The dissertation studied the parallel lives and bequests of the two men. I have chosen to keep the parallel aspect for the poster. My aim in this revised version is to present the difference in the way the two collections were processed.

I went for visual highlights rather than text, using two of the most beautiful bindings, Adams 7.67.14, *Book of Common Prayer*, 1676, a Mearne binding, and Sandars's SSS.42.9 John Barbour, *The Life and Acts of the Most Victorious Conqueror Robert Bruce, King of Scotland*, Edinburgh, 1758, a 19th painted wood and varnished binding. Meanwhile the background evokes the location, as it is *Oppidum Cantebrigiae***,** Cambridge, Richard Lyne, 1574 (Item no. 7 in volume SSS.12.1), which is the earliest known complete map of Cambridge. The poster highlights the main facts about the collections, and illustrates the different ways in which they are kept. Part of the Sandars's donation was his bookshelves, which are not in use today for conservation reasons, but that I have included here.

The Sandars Collection

The collection is arranged by rough order of size, in the Rare Books department book-stacks, which are not in open access. Contrarily to the Adams collection, the Sandars collection is entirely placed in locked shelve and in a controlled atmosphere in a newer aisle of the Rare-books department; this is in part because, contrarily to the Adams collection, the incunabula bequeathed have been kept in. All the books, unless there are size-restrictions, bear a slip from the university Library which indicates that they were "Bequeathed by Samuel Sandars, M.A. of Trinity College". The brief description on the library website insists on the strengths of the collections: "Liturgies, early English printing, books on vellum, fine bindings, 109 incunabula; 15th-19th century."

A brief preliminary survey seems to indicate that many of the book still undistributed in the MS catalogue of the private library are now in the SSS collection, but not all; further research could focus on why and what kind of books have been rejected or excluded from the bequest.

The Library possesses a MS list of books in the Sandars collection, but only for English books from 1501 to1700. The list includes the dates, titles, etc. but no class-marks – it is indicated, in pencil on the first page, as being part of the annual report for 1894. It could have been part of a draft for classification. It is part of an ensemble of Sandars and Jenkinson's manuscript about the private library and bequest originally part of the collection and now increasingly re-classified as manuscripts. Jenkinson's own working notes on the Sandars bequest also used to be part of the collection as SSS.10.20 – (now, add 4172.) It is a list of Sandars' books in August 1894, which also shows that some books were found in drawers and given to the Librarian by Mrs Sandars in July 1915 (although these were not found in the UL collection.) Although the incunabula have been kept together with the rest of the collection, the MS have not, even when they were first classified as "SSS". This is notably the case of Sandars' lists of books (Cambridge University Library, Department of Manuscripts and University Archives, (1869–1915) Lists of books, manuscripts and incunabula belonging to Samuel Sandars, GBR/0012/MS Add. 4170-4172), re-classified as Manuscripts. And SSS.16.6 "Some of the books printed in the fifteenth century in the Cambridge University Library compiled from various sources and notes made on the spot 1869" now Add.4171. The original presence of the documents in the collection, however, shows that it was, from the beginning, important and carefully monitored; it was also more advertised than the Adams collection: Sandars' obituary in the Cambridge Review pointed out to the only Groslier binding and the only two Caxtons added since the death of Bishop Moore in 1715 (*Cambridge Review* 1894, p. 55.) The aesthetics and "treasured" aspect of Sandars' books mean that they are frequently used as a kind of window-display items for the University Library. In 1998, 4 Sandars items feature as illustrations of the great collections of the Library (Fox, 1998) and more recently, a book from the Sandars collection was chosen to be a

CC BY

"featured book" on the University Library rare books website. Sandars also featured in the exhibition "Shelf Lives, Four Centuries of Collectors and their Books". Sandars' books were also used most recently in an incunabula masterclass given in June 2013 by Professor Lilian Armstrong.

The Adams collection:

Along with other rare books named collections, the Adams collection is briefly described on the University Library website as a 'general library of 1500 volumes ranging from the 15th to the 19th century, including 1000 which date from before 1700, and 84 incunabula, with an emphasis on science, especially astronomy'. There is no other general description. The collection is arranged by rough order of size, in the Rare Books department book-stacks, which are not in open access. They are placed under normal conditions of conservation, i.e. not in specifically locked shelves, with the exception of the incunabula which have been classified and kept separately.

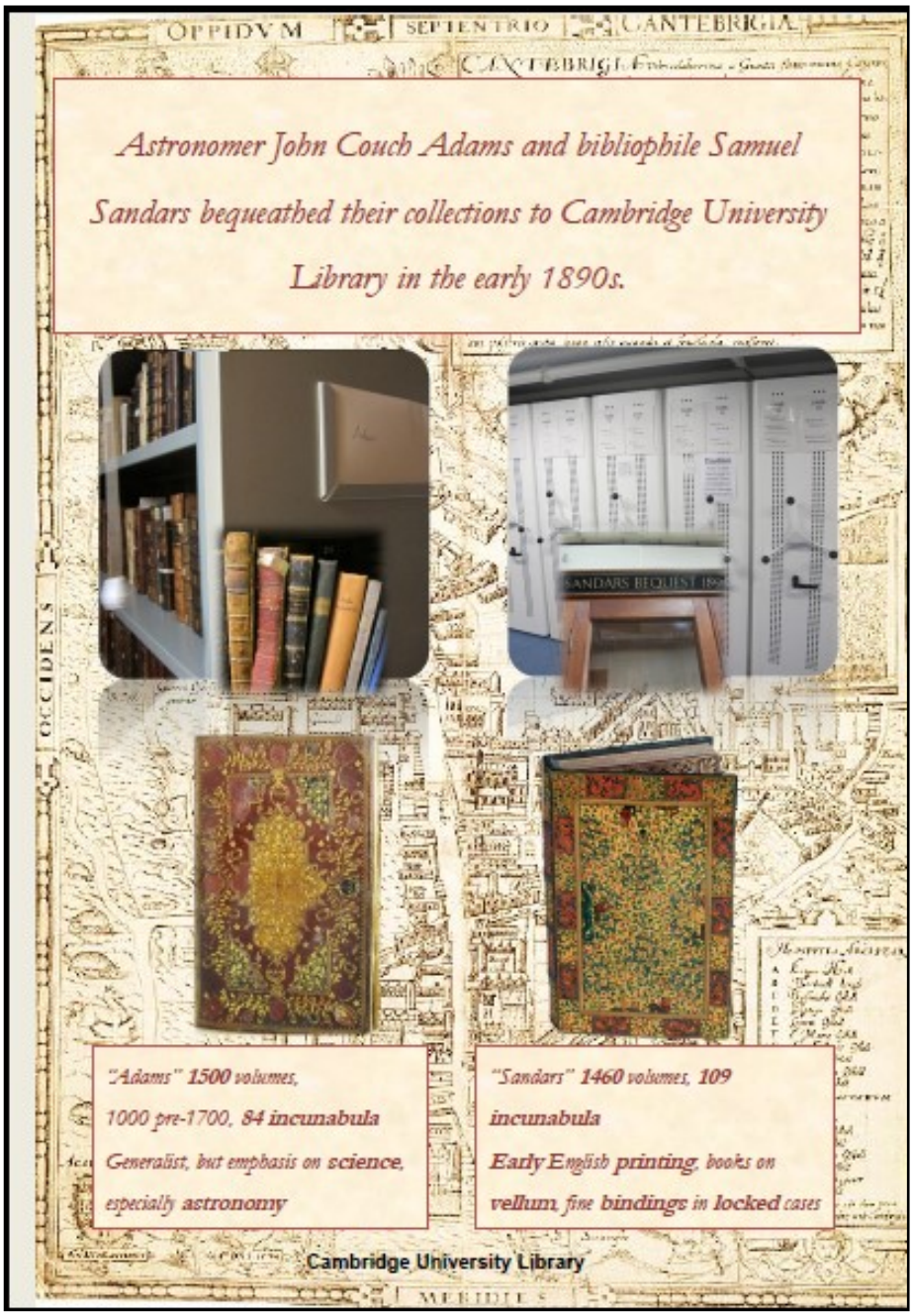
The 10th volume (extra series) of the University Library Bulletin in 1894 consisted of a printed catalogue of the Adams Collection who was published in a bound volume. The printed catalogue contains a printed addendum, and the copy in the UL rare books rooms includes some later additions by hand. This catalogue of the Adams collection also provides an index of authors, titles or subjects. The collection was inspected on January 10, 1902, when a number of addenda and corrections were added by hand to the University catalogue: there was much work on the classification in the first years of the collection, which was first classified whole, before the incunabula re-classification. Some of the items have been added later on, which seems to indicate a bequest/ choice of books in several steps. The striking point is that it has not been kept together; the incunabula have been separated, in sharp contrast with what happened in Sandars' case. The books have had little public exposure since then. "A small exhibition" was organized at the Cambridge University Library in 1996. The booklet published at the occasion proves some brief details on Adams life and collections. Only 3 books only were chosen for that exhibition, including two incunabula, one from the astronomical Library.

Re-cataloguing and complete description of the incunabula in the collection has however taken place thanks to the Incunabula Project, launched in October 2009, highlighting their place and importance in the Adams collection.

I would like to thank my supervisor Anne Welsh for her support.

Poster presented at the LIS DIS conference,

Huddersfield, 14 November 2015



Reproduced by kind permission of the Syndics of Cambridge University Library

Report from the Cataloguing and Indexing conference, "Innovation and discovery" 31st August-2nd September 2016

Emma Booth, London School of Economics

The biennial conference of the Cataloguing and Indexing Group took place between 31st August and 2nd September at Swansea University's Bay Campus. The conference theme was "Innovation and Discovery," with the aim of demonstrating how libraries, archives and museums are striving to improve the quality of their metadata in order to enhance resource-discovery for their users.

Papers and presentations at the conference covered a range of interesting and pioneering metadata enrichment and improvement projects, including collaborations between various libraries, archives and special collections. These projects demonstrate how many institutions are embracing cooperative methods to work together to enhance discoverability and meet the expectations of their increasingly more connected user groups.

The conference explored how refinements in metadata standards and the adoption of Linked Open Data formats, such as BIBFRAME, can enable librarians to acquire new skills in metadata creation and manipulation, whilst simultaneously improving the discoverability of library-resources on external systems via the web. Projects that transform bibliographic metadata into Linked Data are instrumental in unlocking discovery and ensuring that library collections are no longer hidden away in a library's local catalogue or repository; this is because Linked Data standards are compatible with web-data standards and can be indexed by web-based search engines.

Linked Data standards can also present new opportunities for cross-disciplinary research, as they enable users to further explore the relationships and links between different works, individuals, institutions, events and places. This means that library collections can not only expand their discoverability from local to global audiences, but also have a wider impact upon research and learning communities. As such, Linked Data projects can enable an institution to shift towards a more 'user-centric' approach to resource discoverability, acknowledging the fact that researchers often choose to use external systems, tools and platforms to search for information, rather than just using a library catalogue.

Sun, sand and metadata

CC BY

Throughout the conference there were examples of the fundamental work that cataloguers and metadata librarians are doing on a daily basis in order to ensure that collections are made discoverable and accessible to library-users. For example, many libraries are investing time and staff resources in upgrading their legacy metadata records from old standards, and are steadily FRBRising their library catalogue in order to make its content more discoverable. Other institutions are striving to meet the added challenges posed by an expanding number of electronic collections that are hosted on a variety of platforms, with cataloguers working hard to maintain the metadata for these resources due to vendors supplying records of varying quality.

Many of the papers demonstrated that without high quality, standardised bibliographic metadata it is impossible for a library-user to know what resources are in a library's collections, whether they are relevant to the their research, how they relate to materials they have already accessed, or how to gain physical or electronic access to those resources. As a consequence, cataloguing and metadata practices, together with library systems and discovery layers, ultimately determine the user's experience of a library. Thus, whilst the work of the metadata team is done "behind the scenes" and is not directly visible to the end-user, it is fundamental to a library's functionality and, ultimately, its reputation.

The overall feeling of the conference was that cataloguing and metadata librarianship is in an exciting place, with great opportunities for development and innovation opening up through projects involving Linked Data. However, there was a feeling that cataloguers and metadata specialists need to be more vocal advocates for the work that they do, particularly when demonstrating to their stakeholders the importance of metadata enrichment projects as a means of enhancing the user-experience and improving the discoverability of collections.

Slides, workshop materials and posters from the conference can be found at:

http://www.cilip.org.uk/cataloguing-indexing-group/presentations/conference-2016-innovation-discovery.

Catalogue and index 185 will include a large selection of papers from the conference.

Managing metadata in web-scale discovery systems (2016) edited by Louise F. Spiteri. London: Facet Publications, ISBN 9781783300693 (paperback)

Web-scale discovery systems are becoming increasingly common in libraries and this book aims to show librarians how to get the most out of these systems through the management of metadata. Products such as Primo (Ex Libris), Summon (ProQuest), EBSCO Discovery Service and OCLC WorldCat are highlighted, but most of the discussions focus on discovery systems in general. The contributors have a range of expertise meaning the book covers many different aspects. Each chapter is designed to be read independently so there is a certain amount of crossover and repetition but this often serves to emphasise important points and arguments.

The introduction provides a useful brief history of the development of library catalogues towards web-scale discovery systems which sets up the rest of the book well. It explains how the "Google effect" has changed the way people search for information on the internet and affected user expectations of search and discovery in the library. The scale of managing the huge variety of resources held in discovery systems and their differing metadata is a major challenge and I felt this was discussed successfully.

Some technical aspects of managing and sharing metadata are explored, with the volume of resources available today meaning sharing is now crucial. After covering metadata mapping and systems such as FTP, the conclusion was for a focus on collaboration and decentralisation rather than local enhancements benefiting only a single catalogue. This is followed by a chapter about linked open data and the opportunities it provides for libraries. It was interesting to learn about the leading projects, primarily from national libraries, and see some real examples of linked open data sets. Big data, the Semantic Web and BIBFRAME are also covered but nothing gets too complicated, making it a great introduction to this area.

One chapter focuses on academic libraries and how they handle the amount of content and metadata in their discovery systems. This is then compared with Google Scholar, arguably their biggest competitor, creating a fascinating look at how library systems compare with the corporate giant used by so many. There is then a detailed discussion of the relationship between libraries and vendors, and in particular the difficulties that often occur.

The final chapter focuses on the social features of discovery systems, especially user-generated metadata such as tags and reviews. It goes through a number of studies showing the possibilities of user tags and suggests that these can enhance the existing metadata in library catalogues by adding natural language alongside controlled vocabularies. The chapter concludes that user-generated metadata in discovery systems has not yet reached its full potential. This was one area that I had not previous considered but this chapter caught my attention and left me intrigued as to how it will develop.

This book was an interesting read and made me think in more detail about the metadata I create. Altogether it was clear, as obvious as it may sound, that metadata is crucial to discovery. The performance of discovery systems and the accuracy of their results depends on the available metadata, but there can be huge differences in its structure, quality and completeness. By looking at a range of aspects the book showed the bigger picture of what happens in web-scale discovery systems and ultimately how users find information.

Catherine Smith, Cataloguer, ProQuest/Coutts Information Services

Book review: RDA Essentials (2016) Thomas Brenndorfer. London: Facet, 256 pages, ISBN 99781783300563

RDA Essentials bills itself as "a concise guide to cataloguing with RDA". As such, anyone feeling rather daunted by RDA and/or expecting a neat volume along the lines of the concise guide to AACR2 may not feel encouraged by this rather hefty tome. But fear not and stick with it. RDA Essentials delivers pretty much as promised, clearly and systematically, and without getting bogged down in any long-winded explanations.

The book begins with a potted history of the development of RDA which helps provide some context for its use and what RDA is trying to achieve which is, ultimately, greater discoverability of resources. The introduction is the most discursive section, and teeters on providing a summary, step-by-step guide to the whole process of RDA cataloguing...but not quite, as it loops around and repeats itself. Given the straightforward and plainly structured format of the following chapters, I felt this could have been done better. That said, various FRBR concepts at the core of RDA - content v. carrier, user task, relationships - are all described neatly and reiterated throughout the rest of the book.

The work concerns itself firstly with the elements of RDA (the bulk of the book), followed by guidelines (the 'how' rather than the 'what' to record), construction of access points and a short section on "additional instructions" which addresses the kind of "what if…" scenarios not yet covered by the main body of instruction. The elements themselves are worked through systematically, organised into chapters which address particular attributes or aspects of a resource. Each chapter opens with a table of the elements covered and then proceeds to drill through each in turn, with a concise description, example and anything else pertinent to that element, notably options, alternatives, common vocabulary (where appropriate) and a pointer to related elements. And that is more or less it. It is, essentially, exactly as it says; a concise guide, a distilled and more accessible version of the full RDA Toolkit. And while comprehensive, it is not exhaustive. There are caveats, for example, that it is not intended for more complex resources, especially musical, religious and legal works.

All considered, my only real criticism would be that some of the instructions and examples are a little too succinct. For a cataloguer completely new to RDA, this might better sit alongside an introductory text which provides more in the way of context for each element, by showing how they are applied to a few example resources as a whole, say, rather than in isolation. But as an easy-to-use reference for cataloguers already getting to grips with RDA, this is an excellent resource and a worthwhile investment.

Helen Burns, Edinburgh University

Catalogue & Index is electronically published by the Cataloguing and Indexing Group of the Chartered Institute of Library and Information Professionals (CILIP) (Charity No. 313014)

Advertising rates: GBP 70.00 full-page; GBP 40.00 half-page. Prices quoted without VAT.

Open access statement: http://cilip.org.uk/cataloguing-indexing-group/catalogue-index/open-access-statement

Submissions: In the first instance, please contact the Co-editors: Helen Garner, (h.j.garner@shu.ac.uk) and Karen Pierce (PierceKF@Cardiff.ac.uk)

For book reviews, please contact the Book Reviews Editor: Neil T. Nicholson, Cataloguing & Metadata Services Team Leader, National Library of Scotland, e: n.nicholson@nls.uk

ISSN 0008-7629

CIG website: http://www.cilip.org.uk/cataloguing-and-indexing-group/catalogue-index

CIG blog: http://catandindexgroup.wordpress.com/

CIG twitter @CILIPCIG

Tags from the CIG blog on specific areas of interest:

authority control book reviews Catalogue and Index cataloguing CIG activities CIGS classification committees and working groups conferences Dewey digitised material Dublin Core events folksonomies linkblog MARC metadata news RDA Semantic Web social software standards taxonomies UDC