

UC Irvine

UC Irvine Previously Published Works

Title

Exploring Bibliographic Records as Research Data

Permalink

<https://escholarship.org/uc/item/0fc560xq>

Authors

Wallbank, Sarah

Kane, Danielle

Dickerson, Madelynn

et al.

Publication Date

2019-12-01

Peer reviewed

# Catalogue and Index

Periodical of the Cataloguing and Indexing Group,  
a Special Interest Group of CILIP, the Library and Information  
Association



**Editorial**

**December 2019, Issue 197**

Welcome to issue 197 of C&I,

In this issue, we offer a sample of the wide and varied range of research being conducted into – and making use of – bibliographic metadata.

Sarah Wallbank, Danielle Kane, Madelynn Dickerson, and Joshua Hutchinson present early insights from their investigations into using the University of California's bibliographic data for digital humanities research. They look at some of the questions that they might attempt to answer, and take us through how they aim to clean the data. Their work so far has left them hopeful that the catalogue might provide them with fertile ground for future research.

Annick Stein compares data from large-scale RDA- and AACR2-based catalogues (one in Switzerland, the other in Luxembourg) in order to assess whether moving to RDA really does help create a more FRBRised catalogue. Her findings – that the RDA-based catalogue does not present to end users the structured, linked experience we might hope – reinforce this editor's view that the job of the metadata specialist is not complete once the record is created, and engagement with the mechanics of the discovery layer is essential to ensure that users can make best use of our hard work.

Mackenzie Johnson and Carlie Forsythe's article, on disability and accessibility language in subject headings and social tagging, stresses the importance of involving subject experts in the creation of subject headings, and of getting the headings right to allow effective information retrieval. The authors also assess the 'third way', of semi-structured, moderated social tagging systems, that lies between fully controlled vocabularies and free social tagging.

## Contents

**3-9:** Exploring bibliographic records as research data, by Sarah Wallbank, Danielle A. Kane, Madelynn Dickerson, & Joshua Hutchinson

**10-15:** Has FRBR revolutionised our catalogues? A comparative analysis of AACR2 and RDA-formatted records to the FRBR model, by Annick Stein

**16-26:** Disability and accessibility language in subject headings and social tags, by Mackenzie Johnson, & Carlie Forsythe

**27-37:** BIBFRAME-ing the rare books catalogue: The Art and Rare Materials (ARM) ontology and the transformation of special collections metadata, by Argula Rublack



Finally, Argula Rublack looks at how special collections cataloguers of the future might use linked data. Using a copy of Copernicus's "De revolutionibus orbium caelestium" as an example, she models work, edition, and item descriptions in the Art & Rare Materials (ARM) BIBFRAME Ontology Extension, taking us through the mapping challenges, and comparing the results with the MARC original. She analyses the viability of automating such a process, and presents recommendations for further development.

Enjoy!

The Editors

Karen F. Pierce: [PierceKF@Cardiff.ac.uk](mailto:PierceKF@Cardiff.ac.uk)

Philip Keates: [P.Keates@Kingston.ac.uk](mailto:P.Keates@Kingston.ac.uk)

## Exploring Bibliographic Records as Research Data

**Sarah Wallbank**, Electronic Resources & Serials Cataloging Librarian,

**Danielle A. Kane**, Digital Scholarship Services Emerging Technologies Librarian,

**Madelynn Dickerson**, Research Librarian for Digital Humanities and History

**Joshua Hutchinson**, Cataloging and Metadata Librarian, University of California, Irvine

### Exploratory Origins

This article describes the way in which a group of four librarians at the University of California, Irvine is exploring potential uses for bibliographic data from the library catalog for digital humanities (DH) research. The project started when the Research Librarian for Digital Humanities and History (Madelynn Dickerson), who has a background in collections and technical services, reached out informally to colleagues in the Cataloging and Metadata Services department. In a December 2018 email she wrote, “There are so many interesting intersections between DH and ‘technical services’ work. I would be really interested to work with you on something, and I’m particularly interested in ideas for demonstrating the value of library data for scholarly research if that’s something you’d ever be interested in working on too.” The Cataloging and Metadata Librarian (Joshua Hutchinson) heeded the call, and after a few exploratory conversations, the group ultimately included the Electronic Resources & Serials Cataloging Librarian (Sarah Wallbank) and the Digital Scholarship Services Emerging Technologies Librarian (Danielle Kane).

From these casual beginnings, our group developed the dual purposes of (1) showing campus researchers the possibilities of using the catalog metadata for their work, and (2) providing a hands-on educational project for the librarians involved. This project gave us an opportunity to practice managing and cleaning large datasets (using tools like C# MARC Editor and OpenRefine), and to practice using digital humanities research methods such as text analysis (using tools like Voyant Tools).

### Exploratory Procedures Scoping

Starting in December 2018, our group began regularly scheduled monthly meetings where we discussed our scope and planned procedures. We decided to focus on an analysis of women authors of print monographs in History (all monographs with Library of Congress call numbers within the C-F range). With this scope, we hoped to get a better understanding of how bibliographic data from UC Irvine’s library catalog could be used by researchers, what skills would be needed in order to successfully complete a similar project, and complete a use-case as a model for future work.

Our initial research questions included:

- Of all the monographs in our catalog with the call numbers C-F, how many were written by women?
- Are women historians likely to write about a particular topic within the discipline?
- Is it possible to accurately and ethically identify an author as a woman based on their name alone? How would one go about doing this for the purpose of scholarly analysis?

We chose these initial research questions because we found them interesting and relevant to our work, and because they helped us define a reasonable scope for the project while simultaneously forcing us to engage with bigger critical questions beyond solely number crunching. In addition, as all members of the team were somewhat familiar with the metadata included in bibliographic records, the scoping exercise for this project involved thinking about how this bibliographic data could be used to achieve interesting research results. For instance, because all records include publication information (generally place of publication, publisher and date) additional research questions related to the diversity of the collection in terms of place of publication and publisher— does the UC Irvine print collection focus primarily on the Anglo-American world? Is history primarily published/collected from university publishers in the US and the UK? Has the library been collecting broadly across the decades, or are there interesting patterns that might be gleaned from studying the date of publication?



## Downloading the Data

Once we had decided on a general scope, our next step was to download the bibliographic data from our library management system, Ex Libris' Alma. Our first data download from the Alma Analytics module took place on January 11, 2019 as a Binary MARC file that included records with LC Classification that began with C, D, E, F and had a location of Langson Library (the building that houses our humanities monographs, including history). The intention was to capture all history monographs with a physical copy. We excluded serials by accounting for the location within the library, and the bibliographic leader byte 7. We expected that some level of serials and electronic books would come into our data, but for the purposes of this exercise, a small amount of imperfection was deemed acceptable. Exporting from Alma Analytics had to be completed in four exports (one per letter) because it appeared that Alma Analytics was unable to export more than 65,000 records at a time. We then ran the MARC records through the C# MARC Editor program (<https://csharpMARC.net/>) in order to create a .csv file. This data file was 306 MB and had 184,105 rows (this number was later reduced as the data were refined and additional serials were removed based on the Leader byte 07) and 220 columns, with each row representing a single physical book and each column indicating an individual MARC field.

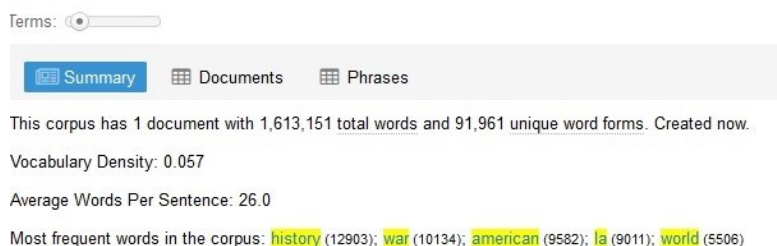
## Playing with the Data

Once we downloaded the data, the first thing we did was simply play with it as a means of becoming familiar with it and exploring the possibilities of what we could learn. This dedicated “play time” informed later decisions about the direction of our work and produced some preliminary statistics that, while not rigorous enough to draw scholarly conclusions from, gave us a sense of the real potential of the project.

We uploaded a spreadsheet with just the title data from our dataset, uncleaned, into Voyant Tools (<https://voyant-tools.org>), an open source, browser-based platform for text analysis. Voyant provides a default display featuring a visual word cloud of frequently occurring terms, and a corpus summary that includes statistics such as the total number of words in a document.

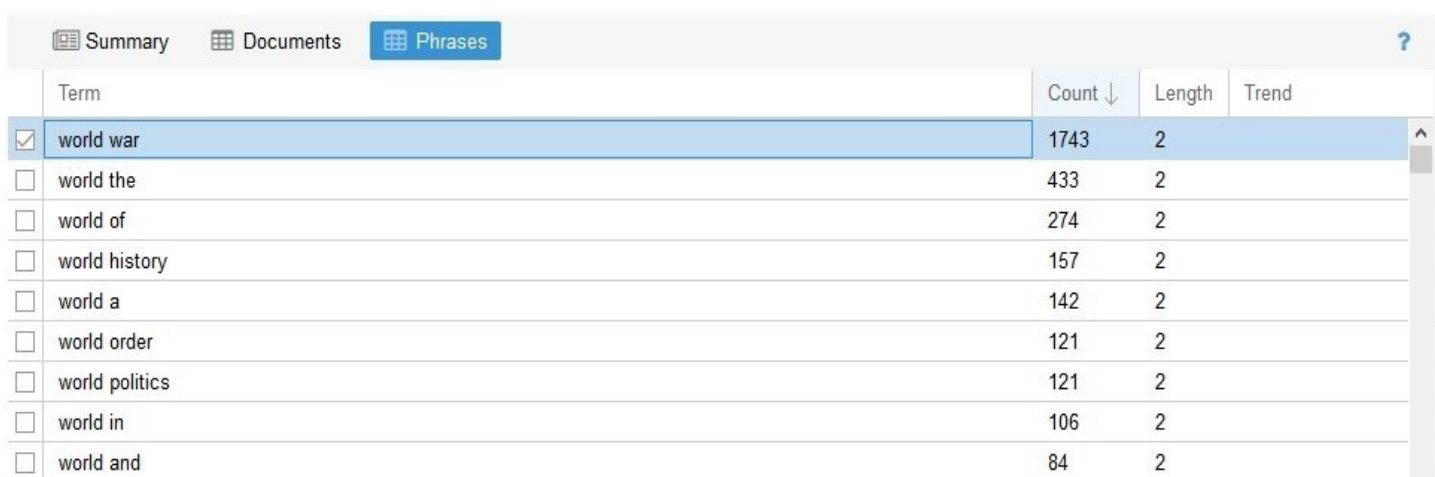


Figure 1: Screenshot of the Voyant Tools dashboard showing a word cloud and summary table of the raw, uncleaned bibliographic data for monographs in call numbers C, E, D, and F.



According to Voyant, our (uncleaned, imperfect) corpus of book titles had 1,613,151 total words and 91,961 unique word forms. The most frequently occurring words in the corpus were *history* (12,903 occurrences), *war* (10,134 occurrences), *american* (9,582 occurrences), *la* (9,011 occurrences), and *world* (5,506 occurrences) (see Figure 1, on previous page).

While unscientific, these exploratory results provided general insight into the makeup of UC Irvine’s collection of history monographs. It is no real surprise that among history book titles, “history” is the most frequently occurring term. It is, however, a bit sad that “war” is the second most frequent. A term like “la” is likely a definite article appearing in multiple romance languages, and could potentially also be a reference to Los Angeles. The word “world” is interesting, especially in that via Voyant’s “phrases” tool view, we see that the most frequent two-word phrase by far is “world war,” which occurs 1,743 times (see Figure 2). Voyant’s “phrases” view, as well as its separate “Contexts” widget are helpful in identifying the particular meaning of frequently occurring words by enabling researchers to see the other terms that appear in proximity.



Term	Count ↓	Length	Trend
<input checked="" type="checkbox"/> world war	1743	2	
<input type="checkbox"/> world the	433	2	
<input type="checkbox"/> world of	274	2	
<input type="checkbox"/> world history	157	2	
<input type="checkbox"/> world a	142	2	
<input type="checkbox"/> world order	121	2	
<input type="checkbox"/> world politics	121	2	
<input type="checkbox"/> world in	106	2	
<input type="checkbox"/> world and	84	2	

Figure 2: The “Phrases” tool in Voyant Tools provides insight into term frequency in context. Here we can see the phrase “world war” is the most frequent phrase in the corpus.

We also wanted to look at author data. We had done some preliminary reading on similar projects, such as Peng et al’s 2014 article, “Author Gender Metadata Augmentation of HathiTrust Digital Library,” which explained the way that team had determined author gender using metadata available in the HathiTrust Digital Library. Peng et al used a range of name matching techniques, including Virtual International Authority File (VIAF) lookup, and matching data to baby name websites. We were not ready for that. What we did want to experiment with, however, were ways to simply tell which names appeared more frequently than others. As with the title data, we therefore uploaded a spreadsheet including only (uncleaned, imperfect) author name data to Voyant Tools. We’re not prepared to make definitive declarations of author gender at this time, however Voyant Tools did display the most frequently occurring terms in our author data and they were *john* (5,312 ), *robert* (3,546 occurrences), *david* (3,423 occurrences), *william* (3,228 occurrences), and *james* (2,886 occurrences) (see Figure 3, on following page). Our author name dataset included both first names and last names, so it is likely that some of these frequently occurring names appeared as last names.

Rank	Term	Count
1	john	5,312
2	robert	3,546
3	david	3,423
4	william	3,228
5	james	2,886

Figure 3: A list of the 50 most frequent terms from the author data column, truncated for viewability. The first ostensibly female name is “Mary”, which appears 636 times and is the 40th most frequently appearing name.

Text analysis is fairly unforgiving work, and uploading our uncleaned data highlighted many errors and discrepancies in our dataset that need to be cleaned. Playing around with the title and author data in this way helped us to see these errors so that we could target them in our next step, which was data cleaning. It also helped us to feel comfortable working with the dataset and to learn about the various features of Voyant Tools.

## Cleaning the Data

Our first attempts at data cleaning were to discuss each of the columns contained in our dataset and to determine what data needed to be kept going forward. Our main focus was to start working with the publisher, location, and publication date fields. First, we created a file naming convention for versioning while cleaning the data. After reviewing the dataset, we determined that some columns of data were unnecessary for the current project and we removed these in order to reduce the size of the dataset before uploading the .csv into OpenRefine (<http://openrefine.org/>). OpenRefine runs in a browser window and the larger the dataset, the more memory OpenRefine will need to be able to work with it effectively. We would have run into memory errors without reducing the size of our dataset. As it was, we experienced lag on a number of clustering and merge operations.

OpenRefine was used to split the publication information (from the MARC 260 and 264 fields) into separate columns for place of publication, publisher, and date of publication by splitting multi-valued cells using the \$ sign as a separator (see Figure 4, on following page). The results were reviewed and we made adjustments as necessary using Excel. It was important to make sure that the same information appeared in the same column and, due to differences in the amount of information contained in the 260 field, columns needed to be adjusted. Information that was originally contained in 1 column was split into and organized by hand using Excel into a total of 33 columns. We then re-uploaded the file, now with the publication information from the MARC subfields in separate columns, into OpenRefine. By using the facet tool, we first turned columns into a text facet. Then, using the cluster feature, we were able to remove extraneous information surrounding the data we were hoping to clean. We removed the subfield delimiters “a,” “b,” and “c”, and also extra punctuation such as periods, commas, colons, semicolons, and “less than” symbols. While we focused on standardizing the publication year from the MARC subfield ‘c’, we ultimately plan to standardize place of publication and publisher (MARC subfields ‘a’ and ‘b’).



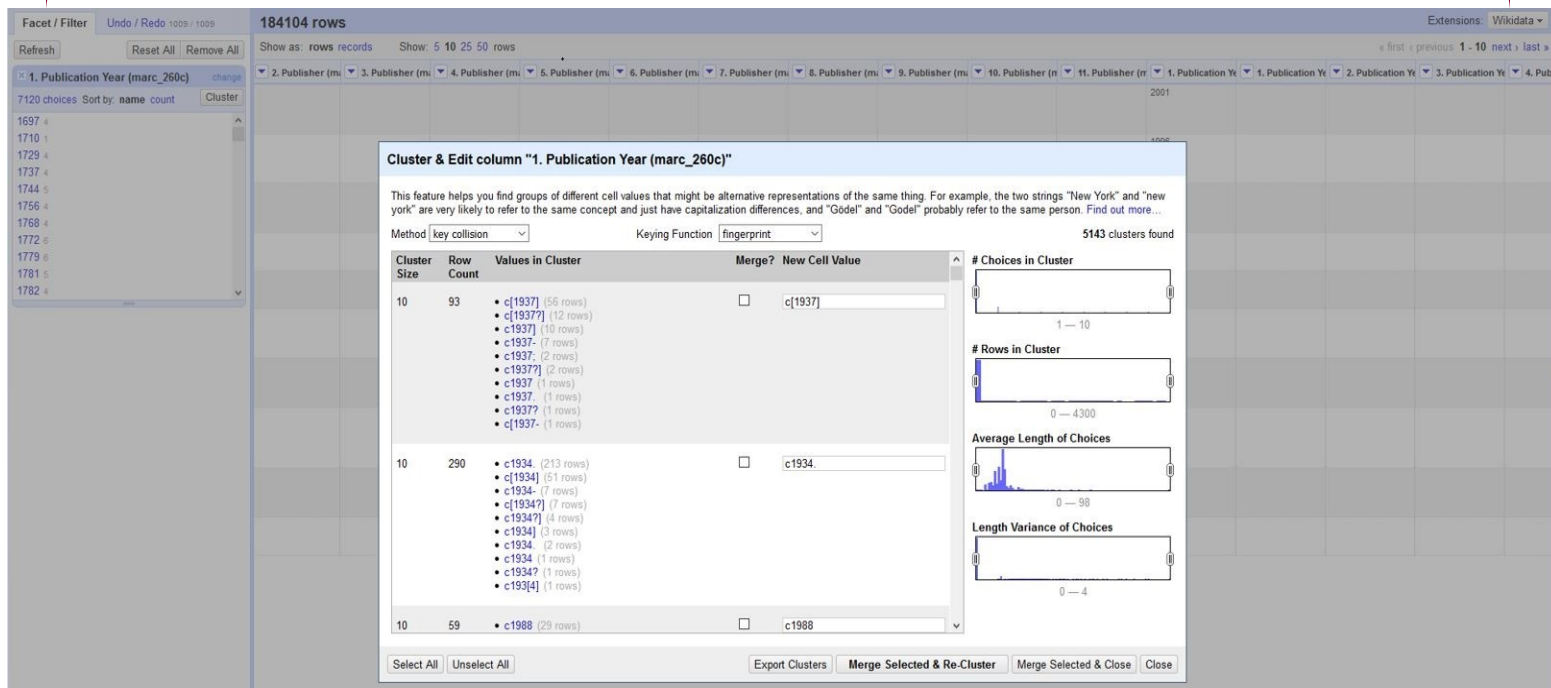


Figure 4: Cluster and edit feature in OpenRefine. The clustering function can help identify inconsistencies in data from misspellings, to non-standardized value formatting, or input error.

Using the publication year, we considered splitting the single large data file into multiple smaller files by decade in order to make data cleaning and processing easier with limited computing power, and as a way to split the work between members. We tested sorting and splitting the data up by decade, but determined that this might introduce too many discrepancies between files so decided to keep all data together throughout the cleaning process. The fact that we had a file naming convention, and saved our files at each stage of the cleaning process, made it easy to take a step back in our process and proceed in a new direction.

Further, we made an effort to connect the MARC language and country codes with the corresponding English language term (e.g., *eng* would become English and *enk* would become England). We did this by querying the Library of Congress linked open data system, parsing the HTML for the page title, and then extracting the name of the language and country from that HTML. While this procedure worked in small samples, it is as yet unsuccessful for the full dataset, presumably due to the large size of the dataset and the computing capacity required.

## Looking to the Future (Exploratorily)

This is only the beginning for the project. Our biggest questions remain unanswered and we look forward to taking the first steps towards our original goal of determining and evaluating authors' gender. Our next steps might include:

- Substituting MARC codes with text for human readability;
- Applying what we learned cleaning location and publication data fields to title and author data;
- Dividing the spreadsheet by decade for individual team members to do further data cleaning and temporal analysis;
- Performing a reconciliation of names in our dataset against databases that provide "name registries," such as Wikidata.



## Conclusion

Playing with the data before actually cleaning it may seem out of order, but playing with it has helped us to know what we want to do, or need to do, for cleaning. Because one of the goals of this project is to use this as a demonstration of the potential of bibliographic metadata for research, we thought it was important to spend some time with our data, thinking about what trends and patterns we could glean from it. In addition, we were interested in challenging ourselves to learn new tools, and to make sure that all members of the team gained new skills. We are looking forward to tackling the next steps of the project in order to more definitively address our initial research questions.

## References

- “C# MARC Editor.” Accessed November 8, 2019. <https://csharpMARC.net/>
- Doboš, Ján, and Tibor Csóka. “CRISP-DM Data Evaluation Phase of Simple Bibliographic Data Evaluation Platform Design.” *Innovative Methods in Education and Research*, n.d.
- Lahti, Leo, Jani Marjanen, Hege Roivainen, and Mikko Tolonen. “Bibliographic Data Science and the History of the Book (c. 1500–1800).” *Cataloging & Classification Quarterly* 0, no. 0 (January 7, 2019): 1–19. <https://doi.org/10.1080/01639374.2018.1543747>.
- Marciano, Richard J., Robert C. Allen, Chien-Yi Hou, and Pamella R. Lach. “‘Big Historical Data’ Feature Extraction.” *Journal of Map & Geography Libraries* 9, no. 1–2 (January 1, 2013): 69–80. <https://doi.org/10.1080/15420353.2012.732020>.
- Muñoz, Trevor. “Recovering a Humanist Librarianship through Digital Humanities.” In *Laying the Foundation*, edited by John W. White and Heather Gilbert, 7:3–14. Digital Humanities in Academic Libraries. Purdue University Press, 2016. <https://doi.org/10.2307/j.ctt163t7kq.4>.
- O'Dell, Allison Jai. “Book Artists Unbound: Providing Access to Creator Metadata with EAC-CPF.” *Art Documentation: Journal of the Art Libraries Society of North America* 33, no. 2 (2014): 267–78. <https://doi.org/10.1086/678527>.
- “Openrefine.Github.Com.” Accessed November 8, 2019. <http://openrefine.org/>.
- Pattuelli, M. Cristina, Karen Hwang, and Matthew Miller. “Accidental Discovery, Intentional Inquiry: Leveraging Linked Data to Uncover the Women of Jazz.” *Digital Scholarship in the Humanities* 32, no. 4 (December 1, 2017): 918–24. <https://doi.org/10.1093/llc/fqw047>.
- Peng, Zong, Miao Chen, Stacy Kowalczyk, and Beth Plale. “Author Gender Metadata Augmentation of Hathitrust Digital Library.” *Proceedings of the American Society for Information Science and Technology* 51, no. 1 (2014): 1–4. <https://doi.org/10.1002/meet.2014.14505101098>.
- Prescott, A. “Bibliographic Records as Humanities Big Data.” In *2013 IEEE International Conference on Big Data*, 55–58, 2013. <https://doi.org/10.1109/BigData.2013.6691670>.
- Royal, Peter. “Untangling Text: Voyant Tools’ Knots for Text Analysis.” *Medium* (blog), November 10, 2015. <https://medium.com/dh-tools-for-beginners/voyant-tools-2-0-less-common-tools-for-text-analysis-a922cfc85cb>.
- Sheffield, Philip, and Sam Saunders. “Using the British Education Index to Survey the Field of Educational Studies.” *British Journal of Educational Studies* 50, no. 1 (2002): 165–83.

Suarez, Michael. "Towards a Bibliometric Analysis of the Surviving Record, 1701–1800." *The Cambridge History of the Book in Britain*, October 2009. <https://doi.org/10.1017/CHOL9780521810173.003>.

Helsinki Computational History Group. "Supplementary Material." Accessed January 14, 2019. [https://comhis.github.io/2019\\_CCQ/](https://comhis.github.io/2019_CCQ/).

*Tools for Bibliographic Data Analysis. Contribute to COMHIS/Bibliographica Development by Creating an Account on GitHub*. R. 2015. Reprint, COMHIS, 2018. <https://github.com/COMHIS/bibliographica>.

Tuppen, Sandra, Stephen Rose, and Loukia Drosopoulou. "Library Catalogue Records as a Research Resource: Introducing 'A Big Data History of Music.'" *Fontes Artis Musicae* 63, no. 2 (2016): 67–88. <https://doi.org/10.1353/fam.2016.0011>.

"Voyant Tools." Accessed November 8, 2019. <https://voyant-tools.org/>.

The authors can be reached via email:

Sarah Wallbank— [wallbank@uci.edu](mailto:wallbank@uci.edu)  
Danielle A. Kane— [kaned@uci.edu](mailto:kaned@uci.edu)  
Madelynn Dickerson— [mrosed@uci.edu](mailto:mrosed@uci.edu)  
Joshua Hutchinson— [jchutchi@uci.edu](mailto:jchutchi@uci.edu)

In the early 1990s, the International Federation of Library Associations and Institutions (IFLA) decided to revise the Anglo-American Cataloguing Rules, second edition (AACR2), a standard that was successfully implemented not only in the cataloguing communities in Britain and the United States, but also in Canada, Australia and many other English and non-English-speaking countries. Multiple factors influenced this decision: new technologies emerged, there was a remarkable increase in publishing output and electronic publishing, whereas economic realities pressured libraries to record only a minimum set of elements in a record in order to save time and to reduce the cost of cataloguing by simplifying its process (IFLA, 2009).

In 1998, the results of IFLA's study were published in a report entitled *Functional Requirements for Bibliographic Records*, a model that is now mainly known by its acronym FRBR. This conceptual model revolutionised the perception of the bibliographic universe as it was known until then by dividing it into entities, attributes and relationships, placing the user in the centre of its design.

The idea behind FRBR was to approach cataloguing not from the type-of-resource perspective the way AACR2 did, but from the entities-attributes-relationships perspective that would enable to make relationships between entities not only within a record but also beyond. FRBR was believed to better respond to the fast-changing digital environment, as it was developed to embrace all types of resources even those that were yet to come. Besides the technical aspect, FRBR became a very much user-centred model defining four main user tasks: find, identify, select and obtain.

Nonetheless, FRBR was met with scepticism. LeBoeuf (2001, p.18) perceived the publication of FRBR as an "earthquake" because it questioned former cataloguing codes, the International Standard Bibliographic Description (ISBD) and even the MACHine-Readable Cataloguing (MARC) format. Furthermore, the term 'navigation' was used multiple times throughout the report, which was a hint that the catalogue turned into "an entirely electronic device" (LeBoeuf, 2001, p.17). In their article, Pisanski and Žumer (2012, p.583) pointed out that no user studies were conducted neither prior to the creation of FRBR nor later, and that there was thus no proof that FRBR would "enhance user experience". O'Neill and Žumer's study (2018) revealed that the FRBR model lacked a framework of implementation as well as a clear and comprehensive terminology. Strader (2017) agreed with O'Neill and Žumer's views and continued that no progress could be made unless the terminology issues were resolved, and extensive user studies were conducted that would help to better understand the users and their needs.

Although no extensive user studies were conducted that investigated the benefits such a model would bring to the cataloguing communities and the users of the catalogue, it was soon agreed that FRBR would henceforth inform new cataloguing standards. The first standard based on FRBR was the Resource Description & Access (RDA) standard, implemented by the Library of Congress (LC) in 2013. Denton (2007, p.35), however, pointed out that, even though future cataloguing codes would rest on FRBR, the model should not be "*the* end point" but "*an* end point", leaving room for future studies, research and developments. Nonetheless, when RDA was officially implemented in 2013, not all the flaws already detected in FRBR were resolved in RDA. Tosaka and Park (2013), for instance, observed that the user was also the least studied component of RDA.

More than twenty years have passed since the publication of FRBR and many questions regarding FRBR, RDA and the user-centred approach still need to be investigated. Critical voices have been raised ever since that, even with the adoption of RDA, the library catalogues have not yet been FRBRised.

This research explored to what extent an RDA-formatted catalogue was FRBRised, to what extent the underlying FRBR structure became apparent when RDA-formatted records were compared to their AACR2-formatted counterparts, and to what extent these findings would impact FRBR's four user tasks.

## THE NETWORKS

The data analysis was based on a comparative analysis of AACR2 and RDA-formatted records to the FRBR model. Data was gathered from two different Online Public Access Catalogues (OPACs): NEBIS and a-z.lu.

NEBIS is a member of the Network of German-speaking Switzerland (IDS), and it is the largest network of academic and research institutions in Switzerland (Küssow and Märchy, 2017). Almost 150 libraries from across Switzerland collaborate in NEBIS in order to make records of approximately 10.5 million resources online available and accessible (NEBIS, 2019). Together with Germany and Austria, the library communities of German-speaking Switzerland were the first non-English-speaking cataloguing communities to adopt RDA.

The union catalogue a-z.lu is operated by the Luxembourgish library network bibnet.lu and gives access to the resources of 88 Luxembourgish libraries including the National Library, public libraries, research institutions, state cultural institutions, higher education and academic libraries, school libraries, and special libraries (bibnet.lu, n.d.; Kieffer, 2012). Libraries of the Luxembourgish network use the French version of the Swiss-German cataloguing rules IDS for descriptive cataloguing (KIDS), which are based on AACR2 (BnL, 2017). IDS was in charge of maintaining and regularly updating KIDS, while the Luxembourgish network adopted those. However, as IDS adopted RDA in 2016, no major updates have since been made to KIDS.

It was hoped that the comparative analysis would show significant differences between the MARC records and provide evidence of a different search experience in the two catalogues.

## METHODOLOGY

A few criteria had to be met in order to do the comparative analysis. First of all, it was essential to have access to the MARC records of the two chosen online catalogues. Both NEBIS and a-z.lu had their records freely available. It was then necessary that the records were available in both the NEBIS catalogue and the a-z.lu catalogue. Furthermore, the items had to be published after 2016. Choosing records that were catalogued before 2017 would increase the probability that the NEBIS records were catalogued according to KIDS, and it would not make sense within the framework of this project to compare NEBIS records with a-z.lu records that were catalogued according to the same standard. It was also required that the chosen items covered at least five different topics and themes, and at least five different types of materials.

The AACR2 and RDA-formatted record of an identical resource were compared, and the similarities and differences were recorded in spread sheets. This procedure allowed to identify, which MARC fields were most commonly used by both cataloguing networks, which were used by one rather than by the other network, which fields were infrequently used, which fields were frequently used, or which fields were more likely to be used for cataloguing one type of resource but not used for cataloguing other types of resources.

## FINDINGS

When starting the project, it was assumed that the RDA-formatted records would significantly differ from the AACR2-formatted records because of the former's underlying FRBR structure. These expectations were not met.

### A. The MARC records

The comparative analysis revealed that the records were not that different after all. Most of the records in NEBIS and a-z.lu were catalogued at core level (95.35% and 76.74% respectively), with the MARC fields 000, 008, 1XX and/or 7XX, 245 and 260/264 appearing without exception in all the records, while fields 300, 5XX and 6XX appeared with vast majority. Both catalogues thus acknowledged that these fields could not be omitted as they contained important information (title of a work, creator, contributor, year of publication) that would support users in finding, identifying, selecting and eventually obtaining materials, while they also recognised that other fields like the physical description or the note fields would benefit users.

The values encoded in the MARC fields were, however, not always identical with one network being either more specific in describing or interpreting data. Variation between the records mainly happened in the title statement (245) on the level of subfields b and c. The use of subfield b depended on how the cataloguers interpreted the title. For the eBook *Babel*, for instance, a-z.lu put the title in 245\$a and the remainder of the title in 245\$b, whereas NEBIS put all title information in 245\$a. 245\$c greatly differed in the records of the audio-visual materials depending on who the cataloguers judged responsible for a work. NEBIS tended to encode the director in 245\$c, while a-z.lu generally added subsequent statements of responsibility, such as screenwriter or cinematographer.

In some cases, both networks recorded the same information but used different MARC fields. When the resource was a translation, NEBIS recorded the original title in 240, while this information was recorded in 509 in the a-z.lu record. The uniform title for audio-visual materials was recorded in 130\$a by NEBIS and in 245\$d in the a-z.lu record. 245\$d no longer exists in RDA, while 130 is a field not used in KIDS. In the records of both networks, language information was always recorded in 008/35-37 and 041 in the case of a translation. In the a-z.lu records, language information for audio-visual materials was also recorded in the general note field (500), while NEBIS recorded this information in the RDA-field 546.

The 5XX-fields were used inconsistently across the records. The 500-field or the summary and abstract field 520 were used contradictorily among the records. By analysing the 500-fields in comparison across both catalogues, one noticed that hardly ever the same information was recorded. In some instances, 500 was used by one network to say, for example, that it was the unabridged version of a work, while the other network recorded this information in 250. One of the printed book resources had stated on its title page that it was the 26<sup>th</sup> instalment of the series. a-z.lu interpreted this information as the remainder of the title, therefore, they recorded this piece of information in 245\$b, while NEBIS considered it to be rather general information, therefore, recorded in 500. The bonus content of the audio-visual materials was alternately recorded by both networks either in 500 or in 520.

Even though the subject access fields (6XX) were used by both networks, they occurred more frequently in NEBIS. Another particularity was that some of the NEBIS subject access fields allowed multilingual access to a resource. NEBIS recorded multilingual subject access terms in the MARC field 691E1. They were recorded in German, French and English, while only one term (generally in the language of the browser) was displayed in the online catalogue. As a consequence, users searching the catalogue for 'libraries cultural administration' would find the DVD *Ex libris: the New York Public Library* among their search results, the same as users searching for its French equivalent 'bibliothèques administration de la culture' or for its German equivalent 'bibliotheken kultusverwaltung'.

## **B. Bibliographic display in the catalogues**

The NEBIS and a-z.lu catalogues were constructed in similar ways with a top section, the send-to action section, the main labelled record, a link section and the availability and location information section. NEBIS chose a different sequential arrangement for these sections than a-z.lu, but the content was similar. The depth of information displayed in the main body of the record depended on the data the networks recorded in their MARC records. As NEBIS tended to encode more data in their MARC records, the information displayed in the catalogue was also more detailed.

Differences between the catalogues were, in turn, perceived on the level of the title statement. a-z.lu provided a title and a complete title. The title referred to 245\$a, 245\$d and 245\$b in the MARC record and the complete title was all information recorded in the title statement (245\$a, 245\$d, 245\$b and 245\$c). The choice of displaying information like this resulted in a repetition of the title, the parallel title and the remainder of the title in the online catalogue. NEBIS, in turn, provided a title and additional title information. Under the title heading, NEBIS displayed all information regarding title (245\$a) and remainder of the title (245\$b), whereas all information recorded in 245\$c was displayed under the additional title information heading in the catalogue. As a consequence, no information was repeated, which resulted in a better readability of the record.



## C. FRBR mapping

The labels attributed to the 1XX and 7XX entities in the RDA-formatted catalogue drew particular attention as the standard's underlying FRBR influence emerged. While these entities were displayed as 'Author' (1XX) and 'Contributor(s)' (7XX) in the AACR2-formatted catalogue creating a hierarchy between entities responsible for the intellectual creation of a work and entities participating in the realisation or production of a work, the term 'Creator' was used in the RDA-formatted records to refer to all entities (1XX and 7XX) responsible for a work, expression or manifestation. Even though creator was technically not a term used in FRBR as the model's preferred terminology was 'Person, family or corporate body' (PFC) (Group 2 entities), no hierarchical distinction between the creator of a work, an entity responsible for the realisation of an expression and an entity responsible for the production of a manifestation was made. They were equalised although they performed different functions and had different responsibilities.

One could argue, for instance, that displaying PFCs under the same heading created difficulties, as some users might be interested in the different functions these PFCs exercised. A user of a public library might want to learn from a catalogue record who was the creator and who was the translator. It is after all the purpose of the catalogue to provide all information that users need in order to find, identify, select and obtain a document. NEBIS and a-z.lu both linked to the MARC records, which means that users could use those as information source as well, provided that they knew how to read them. NEBIS' MARC records provided indeed more detailed information about the PFC entities than the catalogue records. Along with the name (1XX\$a) and the associated dates (1XX\$d), NEBIS recorded a relator term (1XX\$e), a relationship (1XX\$4) and a real-world object URI (1XX\$1). These relationship designators non-existent in KIDS, explained the role and function of the PFC in the creation, realisation or production of a work, expression or manifestation. As these relator terms and relationships were recorded in NEBIS' MARC records, the question arose why they did not appear in the catalogue except for the information regarding any related dates (1XX\$d), since these relator terms would help users identify PFCs. This is in line with Cotterman's (2017) statement that "much work [cataloguers] do is hidden but is also essential for access".

Making relevance assessments and relevance judgments of the results returned was not within the scope of this project, but it was still noticed that the number of search results returned by NEBIS was generally lower than that returned by a-z.lu. FRBR's entity-relationship (E-R) model is based on the idea that entities and attributes making up the bibliographic universe are in complex relationship with one another. A FRBR-based catalogue should, therefore, be able to identify related resources. The Swiss-German catalogue was able to make these relationships between expressions and manifestations of a same work, provided that they belonged to the same type of material. NEBIS, for instance, retrieved an audio version and multiple print versions of a work bearing the title *Stille Wasser*. While the audio version had its own entry in the catalogue, the 2017 printed book version and the 2018 printed book version of *Stille Wasser* were clustered and considered as related resources.

Besides identifying related resources, the RDA-formatted catalogue was also able to identify and cluster related information under the same heading in its catalogue. The system recognised that all 5XX-fields, for example, were note fields describing a resource, and hence, it gathered that information under the heading 'Description' in the catalogue. The choice of displaying information like this contributed to a better readability of the record, as related information was gathered together, and the record was not overloaded by too many headings.

## CONCLUSION

A comparative analysis of AACR2 and RDA-formatted records to the FRBR model was conducted in order to highlight similarities and differences between the records, and to investigate to what extent these resulted from, were in line with or contradicted the FRBR model. The analysis revealed that the AACR2 and RDA-formatted records did not significantly differ from one another. Most of the records were catalogued on core level, which means that they contained many elements that were relevant for users searching the catalogues.

The analysis showed that the data recorded in the MARC records often differed between the catalogues. As a consequence, one can conclude that there is no uniform way of understanding MARC fields. Even though information is provided by the standards and on the internet that is targeted at helping cataloguers to understand the meaning of the individual MARC fields, it is left to the cataloguers' subjective interpretation of the information that makes them record data in one field rather than the other. The role of the standards is to guide cataloguers through the cataloguing process, to give instructions, but they are not binding. Therefore, it is crucial for cataloguers to be consistent in their cataloguing practice, because as Hider (2018, p.129) said, catalogues can only be effective when they are constructed in a consistent way.

Although RDA's underlying FRBR structure became apparent by the way information was displayed in the NEBIS catalogue, FRBR's influence on the RDA-formatted catalogue was minor. More than 20 years have passed since FRBR was first published, and its potential is neither fully explored nor exhausted. Many questions remain especially when it comes to the users. User studies should be conducted in order to better understand what information users need, what users expect from a library catalogue in the digital age, the age of big data and the semantic web, and then, standards and frameworks need to be revised and adapted.

## REFERENCES

- bibnet.lu (n.d.) *Réseau des bibliothèques luxembourgeoises bibnet.lu*. Available at: <https://www.bibnet.lu/blog/> (accessed 23 November 2019).
- BnL Bibliothèque nationale de Luxembourg (2017) *Systèmes et technologies*. Available at: [https://bnl.public.lu/fr/services-professionnels/innovation/syst\\_technologies.html](https://bnl.public.lu/fr/services-professionnels/innovation/syst_technologies.html) (accessed 11 June 2019).
- Cotterman, B. (2017) 'CaMMS Forum: Power that is moral: creating a cataloguing code of ethics'. *CaMMS Forum*, 25 June 2017. Available at: <https://alcts.ala.org/news/2017/ac-camms-forum/> (accessed 8 August 2019).
- Denton, W. (2007) 'FRBR and the history of cataloging'. In Arlene G. Taylor (ed.) *Understanding FRBR: What it is and how it will affect our retrieval*, pp.35-57. Westport, CT: Libraries Unlimited. Available at: <https://yorkspace.library.yorku.ca/xmlui/handle/10315/1250> (accessed 29 May 2019).
- Hider, P. (2018) *Information Resource Description: Creating Managing metadata*. 2<sup>nd</sup> edn. London: Facet Publishing.
- IFLA Study Group on the Functional Requirements for Bibliographic Records (2009) *Functional Requirements for Bibliographic Records*. Available at: <http://www.ifla.org/VII/s13/frbr/> (accessed 20 May 2019).
- Kieffer, M. (2012) 'The National Library of Luxembourg'. *Alexandria*, 23(1), pp.24-35. Available at: <https://doi.org/10.7227/ALX.23.1.4> (accessed 1 March 2019).
- Küssow, J. and Märchy, S. (2017) 'Regelwerke im multilingualen Kontext – ein Erfahrungsbericht aus einem multilingualen Verbund'. *O-bib Das offene Bibliotheksjournal*, 4(3), pp.16-26. Available at: <https://doi.org/10.5282/o-bib/2017H3S16-26> (accessed 13 June 2019).
- Le Boeuf, P. (2001) 'FRBR and further'. *Cataloging & Classification Quarterly*, 32(4), pp.15-52. Available at: [https://doi.org/10.1300/J104v32n04\\_03](https://doi.org/10.1300/J104v32n04_03) (accessed 28 May 2019).
- NEBIS (2019) *The NEBIS Network*. Available at: <https://www.nebis.ch/en/about/> (accessed 9 June 2019).
- O'Neill, E. and Žumer, M. (2018) 'FRBR: application of the model to textual documents'. *Library Resources & Technical Services*, 62(4), pp.176-187. Available at: <https://doi.org/10.5860/lrts.62n4.176> (accessed 28 May 2019).



Pisanski, J. and Žumer, M. (2012) 'User verification of the FRBR conceptual model'. *Journal of Documentation*, 68(4), pp.582-592. Available at: <https://doi.org/10.1108/00220411211239129> (accessed 27 May 2019).

Strader, C.R. (2017) 'From user tasks to user services: placing the Functional Requirements for Bibliographic Records models into a larger framework'. *Technical Services Quarterly*, 34(4), pp.347-368. Available at: <https://doi.org/10.1080/07317131.2017.1355147> (accessed 28 May 2019).

T

osaka, Y. and Park, J.-r. (2013) 'RDA: Resource Description & Access – a survey of the current state of the art'. *Journal of the American Society for Information Science and Technology*, 64(4), pp.651-662. Available at: <https://doi.org/10.1002/asi.22825> (accessed 22 May 2019).

### Introduction

People with disabilities face innumerable challenges in fighting for equal rights and equity, and one such challenge is regarding access to libraries and information. Although libraries are often welcoming, underlying systems such as subject headings can perpetuate harmful stereotypes, particularly towards marginalized populations. This study examines disability language in two controlled subject vocabularies – the Library of Congress Subject Headings (LCSH) and the Education Resources Information Centre (ERIC) Thesaurus – as well as in two social tagging systems – the Steam computer game client and the fan works repository Archive of Our Own (AO3).

### Literature review

#### *Disability language in subject access systems*

Previous research on disability language in subject access systems has indicated an ongoing struggle to reflect language used in disability communities. Historically, disability language in subject access was derived from the medical field (Adler, Huber, & Nix, 2017). The so-called “medical model of disability” views disability as pathological; descriptors include “defective”, “crippled”, and “handicapped”, perpetuating stigma (2017). This contrasts with the social model – preferred by disability scholars – which frames disability in terms of socio-cultural barriers (Koford, 2014).

Changes in language have previously been made – LCSH revised “Handicapped” to “People with disabilities” in 2002 (Adler, Huber, & Nix, 2017) – but there have not been accompanying changes in hierarchical structures. In the Library of Congress Classification (LCC), works on “People with disabilities” are classified under HV, which is titled “Social Pathology. Social and Public Welfare. Criminology”. Similarly, Dewey Decimal Classification (DDC) uses the term “People with illnesses and disabilities” (2017). Thus, despite changes to wordings, disability is still being explicitly linked to pathology (the medical model) instead of impairment (the social model).

#### *The effects on disability studies*

The ongoing linguistic and hierarchical issues surrounding disability have naturally impacted disability studies. Koford (2014) found that disability scholars struggle to find search terms that yield the information they seek. Oftentimes, results are too broad or they are forced to use non-preferred terms (medical model instead of social) to conduct their search. Some of the researchers Koford interviewed also reported being more likely to use multidisciplinary databases over subject-specific databases, as they tend to yield more results. Koford’s (2014) key takeaway was that “the way documents are organized in libraries and databases has a profound impact on what information is retrieved and what remains unseen” (p. 388), and that while no subject access system can meet every user’s needs, consultation with subject experts is a good first step towards improving language and hierarchies.

#### *Social tagging & Archive of Our Own*

Within the literature, researchers suggest the existence of two major forms of tagging: controlled vocabulary and social tagging. A controlled vocabulary uses predetermined and standardized language in order to organize information. Controlled vocabulary tends to be rigid, with language that becomes outdated easily (McCulloch, 2019). Social tagging, also known as user-tagging, allows users to create their own tags with no specific guidelines or vocabulary (2019).

The most common example of social tagging is through social media tagging, such as on Flickr or Twitter. Since very few rules apply within social tagging, users often find themselves listing multiple variations of the same tag, thereby creating long, conflicting, and cluttered links to information.

AO3 is a non-profit fanfiction repository. It uses a system of social tagging combining controlled vocabulary with user-tags (see Figure 1). Users are encouraged to create their own tags relating to specific fandoms with some help from an autocomplete function which suggests established canonical terminology, which is officially accepted terminology derived from individual fandoms (Dalton, 2012). Resembling LCSH, a user may select a tag to see its parent tags, synonymous tags, and subtags (see Figure 2). AO3 also uses tag wranglers: volunteers who look through tags and match non-canonical tags with canonical tags to ensure some level of control over terminology (McCulloch, 2019). Tag wranglers operate using a set of rules, which Dalton (2012) likens to the “use” and “used for” devices “in authority records of the Library of Congress” (p. 3). Overall, the combination of social tagging and controlled vocabulary exhibited through tag wranglers and AO3’s autocomplete functionality ensures that information organization and retrieval is more useful and insightful than strictly controlled vocabulary or user-tagging (McCulloch, 2019).

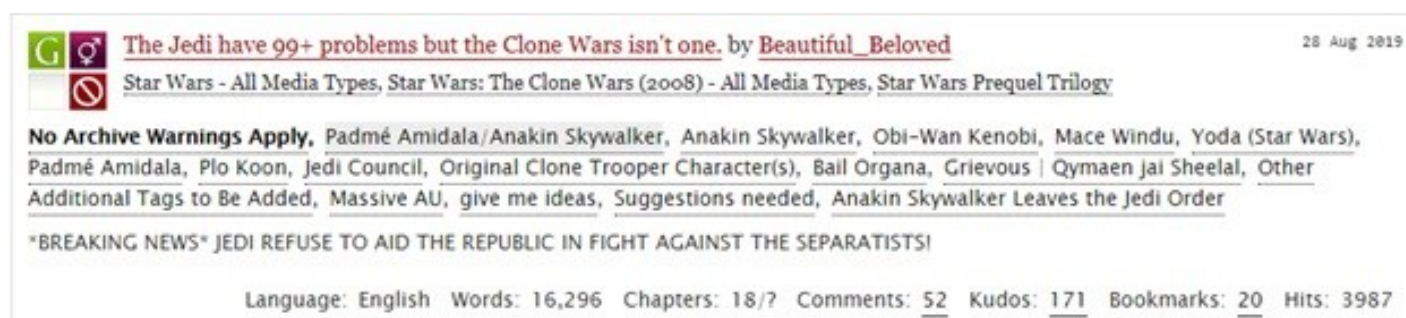


Figure 1: AO3 tags on a work of fanfiction

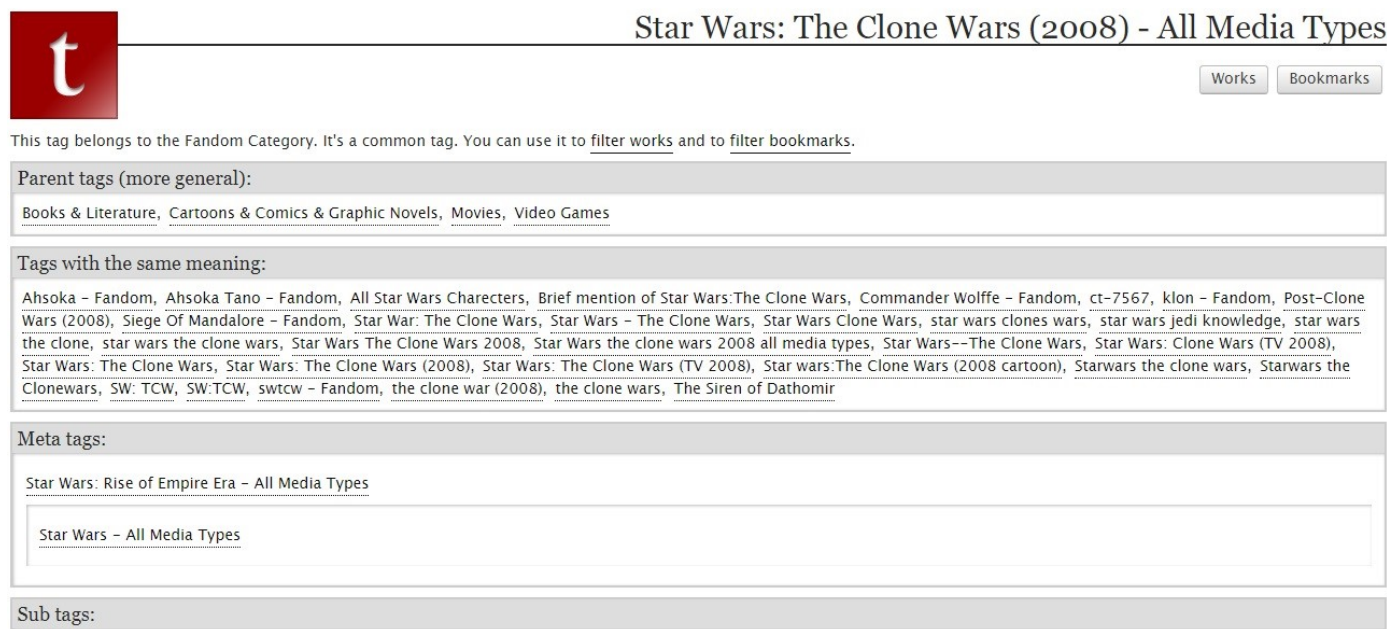


Figure 2: AO3 tag structure

## Tagging on Steam

Steam is a digital store and client program for computer games. Users and game developers can apply tags to games, and the most frequently used tags are compiled into a public list of “global tags” (Valve Corporation, n.d.a). Valve seems to exert some control over this list, however, as profanity is filtered out (Valve Corporation, n.d.a), and they only recently officially included an LGBTQ+ tag (Good, 2019a).

An analysis of Steam tags found that they can be organized similarly to the Video Game Metadata Schema (VGMS) developed by the Game Metadata Research Group and the Seattle Interactive Media Museum (Windleharth, Jett, Schmalz, & Lee, 2016). Categories included genre, visual style, pacing and customization options; however, some categories – including mechanics and input types – were not part of the VGMS at the time (Windleharth et al., 2016). The study concluded that tags are mostly used to denote genre, mood, and characteristics that wouldn’t otherwise be listed in promotional materials.

## Digital games & accessibility

Accessibility has been coming to prominence as a topic in the games community, notably in early 2019 after the release of the game *Sekiro: Shadows Die Twice* (Thompson, 2019). Yuan, Folmer, and Harris (2010) grouped impairments as they affected gameplay into four categories – visual, hearing, motor, and cognitive – and modeled gameplay interaction into three steps with which impairment might interfere: receiving stimuli, determining a response, and providing input. Following that, in 2012 a living document was published with guidelines for developers for what accessibility features they should consider including, and the level of time and effort it would take to implement each of them (Ellis et al., 2012). Improvement is uneven, and accessible hardware traditionally cost-prohibitive, but more and more developers are committing to accessibility, comparatively-affordable adaptive controllers have been released in the past few years, and organizations like The AbleGamers Charity can help offset remaining costs (Good, 2019b; Gwaltney, 2019).

## Methodology

We examined four subject access systems altogether: two fully-controlled taxonomies – LSCH (2006) and the ERIC Thesaurus (Institute of Education Sciences, n.d.) – and two folksonomies – AO3’s social tags (Organization for Transformative Works, n.d.) and Steam’s Global Tags (Valve Corporation, n.d.b). A list of terms related to disability and accessibility was selected from each taxonomy. To reduce bias, we tried to collect terms that were explicitly connected to accessibility or disability, either based on the term itself or, in the case of ERIC, collecting terms from their pre-established Disability category. An exception was made for Steam, as no terms in their global tags explicitly referred to disability language. Terms from Steam were instead selected based on whether they might indicate the inclusion of accessibility features per the Game Accessibility Guidelines (Ellis et al., 2012) or features that would make the game inaccessible for certain disabilities.

Collected terms were sorted into sub-categories based on prominent similarities. Most sub-categories were centred around different types of disabilities or in relation to particular people groups, as well as accessibility and accommodation. However, the categories for the Steam tags drew greater inspiration from the VGMS and the categories used by Windleharth et al. (2016) in their sorting of Steam tags. The collected and sorted lists of terms can be found in the appendices.

## Discussion

### LSCH & ERIC

As controlled vocabularies, LCSH and ERIC share some similarities in regards to terminology. While both vocabularies focus on disability and accessibility language, LCSH (n=76) tends to consist of more age-group headings and more terms pertaining to accessibility than ERIC. An examination of ERIC terms (n=92) indicates that the language is directly focused on disabilities and accessibility, primarily in an educational context. ERIC is also the only thesaurus in this study to highlight dead terms.

Interestingly, “Architectural barriers” has been marked dead, even though the term “barriers” is commonly used throughout disability studies, the social model of disability, and in major thesauri, such as LCSH. Both thesauri have shifted towards using people-first language, which demonstrates a willingness to update the language used to describe disability; however, ERIC still contains instances of “[the] disabled”.

### *AO3 & Steam*

Despite the large contributing communities of both AO3 and Steam, the content of their respective tags vastly differs. The Steam tags (n=35) are functionally descriptive: describing a component of gameplay or visual aesthetic. Even still, only one term concretely suggests an accessibility feature (tutorial), nor are there many terms that clearly indicate inaccessibility. On the other hand, the AO3 tags (n=102) are narratively descriptive: describing character or story elements. Given the works the tags respectively describe, this makes some sense. But, games often have narratives as well, yet there is only one narrative Steam tag that can apply to disability: transhumanism, which is weighted with controversy (Bradshaw & Ter Meulen, 2010; Hilvoorde & Landeweerd, 2010; Koch, 2018).

### *General Discussion*

Although controlled vocabularies, such as LCSH and ERIC, have undergone updates to modernize disability terminology, Koford’s (2014) research suggests disability scholars still disregard controlled vocabulary as it often provides an inaccurate portrayal of the disability community. Folksonomies, such as Steam’s global tags and AO3’s social tags, provide an interesting contrast as they combine aspects of controlled vocabulary with user-generated tag structures. Within these folksonomic systems, users are encouraged to provide their input while organizational authorities ensure consistency of tags (i.e., AO3’s canonical and Steam’s global tags). With AO3, we can see this system working most clearly. The combination of tagging freedom and supportive structure has allowed for tags that describe social and activist dynamics of disability: intersectionality, confrontation of ableism, and sexuality – the latter of which contrasts with consistent desexualization of people with disabilities elsewhere. This does not mean, however, that user-tagging systems are the only means of including activist description. ERIC has a heading for “Disability Discrimination”, though it falls within the “Bias and Equity” category (and thus was not included in our list), and said category also includes terminology to describe activist work. On the other hand, we have already discussed Steam’s lack of disability language. Whether or not these subject access systems have accomplished inclusivity in their disability language has largely been dependent on whether their respective central communities have cared to make changes.

How, then, to proceed? Deodato (2014) proposed introducing social tagging to exist alongside official bibliographic records and subject headings, which has been implemented in some discovery platforms such as BiblioCommons. But these tags seem to lack control in their vocabulary, and there does not seem to be any indication that they do, or even could, influence updates to official subject headings. At its core, any changes that are made require input from members of the disability community (including scholars) to determine preferred terminology. But a new hybrid approach can look to some of the strengths of the systems we have observed. Elements like LCSH’s use of the term “barriers”, AO3’s strong social language, ERIC’s categorization, and Steam’s use of mechanically-descriptive tags would all be useful in an updated subject access system.

### **Conclusion**

Currently, disability scholars and members of the disability community experience difficulty when searching databases and catalogues, because much of the language in those systems has not completely caught up to that which is used by these communities, instead remaining medically-derived and negatively-toned. Social tagging systems, especially if some measure of control is applied, provide a potential remedy to this issue, but success is largely contingent on their contributing communities and the potential for inducting tags into official terminology. Neither fully controlled vocabularies such as LCSH or ERIC, nor social tagging systems like AO3 or Steam have all the answers. Future research in this field could explore the effectiveness of other systems, including the hybrid-structure discovery systems already in place.



## References

- Adler, M., Huber, J. T., & Nix, A. T. (2017). Stigmatizing disability: Library classifications and the marking and marginalization of books about people with disabilities. *The Library Quarterly*, 87(2), 117-135. doi:10.1086/690734
- Bradshaw, H. G., & Ter Meulen, R. (2010). A transhumanist fault line around disability: Morphological freedom and the obligation to enhance. *Journal of Medicine and Philosophy*, 35(6), 670-684. doi:10.1093/jmp/jhq048
- Dalton, K. L. (2012). *Searching the archive of our own: The usefulness of the tagging structure* (Order No. 1531883). Available from ProQuest Dissertations & Theses Global. (1283371530).
- Deodato, J. (2014). The patron as producer: Libraries, web 2.0, and participatory culture. *Journal of Documentation*, 70(5), 734-758. doi:10.1108/JD-10-2012-0127
- Ellis, B., Ford-Williams, G., Graham, L., Grammenos, D., Hamilton, I., Headstrong Games, & Westin, T. (Contributors). (2012). *Game accessibility guidelines: A straightforward reference for inclusive game design* [Web page]. Retrieved from <http://gameaccessibilityguidelines.com/>
- Good, O. S. (2019a, June 22). Steam adds official LGBTQ+ tag for games. *Polygon*. Retrieved from <https://www.polygon.com/pc/2019/6/22/18713367/steam-lgbtq-tag-storefront-pc>
- Good, O. S. (2019b, November 18). Xbox Adaptive Controller gets a toolkit full of assists from Logitech. *Polygon*. Retrieved from <https://www.polygon.com/2019/11/18/20970350/logitech-g-adaptive-gaming-kit-xbox-adaptive-controller-release-date-price-where-to-buy>
- Gwaltney, J. (2019, May 15). The fight to make games accessible for everyone. *Game Informer*. Retrieved from <https://www.gameinformer.com/2019/05/15/the-fight-to-make-games-accessible-for-everyone>
- Hilvoorde, I. V., & Landeweerd, L. (2010). Enhancing disabilities: Transhumanism under the veil of inclusion? *Disability and Rehabilitation*, 32(26), 2222-2227. doi:10.3109/09638288.2010.491578
- Institute of Education Sciences. (n.d.). Purpose and scope [Web page]. Retrieved from <https://eric.ed.gov/?ti=all>
- Koch, T. (2018). Disabling disability amid competing ideologies. *Journal of Medical Ethics*, 44(8), 575-579. doi:10.1136/medethics-2017-104253
- Koford, A. (2014). How disability studies scholars interact with subject headings. *Cataloging & Classification Quarterly*, 52(4), 388-411. doi:10.1080/01639374.2014.891288
- Library of Congress Subject Headings* (29th ed., Vols. I-V). (2006). Washington, DC: Library of Congress.
- McCulloch, G. (2019, June 11). Fans are better than tech at organizing information online. *Wired*. Retrieved from: <https://www.wired.com/story/archive-of-our-own-fans-better-than-tech-organizing-information/>
- Organization for Transformative Works. (n.d.). Tags [Web page]. Retrieved from <https://archiveofourown.org/tags>
- Thompson, C. (2019, April 5). Sekiro: Accessibility in games is about far more than 'difficulty'. *IGN*. Retrieved from <https://www.ign.com/articles/2019/04/05/sekiro-accessibility-in-games-is-about-far-more-than-difficulty>

Valve Corporation. (n.d.a). Browse Steam your way: Introducing Steam Tags, a powerful new way to shop for games [Web page]. Retrieved from <https://store.steampowered.com/tag/>

Valve Corporation. (n.d.b). Popular tags [Web page]. Retrieved from [https://store.steampowered.com/tag/browse#global\\_492](https://store.steampowered.com/tag/browse#global_492)

Windleharth, T. W., Jett, J., Schmalz, M., & Lee, J. H. (2016). Full Steam ahead: A conceptual analysis of user-supplied tags on Steam. *Cataloging & Classification Quarterly*, 54(7), 418-441. doi:10.1080/01639374.2016.1190951

Yuan, B., Folmer, E., & Harris, F. C., Jr. (2010). Game accessibility: A survey. *Universal Access in the Information Society*, 10(1), 81-100. doi:10.1007/s10209-010-0189-5



## Appendix A: Terms Selected from the Library of Congress Subject Headings

### Accessibility & Benefits

Accessible Web sites for people with disabilities  
Animals as aids for people with disabilities  
Assistive computer technology  
Automobile parking for people with disabilities  
Barrier-free design  
Barrier-free design for children with disabilities  
Barrier-free design for children with mental disabilities  
Barrier-free design for older people  
Barrier-free design for people with mental disabilities  
Barrier-free design for people with visual disabilities  
Barrier-free design for students with disabilities  
Disability awareness  
Disability evaluation  
Disability insurance  
Disability retirement  
Disabled veterans  
Drug accessibility  
Health services accessibility  
Hearing ear dogs  
Hearing impaired - Means of communication  
Hearing impaired - Services for  
Local transit accessibility  
Meetings - Accessibility for people with disabilities  
Passenger lifts for people with disabilities  
Television - Accessibility for people with disabilities

### Physical Disabilities

Deaf  
Deafness  
Disabilities  
Hearing disorders  
Hearing impaired  
Movement disorders  
People with disabilities  
People with disabilities - Abuse of  
People with visual disabilities  
Perceptual disorders  
Sensory disorders  
Vision disorders  
Visual agnosia

Visual discrimination

Word deafness

### Mental/Cognitive/Neurological Disabilities

Adjustment disorders  
Affective disorders  
Communicative disorders  
Developmental disabilities  
Developmentally disabled  
Dissociative disorders  
Dyslexia  
Impulse control disorders  
Learning disabilities  
Mental illness  
People with mental disabilities  
People with social disabilities

### Children

Children with disabilities  
Children with disabilities - Development  
Children with disabilities - Education  
Children with mental disabilities  
Children with mental disabilities - Education  
Children with perceptual disabilities  
Children with social disabilities  
Children with visual disabilities  
Children with visual disabilities - Orientation and mobility  
Children, Blind  
Children, Blind - Orientation and mobility  
Deaf children  
Hearing disorders in children  
Hearing disorders in infants  
Hearing impaired children  
Hearing impaired infants  
Hearing impaired students  
Students with disabilities  
Youth with disabilities  
Youth with mental disabilities  
Youth with social disabilities

### Older Adults

Developmentally disabled older people  
Hearing impaired middle-aged persons  
Older people—Mental health  
Older people—Orientation and mobility  
Older people with disabilities  
Older people with disabilities—Long term care  
Older people with mental disabilities  
Older people with visual disabilities

## Appendix B: Terms Selected from the ERIC Thesaurus

Note: Dead terms are denoted with strikethroughs and the years in active use.

### Accessibility

Academic Accommodations (Disabilities)  
Accessibility (for Disabled)  
Adapted Physical Education  
Assistive Technology  
~~Architectural Barriers (1970-1980)~~  
Augmentative and Alternative Communication  
Individualized Family Service Plans  
Sensory Aids  
Travel Training

### Mental/Cognitive/Neurological Disabilities

Alzheimers Disease  
Aphasia  
Asperger Syndrome  
Attention Deficit Disorders  
Attention Deficit Hyperactivity Disorder  
Autism  
Cerebral Palsy  
Developmental Disabilities  
Down Syndrome  
Dyslexia  
~~Educational Retardation (1966-1980)~~  
Epilepsy  
Fetal Alcohol Syndrome  
Head Injuries  
Hyperactivity  
Intellectual Disability  
Learning Disabilities  
Mild Intellectual Disability  
Minimal Brain Dysfunction  
Moderate Intellectual Disability  
Neurological Impairments  
Pervasive Developmental Disorders  
Seizures  
Severe Intellectual Disability  
Slow Learners

### Physical Disabilities

Articulation Impairments  
Physical Disabilities  
Physical Mobility

### Visual Disabilities

Blindness  
Braille  
Large Type Materials  
Partial Vision  
Perceptual Impairments  
Tactile Adaptation  
Visual Impairments  
Visually Impaired Mobility

### Communication Disorders

American Sign Language  
Communication Disorders  
Cued Speech  
Deaf Interpreting  
Deafness  
Finger Spelling  
Hearing Impairments  
Hearing Therapy  
Language Impairments  
Lipreading  
Manual Communication  
Oral Communication Method  
Sign Language  
Speech Impairments  
Speech Language Pathology  
Speech Therapy  
Stuttering  
Total Communication  
Voice Disorders

### Sociocultural Elements

Disability Identification  
Institutionalized Persons  
Mainstreaming  
Normalization (Disabilities)  
Parents with Disabilities

### Miscellaneous

Adventitious Impairments  
Community Based Instruction (Disabilities)  
Congenital Impairments  
Deaf Blind  
Daily Living Skills  
Disabilities  
~~Exceptional Child Education (1968-1980)~~  
~~Exceptional Child Services (1968-1980)~~  
Gifted Disabled  
Group Homes  
~~Handicapped Children (1966-1980)~~  
~~Handicapped Students (1967-1980)~~  
Homebound  
Mild Disabilities  
Multiple Disabilities  
Residential Care  
Respite Care  
Severe Disabilities  
Severity (of Disability)  
Sheltered Workshops  
Special Education  
Special Education Teachers  
Students with Disabilities

## Appendix C: Terms Selected from Archive of Our Own Tags

### Canonical Tags

Disability  
Disability Fest  
Physical Disability

### Accessibility & Accommodation

accessibility  
accessibility and cultural  
accommodations [sic]  
Accessibility Issues  
Accommodations of Disability  
adapting to disability  
adjusting to a disability  
Americans With Disability Act  
Disability Accommodation [sic]  
Disability and accessibility  
disability rights  
Lack of Accessibility is Real  
Magic used for accessibility  
misuse of accessibility devices

### Struggle, Learning to Cope, & Acceptance

(poorly) coping with disability  
Acceptance of disability  
Acknowledge of character's disability  
basically struggling to come to terms with a new disability  
characters learn how to handle disability  
confluence between disability shaming and self-esteem issues  
Conversations About Disability  
Coping with a Disability  
Coping with a Long-Term Disability  
dealing with a new disability  
Dealing with disability  
disability acceptance  
Disability Curse  
Disability Difficulties  
disability feelings  
Disability Frustrations  
Struggle with Disability  
Struggles with disability acceptance/identity

### Intimacy

(people who see my disability as a kink dni [do not interact])  
Adaptive Sex Practices or Sexual Disability  
Disability Intimacy  
disability sex  
Intimacy and disability  
non-fetishized disability  
Pregnancy with disability

### Sex and Disability

### Physical Disabilities

Body Disability  
Canon Physical Disability  
character with a hearing disability  
Childhood Physical Disability  
chronic disability  
Disability (blindness)  
disability (loss of limb)  
Disability and injuries  
disability and terminal illness  
Hearing disability  
Injury-related physical disability  
Mute Disability  
permanent disability due to injury  
Temporary Physical Disability  
Vision disability  
visual disability

### Mental/Cognitive Disabilities

Alien with a developmental disability  
alluded! [sic] intellectual disability  
Brief Mentions of Mental Disability  
Canon reading disability  
cognitive disability  
Developmental Disability  
Intellectual Disability  
Psychiatric Disability  
Psychological Disability  
reading disability  
Verbal Disability

### Identity & Activism

adequate portrayal of disability  
bamf female character with disability  
Disability Bechdel Test Pass  
Disability Community  
disability fashion  
disability identity  
Disability Pride  
disability slur  
disability' does not equal 'disadvantage'  
queer character with a disability  
queer female character with a disability  
Realistic depictions of disability  
talks of disability and ableism and deafness

[Continued on next page]



## **Miscellaneous**

(spoiler) disability  
acquired disability  
and ableism from the man with a disability  
At-Birth Disability  
because not all disability stories have to be  
about pure angst  
can disability be used to justify abusive  
tendencies and behaviors?  
canon character with disability  
canon disability  
Character feigning disability  
character with a disability  
character with permanent disability  
Child with disability  
convenient disability  
Disability in Middle-earth  
Grieving a disability  
hidden disability  
Invisible disability  
Parental illness/disability  
Powers as Disability  
Temporary Character Disability  
temporary disability

## Appendix D: Terms Selected from the Steam Global Tags

### Game Feature

Asynchronous Multiplayer  
Mod  
Moddable  
Quick-Time Events  
Tutorial

### Gameplay Genre

Bullet Hell  
Point & Click  
Rhythm  
Rogue-like  
Rogue-lite  
Souls-like  
Spectacle Fighter  
Spelling  
Strategy  
Word Game

### Skill Level

Casual  
Difficult  
Unforgiving

### Visual Descriptors

Colourful  
Minimalist  
Pixel Graphics  
Psychedelic  
Text-based

### Game Inputs

Controller  
Intentionally Awkward Controls  
Mouse only  
Touch-Friendly  
Voice Control

### Pace

Fast-Paced  
Real-Time  
Real-Time with Pause  
Relaxing

### Sound

Dynamic Narration  
Narration

### Narrative elements/genres

Transhumanism

In “Five Theses on the Future of Special Collections”, John Overholt (2013) presents five objectives he sees as central to the future of special collections libraries: distribution, openness, disintermediation, transformation and advocacy. A lot of the barriers hindering the fulfilment of these goals, particularly the first four, arguably centre on access to special collections metadata. Vast repositories of metadata describing special collections items remain siloed within local library management systems without being exposed to web-scale discovery (Cullingford, 2016, p. 95). There are currently wide-ranging debates on how best to facilitate greater physical and digital access to special collections (Tam, 2017; DeZelar-Tiedman, 2017, p. 121). One possibility is the adoption of Semantic Web technologies and Linked Data to make special collections metadata more discoverable.

This article will present the results of a dissertation written as part of a Master’s degree in Library and Information Studies at University College London. It will analyse how well suited BIBFRAME, a Linked Data model developed by the Library of Congress, and its extensions are for representing special collections metadata. The focus is the Linked Data for Production’s (LD4P) Art and Rare Materials (ARM) Ontology, an extension to BIBFRAME 2.0. To perform an assessment of the ontology’s capabilities existing MARC-encoded special collections were modelled with the ARM ontology through the ontology editor software Protégé. With a future outlook that the library profession will be highly involved in Linked Data and the Semantic Web, this investigation hopes to be part of an on-going discussion on how cataloguers, and special collections cataloguers in particular, can implement, enhance and revise BIBFRAME and its extensions in the coming years.

## **BIBFRAME and special collections cataloguing**

In the process of creating special collections metadata each item is treated as a unique artefact. Its unique characteristics, or its “copy-specific information”, must be described to help users make decisions about which materials they wish to consult (Cullingford, 2016, p. 91-92). The current standard to describe special collections materials is the “Descriptive Cataloguing of Rare Materials” (DCRM). Although most special collections cataloguing still relies on data encoded in MARC, the Rare Books and Manuscripts Section lists in its “Guidelines: Competencies for special collections professionals” the ability to apply “knowledge of data management platforms and data models, such as ... linked open data ... in order to facilitate information retrieval and intellectual access to special collections materials” (American Library Association, 2017). Cataloguing data made available as Linked Data is already an established option for librarians in support of increasing discovery and use of special collections.

So what is Linked (Open) Data and the Semantic Web? The Semantic Web’s purpose is to create a “Web of Data”, a World Wide Web populated with data that is machine-readable and -actionable (DeWeese & Segal, 2015, pp. 1-2). To achieve this, structured data is made available in standardized formats. Individual data elements and their relationships with each other are made explicit and are organized into a formal semantic structure to determine how they should be processed by machines (Antoniou et al., 2012). In practice this translates to the construction of triples for objects and their relationships to each other expressed in RDF (Resource Description Framework). Each part of the triple, the two objects and the connector expressing their relationship, are given URIs (Unique Resource Identifiers) to identify them as data elements. The final RDF statement takes the form of a simple sentence or assertion with a subject, predicate and object. Ontologies, “vocabularies [that] define the concepts and relationships ... used to describe and represent an area of concern” (W3C, 2015), are constructed with an ontology language, such as OWL (Web Ontology Language), to model the data. An ontology governs the rules of how data elements can relate to each other to prevent contradictory statements and other errors. When this Linked Data is then published on the Web, it enables the data points to be explored as a network (O’Dell, 2017, p. 131).

The Library of Congress uses Semantic Web technologies to build its new model of bibliographic description, BIBFRAME. It consists of RDF classes and properties describing the characteristics of a resource and relationships between resources which are labelled with unique identifiers (Library of Congress, 2014; see Figure 1). This stands in contrast to a record encoded in MARC which identifies a resource with strings of texts containing bibliographic description. The current iteration of BIBFRAME, BIBFRAME 2.0, has three core classes – Work, Instance and Item – and additional key concepts relating to the core classes – Agents, Subjects and Events – which form the main structure of the model (Library of Congress, 2019a; see Figure 2).

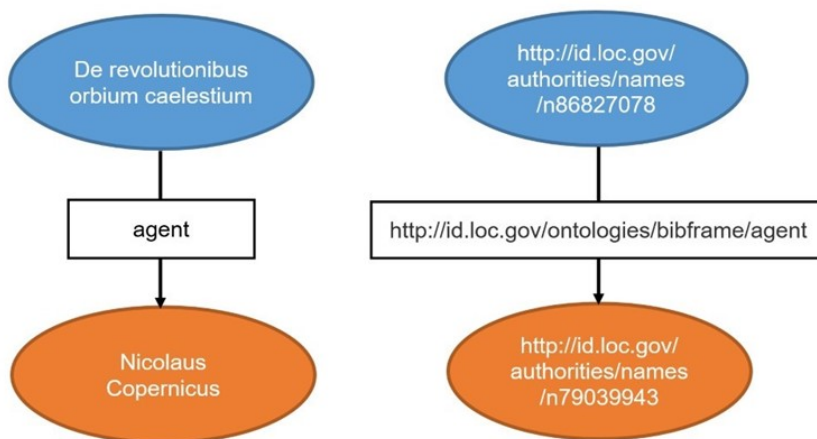


Figure 1: An example of an RDF triple expressing a subject-predicate-object relationship with URIs: “De revolutionibus orbium caelestium” has the agent (author) Nicolaus Copernicus

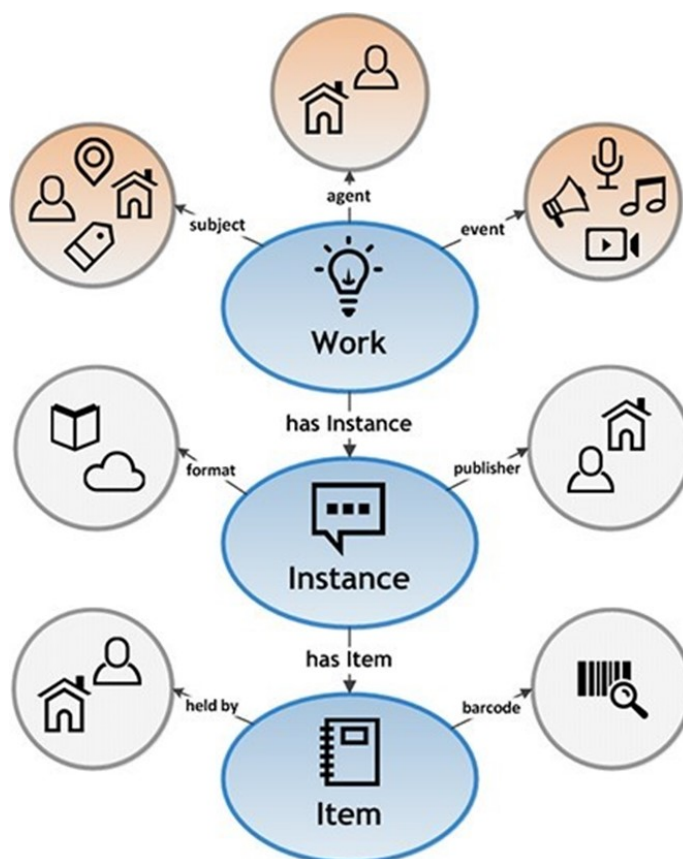


Figure 2: BIBFRAME 2.0 Model (Library of Congress, 2014)



BIBFRAME has been criticized for not providing enough vocabulary to adequately describe special collections items (Schreur, 2018, p. 7). To extend BIBFRAME beyond its core model the project Linked Data for Production (LD4P) was founded in 2016 (Rayle, 2016). As part of the project the Art & Rare Materials (ARM) BIBFRAME Ontology Extension was developed “to provide specialized modeling in the art and rare materials domains” (Futornick & Kovari, 2018). With its more granular classes and properties available for description, the ARM ontology is geared towards helping special collections cataloguers gain more precise descriptive powers by paying closer attention to the data created at Instance and Item level (Futornick & Kovari, 2018). In June 2018, it was announced that the Bibliographic Standards Committee is committed to developing and maintaining the ARM extension of BIBFRAME (American Library Association, 2018).

The current version (Version 0.1.0) of the [ARM ontologies](#) were released on 27 April 2018 (Folsom et al., 2018a). The developers have created a Core ontology, an Activity ontology, an Award ontology, a Custodial History ontology and a Measurement ontology. In its current form the ARM Core ontology has 56 classes and 51 object properties. The main classes are:

Arrangement Attribution Citation ConditionAssessment Enclosure Entry Exhibition	Font FontStyle Handwriting HandwritingStyle Marking Mount Page	PhysicalCondition SignatureStatement StylePeriod TextBlock Typeface
---	--	---

The classes Enclosure, Marking and Mount have subclasses to specify them further. For example, “Enclosure” can be typed as “Binding” or “DustJacket”. The ontology re-uses classes already present in the BIBFRAME: “Activity”, “Agent”, “Extent”, “Identifier”, “Item”, “Note” and “Title”. In some cases, ARM includes subclasses to these for further description, such as the class “Extent” which is subclassed with “PaginationFoliation”. ARM’s main object properties are:

agent atLocation carries cites covers describes hasArrangement hasAttribution hasCitation hasExhibition	hasHandwritingType hasInput hasMaterial hasOrigin hasPhysicalCondition hasSignatureStatement hasSource hasStyle hasStylePeriod hasTypeface	isAppliedMaterialOf isBaseMaterialOf isColorContentOf marks notates subjectOf title writtenIn
--	---	--

The ARM Activity ontology “provides classes and predicates to model agents and roles for the ARM Core Ontology” (Folsom et al., 2018c). It includes 66 classes and two object properties. The main class is “Activity”, which is defined as “an activity or contribution by a single agent that affects or alters the existence or state of a resource” (Folsom et al., 2018c). The “Activity” class is extended by subclasses defining different activities such as “Author”, “Binder” or “Printer” (see Figure 3). All of these Activities are then linked to their respective resource with the object property “hasActivity” or its inverse “isActivityOf” which “[r]elates an activity to the affected resource” (Folsom et al., 2018c).



Figure 3: The Art and Rare Materials (ARM) Activity ontology's "Activity" class and its subclasses (Folsom et al., 2018c)

## Methodology

A task-based evaluation of ARM was performed aided by the ontology editor software Protégé. New RDF data expressed in Turtle (Terse RDF Triple Language) was created using BIBFRAME and ARM on the basis of existing MARC-encoded metadata. The goal of this was to create a use-case scenario mimicking the task of a future special collections cataloguer creating new and converting existing metadata. This was to provide a demonstration of how easily and effectively ARM could be used by a special collections cataloguer as well as how well the ontology can express all the information a special collections cataloguer requires and a special collections user would want to access. Here one example of the modelling process of a record will be presented.

## Description of modelling process

Figure 4: The MARC display for the record of "Nicolai Copernici Torinensis astronomia instaurata, libris sex comprehensa, qui de revolutionibus orbium coelestium inscribuntur..." from the Senate House Library catalogue (Persistent link: <https://catalogue.libraries.london.ac.uk/record=b3191202>)

```
LEADER 00000cam 2200361 a 4500
008 110512s1617 ne a 000 0 lat d
100 1 Copernicus, Nicolaus,|d1473-1543.
240 10 De revolutionibus orbium coelestium.|f1617
245 10 Nicolai Copernici Torinensis astronomia instaurata, libris
sex comprehensa, qui de revolutionibus orbium coelestium
inscribuntur :|bnunc demum post 75 ab obitu authoris annum
integritati suae restituta, notisque illustrata /|copera &
studio D. Nicolai Mulerii medicinae ac matheseos
professoris ordinarij in nova academia quae est Groningae.
246 3 Astronomia instaurata
246 30 Astronomia instaurata
260 Amstelrodami :|bexcebat VVilhelmus Iansonius, sub Solari
aureo,|cA[n]no MDCXVII [1617]
300 [28], 487, [1] p. :|bill. ;|c23 cm. (4to)
500 Previous eds. were published under title: De
revolutionibus orbium coelestium libri VI.
500 Signatures: [sec.]4 (3*)-(4*)4 5*2 A-3P4.
500 Printer's device on title page; engraved initials.
500 Illustrated with diagrams.
510 4 Grässe,|cv. 2, p. 260
510 4 Houzeau & Lancaster. Astronomie (1964 ed.),|c2503
561 ULL copy is from the library of Augustus de Morgan
(bookplate); note by de Morgan dated 24 Oct. 1852 facing
title page.|5UK-LoURL
563 ULL copy is bound in early 17th-cent. blind-tooled calf
over boards; initials "R.W." in stamps i.e. Robert Way;
remains of ties; rebounded.|5UK-LoURL
590 ULL copy has notes relating to Copernicus in French and
English on front flyleaf; price on title page "pret 8s -
3d".|5UK-LoURL
591 99 HPB Sept 2013
650 0 Astronomy|vEarly works to 1800.
651 0 Solar system|vEarly works to 1800.
700 1 Mulerius, Nicolaus,|d1564-1630.
700 1 Blaeu, Willem Janszoon,|d1571-1638,|eprinter.
700 1 De Morgan, Augustus,|d1806-1871,|eformer owner.|5UK-LoURL
700 1 Way, Robert,|ebinder.|5UK-LoURL
752 Netherlands|dAmsterdam.
907 ur
981 |bDM
```

This MARC record, catalogued in accordance with DCRM(B), represents a third edition of Nicolaus Copernicus's "De revolutionibus orbium caelestium", printed in Amsterdam, the Netherlands, by the printer Wilhelm Jansz Blaeu in 1617, from the De Morgan Library, held at Senate House Library (Senate House Library, 2019).

## Work

The BIBFRAME Work level "reflects the conceptual essence of the cataloged resource: authors, languages, and what it is about (subjects)" (Library of Congress, 2014). The MARC 240 field for Uniform Titles, the 100 field designating the Main Entry for personal names and the 65X fields for Subject Added Entries for Topical Terms and Geographic Names apply to Work level.

The *bibframe:Work* entity was created by reusing a Virtual International Authority File (VIAF) URI. The subject terms were added as individuals of *bibframe:Topic* by re-using URIs from the Library of Congress Linked Data Service and attached to *bibframe:Work* using the object property *bibframe:subject*. To convert the 100 field the new ARM Activity Ontology was used. BIBFRAME expresses the relationship between author and work by defining the author as a *bibframe:Agent* and connecting them with *bibframe:AgentOf*. ARM conceives of this relationship differently – rather than defining a person as an agent of a resource, it defines an agent whose activity is being applied to a resource. To attach the author to the *bibframe:Work* entity, ARM defines "Copernicus, Nicolaus, 1473-1543" as an individual of a subclass of *arm:Activity*, *arm:AuthorActivity*. This is then connected to the *bibframe:Work* class through the object property *arm:isActivityOf*.

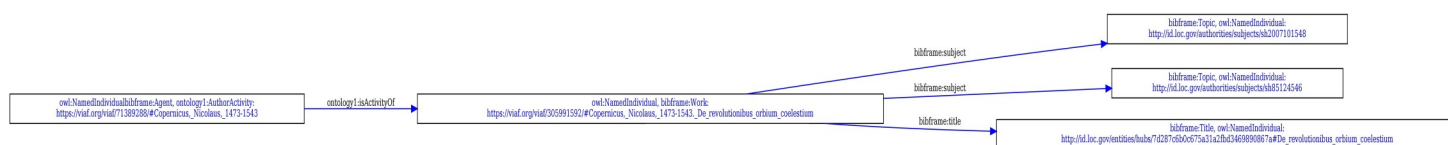


Figure 5: Visualization of Work level of Senate House Library's copy of Nicolaus Copernicus, De revolutionibus orbium caelestium (Amsterdam: printed by Wilhelm Jansz Blaeu, 1617), created with <http://rhizomik.net/html/redefer/rdf2svg-form>

## Instance

The BIBFRAME Instance level represents that "[a] Work may have one or more individual, material embodiments, for example, a particular published form. An Instance reflects information such as its publisher, place and date of publication, and format" (Library of Congress, 2014). At Instance level the 245 field for the Title Statement was included to designate the title of this edition as well as the 246 fields for Varying Form of Title, the 260 field for the Imprint, the 300 field for Physical Description, the 500 Note fields and two 700 Added Entry fields ("Mulerius, Nicolaus, 1564-1630" and "Blaeu, Willem Janszoon, 1571-1638").

The information provided in the 245 field was connected to the *bibframe:Instance* entity using the object property *bibframe:title* to attach the full title as an individual of the class *bibframe:Title*. To define the main title, subtitle and responsibility statement portions of the 245 the datatype properties *bibframe:maintitle*, *bibframe:subtitle* and *bibframe:responsibilityStatement* were used. The 246 fields for the variant forms of the title were added as individuals of the subclass of *bibframe:Title*, *bibframe:VariantType*, and connected through *bibframe:title*. The imprint was represented with the datatype property *bibframe:provisionActivityStatement* and the date with the datatype property *bibframe:date*. The place of publication was added as an individual of the class *bibframe:Place* and linked with the object property *bibframe:place*. To describe the pagination ARM recommends the use of its class *arm:PaginationFoliation*, a subclass of *bibframe:Extent*, which is then connected to *bibframe:Instance* with the property *bibframe:extent*. The dimensions of the resource are attached with the datatype property *bibframe:dimensions*.

To model the first 500 field ("Previous eds. were published ...") the new subclass of *bibframe:Note* that ARM created came into use, the *arm:RelatedMaterialNote*, which can be used for "providing information on materials ... related to the described resource" (Folsom et al., 2018d).

The second Note field describing the signature statement was added with *arm:SignatureStatement*, and then connected to the Instance with its corresponding property *arm:hasSignatureStatement*. The rest of the Note fields were modelled with the help of BIBFRAME to add the illustrative contents described with *bibframe:Illustration* and connected with *bibframe:illustrativeContent*. The two 700 fields describing annotator “Mulerius, Nicolaus, 1564-1630” and the printer “Blaeu, Willem Janszoon, 1571-1638” of the edition were added as above with the ARM Activity Ontology.

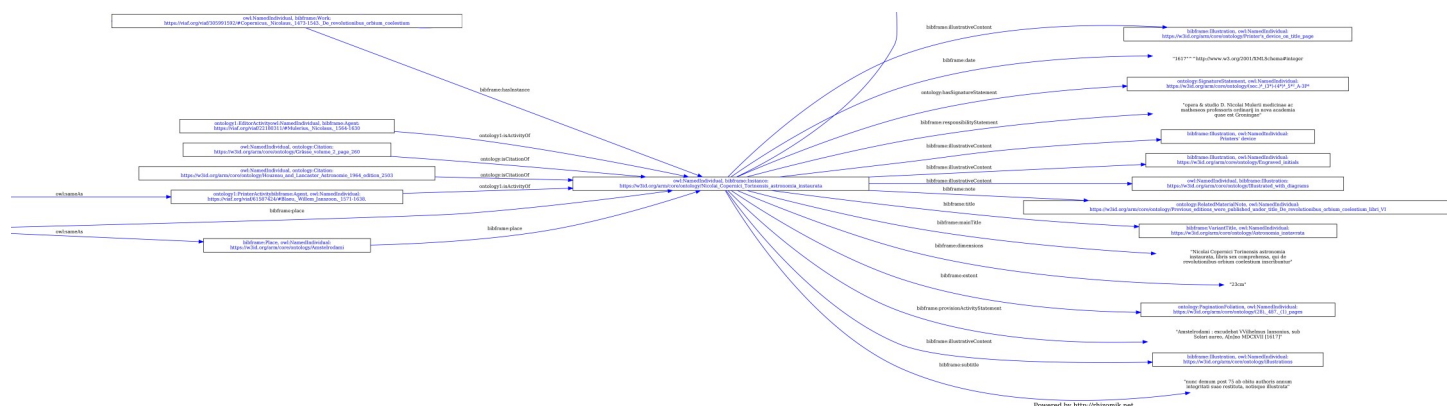


Figure 6: Visualization of Instance level of Senate House Library's copy of Nicolaus Copernicus, *De revolutionibus orbium caelestium* (Amsterdam: printed by Wilhelm Jansz Blaeu, 1617), created with <http://rhizomik.net/html/redefer/rdf2svg-form>

## Item

The rest of the record's information can be ascribed to the Item level which according to BIBFRAME terminology is “an actual copy (physical or electronic) of an Instance” (Library of Congress, 2014). This includes the 561 field for Ownership and Custodial History, the 563 field for Binding Information, the 590 Local Note field and the remaining 700 fields for “De Morgan, Augustus, 1806-1871”, the former owner of the book, and its binder “Way, Robert”.

With the *bibframe:heldBy* property the item and the De Morgan Library, an individual of *bibframe:Collection*, can be connected to Senate House Library, an individual of *bibframe:Organization*. It was possible to model the provenance information provided in the record with another *arm:Activity* subclass, *arm:CollectorActivity*, to describe “De Morgan, Augustus, 1806-1871” in his role as the collector of his library. The ARM Activity ontology, however, does not offer an equivalent of the existing relator term “former owner” that the MARC record utilizes, making it only an indirect translation of the original MARC field. Augustus de Morgan's bookplate can be described through an *arm:Marking* subclass, *arm:Label*. Augustus De Morgan himself can then also be connected to his inscription in the book and the bookplate as an *arm:Activity* subclass, *arm:InscriberActivity* (see Figure 7).

```
### https://viaf.org/viaf/29575875/#De_Morgan,_Augustus,_1806-1871
<https://viaf.org/viaf/29575875/#De_Morgan,_Augustus,_1806-1871> rdf:type owl:NamedIndividual ,
    <http://id.loc.gov/ontologies/bibframe/Agent> ,
    <https://w3id.org/arm/activity/ontology/0.1/CollectorActivity> ,
    <https://w3id.org/arm/activity/ontology/0.1/InscriberActivity> ;
    <https://w3id.org/arm/activity/ontology/0.1/isActivityOf> <https://catalogue.libraries.london.ac.uk/record=b3191202> ,
    :Bookplate_of_Augustus_De_Morgan ,
    :De_Morgan_Library ,
    :note_by_Augustus_de_Morgan_dated_24_October_1852_facing_title_page .
```

Figure 7: Extract from Turtle code for individual of *arm:CollectorActivity* and *arm:InscriberActivity* “De\_Morgan,\_Augustus,\_1806-1871”



The 561 field and 590 fields had to be split up into more granular elements of description to be mapped. The item features several other inscriptions, one inscription of unclear origin in English and French on the front flyleaf as well as one book price. They were defined as individuals of *arm:Inscription*, a subclass of *arm:Marking*, and were related to the item through their object property *arm:markedBy* (see Figure 8).

```
### https://catalogue.libraries.london.ac.uk/record=b3191202
<https://catalogue.libraries.london.ac.uk/record=b3191202> rdf:type owl:NamedIndividual ,
    <http://id.loc.gov/ontologies/bibframe/Item> ;
    <http://id.loc.gov/ontologies/bibframe/heldBy> <http://id.loc.gov/authorities/names/nb2006009415> ;
    dcterms:hasPart <http://www.rbms.info/vocabularies/binding/tr119.htm> ,
        <http://www.rbms.info/vocabularies/binding/tr73.htm> ,
        :early_17th-century_blind-tooled_calf_over_boards ;
    dcterms:isPartOf :De_Morgan_Library ;
    <https://w3id.org/arm/core/ontology/0.1/markedBy> <https://rbms.info/vocabularies/printing-publishing/tr47.htm> ,
        <https://rbms.info/vocabularies/provenance/tr19.htm> ,
        <https://rbms.info/vocabularies/provenance/tr41.htm> ,
        <https://w3id.org/arm/core/ontology/R.W.> ,
        :note_by_Augustus_de_Morgan_dated_24_October_1852_facing_title_page ,
        :notes_relating_to_Copernicus_in_French_and_English_on_front_flyleaf ;
    <http://id.loc.gov/ontologies/bibframe/custodialHistory> "ULL copy is from the library of Augustus de Morgan" .
```

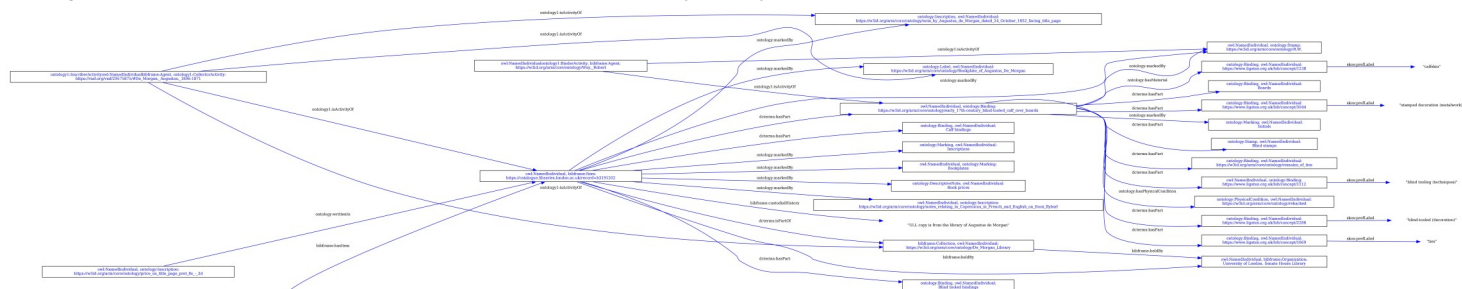
Figure 8: Extract from Turtle code for the individual of *bibframe:Item*

The 563 field was mapped by defining it as an individual of *arm:Binding*. The current modelling recommendations for bindings suggest the use of *dcterms:hasPart* to connect a binding with its item (Folsom et al., 2018b). Another *arm:Activity* subclass, *arm:BinderActivity*, can be used to express that “Way, Robert” is the binder. His stamp marks the binding with the initials “R.W.”. In the model this translates as an individual of *arm:Stamp*, which was added as a description of the binding with the property *arm:isMarkedBy*. Making use the RBMS Vocabularies and the Ligatus Languages of Bindings (LOB) and re-using *dcterms:hasPart* as well as *arm:hasMaterial*, the binding can be described in further detail. For example, using a LOB URI we can express that the material of the binding is calfskin and that it features stamped decoration and blind tooling. This helps the binding to, at least in part, become a machine-readable and -understandable entity (see Figure 9).

```
### https://w3id.org/arm/core/ontology/early_17th-century_blind-tooled_calf_over_boards
:early_17th-century_blind-tooled_calf_over_boards rdf:type owl:NamedIndividual ,
    <https://w3id.org/arm/core/ontology/0.1/Binding> ;
    dcterms:hasPart <https://rbms.info/vocabularies/binding/tr92.htm> ,
        <https://rbms.info/vocabularies/provenance/tr55.htm> ,
        :remains_of_ties ,
        <https://www.ligatus.org.uk/lob/concept/1212> ,
        <https://www.ligatus.org.uk/lob/concept/1669> ,
        <https://www.ligatus.org.uk/lob/concept/2288> ,
        <https://www.ligatus.org.uk/lob/concept/3044> ;
    <https://w3id.org/arm/core/ontology/0.1/hasMaterial> <https://www.ligatus.org.uk/lob/concept/1238> ;
    <https://w3id.org/arm/core/ontology/0.1/hasPhysicalCondition> :rebacked ;
    <https://w3id.org/arm/core/ontology/0.1/markedBy> <https://rbms.info/vocabularies/provenance/tr83.htm> ,
        <https://w3id.org/arm/core/ontology/R.W.>
```

Figure 9: Extract from Turtle code for the individual of *arm:Binding*

Figure 10: Visualization of Item level of Senate House Library's copy of Nicolaus Copernicus, *De revolutionibus orbium caelestium*



(Amsterdam: printed by Wilhelm Jansz Blaeu, 1617), created with <http://rhizomik.net/html/redefer/rdf2svg-form>

## Evaluation

Overall, the modelling of the record was accommodated well by BIBFRAME and ARM. It was possible to map all the existing MARC encoded information onto the new classes and properties without a loss of information. Instance and especially Item level description was better accommodated. At Item level almost all modelling relied on ARM to be successful with no BIBFRAME equivalent being as precise. The integration of existing Linked Data from vocabularies and thesauri to describe the item showed that special collections cataloguers are already able to use an existing wealth of web data, one of the main advantages of Semantic Web technologies (Cole et al., 2013, p. 164).

However, the translation process was not always straightforward. A lot of 5XX fields needed to be split up to be translated into multiple individuals of separate classes which represented a significant shift in how the item was described. Much of the conversion relied on trying to choose the most appropriate translation of concepts and terms, demonstrating that “cataloguers’ judgement” was still vital to model the information. Cole et al. (2013, p. 167) have argued that one of the major questions in the transition from MARC to BIBFRAME will be how to convert the wealth of existing MARC data into Linked Data. Automation will likely have to be supplemented by human intervention.

## Recommendations for further development

Generally, it would be advisable to improve the ease of use of the ontology with better documentation. To clarify the use of the vocabularies it would be useful if the domain and range definitions were added for each class and property as they are in the BIBFRAME vocabulary (Library of Congress, 2019a). Practical examples for how to use each class and property would be very useful too.

Another recommendation would be to synergise the Activity ontology with the existing relator terms the Library of Congress provides in Linked Data form (Library of Congress, 2019b). The Activity ontology has mostly proved to be an effective device to describe the influence of agents over resources but gaps remain. The developers of ARM have stated a preference for adopting the BIBFRAME Contribution model as an alternative to make use of the existing role vocabularies and to more closely align to BIBFRAME modeling standards (Kovari, 2018).

Above all, if the benefit of these new technologies is to be for our users, we must create and test interfaces that base the discoverability of information on Linked Open Data and build systems which exploit its advantages for our users. As McCulloch (2019) has argued creating metadata is “not directly user-facing ... [b]ut it is *user-focused*” (her italics). It is the user who should gain the most significant benefits from this shift in bibliographic format, not the cataloguer. With a new editor, Sinopia (<https://sinopia.io>), for libraries to create metadata in a Linked Data environment already in the works as part of the second phase of Linked Data for Production (LD4P2), more testing is already on the horizon.

## Conclusion

With this study we have got a glimpse into how the application of BIBFRAME to special collections description will transform special collections data. ARM projects a fundamental shift in cataloguing practice. Moving to the new bibliographical standard of BIBFRAME and its extension ARM will profoundly change how a special collections item is represented on the web and how its cataloguer approaches its description to bring it to the increasingly digital end user. By dissolving the traditional catalogue record into multiple data points the cataloguer's output becomes more distributive, easier to explore and re-use. With the creation and development of ARM, a solid foundation for more structured description of special collections as Linked Data is starting to be laid, even though further development work will be necessary before the special collections catalogue is fully integrated into the Semantic Web.

At present we are in a transitional phase where we continue to live in a library data world with many parallel encoding languages and data formats with librarians engaging in a lot of cross-walking between different standards and technologies. MARC and the record-view are not yet completely redundant for the majority of libraries as they form the main basis for their data pool. BIBFRAME, ARM and the other extensions will likely form a new layer of discovery on top without replacing previous data structures until this is fully feasible and sustainable. We are still only just starting to form new principles of knowledge organization and form a "new universe" of bibliographic data. The path towards increasing distribution, openness, disintermediation and transformation of special collections services supported by Linked Data is still a work in progress. If ARM is to be the future standard of the special collections cataloguing community, this is the time for cataloguers to join the discussion. Is this the language we want to use to describe our items? Can triples really express the complexity of a unique item? How well will it accommodate non-book formats? Are there alternatives? Hopefully, this article has created an interest in engaging with one of the many possible futures of cataloguing and will encourage the wider community to join the discussion on what these futures may look like.

## References

- American Library Association, 2017. Guidelines: Competencies for Special Collections Professionals. [online]. Available at: <http://www.ala.org/acrl/standards/comp4speccollect> [Accessed 21 June 2018].
- American Library Association, 2018. RBMS-ARLIS/NA-SAA Joint Task Force on Development of the Art and Rare Materials BIBFRAME Ontology Extension. [online]. Available at: <http://www.ala.org/acrl/rbms/acr-rbmtfbib> [Accessed July 13, 2019].
- Antoniou, G. et al. (eds.), 2012. A Semantic Web primer (3rd edition). Cambridge, Mass.: MIT Press.
- Cole, T.W., Han, M.-J., Fletcher Weathers, W. & Joyner, E., 2013. Library Marc Records Into Linked Open Data: Challenges and Opportunities. In: Journal of Library Metadata, volume 13, nos. 2-3, pp. 163-196. DOI: <https://doi.org/10.1080/19386389.2013.826074>.
- Cullingford, A., 2017. Special collections handbook. London: Facet Publishing.
- DeWeese, K.P., Segal, D., 2015. Libraries and the semantic web, Synthesis lectures on emerging trends in librarianship. San Rafael, California: Morgan & Claypool.
- DeZelar-Tiedman, C., 2017. Discovery of Special Collections Resources at Web Scale: Opportunities and Challenges for Metadata Librarians. In: Thomas, L. M., & Whittaker, B. M., 2017. New Directions for Special Collections: An Anthology of Practice: An Anthology of Practice. Santa Barbara, California: Libraries Unlimited.
- Folsom, S., et al., 2018a. Art and Rare Materials (ARM) BIBFRAME Ontology Extensions. [online]. Available at: <https://github.com/LD4P/arm> [Accessed 18 May 2019].



- Folsom, S., et al., 2018b. Art and Rare Materials (ARM) BIBFRAME Ontology Extensions: Modeling Recommendations. [online]. Available at: [https://github.com/LD4P/arm/tree/master/modeling\\_recommendations](https://github.com/LD4P/arm/tree/master/modeling_recommendations) [Accessed 18 May 2019].
- Folsom, S., et al., 2018c. Art and Rare Materials Activity Ontology. [online]. Available at: <https://ld4p.github.io/arm/activity/ontology/0.1/activity.html#Activity> [Accessed 18 May 2019].
- Folsom, S., et al., 2018d. Art and Rare Materials Core Ontology. [online]. Available at: <https://ld4p.github.io/arm/core/ontology/0.1/core.html> [Accessed 18 May 2019].
- Futornick, M., Kovari, J., 2018. Rare Materials Ontology Extension. [online]. Available at: <https://wiki.duraspace.org/display/LD4P/Rare+Materials+Ontology+Extension> [Accessed 6 June 2019].
- Kovari, J., 2018. Areas for Future Research and Development in the Art and Rare Materials BIBFRAME Ontology Extension. [online]. Available at: [https://github.com/LD4P/arm/blob/master/future\\_research.md](https://github.com/LD4P/arm/blob/master/future_research.md) [Accessed 18 May 2019].
- Library of Congress, 2014. Overview of the BIBFRAME 2.0 Model. [online]. Available at: <https://www.loc.gov/bibframe/docs/bibframe2-model.html> [Accessed 26 August 2018].
- Library of Congress, 2018. Bibliographic Framework Initiative. [online]. Available at: <https://www.loc.gov/bibframe/> [Accessed 20 May 2018].
- Library of Congress, 2019a. Bibframe Ontology: BIBFRAME 2 List View. [online]. Available at: <http://id.loc.gov/ontologies/bibframe.html> [Accessed 27 August 2019].
- Library of Congress, 2019b. ID.LOC.GOV – Linked Data Service. [online]. Available at: <http://id.loc.gov/> [Accessed 3 August 2019].
- Ligatus Research Centre (University of the Arts, London), 2019. Language of Bindings. [online]. Available at: <https://www.ligatus.org.uk/lob/> [Accessed 3 August 2019].
- McCulloch, M., 2019. Cataloguing as a (customer) service. [online]. Available at: <https://lissertations.net/post/1205> [Accessed 31 July 2019].
- O'Dell, A.J., 2017. Special Files on the Semantic Web: Using Linked Data to Revitalize Special Collections Catalogs. In: Thomas, L. M., & Whittaker, B. M., 2017. New Directions for Special Collections: An Anthology of Practice: An Anthology of Practice. Santa Barbara, California: Libraries Unlimited.
- Overholt, J.H., 2013. Five theses on the future of special collections. In: RBM: A Journal of Rare Books, Manuscripts, and Cultural Heritage, volume 14, no. 1, pp. 15-20. Available at: <http://nrs.harvard.edu/urn-3:HUL.InstRepos:10601790> [Accessed 19 April 2019].
- Rayle, L. 2016. Summary, Scope, Background, and Rationale (LD4L Labs). [online]. Available at: <https://wiki.duraspace.org/pages/viewpage.action?pageId=76843833> [Accessed 6 June 2019].
- Schreur, P., 2017. Linked Data for Production (LD4P): A multi-institutional approach to technical services transformation. Paper presented at: 2017 European BIBFRAME Workshop, Deutsche Nationalbibliothek, Frankfurt am Main, 26 September 2017. Available at: <https://pdfs.semanticscholar.org/8c7a/faaf1ae61c7dd32775b2e4e80797a814d77d.pdf> [Accessed 6 June 2019].
- Senate House Library, 2019. De Morgan Library. [online]. Available at: <https://london.ac.uk/senate-house-library/our-collections/special-collections/printed-special-collections/de-morgan-library> [Accessed 1 December 2019].

Tam, M., 2017. Improving Access and “Unhiding” the Special Collections. In: The Serials Librarian 73, pp. 179–185. DOI: <https://doi.org/10.1080/0361526X.2017.1329178>.

VIAF, 2019. VIAF: The Virtual International Authority File. [online]. Available at: <https://viaf.org/> [Accessed 18 August 2019].

W3C, 2015. Semantic Web. [online]. Available at: <https://www.w3.org/standards/semanticweb/> [Accessed 2 December 2018].



**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.

**Submissions:** In the first instance, please contact the Co-editors:

Karen Pierce: [PierceKF@Cardiff.ac.uk](mailto:PierceKF@Cardiff.ac.uk)

Philip Keates: [P.Keates@Kingston.ac.uk](mailto:P.Keates@Kingston.ac.uk)

**Book reviews:** Please contact the editors.

ISSN 2399-9667

CIG website: [cilip.org.uk/cig](http://cilip.org.uk/cig)

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](#)