Caroline Winter

Job Talk for Metadata Librarian (Faculty Librarian) Position

Public Presentation

October 2019

The Future is Meta: Five Revolutionary Ideas for Cataloguing and Metadata in
Libraries and Archives

My talk today is about five revolutionary ideas that will shape the future of metadata and
cataloguing in academic libraries (Figure 1). Because I will draw on experiences from my
doctoral program and work I have done as a member of a digital humanities research lab, I will
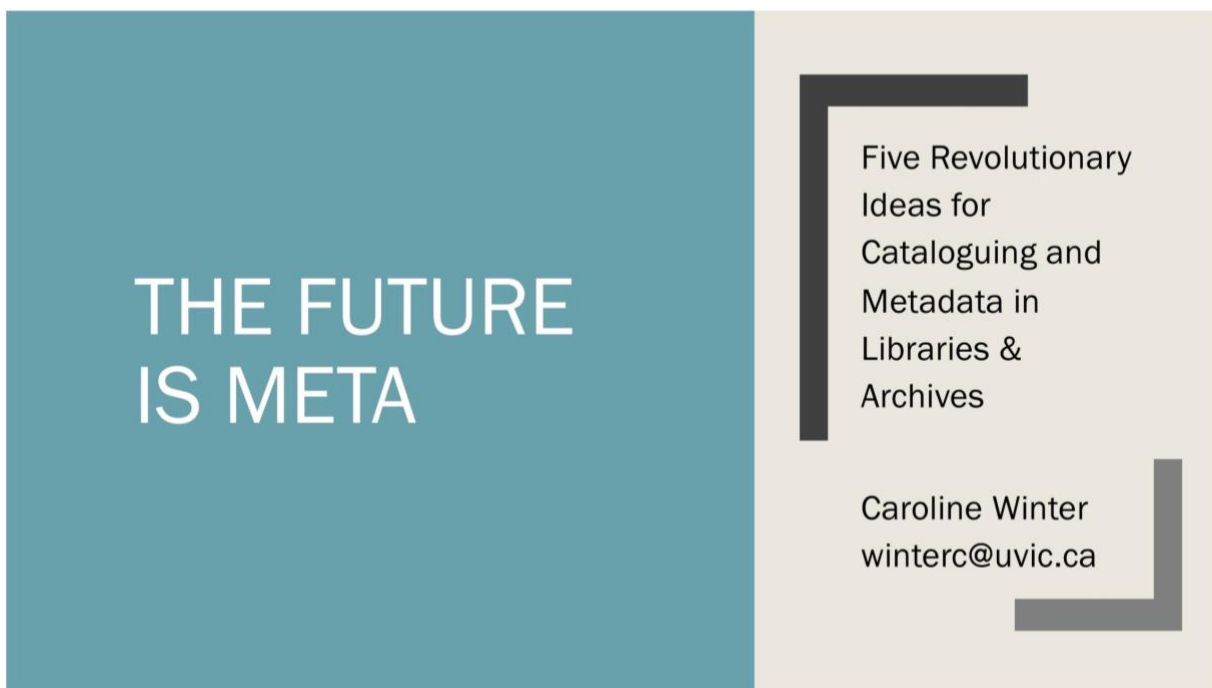start with a bit of background information about me and my work.

Figure 1 The Future is Meta

I am nearing completion of a PhD in English literature, specializing in British fiction
from the Romantic era and digital humanities. I am currently working as the Open Scholarship
Facilitator in the ETCL—the Electronic Textual Cultures Lab—which is located in the Digital
Scholarship Commons and directed by Dr. Ray Siemens. Studying the Romantic age, nearly 250

years in the past, alongside the very current world of digital and open scholarship has highlighted for me how technological and cultural factors affect how we create, use, and share information today and what the future of metadata might look like.

The Romantic age is also known as the Age of Revolution. The political, cultural, and economic revolutions that define that age were driven to a great extent by an information revolution: a rapidly expanding print marketplace made possible by technological innovations, a changing economy, and a growing reading public. Today, we are in the midst of another information revolution, one also driven by technology and economics: a shift from analogue to digital. Both these revolutions have led to a similar problem: an excess of information. Much of my work involves Gothic novels, which were so popular and produced in such numbers in the 1790s that Jane Austen famously referred to this mass of horror in her 1818 novel *Northanger Abbey* as "the trash with which the press now groans" (Austen 23).

The Digital Universe

By 2020:

44,000,000,000,000,000,000,000 bytes

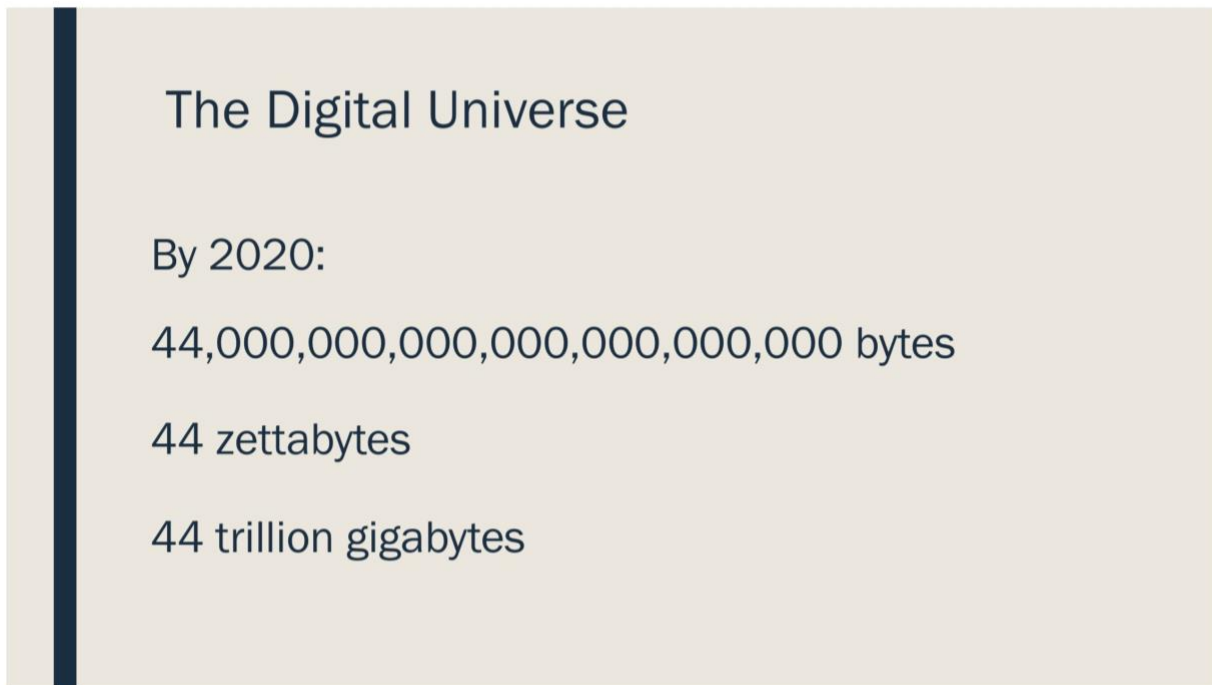44 zettabytes

44 trillion gigabytes

Figure 2 The Digital Universe

Although digital information has the advantage of being weightless, we are still groaning under the burden of mass amounts of digital data (Figure 2). The World Economic Forum estimates that the size of the "digital universe" is expected to reach 44 zettabytes—44 trillion

gigabytes—by 2020 (World Economic Forum). Navigating the vast and rapidly expanding universe of digital information requires a powerful navigational and discovery tool, and that tool is metadata.

Metadata is, essentially, data about data. In a libraries context, metadata is structured data used to locate information items in the library, such as books, articles, maps, and archival materials. Metadata also structures all kinds of digital information. If you have ever Googled something or clicked on a tag in a blog post, you have interacted with metadata. In one study, Jeffrey Pomerantz describes metadata as information "infrastructure," a type of infrastructure as invisible to most of us as electricity grids, but just as vital in the modern digital age (Pomerantz 3). But what is metadata, really, and why does it matter? It matters because, to a greater extent than we may realize, metadata informs how we think about the things it describes.

For example, metadata about gender is gathered in many different contexts, from medical forms to conference registrations to bibliographic records. If the metadata structure offers only two options for describing gender—male and female—any person whose gender falls outside that binary will be misrepresented. And worse, these metadata structures are so pervasive that they tend to limit how we understand the notion of gender itself: if we only ever encounter two options for describing gender, we think that those are the only two options possible.

In a library context, consider the example of Michael Field, a Victorian author of poetry and drama (Figure 3). Michael Field is actually a male gendered pseudonym, though, used by two women: Katharine Harris Bradley and her niece Edith Emma Cooper, who were romantic as well as literary partners. So, how would be describe the creator of Michael Field's poems in a bibliographic record? Is it Michael Field, a name Bradley and Cooper chose deliberately to obscure their individual identifies? Is it Bradley and Cooper? If so, how do we distinguish between their work as Michael Field and their work under other pseudonyms? How would we describe Michael Field's gender? And, crucially, how would each of these decisions affect how Michael Field's work is discovered and engaged with by library users? For instance, would (and

should) Field's work appear in a search for Victorian women writers? This example highlights why a critical approach to creating and using metadata is so important.

Figure 3 Michael Field

Although an evolutionary approach to adapting cataloguing and metadata practices has served users well, adapting to the rapidly changing and expanding digital universe requires a revolutionary approach. The five ideas I outline next will revolutionize the way librarians and library users in the academic community and beyond will conceptualize, create, and use metadata in the next decade or so. These ideas are Digitization, Openness, Connection, Knowledge Creation, and Collaboration.

The idea of digitization is nothing new, so it may seem strange to describe it as a revolutionary idea. To a great extent, though, the history of digitization has involved creating reproductions of print or physical things in digital spaces, such as scanning archival materials or card catalogue records to create digital replications—digital surrogates—of physical objects. Digital surrogates of objects and records are incredibly valuable, of course, for preservation and for accessibility. As essential as this kind of digitization is, though, it remains an extension of the print paradigm. The revolution is to shift our thinking towards a digital paradigm, one that allows us to reimagine—not just reproduce—information objects in digital spaces. As an example,

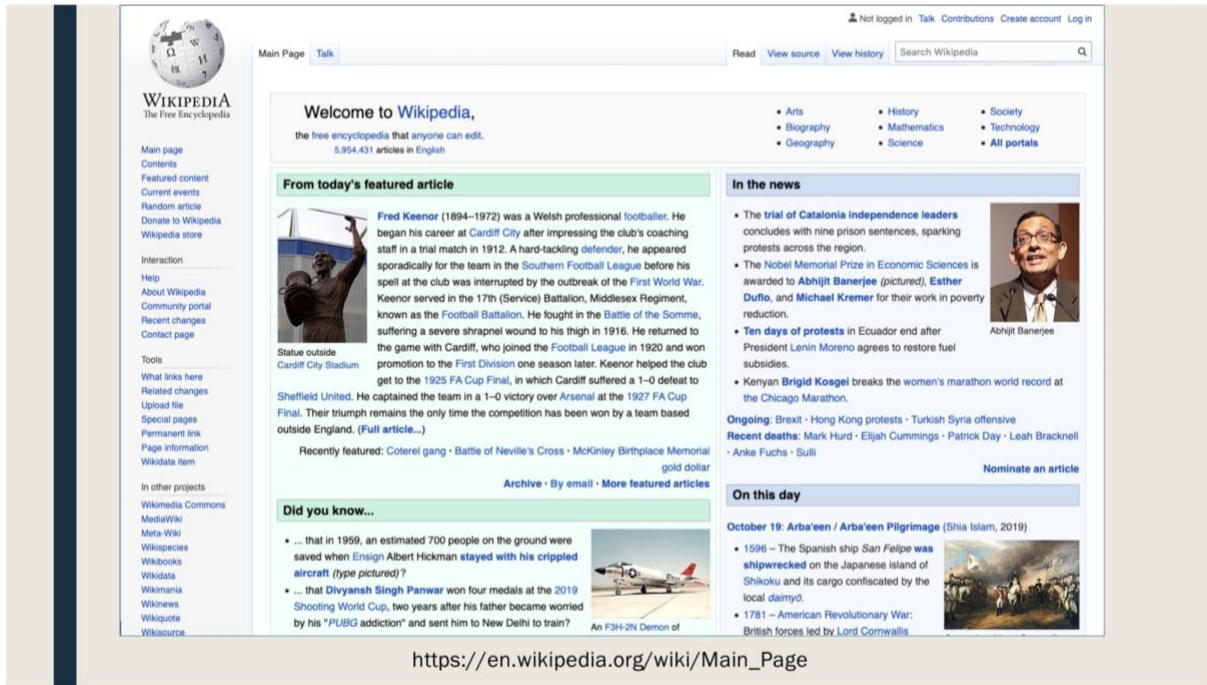consider Wikipedia, the world's largest and most widely used reference work on the internet (Figure 4).

https://en.wikipedia.org/wiki/Main_Page

Figure 4 Wikipedia

Wikipedia is based on the print encyclopedia, a form of reference work that—as it happens—emerged in its modern form during the Romantic-era information revolution. Wikipedia is revolutionary because it reimagines the encyclopedia using the affordances of digital media. Its most contentious feature is its openness: its tagline declares it to be "the free encyclopedia that anyone can edit," and indeed, it is written and edited collaboratively by its community of users. The way we find information in Wikipedia is also radically different from print encyclopedias. Wikipedia has no table of contents, no index, and no page numbers. Instead, it must be navigated with searches and links, foundational elements of the digital paradigm, both of which rely on metadata.

As another example, consider the digitization of the library card catalogue—another invention of the Romantic age (Pomerantz 8). One way to digitize a card catalogue is to scan each card to create a digital surrogate. Using Optical Character Recognition or OCR technology, the text of the card catalogue may even be searchable. The University of Ghent, for example,

digitized its card catalogue this way and made it available online (Figure 5). This is a screenshot of the card for Austen's novel *Northanger Abbey* that I referenced earlier. This interface replicates the browsing experience of the physical card catalogue while also allowing users to search the text. But although this catalogue has been digitized, it is still very much based on the paradigm of print.

Figure 5 An example of a card from the University of Ghent's digitized card catalogue

The MARC (MAchine Readable Catalogue) standards are another evolutionary response to an increasingly digital environment. MARC records are readable to machines, but not particularly human friendly, as this MARC record for *Northanger Abbey* demonstrates (Figure 6). Also, because the standard was designed in the 1960s, it cannot fully realize the potential of current and future information technologies. The need for a new, born-digital format is a matter of "growing consensus," according to studies by Karen Coyle and Kimmy Szeto (Coyle; Szeto 315). BIBFRAME, a metadata format introduced by the Library of Congress in 2012, is a break away from MARC standards, rather than an evolution of them, and therefore seems to be the way forward, but it is still in development. The library catalogue of the future may well reimagine the card catalogue in the way Wikipedia has reimagined the encyclopedia.
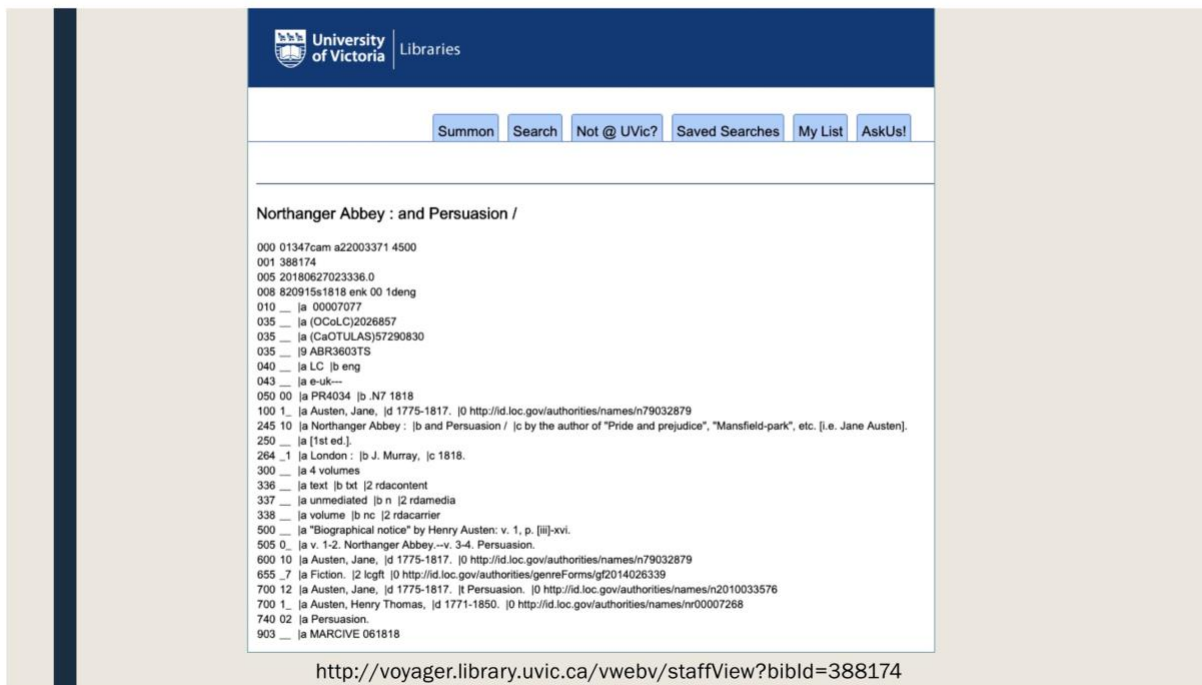
University of Victoria Libraries

Summon Search Not @ UVic? Saved Searches My List AskUs!

Northanger Abbey : and Persuasion /

```
000 01347cam a22003371 4500
001 388174
005 20180627023336.0
008 820915s1818 enk 00 1deng
010 __ |a 00007077
035 __ |a (OCoLC)2026857
035 __ |a (CaOTULAS)57290830
035 __ |9 ABR3603TS
040 __ |a LC |b eng
043 __ |a e-uk---
050 00 |a PR4034 |b .N7 1818
100 1_ |a Austen, Jane, |d 1775-1817. |0 http://id.loc.gov/authorities/names/n79032879
245 10 |a Northanger Abbey : |b and Persuasion / |c by the author of "Pride and prejudice", "Mansfield-park", etc. [i.e. Jane Austen].
250 __ |a [1st ed.].
264 _1 |a London : |b J. Murray, |c 1818.
300 __ |a 4 volumes
336 __ |a text |b txt |2 rdacontent
337 __ |a unmediated |b n |2 rdamedia
338 __ |a volume |b nc |2 rdacarrier
500 __ |a "Biographical notice" by Henry Austen: v. 1, p. [iii]-xvi.
505 0_ |a v. 1-2. Northanger Abbey.--v. 3-4. Persuasion.
600 10 |a Austen, Jane, |d 1775-1817. |0 http://id.loc.gov/authorities/names/n79032879
655 _7 |a Fiction. |2 lcgft |0 http://id.loc.gov/authorities/genreForms/gf2014026339
700 12 |a Austen, Jane, |d 1775-1817. |t Persuasion. |0 http://id.loc.gov/authorities/names/n2010033576
700 1_ |a Austen, Henry Thomas, |d 1771-1850. |0 http://id.loc.gov/authorities/names/nr00007268
740 02 |a Persuasion.
903 __ |a MARCIVE 061818
```

http://voyager.library.uvic.ca/vwebv/staffView?bibId=388174

Figure 6 The MARC record for Jane Austen's Northanger Abbey *in the University of Victoria Libraries catalogue*

The idea of digitization—and particularly reimagining how we create, manage, and use metadata in the digital universe—underwrites the four other revolutionary ideas I will discuss next. The second is openness.

Like digitization, the shift towards openness in the academic world is already underway. The Open Access movement, for instance, works to make information—specifically research published in scholarly journals—freely available to anyone on the web. The various initiatives that collectively form the Open movement—including Open Access, Open Scholarship, Open Data, Open Source, Open Knowledge, and Open Education—are founded upon the idea that information belongs to everyone, and should be openly and freely available to everyone, but also openness to new ways of knowing and of sharing knowledge.

The Open movement's focus on the free movement of information draws attention to how structures of ownership are embedded in existing scholarly practices. An understanding of metadata founded on the principles of Openness promotes knowledge equity by providing a means for giving credit to information creators and specifying how and under what conditions data can be shared. As it happens this week—October 21–27, is international Open Access Week, an opportunity to celebrate the movement's successes and reflect on the way forward. The

Open movement seems to be approaching a tipping point, as funding agencies increasingly stipulate that research they fund must be publicly available and institutional libraries push back against unsustainable subscription models.

This year's theme, Open for Whom?, reflects the community's growing recognition and concern with ensuring that, as the default mode of scholarly communication shifts from closed to open, equity and inclusion are built into these emerging knowledge structures. As mediators between publishers, researchers, students, and the broader community, libraries are at the leading edge of this revolution, and cataloguing and metadata practices have a particularly important role to play.

As Dean Seeman and Heather Dean discuss in a recent paper, libraries and archives have a long history of open and collaborative metadata and cataloguing practices on an international scale, but generally within the library and archives community. While much of this metadata is openly accessible, it is largely closed to contributions from those outside of the library community (Seeman and Dean 6). Libraries recognize the limitations of this open access, closed-contribution model, they note, but shifting that model involves practical, technological, and cultural challenges, not the least of which is the challenge of maintaining the integrity of the metadata itself. I'll return to the idea of open, inclusive metadata in a few minutes. But now, I will turn to the third revolutionary idea, one made possible by open data: connection.

One of the most exciting possibilities for metadata in the coming years is linked open data (LOD). Even though we are "groaning," as it were, under mass amounts of information, our ability to use that information is limited because data tends to be disconnected or siloed, isolated from other data and other users. My experience with linked open data comes from working on Linked Modernisms, a digital humanities project based in the English department here at UVic and led by Dr. Stephen Ross (Figure 7). Linked Modernisms is an open web portal for exploring and visualizing the metadata associated with each entry in the *Routledge Encyclopedia of Modernism*. My colleagues and I—Christine Walde from UVic Libraries and Jana Millar Usiskin from the English Department—developed a custom ontology for structuring this metadata. This schematic shows the various elements of the Linked Modernisms project: the data (The *Routledge Encyclopedia of Modernism*), the metadata associated with it, a machine-reading algorithm that extracts the metadata, and the ontology that structures it. Essentially, this ontology

provides a vocabulary for expressing relationships among the people, places, and things described in the *Encyclopedia*.
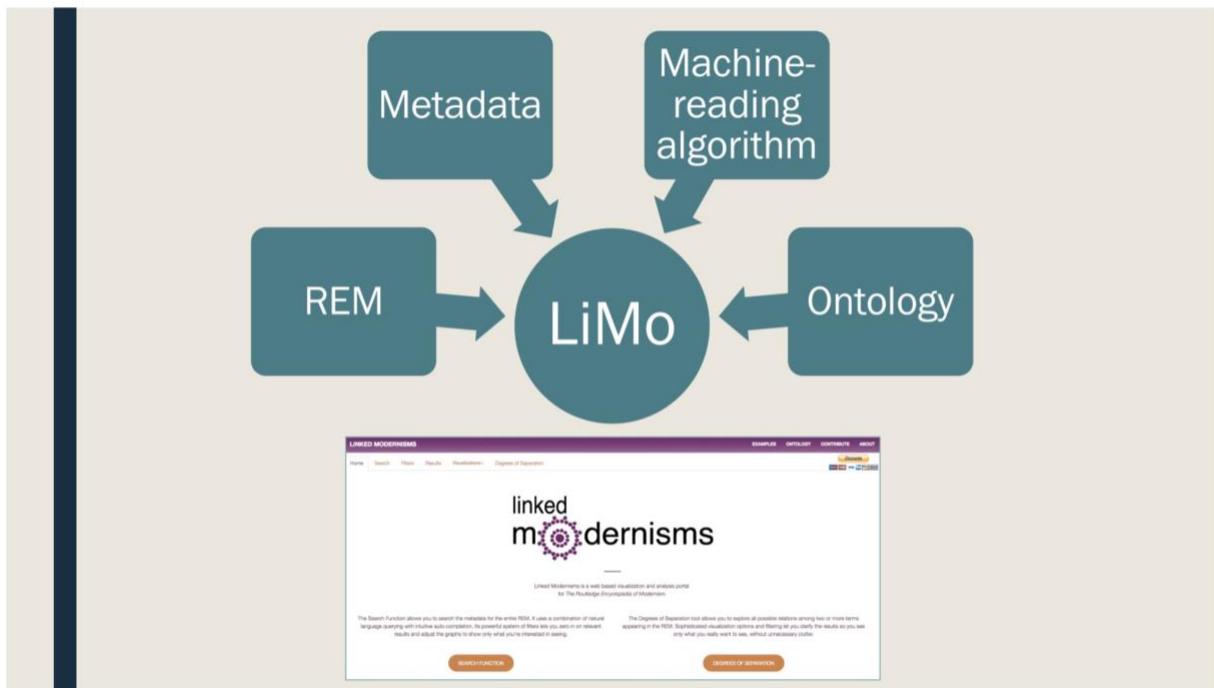

Figure 7 The Linked Modernisms Project

One of our goals in developing this vocabulary was to make it interoperable—to be able to speak with other datasets. For instance, the Linked Modernisms ontology describes a Person as a human being with a name, birthdate, and deathdate. Other vocabularies describe Person differently. The Friend of a Friend (FOAF) vocabulary, for instance, describes social networks, and defines a Person as having properties such as a name, "workplaceHomepage," and "myersBriggs" personality type (Brickley and Miller). Although each vocabulary defines a person differently, according to its purpose, they are essentially describing the same thing. Linked open data allows these projects' datasets to speak to one another by "translating" terms such as Person using shared vocabularies.

Linked open data breaks down data silos, allowing sets of data to interact. If we imagine data silos instead as data islands, each with its own unique language, we can imagine linked open data as networks of islands that can speak to one another—and be read by computers—using shared vocabularies. This network is the Semantic Web (Figure 8). These diagrams produced by the Linked Open Data Cloud illustrate the growth and development of the Semantic Web over

the past decade, from 95 datasets in August 2009 (on the left) to 1239 datasets in March 2019 (on the right). As the digital universe expands over the next decade, so will the Semantic Web, and as Szeto notes, the "convergence" between library information practices and the affordances of the Semantic Web is "a revolutionary moment for library technology" (Szeto 305–06).
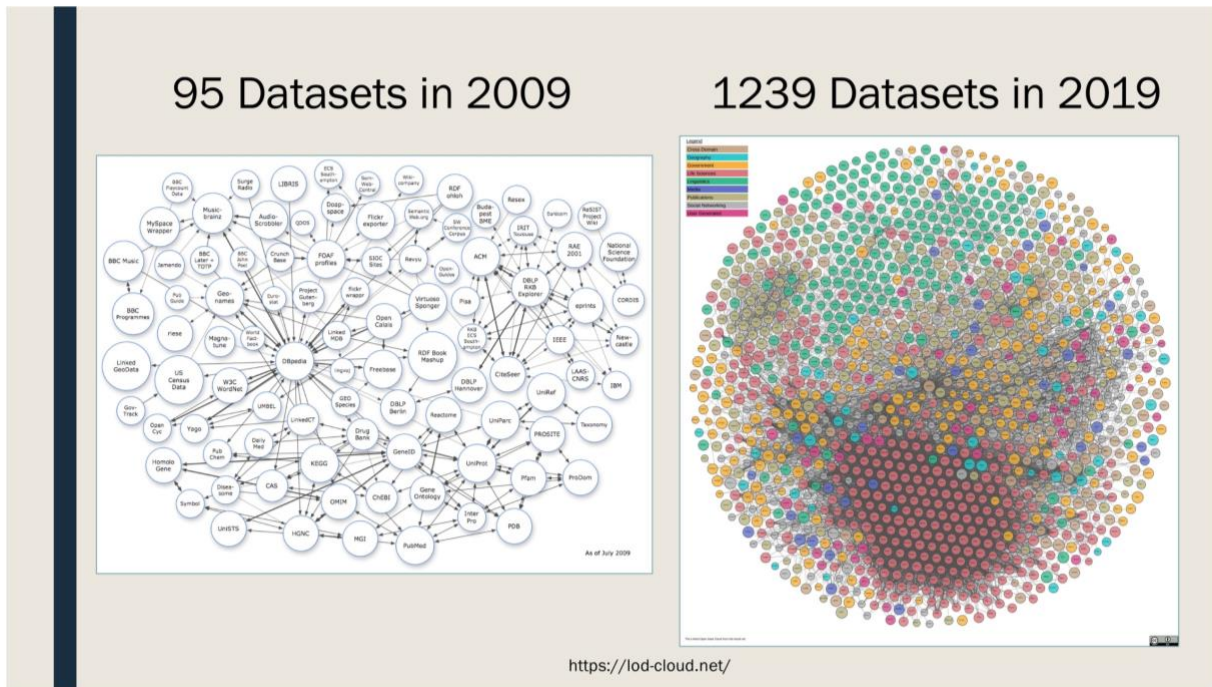

Figure 8 The Semantic Web

Building the Linked Modernisms ontology highlighted for me the interpretive nature of metadata creation. This brings me to the fourth revolutionary idea: knowledge creation.

Subject analysis—the process of capturing an information object in metadata is, at its core, an act of interpretation. Interpreting the identity of Michael Field is one example. Emboldened by the experience of working with metadata on Linked Modernisms, I began developing another digital project, one more closely related to my doctoral research: an experimental digital scholarly edition of six Gothic tales by Mary Wollstonecraft Shelley, all published in the literary annual *The Keepsake* in the 1820s and 1830s (Figure 9). Mary Shelley's Gothic Tales in *The Keepsake*, like other scholarly editions, includes the text of each tale. It also includes digital surrogates, scans of the copies of *The Keepsake* held in UVic's Special Collections. But unlike existing editions of the tales, this one uses metadata to contextualize them and provide pathways for analysis.
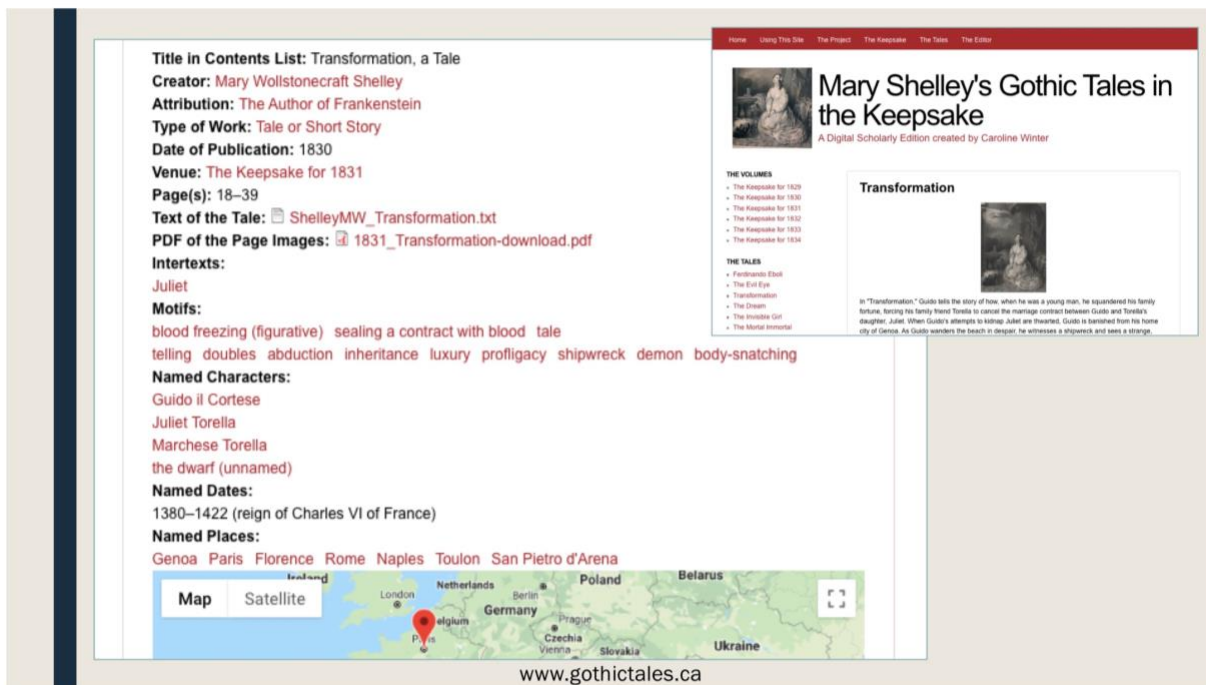
Figure 9 Mary Shelley's Gothic Tales in The Keepsake

In addition to bibliographic metadata for each tale and each text and illustration in these six volumes of *The Keepsake*, I also captured the Gothic motifs employed as well as names of characters, dates, and places. The edition can be navigated through the tales' metadata, which is displayed as clickable links, shown in red. Readers or researchers interested in the Gothic motif of the double, for instance, can click on the link to see all the texts that employ it. Or, they can click on the name of an engraver to see all the illustrations they contributed to these volumes of *The Keepsake*. In this way, the edition leverages its digital platform to create a metadata infrastructure that connects or—more accurately—that reveals connections among seemingly unconnected texts.

Studies by Mjung-Ya Han and Patricia Hswe and by Ivey Glendon have observed that, as metadata practices have evolved alongside digital scholarship, a kind of migration has occurred in metadata creation, from the "back-of-the-house," as Glendon puts it, to the front (Glendon 224). John Chapman notes that cataloguing—particularly cataloguing digital materials— sometimes requires subject matter expertise, and many researchers working on digital projects *have* subject matter knowledge but lack cataloguing and metadata expertise (Chapman 282). My

hope is that the future will see more partnerships between those engaged in knowledge creation—librarians, archivists, faculty, students, and community members—so that we can all benefit from each other's expertise.

The recognition of the shared and complementary practices of librarians and other researchers brings me to my fifth and final revolutionary idea: collaboration. Knowledge, as Peter Burke has argued, is inherently social—we know things because we interact with other people who know other things (Burke). In an academic context, though, and particularly in the humanities context that I'm most familiar with, solitary work is the norm—researchers work alone to produce single-authored papers and monographs. Digital scholarship has challenged this norm, since conducting new kinds of research with unfamiliar tools often requires working with other people with different skills and knowledge, often across disciplines, departments, institutions, and languages (Chapman 281).

Because metadata is a tool for discovery and interpretation, capturing an information object from different points of view creates the most robust description. Although digital tools have made widespread collaboration possible, including collaboration between the research community and the public, enacting these collaborations remains a challenge, as Seeman and Dean have described. One way that libraries can drive collaboration and a shift toward more open contributions is through Wikidata (Figure 10).
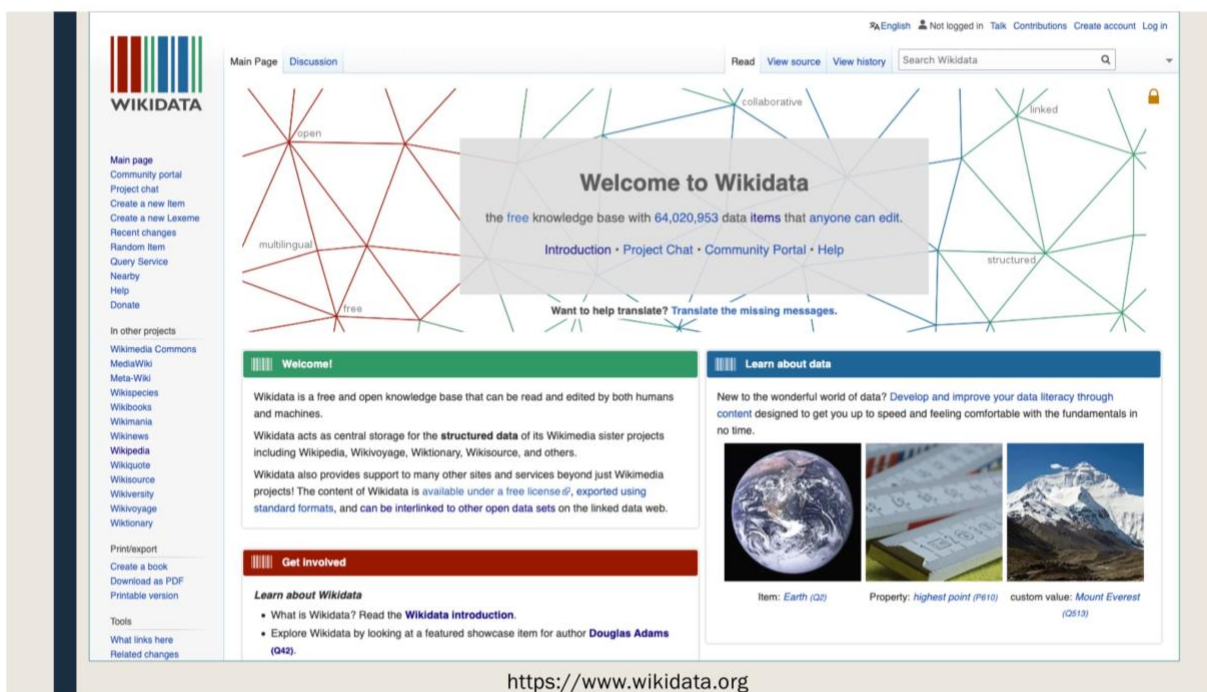
Figure 10 Wikidata

Wikidata is, like Wikipedia, an initiative of the Wikimedia Foundation. It is an open, online, collaboratively edited repository of data, including metadata from other Wikimedia Foundation projects. As Seeman and Dean note, the library community is already experimenting with Wikidata (Seeman and Dean 8). The Association of Research Libraries (ARL), for instance, recently published a white paper on the subject. Although the white paper notes challenges of libraries using data from the repository, particularly relating to the "sustainability and persistence" of the data (Association of Research Libraries and Wikidata 8), it points out that libraries and librarians have a wealth of knowledge and cataloguing expertise to contribute to it. Contributing information about where certain archives are held, for example, is an excellent way to improve discoverability (Association of Research Libraries and Wikidata 9–10). The white paper cites hosting Resident Wikipedians—such as UVic's Honorary Resident Wikipedian, Dr. Erin Glass—and holding Wikipedia Edit-a-Thons—such as the one held in the Digital Scholarship Commons just this past Monday—as ways for libraries to build communities of practice around open data (Association of Research Libraries and Wikidata 9–10).

The ARL describes one of its goals as "creating culturally competent descriptive metadata in collaboration with communities whose lives, collections, and relationships are being described" (Association of Research Libraries and Wikidata 6). Collaborating with communities is one way for librarians to contribute to universities' public missions. Working with communities whose knowledge and experience have traditionally been marginalized, including Indigenous communities, is also an opportunity for academic libraries to improve knowledge equity. The idea of decolonizing metadata, for example, is gaining traction in the academic community. For instance, building on similar events in 2018, Simon Fraser University held a symposium earlier this year on decolonizing classification, and just last week (Oct 17–18, 2019), the [Canadian Research Knowledge Network (CRKN–RCDR)'s Access to Knowledge Conference](#) featured a panel on Decolonizing Metadata in Canada.

Although I have been placing a good deal of emphasis on the digital and on openness, I do not mean to suggest that print will disappear, or that we should adopt open practices uncritically. Privacy, for instance, is an issue that none of using living in this digital age can afford to ignore. My experience, though, is that by prompting us to think of information and

information objects in new ways, digital and open scholarship tends to encourage a more critical approach to our interactions with data.

      The way we create, use, and interact with metadata has evolved in the last 250 years since the Romantic-era information revolution in response to a changing information environment and to users' changing needs. Groaning under the weight of print information in the eighteenth century, the Romantics responded by developing tools for organizing and navigating the complex world of print, including the card catalogue and the encyclopedia. In tandem with this emphasis on order and reason, though, was an increased emphasis on the importance of imagination, the ability to imagine something different than "things as they are." Our challenge over the next decade will be to reimagine our cataloguing and metadata practices for the rapidly changing and expanding digital universe. This is why the metadata of the future will be digital, open, connected, creative, and collaborative.

Works Cited

Association of Research Libraries, and Wikidata. *ARL White Paper on Wikidata: Opportunities and Recommendations*. White paper, 18 Apr. 2019,

    https://apo.org.au/sites/default/files/resource-files/2019/04/apo-nid254221-1380916.pdf.

Austen, Jane. *Northanger Abbey, Lady Susan, The Watsons and Sanditon*. Edited by James

    Kinsley et al., Oxford University Press, 2008.

Brickley, Dan, and Libbey Miller. ″FOAF Vocabulary Specification.″ *FOAF Vocabulary*

    Specification 0.99, 14 Jan. 2014, http://xmlns.com/foaf/spec/#term_Person.

Burke, Peter. *A Social History of Knowledge II: From the Encyclopaedia to Wikipedia*. Polity,

    2012.

Chapman, John W. ″The Roles of the Metadata Librarian in a Research Library.″ *Library*

    Resources & Technical Services, vol. 51, no. 4, Apr. 2011, pp. 279−85. *journals.ala.org*,

    doi:10.5860/lrts.51n4.279.

Coyle, Karen. ″MARC21 as Data: A Start.″ *The Code4Lib Journal*, no. 14, July 2011. *Code4Lib*

    Journal, http://journal.code4lib.org/articles/5468.

Glendon, Ivey. ″Evolution and Revolution in Metadata Librarianship: Identifying New

    Requirements for Professional Qualifications amid Organizational Change.″

    Organization, Representation and Description through the Digital Age: Information in

    Libraries, Archives and Museums, edited by Christine M. Angel and Caroline Fuchs,

    Walter de Gruyter GmbH, 2018, pp. 219−27. *ProQuest Ebook Central*,

    https://ebookcentral-proquest-

    com.ezproxy.library.uvic.ca/lib/uvic/detail.action?docID=5158427.

Pomerantz, Jeffrey. *Metadata*. MIT Press, 2015. *ProQuest Ebook Central*, https://ebookcentral-

proquest-com.ezproxy.library.uvic.ca/lib/uvic/detail.action?docID=4397948.

Seeman, Dean, and Heather Dean. "Open Social Knowledge Creation and Library and Archival

Metadata." *KULA: Knowledge Creation, Dissemination, and Preservation Studies*, vol. 3,

no. 1, Feb. 2019, p. 13. *DOI.org (Crossref)*, doi:10.5334/kula.51.

Szeto, Kimmy. "Positioning Library Data for the Semantic Web: Recent Developments in

Resource Description." *Journal of Web Librarianship*, vol. 7, no. 3, July 2013, pp. 305–

21. *Taylor and Francis+NEJM*, doi:10.1080/19322909.2013.802584.

World Economic Forum. "How Much Data Is Generated Each Day?" *World Economic Forum*,

17 Apr. 2019, https://www.weforum.org/agenda/2019/04/how-much-data-is-generated-

each-day-cf4bddf29f/.