

Hovde, Sarah. "Record added/Page created: leveraging name authority work to highlight women in book history," Women/Gender Minorities in Print/Publishing in the Long 20th Century, July 17, 2019 (remote presentation).

Hello from Washington D.C.! Many thanks to Alice and Anna for organizing this conference, and enabling me to present virtually. I'm a cataloger at the Folger Shakespeare Library, where I work with our modern (i.e. post-1830) printed materials.

I imagine that most of you are familiar with the work of catalogers - we create structured descriptions of library materials in order to facilitate their preservation and access.

Unsurprisingly, it's a great way to come across women and gender minorities in publishing history, as authors and artists, publishers and printers, typesetters and letterers and book jacket designers! Their names can then be added to the catalog record both as transcribed statements and as authorized access points. But today I'd like to suggest that librarians can go even further, by collaborating to update or create Wikipedia and Wikidata items for women and gender minorities as well.

Catalogers and metadata librarians can play a major role in re-writing women and gender minorities back into the historical record by adding them to the knowledge bases that power so much of our present-day information infrastructure. Catalogers aren't the only ones who can do this, of course; and much of this data work is already ongoing. But I will argue that catalogers are uniquely well-placed to document, to research and to use existing research, and to record and maintain data. In this presentation, I'll try to give a brief overview of the extant library authority control landscape, provide some suggestions for how to leverage this landscape to bolster the presence of women and gender minorities within it, and discuss some questions and concerns about the relationship between authority control and underrepresented groups.

Publishers, printers, illustrators, typesetters' names can all be added to a bibliographic record as access points, with terms to specify how they relate to the items being described. In most library systems, this can provide an additional controlled term for users to search on (depending on how the system is configured), and makes it possible to organize records around names. (Ideally, the cataloging system also has a local authority file to provide cross-references for alternative forms of the name.) This enables the connection of items within a library collection or without, linking all the items a particular person or company contributed to; this connective power is strengthened if the name is established in a shared file, such as the Library of Congress Name Authority File. However, names are not always added to records as access points. In the era of card cataloging, limited space necessitated limited information: names were recorded only if they were main entries, that is, authors or sometimes editors. (Libraries might individually maintain local card files of publishers and printers, but this was not the norm.) This lack of information was maintained when card files were transformed into electronic, and then online, catalogs. Current cataloging standards make allowances for recording as many access points as the cataloger desires (or their local best practices dictate), and encourage catalogers to use relator terms. However, many shared cataloging guidelines encourage the use of authorized access points, that is, names that have been established in a shared file such as the Library of Congress Name Authority File. If a name is not established, it can take the cataloger time to research and create the name authority record (and requires specialized training from the Library of Congress, which not all librarians are able or willing to complete).

The main purpose of library authority control, in the era of card catalogs as well as today, is to differentiate people or entities with similar names. Is it the early modern William Shakespeare who wrote all the plays, or one of two later William Shakespeares born in the 1840s? Over the

last two decades, catalog records and databases have had more space for data, and data models such as FRAD ("Functional Requirements for Authority Data," a report published by IFLA in 2009) have emphasized not just disambiguation but also contextualization of entities. Authority control has begun to move towards "identity management," providing more information about entities than just the bare minimum needed to distinguish them from one another.

The MARC record has allowed for increasingly greater amounts of information to be recorded not just about library collection items, but also about their creators and contributors. This allows for more info at a glance to help everyday catalogers: for instance, while a person's established name may be distinguished by their dates of birth and death, the cataloger may not know those dates, and may only know their profession. For programmatic uses, it offers more ways to connect entities to other entities, providing additional ways to explore networks and interpret information.

However, it also comes with several potential pitfalls, which I'll discuss later.

In tandem with the move toward "identity management" has come the increased acknowledgement that traditional library name authority work is time consuming and *hard*, and includes significant barriers to participation that can hinder the participation and representation of marginalized groups in name authority files.

In recent years, there has been a significant amount of discussion of these hurdles in the library community, and a general recognition that something has to give. In about 2015, the Library of Congress-administered Program for Cooperative Cataloging convened the Task Group on Identity Management in NACO to explore ways to lower barriers for participation in the NACO program and other cooperative cataloging efforts. Other PCC task forces have looked at increasing connections between NACO and other authority files such as ISNI (the International

Standard Name Identifier), VIAF (the Virtual Authority File, maintained by OCLC), and Wikidata.

There is also a growing acceptance of Wikipedia, as well as other non-traditional information sources, in the cataloging community. It is not uncommon these days to see Wikipedia cited in a 670 field (i.e., the "Source of information" field in the authority record). There is even a program called the "Authority Toolkit," which works in tandem with the cataloging software OCLC Connexion to automatically populate MARC fields in a name authority record with information from a selected Wikipedia article. Librarians have also begun recognize that Wikidata, even more so than Wikipedia, has a much lower barrier to use.

Wikidata is a cousin of Wikipedia; they're both administered by the Wikimedia Foundation. At its simplest, it is a knowledge base made of statements, or "claims," about entities - people, companies, objects, ideas.

Wikidata is pretty widely acknowledged as a lower-barrier starting point to both Wikipedia and linked data. To edit Wikidata, one finds a topic, and simply adds "claims" to describe its properties. Wikidata's editing interface offers a series of fill-in-the-blank boxes, many with auto-suggestions for their properties. Simply select a property, and fill in the value, to make a claim about the topic; you may also add a URL reference to support your assertion.

For instance, a person named Mary Somerville was born in 1897, worked at the BBC, and was married to Ralph Penton Brown. I can enter all this information into separate boxes, supporting each assertion with a reference. Each property that I'm entering values for is defined and has a unique identifier. Later, the information can be queried by existing programs, or users can write their own instructions to sort through the data.

(When I'm introducing linked data to people, I like to use Wikidata as an example, because it makes it seem so much more approachable than a screen full of JSON or Turtle.) You may notice that this process is significantly easier than creating a name authority record. You don't need any extensive training or certification, you don't need specialized library software, you don't need to know the MARC format, you don't even need to reach a particular level of completeness.

Wikidata's collaborative nature make it more responsive to incorporating different character sets, supporting more non-English content. However, authority work and Wikidata editing ask many of the same questions: When was a person born? Where was a company located? What did someone specialize in?

So, I've spent some time enumerating the Name Authority File's barriers and Wikidata's advantages: how might one put all this together? I'll use an example from my actual cataloging experience to walk through some of the possible outcomes. I do want to underline that I am not necessarily suggesting Wikidata could take the place of a library authority file - just that they can both be complementary when it comes to remembering women and gender minorities in publishing history (and elsewhere).

In 2016, I cataloged a small pamphlet called "The England that Shakespeare knew", written by J.C. Stobart and Mary Somerville, whom I mentioned just now. I had faithfully transcribed both authors in the statement of responsibility, but wanted to include them both as authorized access points as well. Stobart was already established in the LC Name Authority File, but Somerville was not. (Well, several other Mary Somervilles were, but not this one - so I knew that I would need to adequately disambiguate my Mary Somerville.)

I was off to a good start when I found an entry for Mary Somerville in the Oxford Dictionary of National Biography. She was an educator and broadcasting executive with the BBC, who was

born on November 1st, 1897, and died on September 1st, 1963. Those facts alone are enough to identify and disambiguate this Mary Somerville from the others, and to create her NACO record, I didn't need to go any further. However, I was intrigued by Somerville, who got a very early start at the BBC by impressing John Reith while she was still a student at Oxford, and who made her way in her continuing BBC career with spirit. I created a Wikipedia article for her, utilizing other sources I had encountered in addition to her ODNB entry, and then expanded her Wikidata page (which had, unusually, been created almost a year earlier, without a counterpart Wikipedia article).

(Amusingly, when I shared Mary Somerville's draft name authority record with our NACO coordinator, my colleague commented, "There's a Wikipedia page for her, why didn't you cite that?" I responded that there hadn't been when I first started working on the NAR!)

In other situations, creating or updating a name authority record and Wikipedia page for a woman author has led me to create additional pages for other people in her network. In 2015, I cataloged a book of designs "intended to represent visually the basic sound-patterns" of Shakespeare's sonnets, created by a woman named Marjory Bates Pratt. A Marjory Bates Pratt was already established in the name authority file, but with a very sparse record that left me uncertain whether this was the same woman. Researching her further, I discovered that indeed, it was the correct record, and that Marjory Bates Pratt had led a life of fascinatingly varied pursuits. She had received a doctorate in psychology and worked as an instructor, and was married to and friends with other prominent psychologists. Later in life, she was a civil rights advocate, peace activist, and anti-biological warfare protester. In addition to her book of visual sonnet patterns, she was also a haiku author. All this information would probably be excessive in the authority record, but it is great for a Wikipedia article!

In addition to creating a the article for Marjory Bates Pratt, I continued on a few steps further, and also created articles for her husband, Carroll C. Pratt, and friend Lucy May Boring, and then contributed a number of claims to each of their Wikidata pages.

Adapting name authority work to Wikipedia and Wikidata does not always have to involve creating records and articles from scratch, however. It can consist of expanding an existing article (I added a lot of content to the article for the Guild of Women-Binders, including a list of members, while cataloging a book in our collection that they had produced), adding references (I added several 19th-century Shakespeare scholar Teena Rochfort-Smith's Wikipedia article while updating the catalog record for her "Four-Text Hamlet" prototype), or simply adding an Authority Control template to an article (a helpful Wikipedia widget that automatically draws from an article's corresponding Wikidata claims to present the various identifiers for an entity as eye-readable links). In turn, Wikidata and Wikipedia can be starting points for later NACO work. Lest I sound too idealistic, I also want to draw attention to a few potential problems to go with the potential benefits to historical research, concentrating specifically on catalogers' labor and creators' privacy.

Providing increased spaces to record information, between the expanded MARC authority record and public knowledge bases such as Wikidata, also increases the chance of sharing information that creators and contributors may not want available. It can encourage catalogers to go beyond the minimum requirements for a record, not only taking up more of their time, but also possibly infringing on the rights of creators - ironically, interfering with creators' abilities to manage their own identities in the name of shared identity management.

Catalogers have grappled with the privacy problems inherent in authority control with increasing fervor in recent years. This ongoing discussion is particularly relevant in discussing the long 20th

century. Not only are the people whose works you are cataloging or researching more likely to be alive, but zines and other "unpublished" works are becoming more common in library collections.

Expansions to the MARC format for authority records allow catalogers to record additional information about the people and corporations they are naming, from their birthdate to the languages they spoke to their gender. However, these new options (particularly the last option, gender) create two significant and well-documented problems. First, they invite catalogers to guess at demographic features: it's well enough to include languages a person published in, or fields in which they practiced, but is it safe to assume someone's gender from their name? Even if so, is it necessary in every case? (Complicating this even further is the question of pseudonyms, for instance, George Eliot, who is one of the most frequently cited examples.) Secondly, while authority control can be used to manage name changes (much more likely to affect women than men), helping to ensure that searching for a modern scholar or a historical figure brings a user to all their work regardless of the form of their name, the impulse to track a person in the name of information management can be carried too far. Consider a transgender individual who has changed their name, a person fleeing domestic violence who no longer uses their married name, or a pseudonymous creator who wishes to keep their creative work from their professional life: each of these situations might have a different solution, but each requires that the cataloger and/or Wikidata editor act with respect for the person's humanity above their own desire for information proliferation.

For both name authority records and Wikidata claims, we must make an effort to take the creator's preferences into account. It is now much easier than in the past to determine a person's preferred form of their name, by either looking at their self-representation online or by

contacting them directly. Librarians and scholars must bear in mind that not everyone named on a work may want to be named in a public catalog - perhaps they did not mean their work to circulate, perhaps it represents a different period in their lives, or perhaps they do not want it publicly connected with them for reasons of their own.

Libraries have traditionally collected items and credited their creators, embracing a practice of "the more information shared, the better," and it's hard to move away from that. Further, expanding the fields available to be filled in creates higher expectations of authority data that it cannot necessarily meet. On several occasions, I've encountered researchers wanting to use name authority datasets to study women's ownership of books, or authorial work - unfortunately, newer fields are often not universally implemented and may never be, leaving inconclusive data.

However, we may need to be satisfied with an incomplete record.

While I've centered cataloging work in this presentation, many libraries have regrettably seen a trend toward de-centering cataloging in recent years. I am very enthusiastic about the skills that catalogers can bring to non-traditional forms of authority work, but I am wary of encouraging library workers already stretched thin to add another task to their plates (particularly as Wikipedia is not necessarily seen as a high priority for library resources). Instead, I'd like to finish by encouraging collaboration: between catalogers, public-facing librarians, academics, library users, Wikipedia editors, and anyone who exists in the overlapping spaces between those groups. Visit a local edit-a-thon or Wikipedia editors' meetup. Convince your library to host an edit-a-thon, or participate in the Wikimedia Foundation's annual #1lib1ref events! Trade sources and cross-pollinate: check to see if the name you are establishing or editing already exists in Wikipedia or Wikidata. Can you add properties to the Wikidata entity? Adapt information from the Wikipedia article? Add Wikidata URIs to the name authority record, or add an Authority

Control template to the Wikipedia article to connect it with its NACO entry or other identifiers in VIAF? Cite Wikipedia articles in your authority records, cite LCNAF records in your Wikipedia articles, and then add all of their information to Wikidata!