Katrina Cohen-Palacios is a Media Archivist at York University who first edited Wikipedia over a decade ago for a public history undergraduate assignment. In 2014 as a graduate student archivist at the University of Toronto, she resumed her interest in Wikipedia. Since then, her growing interest in structured data led her to Wikidata where she explores relationships of holdings within (and beyond) the Clara Thomas Archives and Special Collections.

Hello all, I am going to try and provide a brief, non-technical introduction to Wikidata and why archivists should be interested in Wikidata.
About Wikidata

- Wikidata is the newest project in the Wikimedia universe
- Wikidata is the youngest Wikimedia project ([see timeline](#)), having launched in 2012, and will be celebrating it’s ninth birthday in a few days.
- Can be described as a “knowledge base”
- Essentially a large linked open data set
On the right is a screenshot of Barack Obama’s Wikipedia article, on the left is his Wikidata item page. In the Wikipedia page one has to scan the article to discover awards received, where as information can be quickly extracted from the Wikidata item.

The wikidata item contains bite sized information that is linked open data because it:
- can be linked with other datasets by following certain standards to be machine readable (structured vocabulary and unique identifiers)
- Is free to use, remix, and publish without copyright restrictions (open)

Essentially the information in the sentences pertaining to awards (highlighted in the Wikipedia page) are broken down into very basic sentences that become machine readable information.

(Note: infoboxes in Wikipedia articles contain structured information.)
The semantic triple formula (subject, predicate, object) follows that of a simple sentence:

- The Toronto Telegram (Q17152683) created (P6241) the Toronto Telegram fonds (Q71342728).
- The Toronto Telegram (Q17152683) employed (P108) John D. Harbron (Q28792876).
- The Toronto Telegram fonds (Q71342728) is part of (P195) the Clara Thomas Archives and Special Collection (Q61467363).
- John D. Harbron (Q28792876) preserved his archives at (P485) the Clara Thomas Archives and Special Collection (Q61467363).

Q = identifier for an item
P = identifier for a property.

Through these simple sentences, we can begin to explore the relationship between two fonds at the Clara Thomas Archives and Special Collections.
Archivists can add links to their finding aids (or simply indicate that material is at an institution) by using the “archives at” property. This example is from Edith Fowke and from it, you can see that her fond is split between multiple institutions across Canada.

It’s very simple to add a link to your institution via this property. If your institution does not pop up while typing it in the text box, you’ll have to create an item. You can add more granular levels of detail with qualifiers such as inventory number (P217), described at URL (P973), level of description (P6224), reference URL (P854), and reference retrieval date (P813).
For a list of properties, I suggest checking out the data structure page of the Wikidata archival description project. It explains which properties are best suited for different types of archival items (institutions, collections/fonds, buildings, individuals, etc). Participants also propose new properties, if one does not exist to fit our needs.

The project aims to “create the world’s most comprehensive high quality database of archival fonds and heritage collections, to represent archival structures within Wikidata where this is deemed useful and to ensure the interlinking between archival finding aids and Wikidata.”

The project’s documentation is written in French since the majority of the archivists working on this project are located in Belgium, France, and Switzerland.
Francophone archives have been early adopters of linking their holdings in Wikidata using the archives at property (as seen in this bubble chart illustrating the total number of “archives at” links per country).

The good news is that Canada is slowly climbing up the ranks with McGill taking the lead with the Clara Thomas Archives & Special Collections on its heels having imported nearly 200 links to our holdings last Friday.

Archivists continue to adopt Wikidata and comparing this chart to that of a few months ago (see tweet). During this time period, Belgium surpassed France, the UK jumped ahead of the Netherlands, and Germany & Canada skipped past Bulgaria.
Why contribute…?
Contributing to Wikidata is a way to automate edits, use shortcuts, and ensure information is up to date in Wikipedia. Certain templates are coded to pull information from Wikidata to Wikipedia.

One such template is {{Authority control}}. By simply placing the words within the double curly brackets, over 30 different name authorities can be added to the webpage (if the information is in Wikidata) -- including the US Congress, NARA, and SNAC (Social Networks and Archival Context) ids. (Note: SNAC id is used on over 125,083 articles!)

{{Official Website}} works in the same way, in that it pulls information from Wikidata… saving you time energy and effort. If the information is not in Wikidata, the code will remain invisible until it is populated.

Now, imagine if there was a template that automatically populated links to finding aids by checking information on Wikidata. Typically archivists manually insert links to finding aids to the external links section of a Wikipedia article. An archival finding aid template would save me an incredible amount of time adding the link to the Margaret Laurence fonds finding aid to the nine other language versions of her page (العربية, Čeština, Deutsch, Français, Հայերեն, Македонски, Română, Русский, Svenska).

(Post presentation update: Biography infoboxes in Wikipedia now pull “archives at” information from Wikidata [1] [2])
Now I have very limited experience with Wikicommons (in fact, I’ve never uploaded an image...), but I’ve discovered that templates in Wikicommons can also pull information from Wikidata. Say for example, the Archives of Ontario has millions of images in Wikicommons. Now imagine that the AO moved, either the information in Wikicommons would have to be manually updated, or it could automatically be updated with one edit in Wikidata, because the information in Wikicommons is linked to Wikidata.

So I experimented Wikidata’fying Wikicommons and replaced free text to wikidata for a photograph from the Toronto Telegram (see edit history).

|institution = {{Institution |wikidata=Q61467363}}
|author = {{Creator|Wikidata=Q17152683}}
|accession number = {{Creator |Wikidata= Q71342728}}

This information will now automatically update if there are any changes.

But most importantly, additional information about the Clara Thomas Archives and Special Collections was added:

- Parent institution York University Libraries
- Location Ontario, Canada
- Coordinates 43° 46’ 21.1” N, 79° 30’ 19.8” W Link to OpenStreetMap Link to Google Maps
If you are planning bulk uploads to Wikidata, I highly recommend considering a preliminary step: editing/upload Wikidata. I would suggest looking at institutions who have relationships with Wikimedia such as:

- BaNQ
- BundesArchiv
- NARA

(Post-presentation update: creating a wikidata item for a category, such as “images from the Clara Thomas Archives and Special Collections” (wikidata item) generates an infobox on the commons page that leads you to statistics of the images used in wikipedia articles)
Even though WikiGLAM includes archives, incorporating archival material to the wiki-universe can often feel like an exercise of square peg, round hole. Our information just does not fit the mould. Templates are often built for/by libraries and museums.

Should archival repository really be under collection?  
Should I place the fonds under accession number? Where else would it go?

The three previously mentioned archival institutions created their own templates for metadata in Wikicommmons (see BaNQ, Bundesarchiv, NARA templates). Shouldn’t we create a single universal archival template that reflects archival arrangement?
Now if you aren’t into the Wiki-universe. There are still reasons why you should contribute information about your holdings to Wikidata.

By adding this data, one can query the information, learn more about your holdings, and discover institution that may contain pieces of a split up fonds. By asking Wikidata (through a [SPARQL query](https://www.w3.org/TR/rdf-sparql-query/)) to pull the location of CTASC fonds held at other institutions.
Linked data may be the way of the future. Wikidata is a really easy, low-risk way to experiment with the concept. AtoM3 plans to incorporate linked data, so the transition will be easier if you already understand linked data.

From the document: “Data aggregation in portal sites is better supported through the use of Linked Data than with traditional hierarchical description. [...] Aggregating data from multiple AtoMinstances to portal sites such as Archeon and ArchivesCanada, for example, is a somewhat cumbersome process of importing updated data via CSV and EAD import and replacing existing content. [...] Linked Data, on the other hand, is designed for dynamic querying across multiple information sources, and for assembling results into web interfaces. [...] Linked Data is also better at representing complex relationships and contextual information in archival description. [...] Linked Data schemas, on the other hand, make it possible to capture rich relationships between individuals and communities that intersect with records as creators, custodians and subjects. Rather than a single dominant narrative, a Linked Data approach also allows for a multiplicity of narratives to be captured as statements about a resource, including a multi-provenantial approach to description and arrangement.”

So, if I've convinced you to give Wikidata a try, there are multiple ways one can edit/add information to the site.
Simply click the edit link, type the information, select from the drop down mention (if a selection does not appear, you must create the item), and hit publish.
Another method is Quick Statements, a tool created to batch edit Wikidata. This is how McGill ROARR bulk uploaded “archives at” properties to items about their holdings (their documentation is available online including a video of the Access2019 presentation).

I suggest checking out this wikimedian-in-residence’s tutorial as it helped me learn how to use Quick Statements in conjunction with Google Sheets.

The problem with this method is that there are few controls to disambiguate the name - though I only attempted this method once. My information was attached to the wrong individual and I had to go back and remove the information in Wikidata. In my opinion this approach is best used for ‘clear cut’ information, such as languages/countries, which requires little, if any, disambiguation.

So instead...I highly recommend my new best friend...
... open refine!

My preferred importation method is Open Refine. Only learned (and experimented) a week ago, I imported nearly 200 “archives at” properties for holdings at CTASC thanks to a tutorial by Emma Carroll at the University of Edinburgh. We already had a list of fonds, identifiers, and URLs from a ‘publish descriptions in AtoM’ project.

I figured it would be simple and an easy morning project for a Friday morning… Well, it took me eight hours to wrangle the data and learn certain function of Open Refine. If I had a clear set of instructions on how to build the schema, etc, the whole thing could have been completed in an hour. So, I hope to write down and share my documentation once I’ve learned a little more… stay tuned. But this was the clearest and easiest way. Highly recommended.
resources
These two hashtags (p485 is the identifier for “archives at” property) on Twitter have been the most informative as I attempt to understand, learn, and edit Wikidata from an archival perspective. It’s helped me join discussions with archivists across the globe, primarily Switzerland and France, but includes Australia and the United States. It’s also a very rewarding, friendly environment that celebrates the small victories.
May be of use.
May be of use.
Feel free to touch base. Happy to answer questions, and brainstorm.