From boxes to AtoM to Wikidata

Frédéric Giuliano frederic.giuliano@mcgill.ca
Anna Dysert anna.dysert@mcgill.ca
Rachel Black rachel.black@mail.mcgill.ca
Eka Grguric eka.grguric@ubc.ca

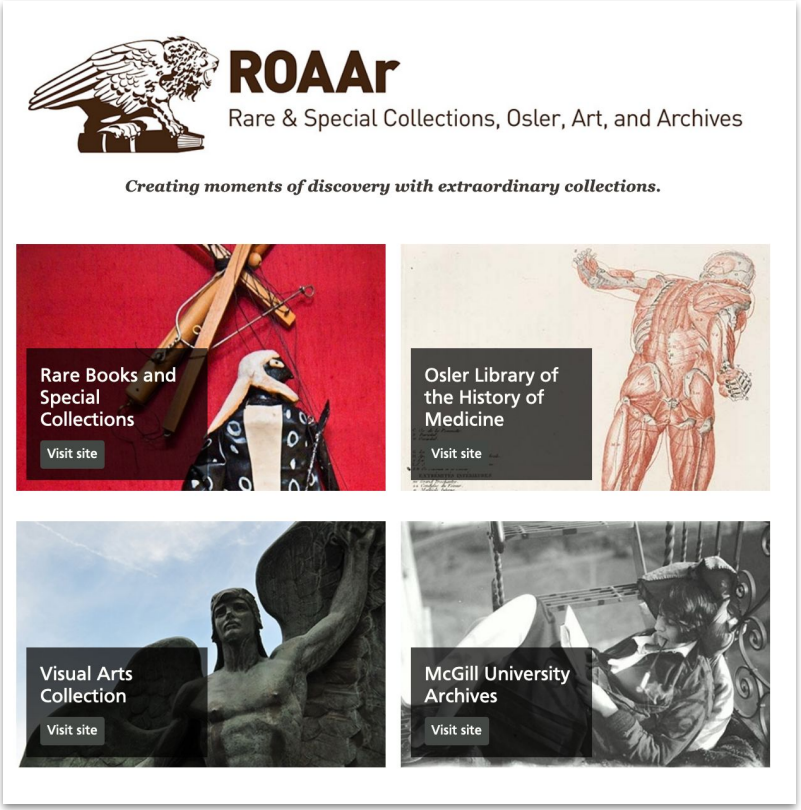
ACCESS 2019 | October 1 | osf.io/9kpws/

Outline

1. Introduction and background
2. Young Canada Works (YCW) Grant
3. Phase 1 - Unboxing the archives
4. Phase 2 - Populating AtoM
5. Phase 3 - Outreach through Wikidata
6. Conclusion

Institutional context

- Academic library (13 branches)
- ROAAr: Rare Books, Osler, Art, and Archives
  - Dedicated Rare materials/archival units
  - Strong collection of rare and archival materials
- Central cataloguing department now integrating archival cataloguing/processing



Project background

Government of Canada | Gouvernement du Canada

Young Canada Works grant to describe and make available a collection of archival fonds and explore ways of incorporating contribution to Wikidata in the process.

- **Timeline:** March 2018 to March 2019
- Collaboration across **three departments** at the McGill University Library collaborated on a project with rotating supervision
    - Archives, Collections, Digital Initiatives
- Added bonus of investigating how this work unfolds in local context.

“Setting yourself up for success**es**”

Conceptualized the project as three distinct phases, and set up many milestones.

- **Phase 1:** Archives, unboxing the data
- **Phase 2:** Collections, “bringing the archives into 2019”
- **Phase 3:** Digital initiatives, exploratory use of a new technology through Wikidata

Each phase aligned with a strategic direction that a respective department wanted to pursue.

Phase 1: Unboxing the Archives

Thinking outside of the boxes

Process

- Compiling resources
- Multitasking to the extreme
- Challenges

Outcomes

- Over 600 entries in the original database
- 576 fonds confirmed
- List of Future Actions
- Options for Outreach

Phase 2: Populating AtoM

Context - adoption of AtoM

- Brief background
  - Desire to implement a centralized archival management system
- Considerations
  - Improved discovery of collections
  - Searching across McGill repositories
  - Getting all many print/legacy finding aids online
- Opportunity to create new workflows
  - Integrating Wiki platforms

Pre-YCW project

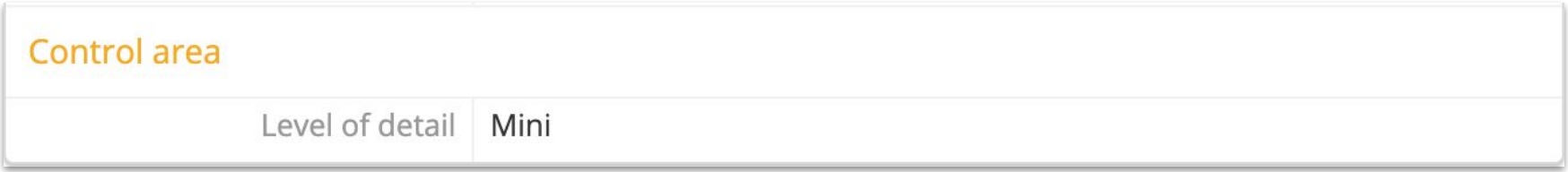
- State of MUA available description: bespoke databases, 3-volume published Guide, paper/PDF finding aids
- Migration accomplished before YCW project: ~358 MUA fonds-level records on AtoM
    - Transcribed from paper finding aids
    - Migrated from the MUA’s online Guide to Archival Resources
    - Imported from one legacy databases

Migration to AtoM: overview

- Crosswalk from MUA DBs to AtoM
  - Straightforward mapping of most fields in legacy DB into AtoM
- Cleaning-up exported data
- CSV import into AtoM

Crosswalk for non-standard fields

- Some custom internal fields, e.g. media codes
  - 01 becomes GMD Textual Record
  - Exception? Media code for “Photographs” = GMD Graphic Materials plus genre access point “Photographs”
- Desire not to lose any internal data -> search capability MUA was used to
  - MUA “Size of fonds” becomes AtoM’s “Level of detail”



Data clean-up for AtoM import

- Data exported in CSV, additional data pulled from MUA accessions database (source of immediate acquisition/custodial history)
- Cleaned-up and standardized
- Creator names standardized (matched to an existing Library of Congress name authority, when possible)

Phase 3: Outreach through Wikidata


WIKIPEDIA
The Free Encyclopedia
encyclopedia
Wiktionary
['wɪkʃənrɪ] n.,
a wiki-based Open
Content dictionary
Wikiquote
WIKIBOOKS
WIKISOURCE
WIKINEWS
WIKIVERSITY
WIKISPECIES
free species directory
MediaWiki
WIKIDATA
WIKIMEDIA
COMMONS
wikivoyage
WIKIMEDIA
META-WIKI
WIKIMEDIA
INCUBATOR
WIKIMEDIA
LABS
WIKIMEDIA
FOUNDATION

Beginning the Exploration

Becoming an “expert”

- ToolForge, Properties
- Wikidata: WikiProject Archival Description

From AtoM to Wikidata

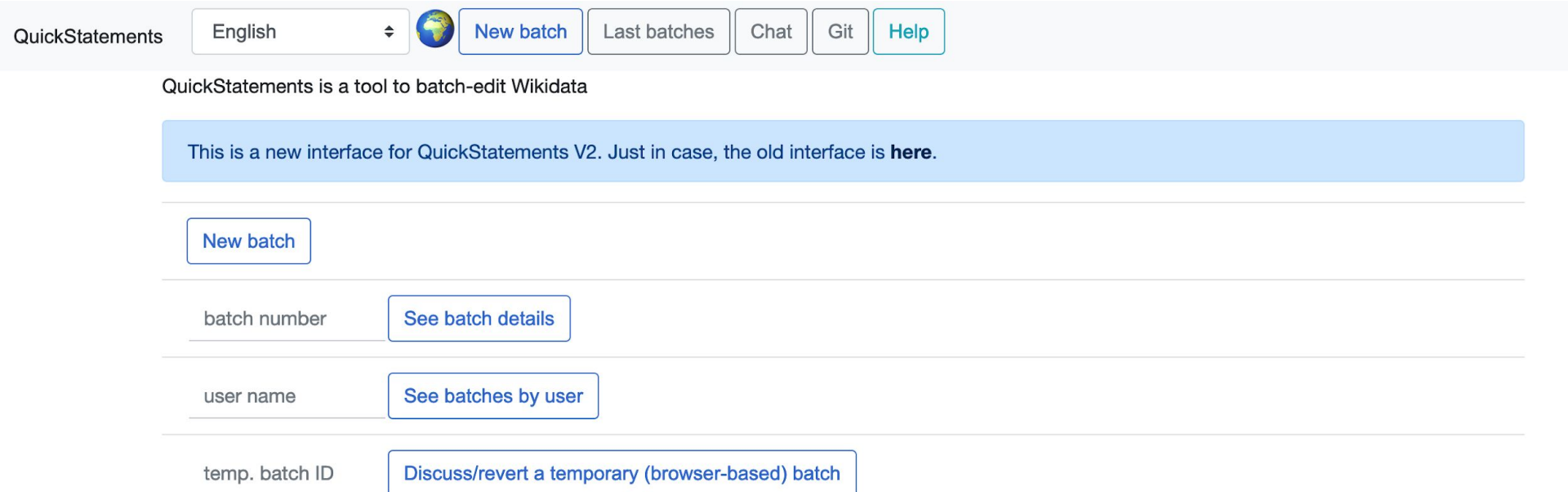
- Crosswalking
- Formatting to CSV

CSV to QuickStatements

QuickStatements

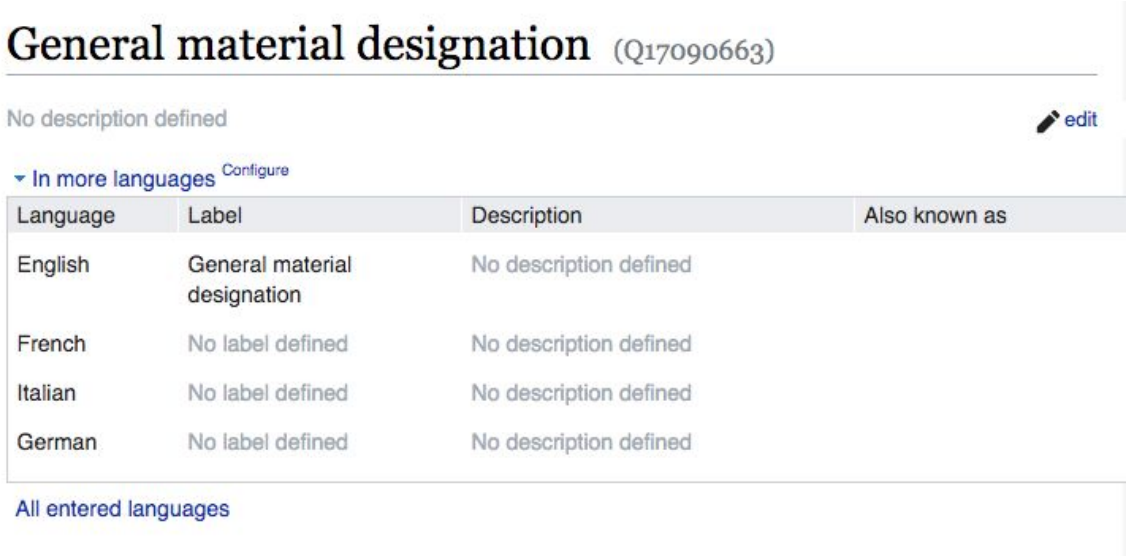
https://tools.wmflabs.org/quickstatements/#/

https://www.wikidata.org/wiki/Help:QuickStatements



Contributing back to the community

- Property Proposals
- Documentation/Workflows



General material designation (Q17090663)

No description defined

edit

In more languages Configure

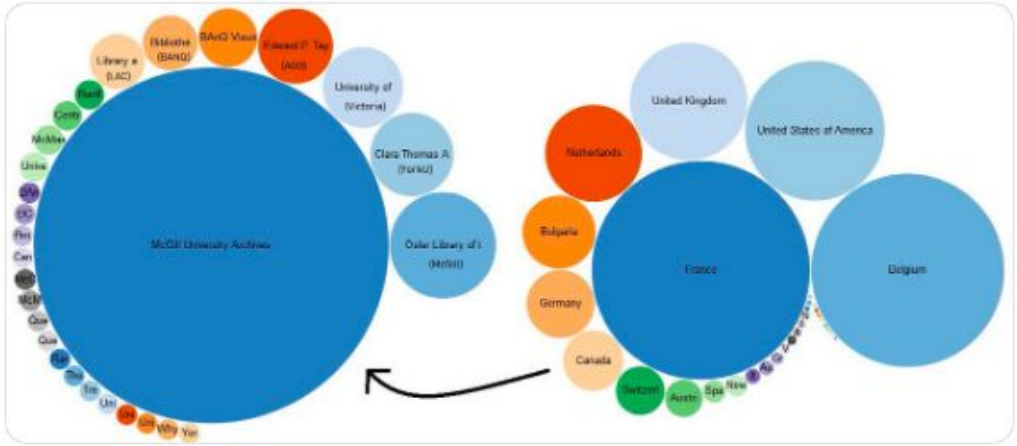
Language	Label	Description	Also known as
English	General material designation	No description defined	
French	No label defined	No description defined	
Italian	No label defined	No description defined	
German	No label defined	No description defined	

All entered languages

katrina c-palacios
@katrina_cp

Follow

#wikidata “archives at” by country and cdn repository.



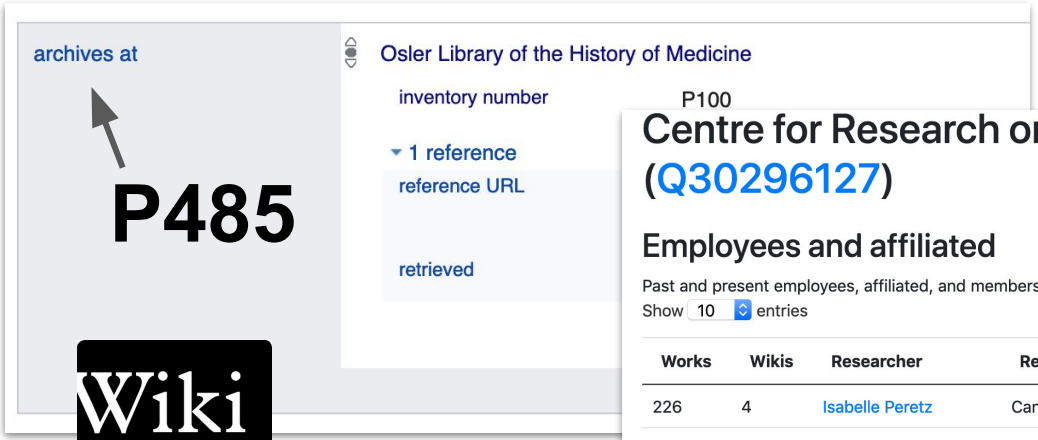
8:31 pm - 28 Apr 2019

2 Retweets **4** Likes

Wikimédia Canada, McGill ROAAr and BAnQ

1 2 4

Wikimedia at McGill



Centre for Research on Brain Language and Music (Q30296127)

Employees and affiliated

Past and present employees, affiliated, and members

Show 10 entries

Search:

Works	Wikis	Researcher	Researcher description	Orcid
226	4	Isabelle Peretz	Canadian psychologist	
172	0	Robert Zatorre	cognitive neuroscientist	
143	0	Pierre Jolicoeur	professor and researcher in experimental cognitive science	
115	0	Bruce Pike	professor and researcher in neuroscience	0000-0001-8924-683X
109	0	Michael Petrides	professor and researcher in neurology	
91	0	Howard Chertkow	cognitive neurologist	

Wiki Edu

ART + FEMINISM

[ ] LIBRARIAN REFERENCE

Lessons Learned

- Pilot to grant is a great model for exploratory projects
- Open Source Tools fail
  - QuickStatements went down during upload
- More time learning the workflow early on avoids repeat labour
- We needed more time for troubleshooting
  - Duplication of identifiers happened and we had a lack of clarity in how issue would resolve
- Wikidata community suffers from lack of documentation around troubleshooting
- Descriptive needs or archives are not widely understood

Next Steps

- A Wikibase instance!
- Increasing discoverability
    - Eg. using schema.org microdata for identifier to wrap around the “alternative identifier” field and include Wikidata Q identifier of each fonds.
- Bulk workflow for future AtoM entries
    - Leveraging tools like OpenRefine to get away from reliance on QuickStatements which, while useful, had moments of downtime during our project.
- Archives are underrepresented on Wikidata, any work to increase their presence has value to the broader archival community
    - Repeat at a different institution

Conclusion

- Wikidata is a powerful tool for outreach and exposure
- Archives are underrepresented and can be a great scope for pilot projects
- Long term maintenance of uploads is a challenge
- This kind of pilot is an effective way of testing internal capacity in supporting exploratory projects around Wikidata and get a sense of resource use
- Document everything (people leave, reproducibility is important)
- Plan for a backup workflow that gets to your MVP

Questions? Thank you!