Digitization @community scale

Loren Fantin
CRKN Digital Heritage
October 18, 2019
OurDigitalWorld is a non-profit organization that supports digital stewardship of community cultural heritage.

We collaborate with libraries, archives, museums, historical societies, and community groups to create sustainable digital collections and bring together for discovery the extraordinary stories that define and shape our communities.
Digitization Strategy

Create opportunities and build capacity, particularly at the grassroots/local level

- Foster equitable participation
- Create reusable cultural heritage data
- Build up the capabilities of GLAM sector
Digital Toolkit (VITA Partners)
"Community" collections

Owner: The McCready Family Address: 655 Goodfellow Road Structural History: Records indicate the following agricultural structures were erected at this address, although some may no longer remain standing. This chronological list provides a built heritage narrative to accompany the image of the remaining barn(s) on the property. 1900 Single Family Detached, 1.75 Storey, 2180 sq. ft. 1900 Shed, 360 sq. ft. 1900 Type I Barn, 2.0 Storey, 972 sq. ft. 1910 Type I Barn, 2.0 Storey, 2172 sq. ft. 1960 Type I Barn, 2.0 Storey, 1182 sq. ft. Owner: The McCready Family Address: 655 Goodfellow Road Structural History: Records indicate the following agricultural structures were erected at this address, although some may no longer remain standing. This chronological list provides a built heritage narrative to accompany the image of the remaining barn(s) on the property. 1900 Single Family Detached, 1.75 Storey, 2180 sq. ft. 1900 Shed, 360 sq. ft. 1900 Type I Barn, 2.0 Storey, 972 sq. ft. 1910 Type I Barn, 2.0 Storey, 2172 sq. ft. 1960 Type I Barn, 2.0 Storey, 1182 sq. ft.
“Local” history = national heritage
Community Digitization Project (2009-2011)

10 Partnerships = over 40 GLAM organizations

Over 18,000 records, 88,000 files, and 500 comments
Fostering an ecosystem for digitization
Discovery as part of digitization

- Discovery is local/thematic/regional/provincial

- Discoverable, online access to millions of digital assets and community stories

- Largest online public archive of Ontario newspapers, from 1810-2018:
  - Ink.ourontario.ca
  - news.ourontario.ca
Newspaper digitization is a priority

- a footprint for nearly every sports team, community event, church social, business opening/closing, birth, marriage, death, celebration, criminal action, expansion, decline, opinions, gatherings, elections, trends, weather patterns and more…
Many libraries, archives, historical societies have been/are are stewarding community newspapers, including digitization

- Factors include: community demand, documenting of local history, phasing out microfilm reader

- Waterloo Public Library - 125,000 pages of The Waterloo Chronicle
- Grimsby Public Library digitized 10,000 pages of the Grimsby Independent
- West Grey Public Library - 29,000 pages of the Durham News to BMDs
- Highland Park Public Library - 30,000 pages of the Highland Park Press
Newspaper Projects

- Bird’s eye perspective
- Expertise
- Copyright
- Advocacy

For each project:

- Identify best copies
- Shepherd the digitization process
- Migrate & clean index records
- Process the digital files: OCR, issue recognition, etc.
- Ingest to toolkit, map data (index or article to newspaper, etc.)
- Custom display site
OCR and historic newspapers

President's Message and Fugitive Slave Law.
Delivered in the U.S. House of Representatives December 9th

Mr. Giddings rose and said: I move, sir, to insert between the resolution last read, and the one succeeding, the following as a distinct resolution:

Resolved, That so much of said message, and accompanying documents as relates to our domestic policy, be referred to the Judiciary Committee.

Mr. G. continued as follows: It has been usual, sir, for this body, while considering the President's Message in Committee of the Whole, to investigate and examine its political character. In accordance with that freedom of discussion which has usually characterized our consideration of that document, I will avail myself of the present moment to examine that portion which relates to our domestic affairs. I did not anticipate, sir, that the message would be under consideration today, and I am not, therefore, prepared at present to give it that examination which I should have bestowed upon it, had I more
OCR and full-text searching
Leveraging digitization funding

• Part of our philosophy (and necessity) = leverage funding from, work with partners (grants preparation)

• MHSO project - OCR’ed 15 multilingual languages, including Inuktitut
Digitization is evolving (1)

- "Best" original = better OCR quality
- Improvements in OCR (machine learning) have been phenomenal

http://news.ourontario.ca/abolition/search
DIY Digitization

• Most expensive part of “digitizing” newspapers is digitization because of the sheer volume

• Experimenting with DIY macro-photography scanner

• Up to 2 reels a day (~2000 newspaper pages)

Newspapers are the richest source of social science data
Re-digitization: Scanning service vs. macro-photography
Re-digitization
Challenges and opportunities ahead

- Interoperable data sets for newspapers - (we are working on this)
- Machine learning and AI for newspapers and digital collections
- Privacy and best practices more critical than ever
- Open, trusted datasets for researchers
- Metrics around use, social and economic benefits
- Climate change!
Collaboration and Partnerships

OurDigitalWorld is proud to be part of a generous community. We thank our sponsors for their ongoing support and advocacy.
Digital collections are *public* and *social assets* - connecting people, connecting research and fostering relationships around the digital assets.

 Merci!

Loren Fantin
 ourdigitalworld.net
 lfantin@ourdigitalworld.org