Digital New France

A global partnership experience in documentary heritage research and data management

Maxime Gohier
Université du Québec à Rimouski

Digital Cultural and Heritage Collections of the Future: Enabling Innovative Access and Expanding Research
CRKN’s Researcher Council
Montreal
November 2nd, 2022
Project goals:

• Scientific:
  • Investigate the production and dissemination of information in the New France era.

• Methodological:
  • Locate, in very large documentary corpora, clues of intertextuality (textual borrowings, recurrent stylistic structures, marks of alterations)
  • Identify actors and the networks that link them

• Technological:
  • Identify the technology needed to allow all researchers on New France to share their research data
  • Facilitate the analysis of this data with new technological tools
DIGITAL NEW FRANCE

• Research partnership (SSHRC)
• Some twenty researchers from different disciplines (history, literature, geography, anthropology, linguistics, art history, heritage studies, archival studies, etc.)
• Institutional partners:
  • BAnQ
  • LAC
  • Museum of Civilization
  • Archives and Rare Book Library, Université de Montréal
  • READ-COOP SCE (Transkribus)
• Citizen Volunteers
Le Monseigneur le Comte

L'Establissement de la Decouverte que l'on a faite du
Mississippi par rapport a l'augmentation des bornes et de la
Grandeur de la France a rejouy la Colonie de Canadas.

A Monseigneur le Comte

de Pontchartrain Ministre Et Secrétaire

d'Etat

à PM

L'Establissement de la Decouverte que l'on a faite du
Mississippi par rapport a l'augmentation des bornes et de la
Grandeur de la France a rejouy la Colonie de Canadas.
FORMING A GLOBAL COMMUNITY

Working protocols
- Standardized transcription
- TEI tagging schema
- Transcription status and project management tools

HTR models training
- Private models
- Public models (open data)

Transkribus Training
- Summer Schools
- Ongoing Literature Review Workshop
- Webinars
- Ad hoc trainings

The « Gardenotes »
- Structuring a crowd sourcing team

Preparing a digital repository
- TNF Plateform
RESULTS

- Over 150,000 pages transcribed
- Two public models
- Dozens of private models
- Wide variety of documents processed:
  - Court records: Bailiwick of Montreal
  - Correspondence from colonial administrators: Louisiana and Canada
  - Ordinance books of intendants
  - Notaries' registries
  - Native language dictionaries
  - Handwritten maps
PUBLIC MODELS: OFFICIAL CORRESPONDANCES

- Variety of writing hands
- ÉSecretaries' and administrators' scripts
  - Round, bastard and running hand
- Real performances on the page below:
- CER: 4.91%  WER: 15.15%

A dozen archival fonds over a century (~1660-1760).
Training set: 1832 pages
Validation set: 197 pages
Total: 2029 pages
PRIVATE MODELS:

17TH & 18TH CENTURIES CANADIAN NOTARIES

- Regularity of writing hands
  - One registry = one sceptor
- Standardized lexicon: recurrent juridical terminology
- Effective error rate on the above page:
  CER: 1.35%  WER: 7.64%
- The quality of the writing varies greatly from one notary to another
- High performance model with little training data

Michel Saindon’s registry (1768-1780)
BAnQ-Rimouski, CN104,S50

Training set: 112 pages
Validation set: 27 pages
Total: 139 pages
NEXT STEPS

- Enlarge partnership
- Develop the TNF digital repository
- Develop AI tools for data processing
- Investigate the possibilities of linking textual data with other data types
- Disseminate the model for other projects