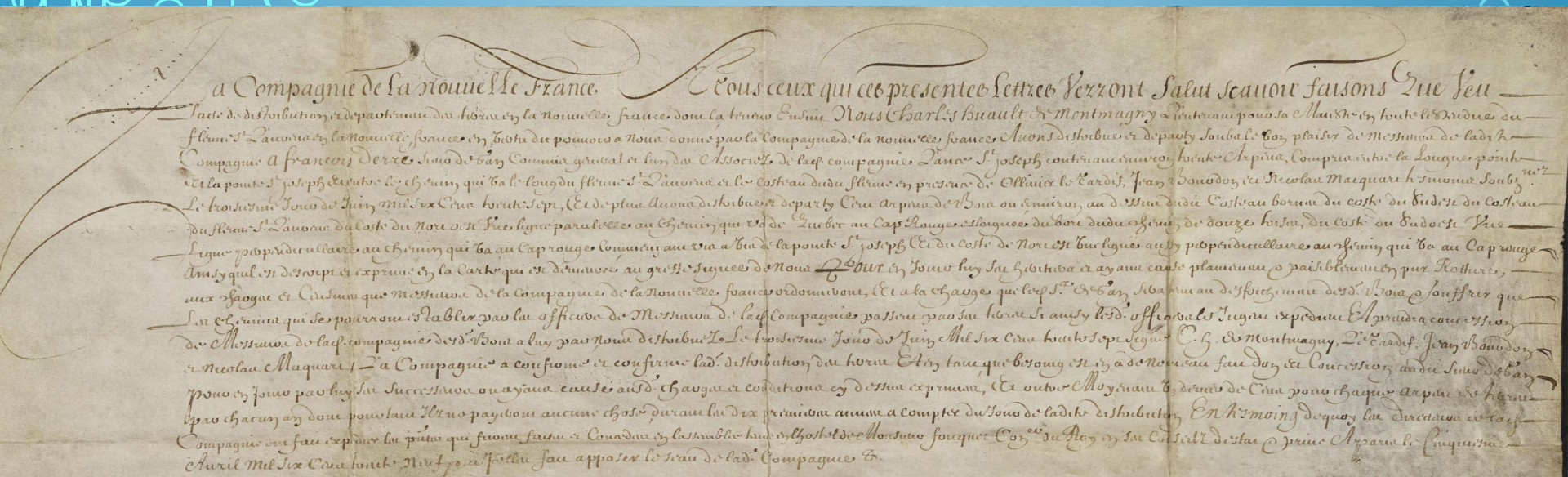


Digital New France

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

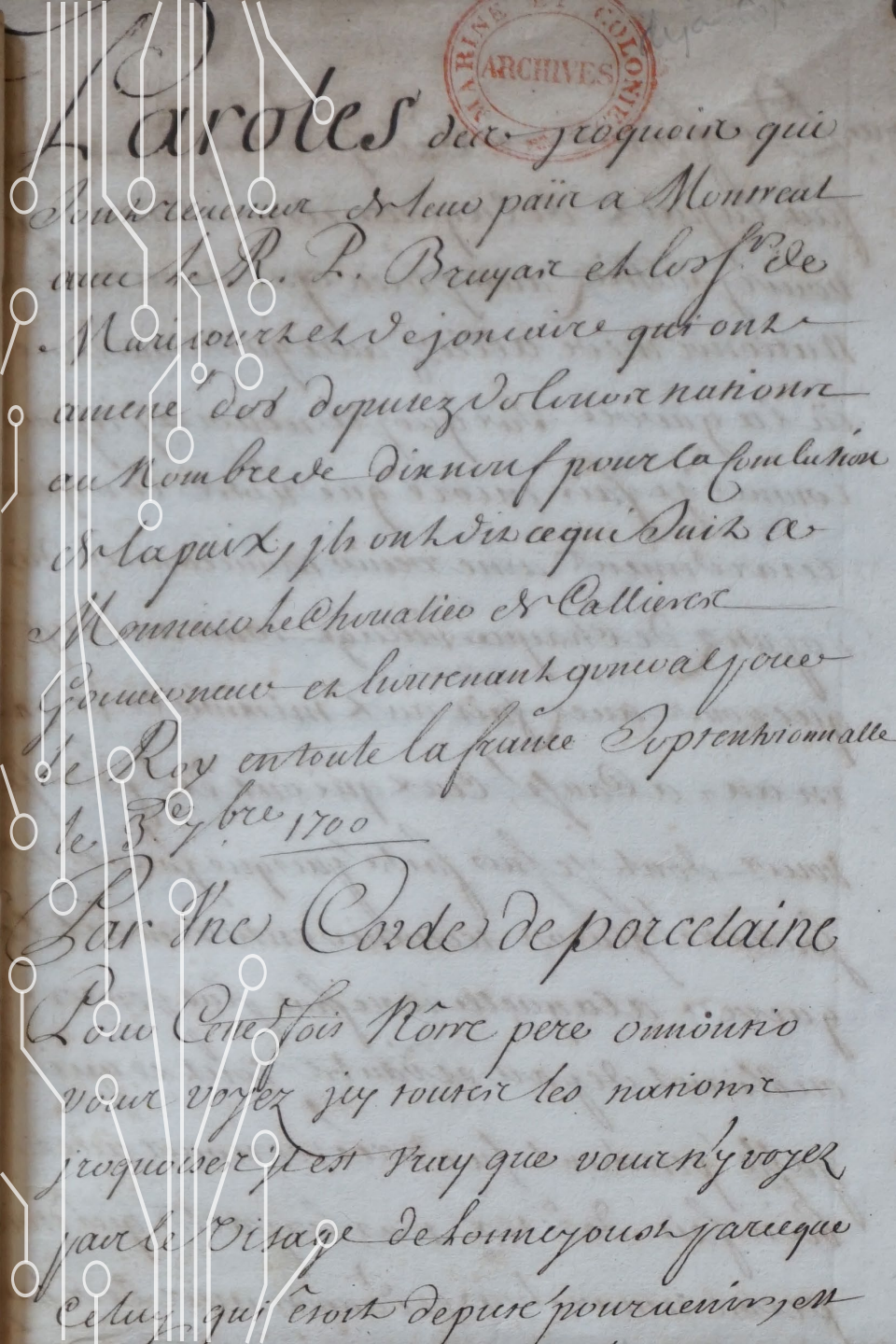


Nouvelle France
NUMÉRIQUE

DIGITAL NEW FRANCE

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

TRANSKRIBUS : USING IA FOR MANUSCRIPT ANALYSIS AND RESEARCH DATA MANAGEMENT

Transkribus Expert Client v1.21.0 (22_09_2022_07:35), Loaded doc: ANOM, C13A, vol. 1, ID: 496877, Page 39, file: FRANOM34_C13_001_0039.jpg [Image Meta Info: (Resolution:200.0, w*h: 2248 * 3152)] [current l...

Search current document... 39 /682 Final

Server Overview Layout Metadata Tools

Logout maxime_gohier@uqar.ca

Document... Find

Document Manager User Manager

Versions Jobs

Recent Documents... User activity

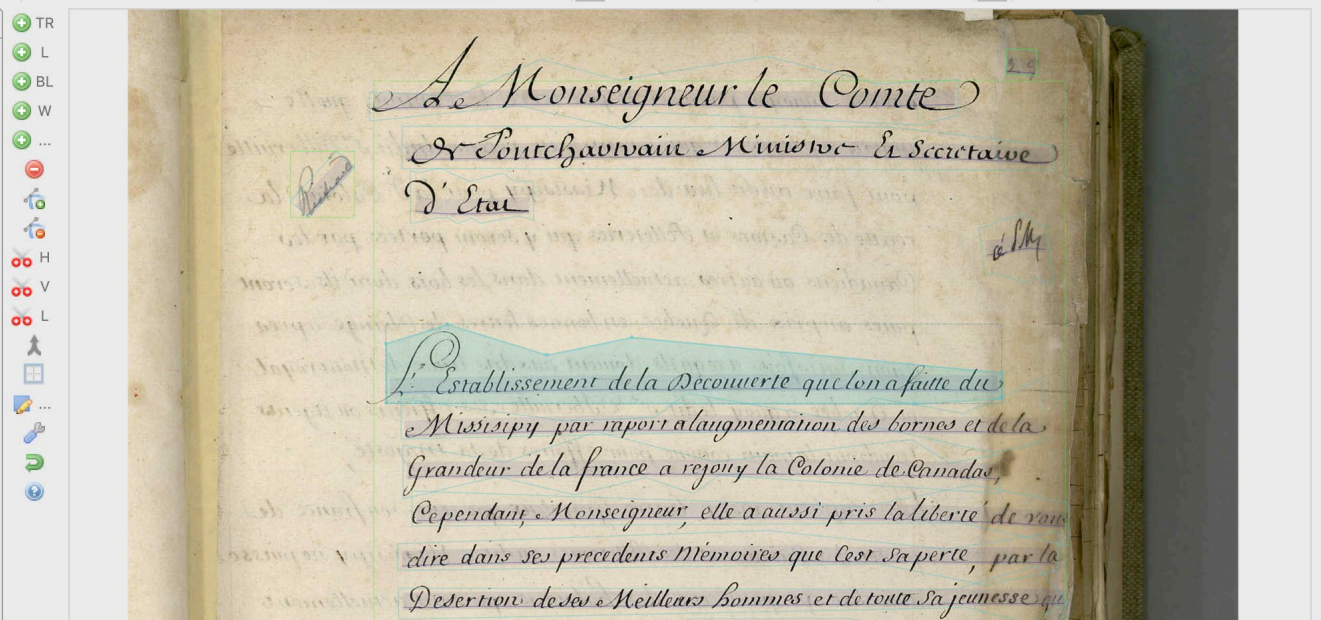
Collections: ANOM, C13A (55686, Owner) Col-ID

Documents Model Data

1-54 / 54

Doc-ID	ID	Title	Pages	Uplo
	49689E	ANOM, C13A, vol. 41	1010	max
	49689E	ANOM, C13A, vol. 40	798	max
	496894	ANOM, C13A, vol. 9	876	max
	496893	ANOM, C13A, vol. 8	956	max
	496892	ANOM, C13A, vol. 7	666	max
	496891	ANOM, C13A, vol. 6	894	max
	49689C	ANOM, C13A, vol. 5	762	max
	49688E	ANOM, C13A, vol. 4	1094	max
	49688E	ANOM, C13A, vol. 3	970	max
	496887	ANOM, C13A, vol. 2	900	max
	49688E	ANOM, C13A, vol. 10	740	max
	496877	ANOM, C13A, vol. 1	682	max

100 Filter



- 3-1 A Monseigneur le Comte
- 3-2 de Pontchartrain Ministre Et Secretaire
- 3-3 d'Etat
- 3-4 à PM
- 3-5 L'Establissement de la Decouverte que l'on a faitte du
- 3-6 Missisipy par raport a l'augmentation des bornes et de la
- 3-7 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

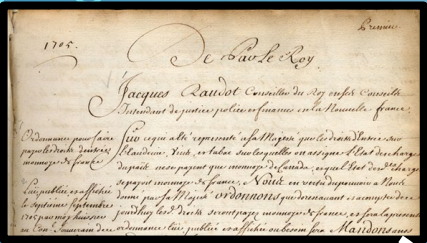
- *TNF* Platform

Cultural heritage institution (GLAM)

Digital collections

Documents digitization

XML-files

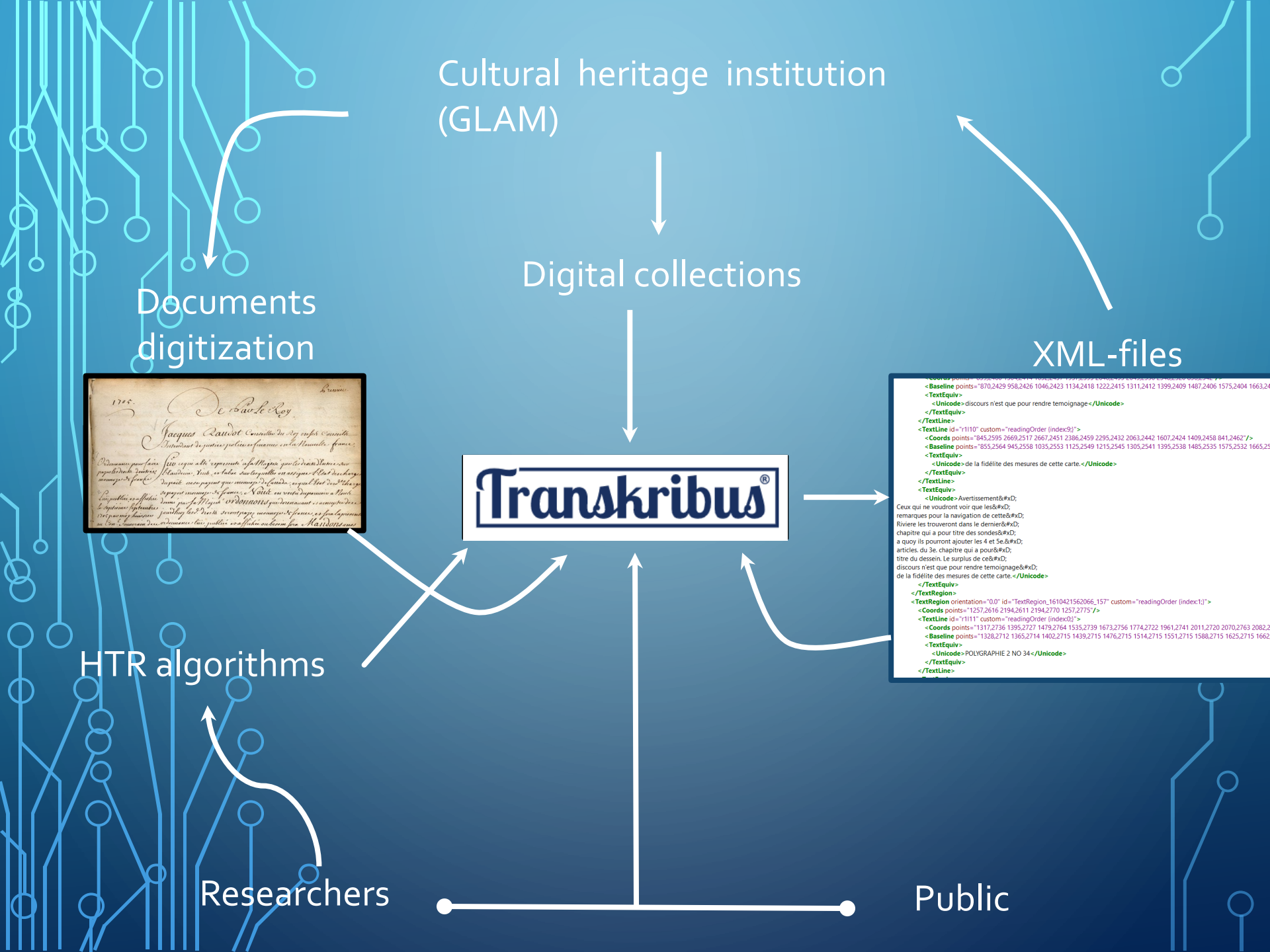


```
<Coords points="870,2429 958,2426 1046,2423 1134,2418 1222,2415 1311,2412 1399,2409 1487,2406 1575,2404 1663,2402" />
<Baseline points="870,2429 958,2426 1046,2423 1134,2418 1222,2415 1311,2412 1399,2409 1487,2406 1575,2404 1663,2402" />
<TextEquiv>
<Unicode>discours n'est que pour rendre témoignage</Unicode>
</TextEquiv>
</TextLine>
<TextLine id="T1110" custom="readingOrder (index:9)">
<Coords points="845,2595 2669,2517 2667,2451 2386,2459 2295,2432 2063,2442 1607,2424 1409,2458 841,2462" />
<Baseline points="845,2595 2669,2517 2667,2451 2386,2459 2295,2432 2063,2442 1607,2424 1409,2458 841,2462" />
<TextEquiv>
<Unicode>de la fidélité des mesures de cette carte.</Unicode>
</TextEquiv>
</TextLine>
<TextEquiv>
<Unicode>Avertissement</Unicode>
</TextEquiv>
Ceux qui ne voudront voir que les<#x>D;
remarques pour la navigation de cette<#x>D;
Rivière les trouveront dans le dernier<#x>D;
chapitre qui a pour titre des sondes<#x>D;
a quoy ils pourront ajouter les 4 et 5e.<#x>D;
articles du 3e. chapitre qui a pour<#x>D;
titre du dessin. Le surplus de ce<#x>D;
discours n'est que pour rendre témoignage<#x>D;
de la fidélité des mesures de cette carte.</Unicode>
</TextEquiv>
</TextRegion>
<TextRegion orientation="0.0" id="TextRegion_1610421562066_157" custom="readingOrder (index:1)">
<Coords points="1257,2616 2194,2770 1257,2775" />
<TextLine id="T1111" custom="readingOrder (index:0)">
<Coords points="1317,2736 1395,2727 1479,2764 1535,2739 1673,2756 1774,2722 1961,2741 2011,2720 2070,2763 2082,2763" />
<Baseline points="1317,2736 1395,2727 1479,2764 1535,2739 1673,2756 1774,2722 1961,2741 2011,2720 2070,2763 2082,2763" />
<TextEquiv>
<Unicode>POLYGRAPHIE 2 NO 34</Unicode>
</TextEquiv>
</TextLine>
</TextRegion>
```

HTR algorithms

Researchers

Public



c. anbedatsin. avoir froid
 c. anbi. bande. ruban.
 c. anbidé. dents.
 c. anbin. vanbin. voir.
 c. anbesépin. jués.
 c. anbi. bouillon.
 e. anbi. moins. peu. petitement.

Archives du Séminaire d



Carte d'une Partie du Lac

Supérieur avec la Découverte de la Rivière depuis le Grand Portage. A
 jusqu'au Lac Quimpigon, ou on a Construit Le Fort Maurepas, Le Fort St.
 Charles, est dans le Lac des Bois, et dans le Lac Tekamamion, un troisième
 Fort qui porte le nom de St. Pierre.

Les Lacs et Rivière Marquis en Couleur Bleu, sont celles qu'on a découvertes

RESULTS

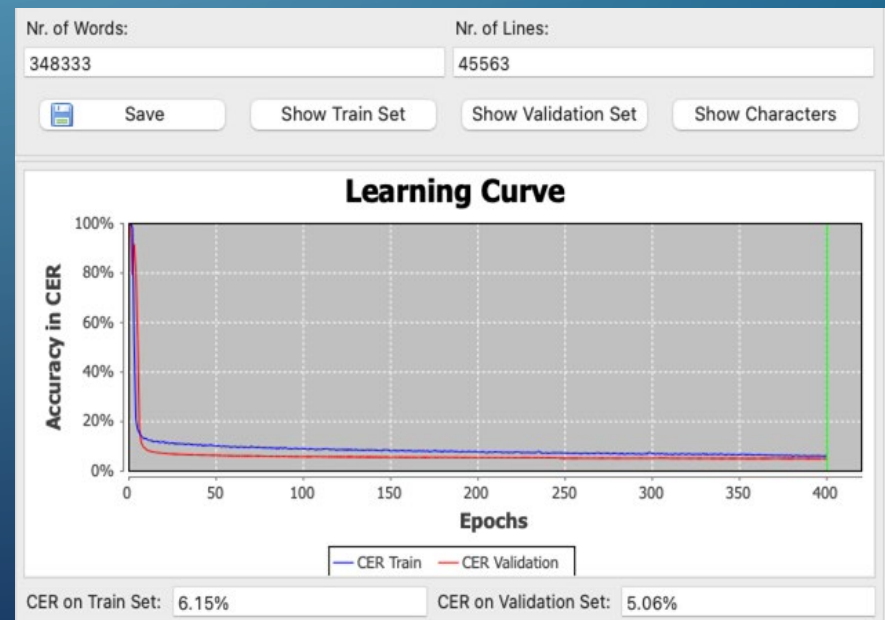
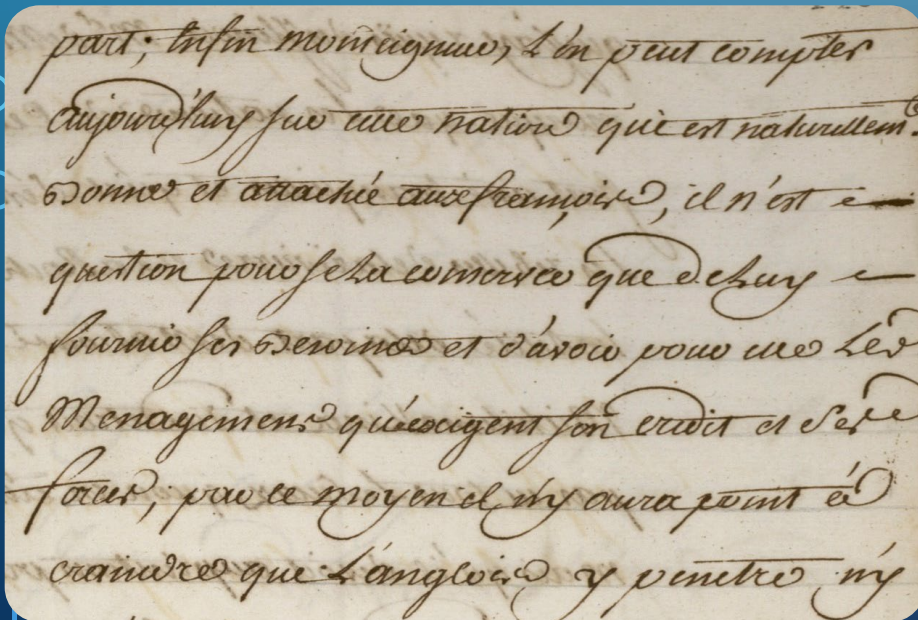
- Over 150,000 pages transcribed
- Two public models
- Dozens of private models
- Wide variety of documents processed :
 - Court records: Bailiwick of Montreal
 - Correspondence from colonial administrators: Louisiana and Canada
 - Ordinance books of intendants
 - Notaries' registries
 - Native language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

Traning set : **1832 pages**
Validation set : **197 pages**
Total: **2029 pages**

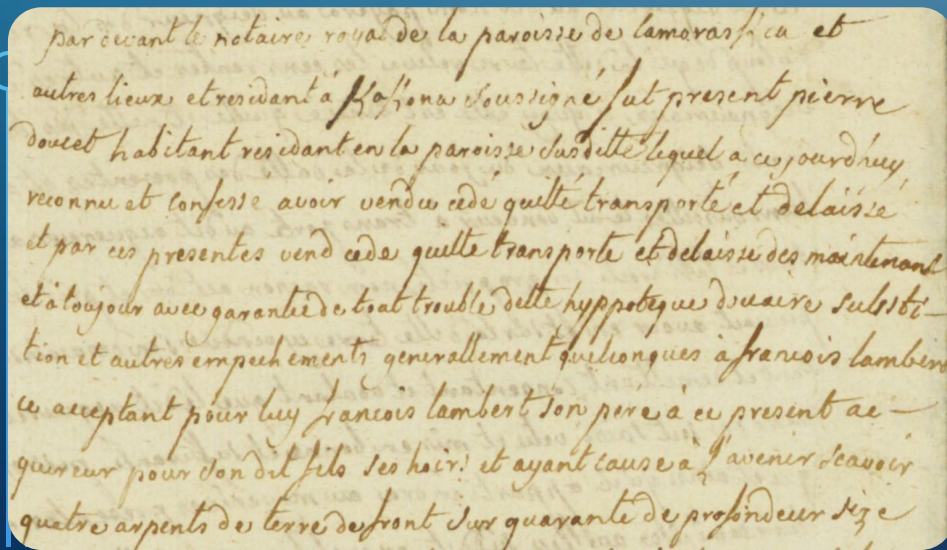


PRIVATE MODELS : 17TH & 18TH CENTURIES CANADIAN NOTARIES

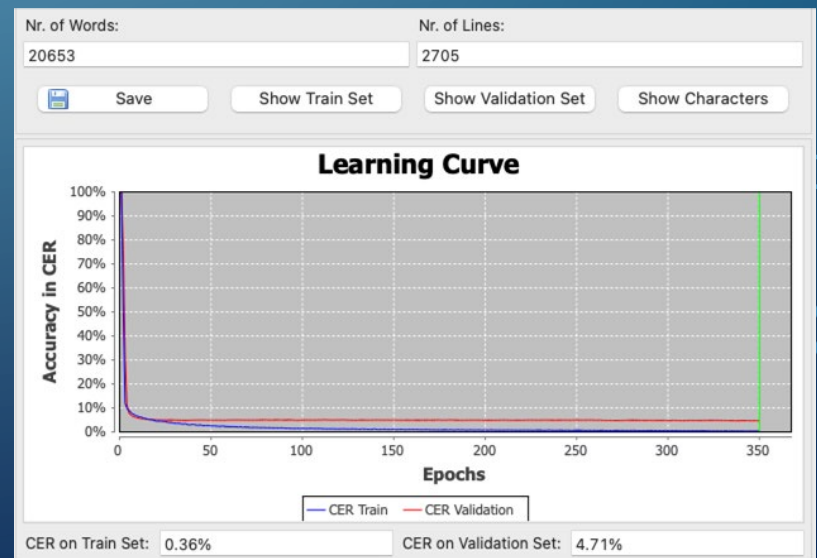
- Regularity of writing hands
 - One registry = one scriptor
- Standardized lexicon : recurent juridicial terminology
- Effective error rate on the above page :
CER : 1,35% **WER : 7,64%**
- The quality of the writing varies greatly from one notary to another
- High performance model with little training data

Michel Saindon's registry (1768-1780)
BANQ-Rimouski, CN104,S50

Traning set :	112 pages
Validation set :	27 pages
Total :	139 pages



par devant le notaire royal de la paroisse de la marine, et
autres lieux et résidant à Halifax sous signé fut present pierre
douce habitant résidant en la paroisse du ditte lequel a ce jourd'hey
reconnu et confesse avoir vendue cede quitte transporte et delaisse
et par ces presentes vend cede quitte transporte et delaisse des maintenant
et à toujours avec garantie de tout trouble ditte hypothèque de uaire substitu
tion et autres empchements généralement quelconques à francois lambert
ce acceptant pour luy francois lambert son pere à ce present ac
querieur pour son dit fils ses heir et ayant cause à l'avenir de avoir
quatre arpents de terre desfront sur quarante de profondeur six e



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