Canadian Research
Knowledge Network

Réseau canadien
de documentation
pour la recherche

Héritage Collection Update: RG-10 ICR Pilot Project

Jason Friedman, Senior Manager, Heritage Services

crkn-rcdr.ca

Canadiana Collections: Content

Canadiana

Digitized monographs, serials, government publications, and maps

Héritage

Digitized microfilm of archival primary-source materials from Library and Archives Canada
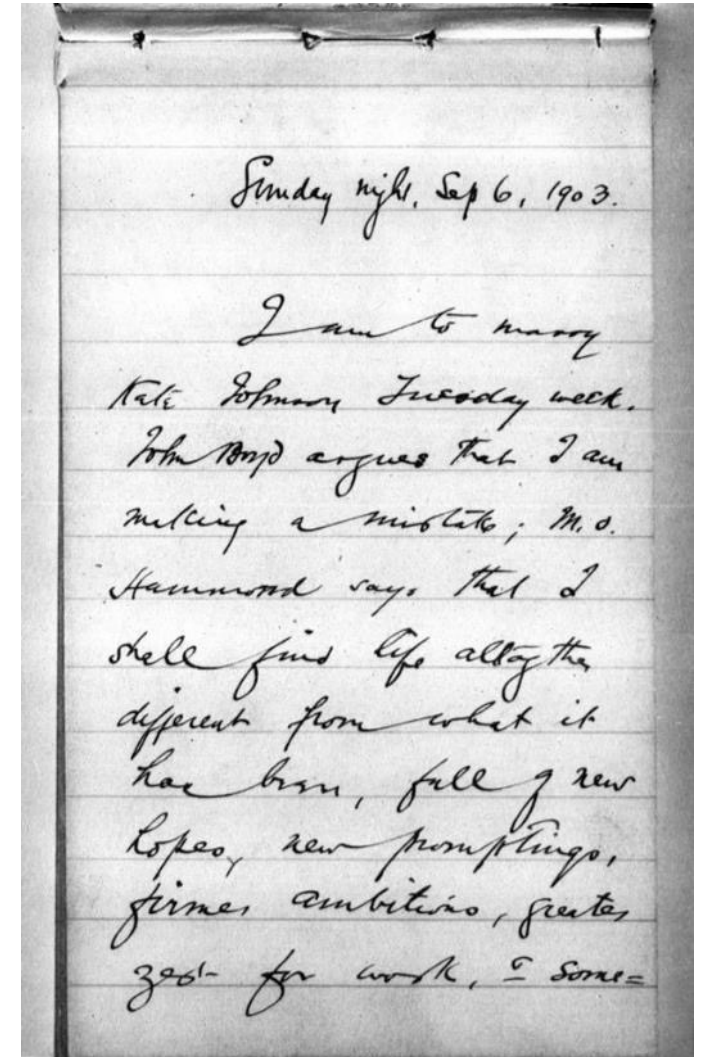

Canadiana

Héritage Collection

- Partnership with Library and Archives Canada (LAC)

- "C", "H" and "T" series of LAC microfilm

- 42+ million pages; 25,000+ reels, 900 collections

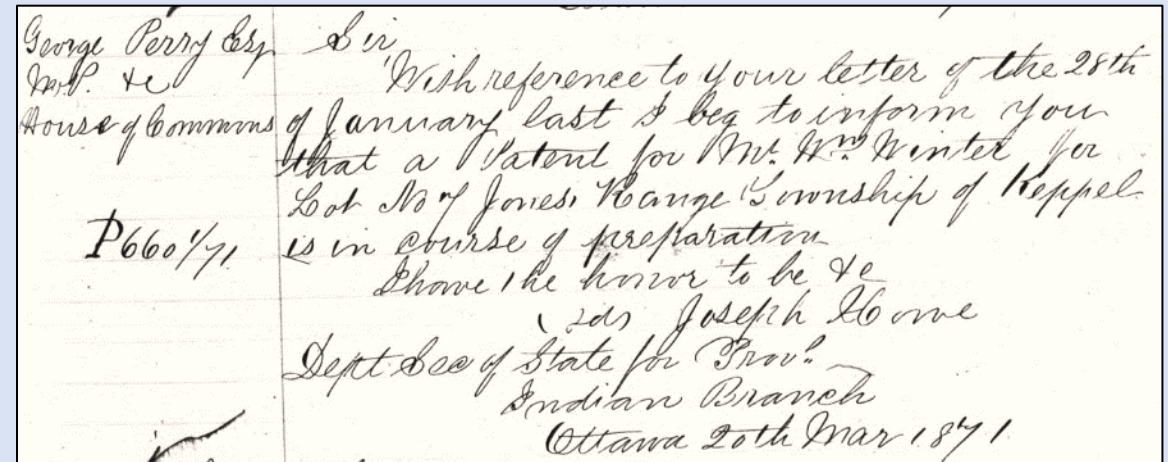- Still going: 482,731 pages added over the past year

Canadiana
Héritage

ICR Pilot Project

Discovery challenges

Digitized microfilm
- Limited finding aids/indexes
- Limited metadata
- Limited search capabilities
- Lots of handwriting means no OCR, no full text search

Solution: Intelligent Character Recognition (ICR)

Uses AI to perform character recognition

Can be used on handwritten materials

Environmental scan conducted in Fall 2022

Solution: Transkribus

Produced by READ-COOP

Used by researchers/historians working with parts of the Héritage collection

Initial sample sent for testing

1-2% error rate for typewritten materials

5-7% error rate for handwritten materials

Next step: pilot project

Implementing: ICR Pilot Project

Partnership with LAC

Full scope: 2 million images

Began in March 2023

Initial batch of 600K+

Additional 500K identified

~1 million images to be identified

Canadiana Héritage

Library and Archives Canada — Bibliothèque et Archives Canada

Why RG-10?

High priority collection ✓

Heavily used ✓

Lots of handwritten content ✓

Supports claims researchers ✓

Variety of document types/quality ✓
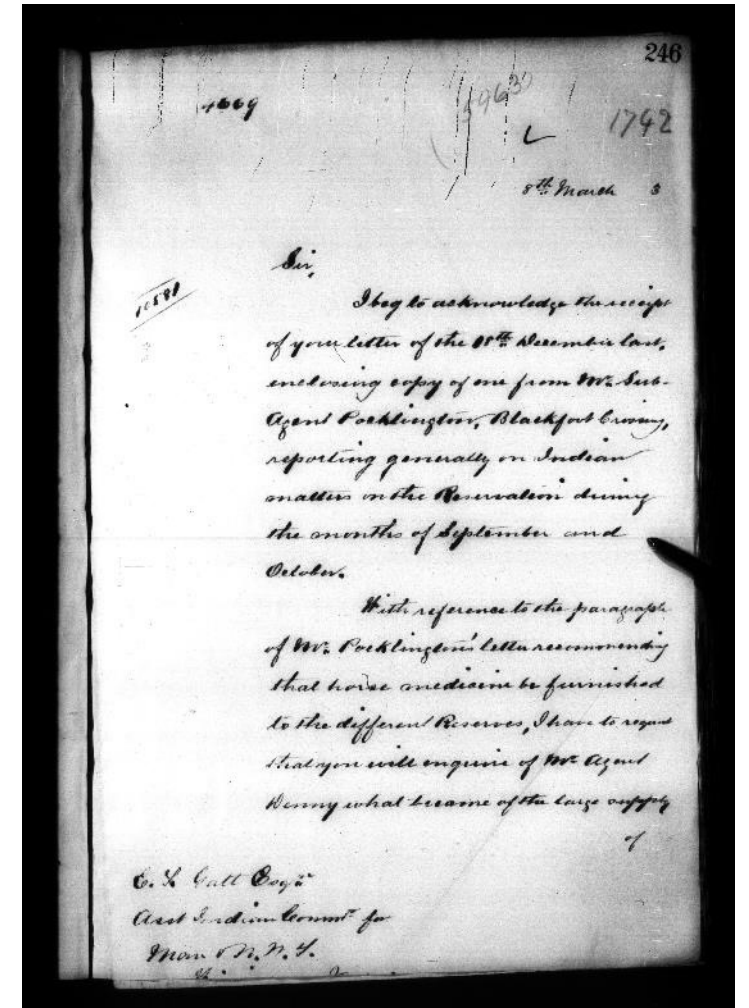
First Batch: Departmental Letterbooks

Handwritten: consistent writing and format

Used for a variety of research topics

Limited metadata

Time period, 1871-1929 (low risk of sensitive content)

Some existing transcription data for model training

be permitted one year to search
for muerals, and may then
select five Square miles, if
satisfied to commence
working. No rent will be
demanded for the land, but
a royalty must be paid on
the clear net returns, not to
(ceed five percent. The
Leose to continue for twenty
one years subject to renewal.

I am
Sir
Your obedient Servant
Joseph Howe
PUBLIC ARCHIVES
ARCHIVES DUBLIQUES
CANADA

Pook LOPY
a 523

Initial results: Out of the box

95

beword, clearing 4 acres on
the part of Section No. 8 in
the Township of Pennefather
pwichased for settlement
by Mr. R. Bannon in the
employ of the Messrs Green
and the erection of the
Codes of two houses and
timber prepared for building
a Dock on wharf, the
renewal will be made
upon the express under-
standing that no transfer

November 10.
PUBLIC ARCHIVES
ARCHIVES DUBLIQUES
CAANADA

Pook LOPY
94

beyond clearnd to acres on
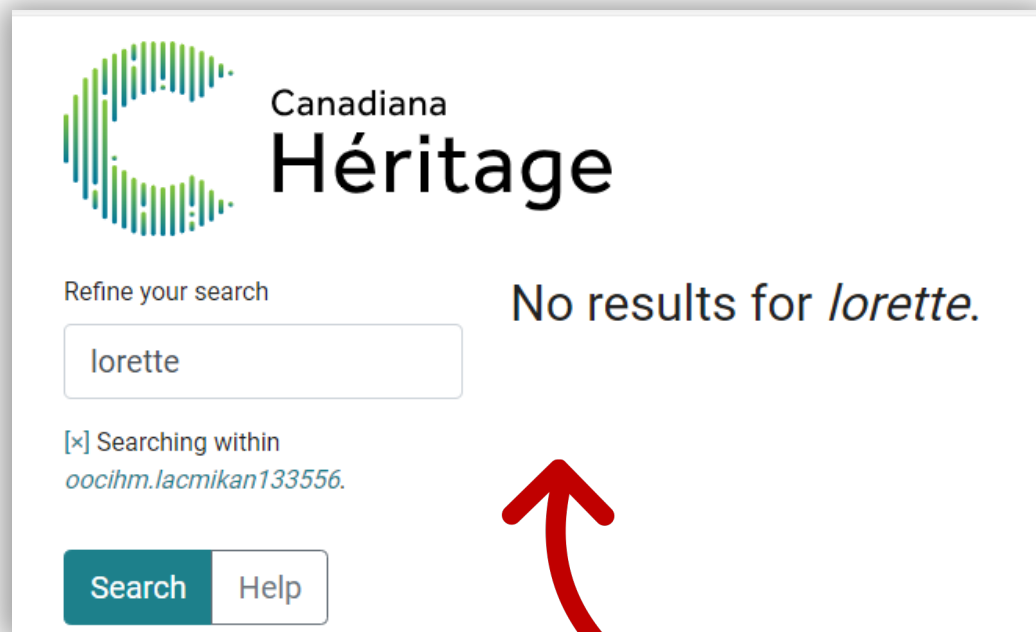the part of Section of in
the Township of Pether
purchased for Settlement
by Mr R. Banon in the
employ of the Messrs Creen
and the creation of the
boodies of to hoses and
timber prepared for bll
a Dock or Wharf the
renewal will be made
upon the expres under.
standing that no transfer

I.
PUBLIC ARCHIVES
ARCHIVES PUBLIQUES
CANADA

POOR COPY
8

Initial results: Comparison of two models

Before and after

CRKN | RCDR
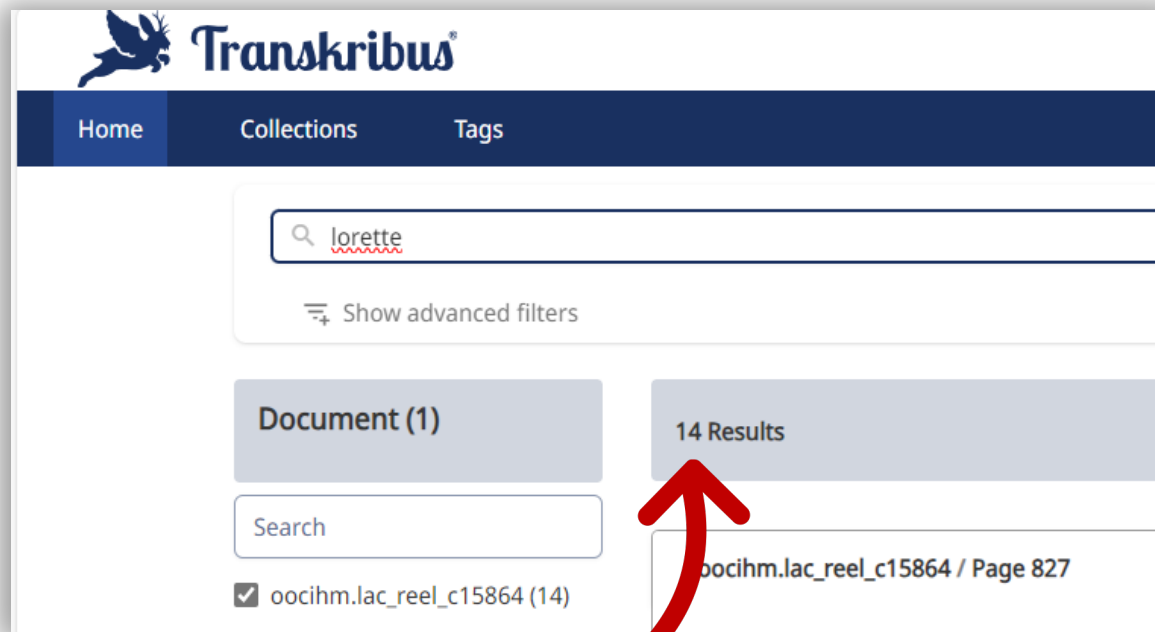
ICR Pilot Project: Next Steps

Training	Complete initial model training for Letterbooks ✓
Receive	Receive initial results for verification/quality control ✓
Review	In partnership with LAC, perform block review
Load	Load ALTO XML/PDFs onto Héritage platform
Transmit	Transmit additional image files to READ-COOP
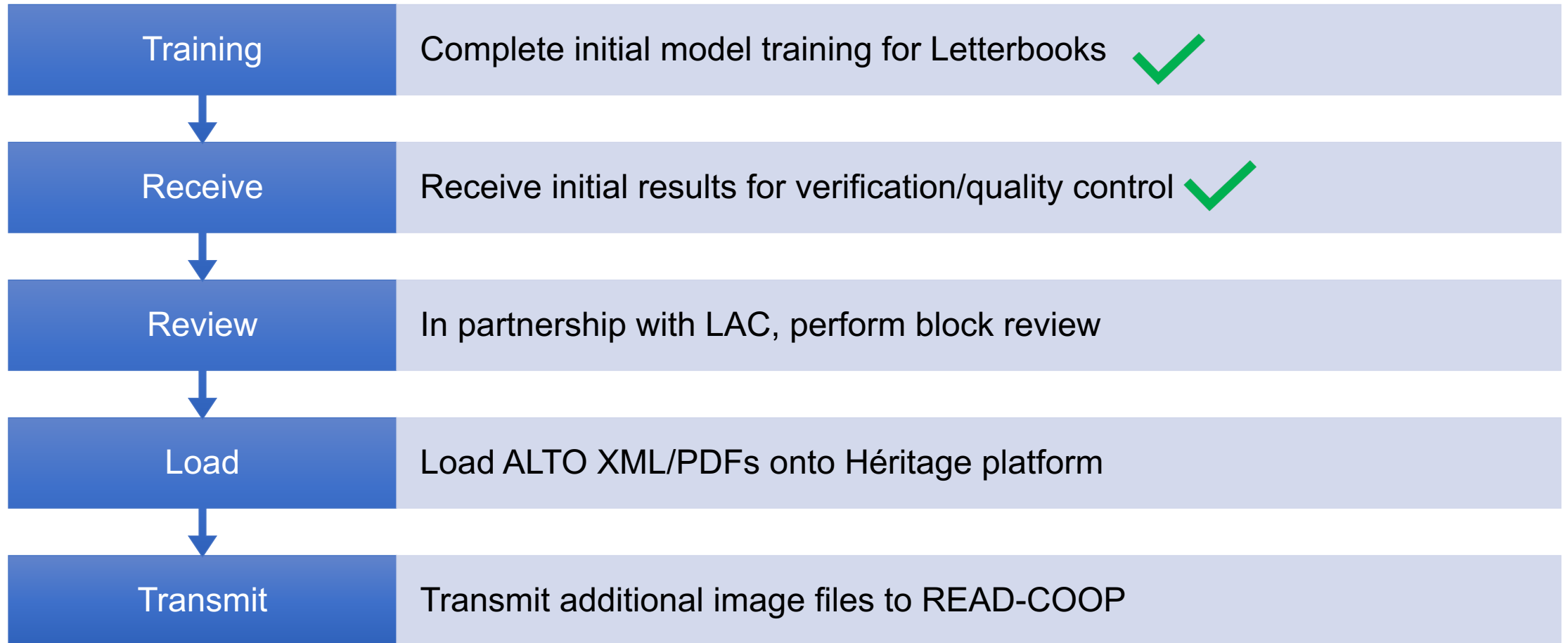
We need your help!

We still have approximately 1 million more images for the pilot, so suggestions are welcome!

Identified series include:

- Annuity paylists
- Census records
- Black series
- Thousand series
- Red series

And if you have transcriptions or are willing to correct transcriptions

Email: jfriedman@crkn.ca

Beyond ICR

PDiiif: Custom PDF downloads from Héritage

- pdiiif.jbaiter.de
- Instructions: crkn-rcdr.ca/en/ navigating-collections
- Images only, no OCR

Coming soon… improved search results display

Department of Indian Affairs : treaties, surrenders and agreements : T-9938

Image 169 | |< | < | > | >| | ↺ | ↻ | ⊖ | ⊕   treaty   Search Clear

Images containing "treaty" in this reel :   Show less ∧

Image 5, Image 6, Image 7, Image 8, Image 9, Image 10, Image 11, Image 12, Image 13, Image 14, Image 15, Image 16, Image 17, Image 18, Image 19, Image 20, Image 21, Image 22, Image 23, Image 24, Image 25, Image 26, Image 27, Image 28, Image 29, Image 44, *Image 169*, Image 286, Image 321, Image 503, Image 722, Image 725

Previous result | Next result   27/32 images containing "treaty" in this reel

Canadian Research
Knowledge Network

Réseau canadien
de documentation
pour la recherche

Thank you! Questions?

jfriedman@crkn.ca

crkn-rcdr.ca