NEW BRUNSWICK HISTORICAL NEWSPAPERS PROJECT

Leah Grandy (UNB Archives & Special Collections)

Mike Meade (Manager, Digital Imaging)

Jeremy McDermott (Senior Web Developer)

Jacob Sanford (Senior Technical Operations Manager)

Jeff Carter (Manager, Systems Group)





NEWSPAPERS: SOMETHING FOR EVERYONE

Students

Academics and researchers

Public



UNB has the largest New
 Brunswick newspaper collection.

 UNB has been digitally preserving newspapers since 2010 from print and microfilm.





JUST RECEIVED,

By the Brig Account. from Lie purcon.
A FEW TONS OF
A FEW TONS OF
For Sale at the Shop of the Sabstriber.
October 10 Company of the Sabstriber.

with a fit.

Atheriting.—Fer an Advertisement of 19 lines and

Atheriting and 18 s.d. for each according

ander, 38.

Atheritine and 18 s.d. for each according

ander, 38.

Atheritinements above 12 levels and

Atheritinements above 12 levels and

Atheritinements above 12 levels and

Atheritinements without contingity.

PREASTANT in and the print contingity.

PREASTANT in and on the most liberal term, at the

thorsest day of the most liberal term.—A1 Or
thorsest day of the most liberal term.—A1 Or
flower and the second of the contingity of the second of the

Roller, must come Part Paul, or they will not be attended. G. BALL, Begs leave to inform his friends and the Public, that he has received by the Warn, from

London, a large supply of
PHRYULEUM, FRANCE:
BANDE: &c.
all of which he oilers for Sais on reasonable terms.

November 30,1824. JUST PUBLISHED,

Miniature Mimanach.

Rises | Sets. R. O. FUL SEA

In 32 pages 8ro. (small type) price One Shilling.
And for Sale at the Star Office, at the Book-Stores of
Messrs. Repealsh and at Milas, and ut
Bronne's Music Salcour-

FULL-LENGTH PORTRAIT

AZER HOYT,
WM. H. RULOFSON.
Segston, March 21, 1825.

All for low Solds.

All Persons having any just demands against the ExA LI Persons having any just demands against the ExLike of the last ileast Waxerers, of Kingston,
Like of the last ileast waxerers of Kingston,
Like of the last ileast the All Like
Like of the last ileast the Like
Like of the Like of Like
Like of the last ileast the Like
Like of the Like of the Like of the Like
Like of the Like of the Like of the Like
Like of the L

A LL Persons having any legal demands against LL Persons having any legal demands against the Estate of Grones Pouss, late of Swans will person the same duly attested, within Six Months from this date; and all those indented to said 15-fate are requested to make immediate payment to the control of the cont

Carleton, March 29.

The Daily Gle.

May 18, 1963 - July 16, 15

(Fredericton, NB)

turns of Maple Island, and scalland to be one of the seal fab-ing stations is the Proteine.

ALSO—despite SSSO areas, where or less, of Wood Land, Island and SSSO areas, the SSSO areas of which are stituated on the Proteiner.

The which are the State of the SSSO areas of which are islanded on the System on the SSSO areas of which are thousand Exement, and will be said on very minkeyed Merchanton and Exement, and will be not on very minkeyed Merchanton and Exement, and will be said on very minkeyed Merchanton and Exement, and will be not on very minkeyed Landers, Day, alkin, S. S. Micke's, Event Point; and Ig. S. Michael, One, South Marinet Wheef.

Watch & Clockmaking, &c. &c.

Carleton, March 29.

Administrator.

A. I., Personshaving any legal demands against the Batate of Rousany Baxtran, late of Dig-thy, Nora-Scotia, Mariner, deceased, are requestly, Nora-Scotia, Mariner, deceased, are requestly and to present the same, duly attented, within Size at the present the same, duly attented, within Size at the present the same, duly attented, within Size at the present the same, duly attented, within Size at the position of the present the same of the present the present the same of the present the same of the present the presen



0 OQ



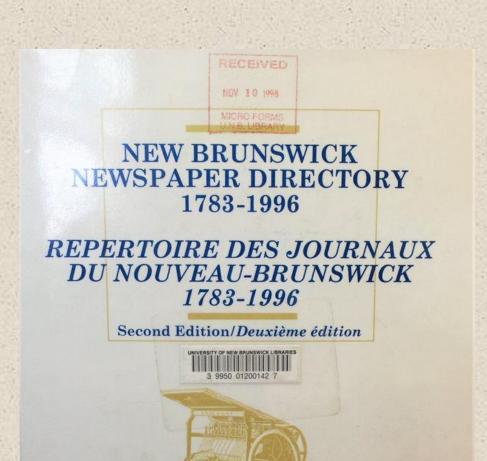




BACKGROUND TO THE PROJECT



• Council of Archives New Brunswick



FREDERICTON
NEW BRUNSWICK
COUNCIL OF HEAD LIBRARIANS
OF NEW BRUNSWICK
1996

FREDERICTON
NOUVEAU-BRUNSWICK
CONSEIL DES DIRECTEURS DE BIBLIOTHEQUE
DU NOUVEAU-BRUNSWICK 1996

Provincial Archives of New Brunswick

Search | Exhibits and Education Tools | Research Tools | About PANB

New Brunswick Newspaper Directory

Foreword Preface Introduction Notes Illustrations Search

This directory contains 697 newspaper listings.

View newspaper

Acadie dimanche, Moncton Co

Or click a link below to view an index:

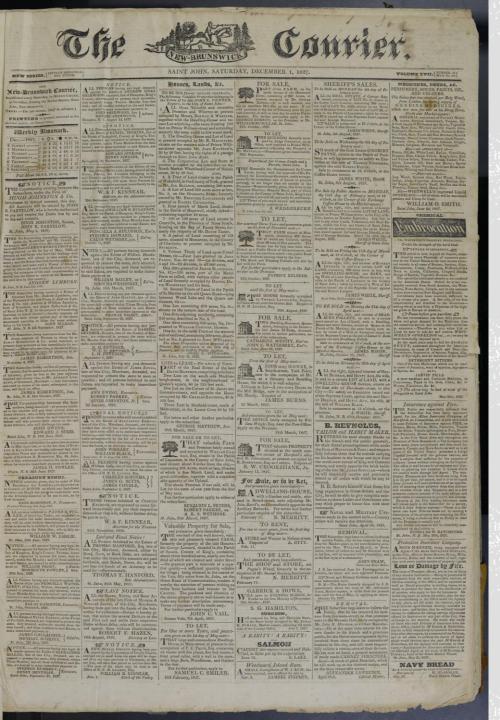
- · View index by Place
- · View index by Publisher
- · View index by Chronological Order

Documents available:

Publishing History

This *PDF file shows the publishing history of the newspapers arranged by place.





PROJECT STATUS

- Currently the most accurate index of New Brunswick newspapers.
- Over 1,050 New Brunswick newspaper titles indexed with linking to free (and pay wall) online newspapers.
- 350 new titles have been discovered and added.
- Approximately 120 locally digitized, searchable titles with digitization is ongoing.
- New partnership with Canadiana/CRKN to share digitized content.

NEWSPAPERS.LIB.UNB.CA



MATERIAL CONDITIONS



LASTERE

1, 1908.

INATE GIRL MPTED SUICIDE

Was Ready to

Will Carry Out Her Soon as She is

one being marin the city, She arning her living as been betrayed the city and to ne decided to kill just as soon as the police station r intention as she

be of unbalanced atly worriedsion, has taken up cing inquiries into nother will be loand made aware licament.

LIQUOR STEALING CHARGE WAS DISMISSED TODAY

SERIOUSLY WOUNDED BY DISCHARGE OF RIFLE

Probably Another Fatality Near Parrsboro -Austin Davis the Victim.

ket Wharf. He was shooting. Durning was some distance | defence. he girl standing on behind Davis when he slipped on some wharf. The officer ice, the gun he was carrying being ac-

JAPANESE BANKER

secretary of the Depositors of Chiyoda Bank Request and a half in the afternoon witness the President to Commit Hari-Kari.

The Three Men, Amburg, Fulton and Campbell, Found Not Guilty of Taking Gin From

rosy cheeked girl, 19 AMHERST, N. S., Aprl 1.—News has The charge of taeft preferred against a captured last night Just been received here this morning George Campbell, Wm. Fulton and res while attempting of what may prove a fatal accident Joseph Amburg, who were arrested on by jumping off the which happened at Five Islands, ten suspicion of stealing gin from the the waters of the miles from Parrsboro yesterday after- steamer Monmouth last Sunday, was noon. Austin Davis may lose his life.

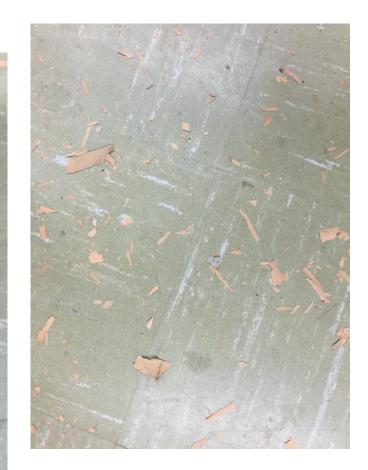
noon and Davis may lose his life.

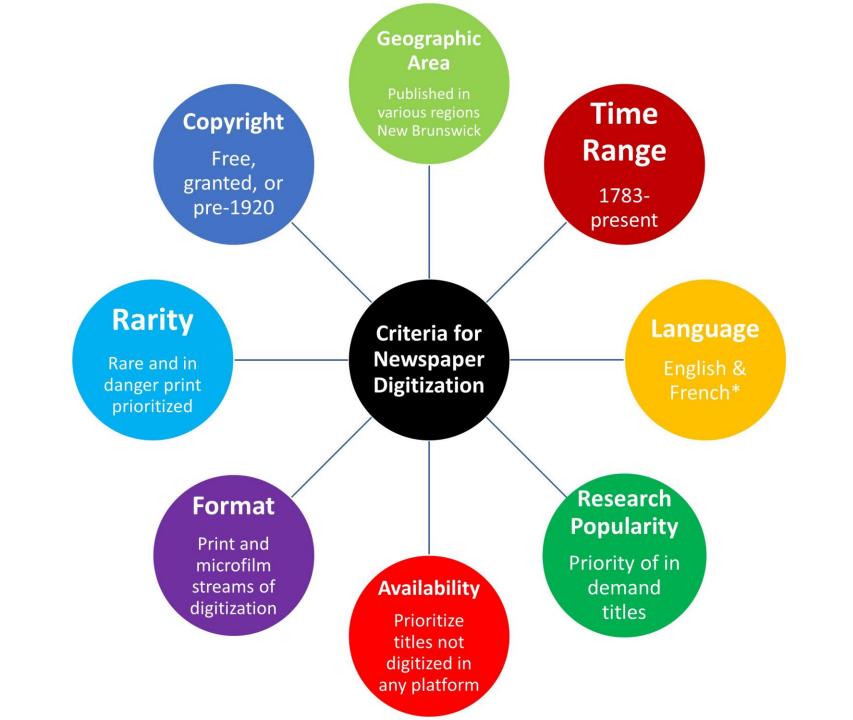
resumed in the police court this morn
nity of 11.30 o'clock

Davis and another young man named ing. F. R. Taylor appeared for the happened to pa- Harold Durning were on the marsh prosecution and J. B. M. Baxter for the

under arrest and cidentally discharged, and the shot was discharged between eleven and andering about and lodging in Davis' back. Davis was twelve o'clock, the officer in charge ctory account of carried home, losing a great quantity thought all the liquor had been disof blood on the way. The bullet has charged when witness told him to be she came from been removed, and reports later say sure and on a search being made anwidowed mother, the patient is doing well as can be ex- other case of gin was found behind a couple of cases of furniture. Amburg did not say that he put the case of gin behind the crates, or that he put the bottle in witnesses' pocket. When witness returned to work at one o'clock Sunday afternoon the refrigerator doors were open and he saw no one around there. The crew go to dinner at one o'clock; some of them were working about the ship's hold all morning but he did not see any of them about the refrigerator. During the first hour was working in the lower hold, and then he returned to the refrigerator to Alex. McGinnis saw that the gin was

discharged in the morning and was





FUN FINDS

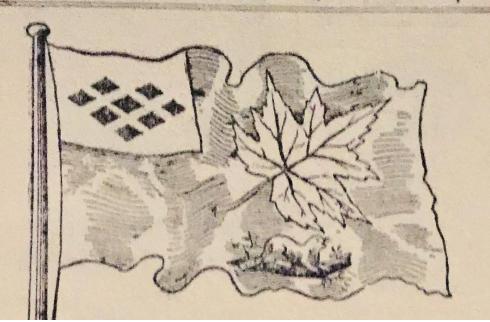


The only Paper in the World, so far as we know, Printed by a Men with the I was

von 1 80 XII

FREDERICTON, N. B. MAY 1901

15 Crs. Pen Year, I Char a Copy



BARKER HOUSE.

F. B. Coleman, Prop

QUEEN SI, FREDERICTON

Thoroughly upsto-date in every particular and fully abreated the Times.

Hoa tquarters for Tourists and Commercial Travellers.

The Frederic



DIGITIZATION PROCESS



DIGITIZING FROM PRINT

- Initially, Atiz Bookdrive Pro
 - two Canon 5D Mark II cameras.
- 2012 purchased a second Atiz unit
 - Nikon D800E cameras to be able to capture newspaper images at 300PPI
- 2021 UNBSJ purchased a Zeutschel Omniscan OS 12000 (A-1 paper size) to assist with newspaper digitization



MACDONALD

SAYS IDEAS

ARE VAGUE

BRITISH PAPERS FEATURE SCANDAL

SCANDALMAY WILL PLAY IN FOREIGN FIELD

London Paper Says It Has Altered Whole Election in the United

ish Haste to Clear Cargoes Before the Blow February 1st - Little Hope of

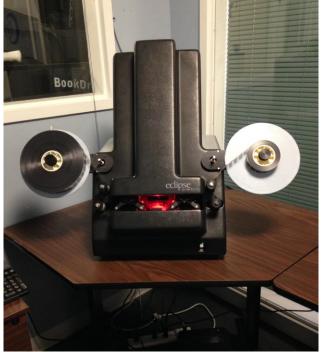
R. B. HANSON, K.C., M.P., DEMANDS RETRACTION FROM GLEANER OR

Dock Laborers in the Old Country Threaten To Tie up Every Port

SUIT FOR LIBEL WILL BEGIN

Editorial Reference on Thursday Charging Him Followed by That Action—P. J. Hughes, K.C., Retained by Mr. Hanson—J. B. M. Baxter,







DIGITIZING FROM MICROFILM

- 2014 Nextscan Eclipse microfilm scanner part of BNI contract
- completed 3 million pages of Brunswick
 News International material in just over 4 years



QUALITY CONTROL

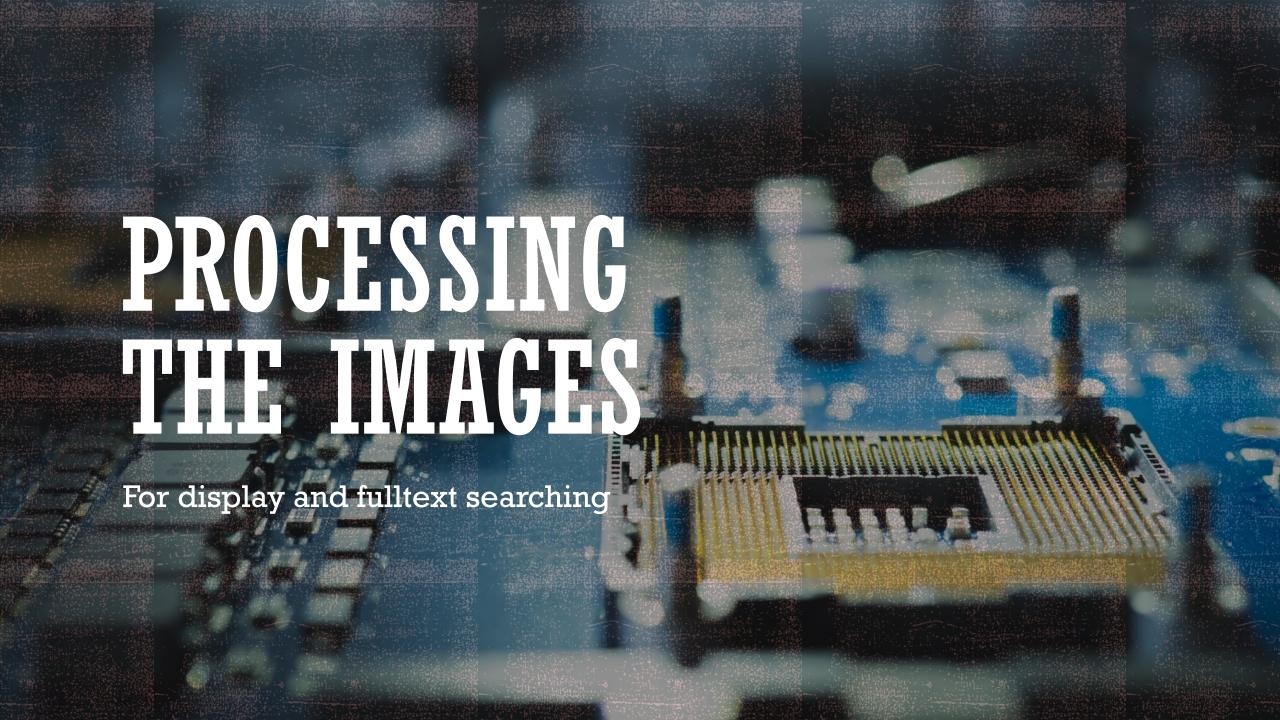
- Image quality enhancements
 - Cropping, rotating and scaling
 - Surrogate generation
 - Can be a bottleneck
- Directory/File structure for ingest
 - Parent folder for titles
 - Unique folder for each issue
 - Images named with page numbers



IMAGE METADATA

```
define('ISSUE_DATE', mktime(0, 0, 0, 20, 7, 1878));
define('ISSUE TITLE','The Penny Dip');
define('ISSUE VOLUME', '1');
define('ISSUE ISSUE', '28');
define('ISSUE EDITION', '');
define('ISSUE SUPPLEMENT TITLE', '');
define('MISSING PAGES', '');
define('ISSUE ERRATA','');
define('ISSUE CONTENT MODEL PID', 'newspaperIssueCModel');
define('PAGE_CONTENT_MODEL_PID', 'newspaperPageCModel');
define('ISSUE LANGUAGE', 'eng');
define('SOURCE MEDIA', 'print');
// ** Save this file as metadata.php
```

- Issue level metadata:
 - Publication date
 - Title
 - Volume/Issue number
 - Alternative titles (supplement or section)
 - Errata



POST-SCAN INGEST PROCESS

Custom Processing Script

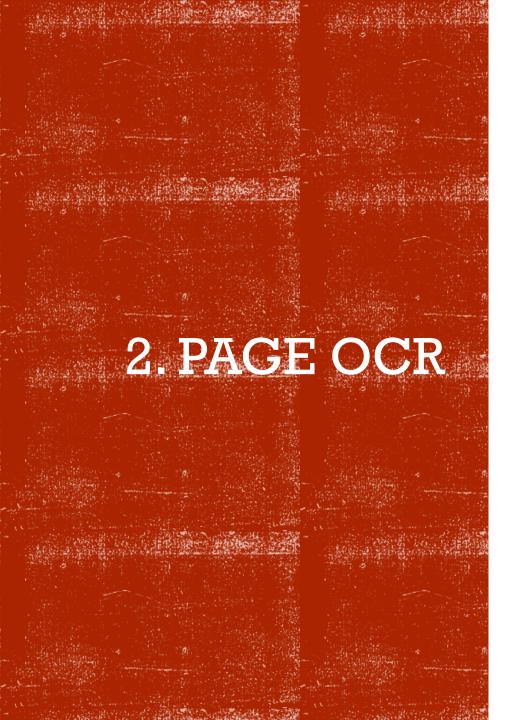
- PHP / Robo Framework
- 'Systems Toolkit'
 - https://github.com/unb-libraries/systemstoolkit

Steps:

- Issue Discovery
- Page OCR
- Issue Creation, Page Ingest
- DZI Tile Generation
- Issue Audit



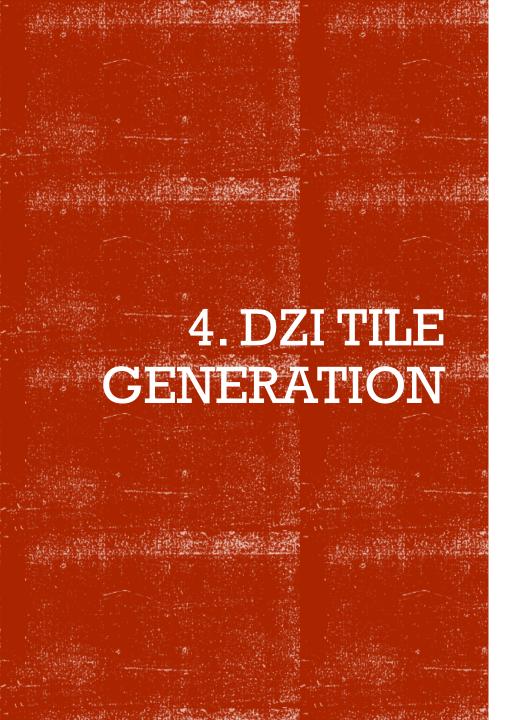
- Search path for all subfolders that contain a metadata file
- Search for all images in that folder and consider them to be each a page from that issue.
- Check if each image has an associated tesseract .hocr file generated. If not, the image is queued for OCR with Tesseract in a later step.
 - This is done to optimize core usage some issues have a small number of pages
- Convert the metadata to JSON format for assertion to the Drupal REST API.



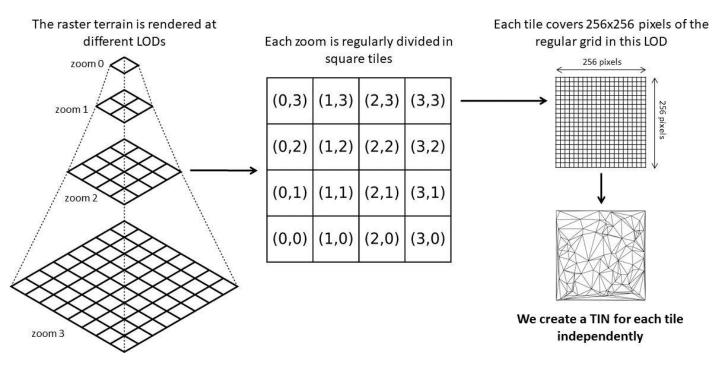
- Operates on all queued images to generate OCR for each page.
- No issue creation is done in the REST API until all issues have OCR complete.
- Tesseract 5.x, run in Docker
 - https://github.com/unb-libraries/docker-tesseract



- Once OCR is complete, the issue is then created.
- Iterate over folders and creates each an issue by POSTing to the Drupal REST API
- After the issue is created, create the pages and attached to the issue (via the REST API)
- Add OCR output to page.
- Writes a marker file inside the subfolder to indicate the issue has been ingested.

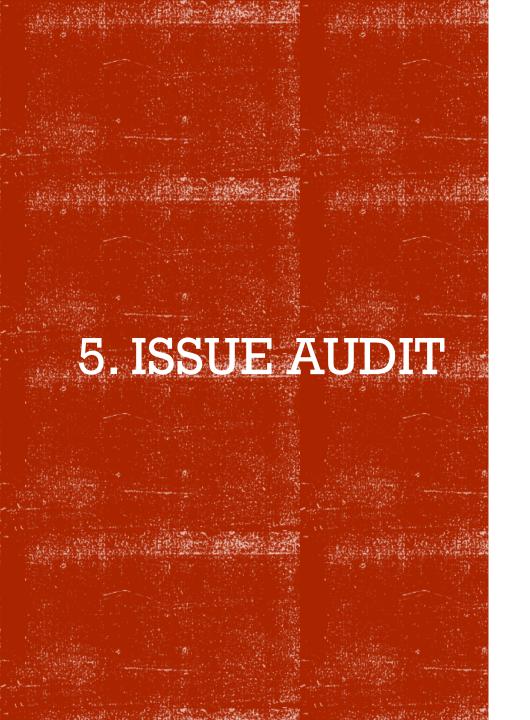


- Page viewer, OpenSeadragon, can leverage a set of 'tiles' to minimize the bandwidth for zooming.
- These tiles are then generated.

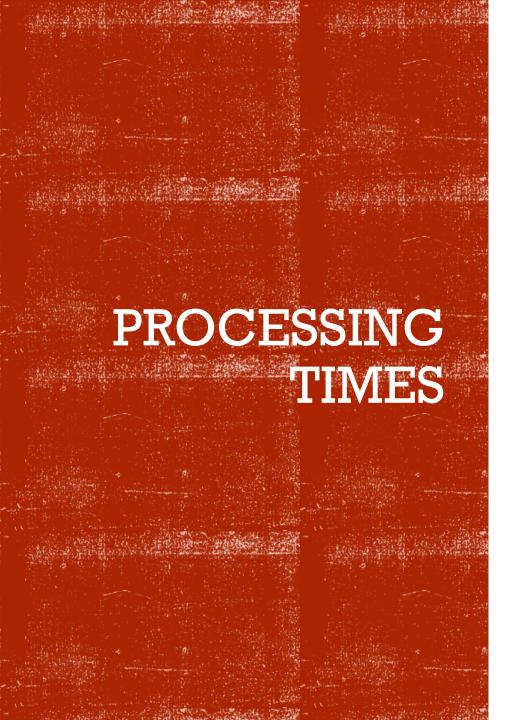


Remote Sens. 2020, 12(3), 437; https://doi.org/10.3390/rs12030437





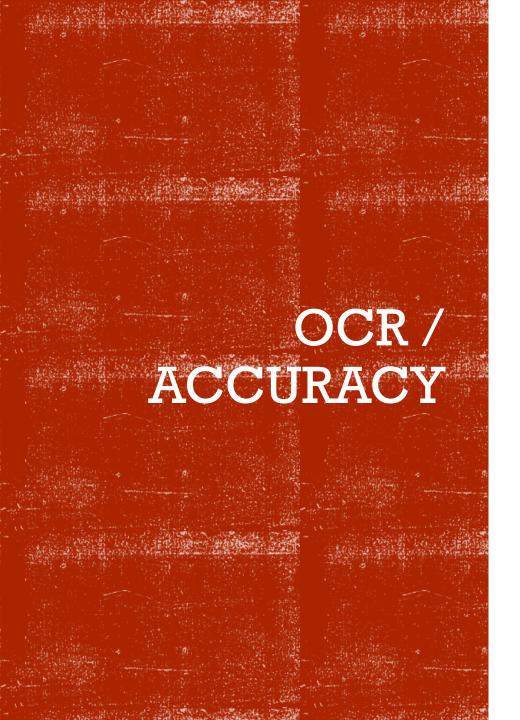
- Audits all issues in the import
 - Verifies metadata
 - Hashes page files and compares to local
 - Checks page numbering



- Processor: Core(TM) i9-12900 (16 Cores)
- Memory: 64GB
- Multi-threaded processes set to use 12 threads.

- Average total processing/ingest/audit time:
 - 30-40s per page

 Can scale horizontally to accommodate volume.



	Delta			Uncropped					
Confidence / Word	Sum Confidence	Word Improvement	Word Count	Confidence / Word	Sum Confidence	Word Count	Confidence / Word	Sum Confidence	Word Count
1.036539299	50380	21.70%	474	85.79716117	187381	2184	89.45109105	237761	2658
0.6851353601	110541	132.18%	1302	83.69746193	82442	985	84.38259729	192983	2287
5.194449735	249046	274.80%	2682	85.77356557	83715	976	90.96801531	332761	3658
46.34609768	73681	16.93%	205	39.2923204	47583	1211	85.63841808	121264	1416
7.880933226	166193	413.64%	2002	73.22727273	35442	484	81.10820595	201635	2486
16.96516512	7562	-15.83%	-420	72.193366	191529	2653	89.15853112	199091	2233
19.1428052	57565	1.19%	34	65.38676934	186810	2857	84.52957454	244375	2891
0.9934245682	35853	13.90%	385	84.98375451	235405	2770	85.97717908	271258	3155
-0.7162238866	100701	155.04%	1107	92.14565826	65792	714	91.42943438	166493	1821
0.09559061486	-2352	-1.05%	-28	93.0060015	247954	2666	93.10159212	245602	2638
1.085431453	-6229	-4.85%	-88	92.09801762	167250	1816	93.18344907	161021	1728
0.173305749	-1436	-0.71%	-21	92.51171477	272447	2945	92.68502052	271011	2924
3.399722158	155127	183.24%	1880	77.25925926	79268	1026	80.65898142	234395	2906
0.9783726319	-40210	-17.54%	-468	90.517991	241502	2668	91.49636364	201292	2200
0.6509815054	28546	19.11%	322	84.59465875	142542	1685	85.24564026	171088	2007
0.7539581905	-3424	-2.31%	-59	89.91738449	229649	2554	90.67134269	226225	2495
0.8304182691	-2262	-1.78%	-54	87.70029674	265995	3033	88.53071501	263733	2979
0.6002796494	13736	3.69%	127	91.26929774	314514	3446	91.86957739	328250	3573
-1.339766911	-323	0.33%	1	89.64820847	27522	307	88.30844156	27199	308
-0.5378180365	-74872	-41.91%	-842	88.17620707	177146	2009	87.63838903	102274	1167
9.070068499	-2069	-11.19%	-350	77.92777245	243836	3129	86.99784095	241767	2779
-0.003966940599	-43637	-19.57%	-504	86.56504854	222905	2575	86.5610816	179268	2071
-1.107757758	-46490	-22.19%	-518	85.86546701	200410	2334	84.75770925	153920	1816
-1.553959378	-97108	-48.25%	-1128	84.42172797	197378	2338	82.8677686	100270	1210
0.9008337778	-6239	-4.99%	-91	85.70597916	156242	1823	86.60681293	150003	1732
1.198319548	-27501	-19.93%	-376	77.95654478	147104	1887	79.15486433	119603	1511
4.685266106	138230	114.21%	1567	79.42565598	108972	1372	84.11092208	247202	2939
1.710086353	85498	38.77%	961	82.846309	205376	2479	84.55639535	290874	3440
-0.3389376291	-1586	-0.82%	-12	91.04223433	133650	1468	90.7032967	132064	1456
1.085037259	149839	REMOVED	1651	89.46774194	27735	310	90.55277919	177574	1961
-1.139011187	883	4.29%	14	90.73312883	29579	326	89.59411765	30462	340
0.4968937136	2499	0.36%	11	90.0843254	272415	3024	90.58121911	274914	3035
-0.2088206353	1998	1.01%	28	92.24666426	255800	2773	92.03784363	257798	2801
22.23458073	178318	REMOVED	1918	70.66666667	424	6	92.9012474	178742	1924
-1.935330655	-60888	-35.36%	-656	89.27978437	165614	1855	87.34445371	104726	1199
0.02192022365	512	0.18%	5	90.14659498	251509	2790	90.16851521	252021	2795
3.87%		33.71%	307.25	Mean Extra Words					

Slight improvement if images are cropped to page content, but can be very labour intensive.





New Brunswick Historical Newspapers Project

New Brunswick Historical Newspapers Project provides researchers with unified access to UNB Libraries' current and historical newspaper collections in all formats, from New Brunswick and across the world. Search and discover print, microform, and selected digital newspaper titles available from UNB Libraries.

Fulltext searching is available for our growing list of digitized titles. Learn more about the New Brunswick Historical Newspapers Project or contact us with questions

For more worldwide digital newspaper content, consult UNB Libraries licensed electronic Newspaper collections &. UNB/STU login required.







NBHNP Partnerships







WEBSITE

- Drupal 9
- Custom entities
- Solr search
- Custom UNB Libraries Bootstrap theme
- newspapers.lib.unb.ca
 - https://github.com/unblibraries/newspapers.lib.unb.ca

