


DIGITIZING YOUR LOCAL NEWSPAPERS

A Step by Step Approach

Tina Beard
tnab@tamarackgenealogy.com

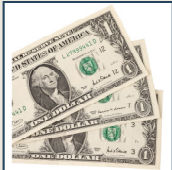
Before You Begin

- How many volumes of newspapers will be digitized?
 - How many pages per:
 - Issue
 - Month
 - Year
- Who will conduct the work?
- Who is your audience?



In-House Scanning

- Assess needs
- Create budget
- Buy equipment
- Test-run to set baseline standards and timeline



Assessing Needs

- How many items/pages will be scanned?
 - An exact count of pages, *paper size, binding*
- Who will be doing the work?
 - Staff or volunteers
 - When estimating staff/volunteer time account for holidays, vacation and sick time.
- Who is your audience?
 - Consider how and where your users will access the pages.

Budgeting

- Where will money come from to support project?
 - Is the money sustainable or grant generated?
 - Are there additional donors/funds/grants for future work?
- Over budget by 10%
 - If grant funded, you never want to give money back!
- Don't just buy online. Contact companies for best deals.

Assessing Needs: Equipment & Software

- What equipment/software do you have? What will you need?
 - Pick scanner large enough to handle your largest item.
 - www.pcmag.com/article2/0,2817,2362752,00.asp
 - OCR software (e.g. Adobe Acrobat Pro or ABBYY FineReader)
 - Scanning software (e.g. Vuescan or Readiris Pro)
- Preservation storage (e.g. physical storage files/external drives)

Assessing Needs: Maintenance and IT

- Who will be maintaining equipment and training users?
 - Do you have an IT department that can support the equipment?
 - Will the added work be covered under current contract?
- Consider up-front and long-term cost factors:
 - *Equipment costs*
 - *Digitization costs (staff and equipment)*
 - *Set-up fees*
 - *Image maintenance and hosting costs (annual increases)*

Copyright Oversimplified



- Copyright guidelines state publications published before 1923 are public domain
- Publications between 1923-1977 MAY be in public domain if items were not registered nor statement of copyright given
- Publications since 1978 are protected for 95 years from date of publication

Copyright Websites



- www.copyright.gov/eco/help-newspapers.html#newspaper
- <https://fairuse.stanford.edu/overview/copyright-research/searching-records>
- Cornell University Copyright Center
<https://copyright.cornell.edu/resources>
- www.legalgenealogist.com/2012/03/19/copyright-the-newspaper-article

Outsourcing

- Interview a minimum of three vendors.
- Ask the vendor for references and examples.
- Look for additional collections (not mentioned) that are available online.
- Ask vendor to provide a free sample using your newspapers.

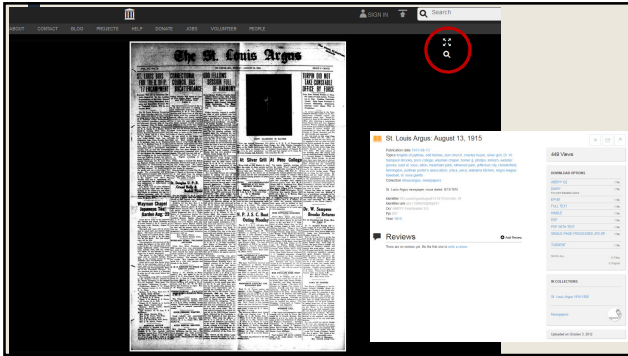


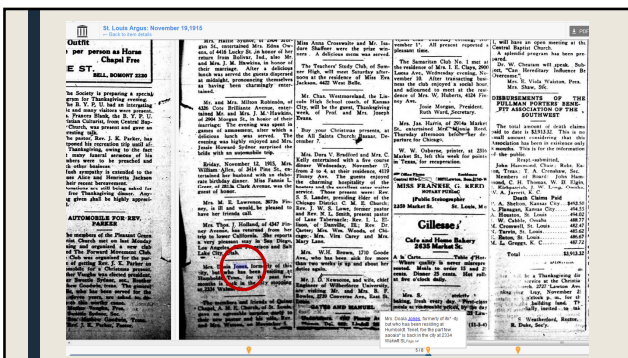
Digitization Vendors

- Internet Archive www.archive.org
- Advantage Microfilm Company www.advantage-companies.com/preservation/markets/libraries-historical-societies
- Veridian www.veridiansoftware.com/collections
- Crowley Company www.thecrowleycompany.com

Internet Archive www.archive.org

- Scans materials as JPG2000 then converts to PDF to run optical character recognition software to create searchable text.
- Uses MARC records or Dublin Core for metadata can connect to library's catalog with Z39.50 but requires .cvs file
- Fees: \$3.00 set-up per ID (issue) then .30 per page for loose pages (11x15) \$2.40 per page for those larger than 11x15. Fees: \$3.00 set-up per ID (issue) then .12 per page for microfilm
 - Prices includes hosting the material on their website





Advantage Preservation Company

- Digitizes files at 300 dpi (B&W) PDFs for \$95.00 per microfilm reel
- Hosts scanned materials and creates a searchable website \$500.00
 - There is no way to search pages while browsing by issue
- Provides an external drive containing scanned images \$200.00
 - Files saved to drive are not in numerical/chronological order
 - For more information visit: <https://bit.ly/219FCxD>

THE ENTERPRISE, WEDNESDAY, OCTOBER 24, 1923 Page 1

GENTRY REALTY

623 LOCKPORT ST.

READY TO SELL? CALL GENTRY. WE'VE GOT A SIGN FOR YOU.

FRUIT TREE OFFERED
 1000-1000-1000-1000
 This beautiful detached house...
 1000-1000-1000-1000

JUST PITCHED SIGN
 1000-1000-1000-1000
 And what chance is there...
 1000-1000-1000-1000

MEMBER'S BENEFIT
 1000-1000-1000-1000
 As an offering to members...
 1000-1000-1000-1000

CALL 436-7552

www.theancestorhunt.com/blog/advantage-preservation-now-hosts-238-free-newspaper-collections-with-over-2000-historic-newspapers

KCs Plan Tootsie Roll Day

Members of the...
 1000-1000-1000-1000
 1000-1000-1000-1000

Class of '27 Enjoys Reunion

The Fairfield High School...
 1000-1000-1000-1000
 1000-1000-1000-1000

In Loving Memory

Mike Evans
 1000-1000-1000-1000

Veridian

- Uses third-party (local) vendors for digitization
 - Scans items as TIF images converts to PDFs
 - Cleans up images to remove de-skewing and cropping
 - Create article segmentation
- Costs:
 - Perpetual licensing of software personal hosting \$10,000
 - Software and Veridian hosting of your digital materials:
 - 1 time set-up fee \$4,800
 - Annual hosting \$400 per month

PAPAKILO
KA LEI MOMI / Ka Lei Momi, Volume 1, Number 14, 6 November 1993

Contents

Page 1

- Hōhōhōhō
- PUA PUAKE NO HOME ALIHI
- KA PUAKE MAKAU A KA PUAKE MAKAU A KA PUAKE MAKAU

Page 2

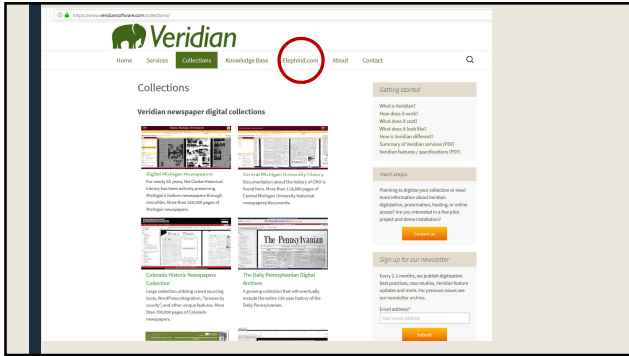
- HE MOUHELE KAU MOUHELE KAU MOUHELE KAU MAUHELE KAU MAUHELE KAU MAUHELE
- KA PUAKE MAKAU A KA PUAKE MAKAU A KA PUAKE MAKAU

Page 3

- KA PUAKE MAKAU A KA PUAKE MAKAU A KA PUAKE MAKAU

KA LEI MOMI

KI MAU MAU OHO ANA



Crowley Company

- Will scan images and send you a sample for free.
- Scans at 300 or 400 DPI grayscale as TIFF or High Res JPG and converts to PDF to run optical character recognition software to create searchable text. Extra fees to create article segmentation.
- Cost for digitization of microfilm reels as low as .03 cents per image capture (page). Cost for hardcopy digitization starts at .50 cents per image.
- **Does not host digital images

Illinois Digital Newspaper Projects

- Illinois Digital Archives www.idaillinois.org
 - ContentDM based metadata
 - Transcription, not segmentation
 - Discoverability through OCLC and DPLA
- Illinois Digital Newspaper Project <http://idnc.library.illinois.edu>
 - Library of Congress Digitization Standards www.loc.gov/ndnp/guidelines/selection.html
 - Veridian software
 - Page level segmentation

Websites For More Information

- www.ifla.org/files/assets/newspapers/2017_iceland/2017-rhyno-en.pdf
- www.michigan.gov/documents/hal/GuidelinesForDigitizingANewspaper_181557_7.pdf
- <https://cdnc.ucr.edu/site/files/BestPracticesforCaliforniaNewspaperDigitization.pdf>
- <http://icon.crl.edu/digitization.php>

Contacts

- Illinois Digital Newspaper Project:
William Schlaac schlaac2@illinois.edu
- Illinois Digital Archives islimg@ilsos.net
- Internet Archive: Jeff Sharpe jsharp@archive.org
- Crowley Company: Meghan McCabe
[Meghan@thecrowleycompany.com](mailto: Meghan@thecrowleycompany.com)
- Advantage Preservation: Michelle Maltas
michelle.maltas@advantage-companies.com



Questions??

OLD-SCHOOL

I ASKED MY GIRLFRIEND IF SHE COULD GET ME A NEWSPAPER
"DON'T BE SILLY," SHE REPLIED, "YOU CAN BORROW MY IPAD."
THAT SPIDER NEVER KNEW WHAT HIT IT

www.tamarackgenealogy.com
tinab@tamarackgenealogy.com
