Digitizing Genealogical Paper Documents

David Devin

devindaca@hotmail.com

You have ever growing stacks of paper in piles, drawers, binders, and boxes relating to your genealogy research. It is becoming impossible to find anything. You know you have to organize it and would like to get rid of most of it before your family starts worrying about your hording tendencies. You also wonder what your heirs will do with all your research material.

Scanning may be the Answer

Reasons to Digitize

- Reduce storage
  Scanning can take a room full of paper and make those thousands of documents available on a space smaller than a paper-back book or package of gum.
- Make computer searchable
  Scanned documents can be searched by computers to various levels of detail. You can find files by filename, by file property details, or by searching file contents (if scanned with optical character recognition)
- Organize
  Scanning gives you the opportunity to organize documents. You can use computer tools to move dozens of documents with a single operation
- Link record for genealogy program
  You can link entries in your genealogy program to the scanned documents the prove the information in the entries. A click in your software can show you the document you need.
- Life Proof
  Disasters Happen! Fire, flood, tornadoes, and pests destroy paper. Worse, our heirs do not know what to do with all the paper and do not have the time or will to figure out what is most important to keep. Your years of research documents end up in a recycling bin. That paper, they just threw away, contained the elusive proof document for your life-long brick wall that you just happened to stumble across. Scanning your research documents make it easy to put the all-important proof documents on-line and inexpensive to distribute your entire set of research notes to family. Finally, computer files are easy to copy to the next file storage technology in just a few minutes.

What to Scan?

- Those all-important documents that prove your family relationships.
- All loose paper – if it is in a file cabinet, 3-ring binder, box, or stacked on a chair
- All donated research – someone gives you boxes or binders because they just "know" you are the family historian.
- When in doubt, SCAN IT!

What Paper Should I Keep?

The short answer is very little! If you have not looked at a document in years, you probably do not need to keep that piece of paper. Not to worry, though; that document is scanned and you can find it on your computer.

You should consider keeping the following documents after you scan the paper:

- Photographs
- Vital records such as birth/death/marriage records
- Wills & Deeds
- Maybe, keep those special heirloom hand-written letters, family stories, and manuscripts

After the above documents are scanned, you can put those few precious important family documents in archival-safe storage for the future family researcher. You will not have to touch them again, because you have it as a scanned document.

Note: Primary proof documents are important! Your research notes not so much. Paper copies of census records of our ancestor's nieces and nephews are not keepers.

Naming Files

There are as many file naming conventions as there are computer users. **The important thing is to use what works for you and be consistent! Put together a cheat sheet on how you want to name various types of documents and keep it next to your keyboard. Use it!**

Using a consistent naming convention saves time trying to think of a name. You do not have to put every tidbit of information in the file name. You need just enough in the file name to make it unique and provide gross sorting and finding aid.

Example: **Smith~John_birtrec_1923-10-24_001.pdf** is the file name for a birth record for John Smith who was born October 23, 1923. The tilde between the surname and given name helps to identify what part is which. The sequence number at the end of the file name shows that there are more than one different birth record available.

Hints:

- Use surname then given name for records about an individual.
- Use book or article titles for such documents
- Determine and document a "slang" language for various type of records
  For example: birth records use different wording depending on the agency. The record might be a "Certificate of Birth," a "Birth Registration," "Record of Live Birth," "Birth Certificate," or "Birth Index." No matter what the official name of the record, it is still a birth record of a person. Using something like "birtrec" in a file name identifies the file, consistently, no matter what the agency's uses on the record. It is unique that you can do a computer file search on that phrase and list all of the records you have identified as a birth record.
- Use a "sequence number" at the end of a file name if you have multiple documents of the same type for the same person or subject
- Library subject, bibliographic, or citation conventions can be the basis for consistent file naming. Saves you from having to think one up.

- If you use dates in your file names, use the format *yyyy-mm-dd* to aid computer sorting.
- Think about using general to specific in the file name convention. Starting with a surname sorts all files with that surname together. Following with the given name and record type organizes the sorted list even further.

File Folder Structures

Just like your paper organization, use a system that works for you!

Hints:

- Have a "root" folder on the storage device. It allows you to make a single folder to backup and transfer your scanned documents. Everything in that root folder are you scanned genealogy documents.
- Try to keep the sub-folder to 2 or three levels. It is easier to browse without a lot of clicking. In addition, it makes the folder path shorter to type or write where needed.
- Consider how Dewey Decimal or library subject labels organizes documents as a possible basis for your folder structure.
- Be consistent!

Examples:

Library environment:

Personal environment

- Genealogy
- Correspondence
- Education
- Forms
- Library
- Pedigree
- Projects
- Subjects
- Surnames

Pedigree
- 00001 Devin, David Carl
- 00002 Devin, Carl Glenn
- 00003 Dawes, Patricia Ann
- 00004 Devin, Ira Holmes
- 00005 Hansen, Oleta Fern
- 00006 Dawes, Walter Arthur
- 00007 Bressler, Gatha Viola
- 00008 Devin, Nathaniel Madison
- 00009 Holmes, Elisa Jane
- 00010 Hansen, Neils Jorgen
- 00012 Dawes, Otis Darwin
- 00013 Tharp, Irene
- 00014 Bressler, George Albert
- 00015 Hitchcock, Ruth Elizabeth

Subjects
- Canada-Waterloo
- France
- USA_ARW
- USA_Census
- USA-CT
- USA-IN-Gibson
- USA-KY
- USA-MO
- USA-OR
- USA-OR_Marion-Pioneer Cemetery
- USA-OR-Marion
- USA-VA
- USA-VA_Library
- USA-VA-Pittsylvania

Surnames
- Arnold
- Beck
- Bennink
- Converse
- Denham
- Devin
  - Devin-AR
  - Devin-IN
  - Devin-Ireland
  - Devin-KY
  - Devin-MO
  - Devin-OR
  - Devin-VA
- Diven
- Hawkins
- Hitchcock

Scanner Settings

Every scanner interface is different. See how different documents work with different settings for your scanner. Document the settings that work the way you want. Remember, after the original sheet of paper is gone, you are relying on the scanned image to reproduce the document, if it is ever needed.

- Resolution: 300 dpi (best for OCR; use 400dpi for really small text or details)
- Image type: Grayscale (unless color is important to the content)
- Brightness: 10-20% above normal (watch for text fading) (use higher if bleed-thru or dark-colored paper)
- Output format: PDF or JPG (see following topic)

PDF vs. JPG

The age-old question. In general, PDF files are better for printed documents, annotation, "out-of-the-box" searchability, and keeping multiple page documents as a single file. JPG files are better for photographs and handwritten documents. In addition, JPG is easier to browse on a computer, better to embed in genealogy programs, but are harder to annotate and searchable content is only available in the file properties detail. The biggest downside to JPG files is the one page per file concern. If a record has multiple pages, they are treated as separate files.

PDF File Format:
- Pros
  - Keeps multipage document together
  - Can be annotated
  - Easier to reproduce
  - OCR & searchable
- Cons
  - Larger file sizes
  - Harder to browse files
  - Metadata needs special app

JPG File Format:
- Pros
  - Smaller file size
  - Better for photographs
  - Easier to browse
- Cons
  - Harder to annotate
  - Single page per file
  - Computer can only search metadata
  - Keeping related documents together is a pain

Always Verify (Proof)

If you are disposing of the paper that you have scanned, **ALWAYS** verify that you have the scanned document, that it looks good and is complete! As soon as you have verified your computer document is good and complete, recycle the paper. Always make sure you have back-up copies of your scanned documents. If is very frustrating to have thrown away the original paper and then, accidently delete your only computer file of that record.

Searchable Text

JPG files are labor-intensive to add computer-searchable keywords to any detail. You can add general keywords to multiple JPG files in a single operation, but detailed information would be lacking without a lot of work. Using the Searchable PDF setting when scanning allows users to search for words within the content of the scanned document without any additional work.

The only text that is searchable in JPG files is the filename and whatever is in the file properties. The entire content (except for handwriting) of PDF is searchable or can be made searchable after it is created. The file properties of PDF files are also searchable, but you need a app for changing the properties content. In addition, Microsoft windows and most PDF viewer apps allow you to add comments to PDF files. The comments are also computer searchable.

Life Proofing

If you go to all the effort to scan your paper documents, safeguard your investment in time and effort.

- Back up your files REGURALY. If you have your files on your computer, purchase an external hard drive or flash drive to make copies of your scanned files.
- Put your genealogy-proof lineage records on a "cloud" drive (GoogleDrive, OneDrive, iCloud). If you lost your research notes or records for colleterial relatives, it is aggravating, but not a disaster for proving your lineage.
- Upload your genealogical-proof lineage records on FamilySearch or Ancestry trees. Share your research that proves your ancestors and relative. Not only does it protect your scanned documents, but you will make contemporary and future researchers happy.
- Invest in mirroring network accessible storage.