Managing the Genealogy Data Monster

Thomas MacEntee, of Abundant Genealogy

Overwhelming Data and Genealogy – Work Smarter Not Harder

Do you feel more overwhelmed than ever when it comes to genealogy and all the data you have access to these days? Remember back 10, 20 or 30 years ago – did it also feel like this or was data just more manageable back then? Well the feeling is new and the reason? BIG DATA and the technologies that make more genealogy data available to us. With advances in technology genealogy (and even non-genealogy) vendors now have the ability to produce more data and make it available to researchers.

The problem of keeping up with data is not going away nor will it be easily solved. Part of the burden is on the providers to make access to data easier and quicker. The other part is on us as researchers: we need to work smarter with the data!

Genealogy Projects

As genealogists it is common for us to have more than one research or writing project going on at the same time. Actually, isn't that just how our daily lives are? Many tasks and projects running parallel to each other.

Project Management Basics

The secret to getting stuff done is this: track, prioritize and document each task and project. Here are the basics of effectively using your own project management system:

- **Projects and tasks**. You can track projects, tasks or both. Some like to take larger, over-arching projects – such as researching a specific ancestor – and break them down into tasks.

- **You need a tracking system**. Simply writing down projects and tasks on a piece of paper, sticky notes or in a Microsoft Word document works for some researchers. However, once you get beyond 10 tasks, you'll have difficulty tracking each item. Create a way to "track" tasks so you can quickly determine priorities and status of each item.

- **Track time spent**. Even if you are not a genealogy professional who bills time to clients, you may find tracking the amount of time spent on a task or project helpful. Over several weeks and months, you can see what you do the most or least in a given time period and make adjustments if necessary.

- **Analyze data**. While none of us want to track every little thing we do, determine what's important, track it and then periodically look at the data. On a monthly basis look at home much time you spend on email, spend on research, spend writing, etc. Is the allotment of time in line with your goals? If not, then commit to spending more time on what matters and consider improving ways to handle those tasks that are not as important.

Tools to Use

Here are some formats for project management currently used by genealogists:

- **A journal or notebook**. It is perfectly acceptable to write down your projects and tasks each day in a journal or notebook. You can develop a system of using colored markers and highlighters to prioritize and categorize tasks. One downside: you may spend quite a bit of time copying uncompleted tasks over to the next day. Also consider the digital alternative to a written journal such as a Microsoft Word document or other document processing/creation platform.

- **A spreadsheet**. One of the best platforms for project management is a spreadsheet maintained in a program such as Microsoft Excel or Google Drive. Spreadsheets allow you to sort tasks and to calculate data points such as hours spent on a task. Most spreadsheet programs allow you to highlight tasks and use color. You can even take data and plot it in a chart, such as a quarterly chart showing how much time you spend on each type of task.

- **A note taking app or program**. If you are more note-oriented, don't forget programs such as Evernote and One Note where you capture information on-screen, add a note, add tags, etc. You can set up a special "To Do" notebook or folder for tasks and projects.

- **A task/project management program or app**. There is a wide variety of task management software programs, websites and apps available. Some of the more well-known are Remember the Milk and Wunderlist.

Tips and Techniques

Project management becomes a breeze and almost second nature if you keep these tips and tricks in mind:

- **Track everything in the beginning**. Over time, you'll get in a groove and know what tasks should and shouldn't be tracked.

- **Track what's important to you**. Some have "Handle e-mail" as a daily task while others don't track it at all. If you want to see "where your day went" track it for a few days and then determine if it is important.

- **Set up common tasks/projects**. If you have tasks or projects that repeat on a daily, weekly or monthly basis, create a special spreadsheet tab called "Common Tasks" and store them here. Then copy and paste to your project sheet when you need them.

- **Prioritize items**. A numbering system such as 1 = High Priority, 2= Medium Priority, and 3 = Low Priority can help you determine which items to tackle first.
- **Color code items**. Consider using colors to indicate whether an item is overdue (red), on schedule (yellow) or completed (green).

Genealogy Research Data

If you're a beginning genealogist you want to research "the right way" and if you're an intermediate one, you probably have learned from your early mistakes. While some would argue as to what is "correct" we would probably all agree on these points as to what a research log could and should do:

- **Create an action plan**. Many genealogists have some sort of research plan whether it is written down or in their heads. A research log can help you fine tune that research plan and convert it into an action plan. For many of us, that is the challenge . . . getting from "What I want to look for" to "Oh, look what I've found!"

- **Keep you organized**. Do you spend hours just bouncing around on Ancestry.com or another website? What about in-person at a repository . . . do you use your time wisely or just go from resource to resource? A research log can help you not only organize your research finds, but also identify areas where you need to do more research.

- **Allow instant recall**. Even if you are successful at finding the records and information you need in genealogy, what good is it if you can't locate that information easily and quickly? Especially if you store your finds on a computer, it helps to have a "map" to where all that stuff is! Use links and entries in a research log to find items fast.

- **Reduce "do-overs."** There's nothing worse than spending time researching only to find out that you already had that particular record. A research log let's you know where you've been and where you should be spending your time.

- **Help analyze data**. If you've located several different records related to an ancestor's marriage, how can you really analyze all the information and come to a conclusion if you can't see the details all at once? A research log lets you sort by fields or column headings and also search for specific keywords so you can easily analyze results.

Research Log Basics – What To Include

Here are some data points that you should consider tracking on a research log or any method you use for your genealogy research:

Important

- **Date**: Enter the date you located the record. Not only will this show the evolution of your research process, but the information can be important for source citations. In addition, you can tell if a new record set available in-person or online was even available when you were researching a specific person or topic.

- **Record Name**: Each record should have a common yet distinct name, such as "Death Certificate for John Leehive." For records that might have several iterations or versions such as an obituary in different newspapers, add qualifiers such as "Obituary of John Leehive, New York Times."

- **Repository**: Identify where the record is physically located. For online records, you decide if you want to list the website, such as Ancestry.com, or drill-down to the actual originating repository such as NARA.

- **Record Type**: This is really a "field of convenience" since it simply gives you an easy way to sort and extract entries and group them by a category. Think "death record," "marriage record," etc.

- **Transcript Extract**: This can be a time consuming entry especially if you are working with a physical record. For online records, many times the item is already transcribed – simply copy and paste the text.

- **Analysis**: For me, this is where I take notes, formulate theories and really do my analytical work. Often I'll enter comments or questions to myself such as "Which silk factory did John Vincent Slattery work at?" or statements such as "Proves Margaret Leehive was the daughter of John Leehive."

- **Link**: Indispensible for online research since it saves time when having to go back to a record. Granted, URLs can change (that is why the source citation will always serve you as the best way to locate a record), but having a clickable link means not having to remember how you found the record.

Optional

- **Record Number**: An internal numbering system to reference entries in your research log. Use a simple number format or perhaps a surname + number format such as **Austin01**.

- **Source Citation**: While we can argue whether or not this should be an optional field, this is where you build your source citation. Stick to the basics such as the name of the record, the repository, date found, etc. The goal is to make it easier to find the record later on, especially if the record is located online and the URL has changed.

- **Result**: This is what I call my "thumbs up, thumbs down" or "plus or minus" column. Basically I use one word like "positive" or "negative" or a phrase like "proves theory" or "disproves theory." Doing so is also important since certain theories can only be proven through negative evidence.

- **File name/location**: Enter the file name of a saved record image or photograph. Why? Most computer operating systems make it easy to search your hard drive contents so all you have to do is copy the text and paste it into your computer's Search field to find the file.

Evidence Evaluation Fields

The following fields can be used to evaluate evidence. They offer an exellent way to help analyze research data.

- **Source Type**: A source is **Original** if it is the first written statement, photograph, or recording of an event. Subsequent copies are **Derivative** and may be reproduced by hand, machine, camera or scanner; they may be reproduced on paper, in microform, as photographs or digital images, or in any other medium that records the image whether transcribed by hand or technology.

- **Clarity**: Use **Clear** If the information can easily be read. Use **Marginal** if information is not clear, is partially obscured and researcher must "guess" at words or letters.

- **Information Type**: Use **Primary** if a piece of information is recorded by a knowledgeable eyewitness or participant in that event, or by an official whose duties require him or her to make an accurate record of the event when it occurs. Use **Secondary** if information is supplied by someone who was not at the event and may include errors caused by memory loss or influenced by other parties who may have a bias or be under emotional stress.

- **Evidence Type**: **Direct** evidence is any fact that is explicitly stated. **Indirect** evidence is inferred from one or more pieces of evidence within the record.

Tools to Use

Here are some formats for research logs currently used by genealogists:

- **Genealogy Database Programs:** Software which lets you track your genealogy research. RootsMagic, with a free version available for download, has a Research Log report built in as a feature. It is organized using a series of questions such as "What were you trying to find?" and "What source did you check?" and "What were the results of your search?" Other programs such as Family Tree Maker and Legacy Family Tree have similar features.

- **Note Taking Programs:** Applications that allow you to create notebooks and capture web pages and documents include Evernote and OneNote. Nice features include the ability to take a photo of a document and send it right to the application, and the ability to tag and label information.

- **Productivity Applications:** Similar to Microsoft Office, Google Drive is a free online productivity suite with Document and Spreadsheet components that can be used to produce a research log.

- **Reference Management Programs**: These programs manage bibliographic data and research materials, usually online. Zotero includes web browser integration and will generate source citations and Mendely offers similar features.

- **Forms**. There are also resources which offer a variety of free forms including research logs such as Ancestry.com's Research Extract and Free Genealogy Forms and Charts forms.

- **Understand the "why" of using a research log**. If you are using a research log only because other researchers are doing so, then you're wasting your time. Understand the benefits of tracking your research journey on paper, in a digital document, or using an online application. You'll have a better appreciation for the research log and the research process.

- **Select a format that you will use**. There's no sense in taking the "square peg, round hole" approach and using a format such as an Excel spreadsheet if you don't like using spreadsheets. You'll only frustrate yourself and abandon all ideas of using <u>any</u> research log. Try different types and stick with what works for you.

- **Spend time setting up headings or categories**. When taking the spreadsheet approach which relies upon column headings, take time to consider which headings to use. And don't be afraid to add or remove headings over time. It is only through constant use of the research log that you'll determine the best headings for your research process.

- **Shoot for a "one pass" goal.** When you find a record or piece of information, note all the information as if you might never find that item again. This means noting the date you found it, location, type of record, etc. You're only kidding yourself if you say, "Oh, I'll come back to that later." Later is often "never" or a time when you need the information right away and it isn't available.

- **Maintaining a research log is a discipline.** A discipline is created through hard work, practice and repetition until it becomes habit. Realize that you'll make mistakes the first few entries. Then you'll become better at entering information accurately and quickly.

- **Source citations matter – take a shortcut!** Sounds like heresy, doesn't it? "There are no shortcuts in source citations," someone once told me. Nonsense. Create a cheat sheet for yourself – a document or a spreadsheet tab where you keep your most commonly used source citation formats. Then copy the format to your record entry and fill in the blanks. Easy peasy!

Genealogy Files

Most of us have had the task of sorting files in manila folders or in a filing cabinet, right? A "system" helps us find what we need during and after research. Yet because we can't "touch" or readily "see" our computer files, we tend not to hold them to the same organizational standards we might use on our paper files.

File Management Basics

Have you ever thought about the amount of time you spending looking for the file you need? And how that time could be better spent on research? Here are some goals to keep in mind when working with a variety of data files on your computer.

- **Save online content NOW not LATER**. Subscribe to the "I may not come this way again" school of thinking and never delay saving an image or file from an online site. Save it now, even if you have to dump it into a general folder for cataloging later. Websites and databases come and go!

- **Create files names that make sense**. A file name should tell you the most important info about a file. "John Austin obit.pdf" is not as helpful as say "19770420 John Ralph Austin obit.pdf"

- **Don't overuse folders; rely on search**. Don't waste too much time creating folders, sub-folders and sub-sub-folders. Today's computers excel at indexing file data and allowing you to search for what you need. Think "Google" but only on your computer, not the Web.

- **Add metadata to files**. Take time to access the file properties of certain files (Microsoft Office, jpg, etc.) and add information. This could be descriptive info as to who is in a photo or copyright information. Some genealogists also embed a source citation in metadata to assist other researchers using the file.

Tools to Use

The best tool to use is the file management system that is part of your operating system (most likely Microsoft Windows or Apple's OS). Take advantage of all features including sorting, customized views, and tags or labels.

Tips and Techniques

Here are some tried and true tips and methods used by genealogists and family historians to get a better handle on computer files:

- **Use YYYY MM DD for easy sorting**. Don't name files "April 12 1970 Smith Mary Obit" but instead use "19700412 Smith Mary Obit." Then you can sort files by date order.

- **Use tags and labels**. Most file system allow you to add a tab or a label such as Death Record, 1930 Census, even surnames and locations. Again, the more you add, the easier it is to find files later.

- **Never rename a file extension**. The file extension – after the last "dot" in a file name – tells your operating system (Windows or OS) what program to use to access the file. If you change it using the file edit feature, you could corrupt and damage the file. Always use a program's export or convert feature instead.

- **Use one file naming system and stick with it**. Again, we all work differently and some will use a surname-based systems, others a date-based system. Don't wobble back and forth between different systems.

Resource List

Genealogy Research Log

- **Ancestry.com's Research Extract**
http://www.ancestry.com/trees/charts/researchext.aspx

- **Genealogy Research Log – Microsoft Excel**
https://abundantgenealogy.com/genreslog

- **Evidence Explained**
https://www.evidenceexplained.com

- **Evernote**
http://www.evernote.com

- **Free Genealogy Forms and Charts**
http://www.genealogysearch.org/free/forms.html

- **Family Tree Maker**
http://www.familytreemaker.com

- **Legacy Family Tree**
http://www.legacyfamilytree.com

- **Mendely**
http://www.mendeley.com

- **Microsoft Excel**
http://www.microsoft.com/excel/

- **Microsoft One Note**
http://www.microsoft.com/onenote/

- **RootsMagic**
http://www.rootsmagic.com

- **Zotero**
http://www.zotero.org

Genealogy Project Management

- **Any.do**
http://www.any.do

- **Asana**
http://asana.com

- **Evernote**
http://www.evernote.com

- **Genealogy Project Management Spreadsheet**
http://hackgenealogy.com/projectmanagement

- **Google Drive**
http://drive.google.com

- **Microsoft Excel**
http://www.microsoft.com/excel/

- **Microsoft One Note**
http://www.microsoft.com/onenote/

- **Microsoft Project**
http://www.microsoft.com/project/

- **remember the milk**
http://www.rememberthemilk.com

- **Toodledo**
http://www.toodledo.com

- **Trello**
https://trello.com

- **Wunderlist**
http://www.wunderlist.com

Genealogy Files

- **Metadata and Digital Images**
http://flip-pal.com/metadata-and-digital-images/