

# 10 Things Genealogy Software Should Do

---

By Mark Tucker

## Abstract

Innovation in genealogy software starts with ideas that lead to better design. This paper discusses 10 things that genealogy software should do but currently doesn't. It is a starting point for discussion among those in the genealogy community: family historians, software developers, and designers. It is a springboard for additional design ideas.

## Introduction

Once we start thinking about the design of genealogy software, the ideas flow. We can see features in other software that can be adapted. We come up with our own ideas. Some ideas have been around for 10 years or more and just haven't been implemented. We will discuss 10 things that genealogy software should do. These 10 things form a "wish list" of design requirements not tied to a specific technology for implementation.

In the course of a conversation with Elizabeth Shown Mills in February 2008, we discussed the question of how people learn to do genealogy. My answer was others walked them through the learning process or they taught themselves through books or courses. Her answer surprised me. She said people learn through the software they use to manage their family history data. As a Software Architect, why wasn't that my first answer?

In 1997 Elizabeth Shown Mills published *Evidence! Citation & Analysis for the Family Historian*. Over the years it has become known as the "Bible" of source citation. In 2007, she published the much anticipated and expanded book called *Evidence Explained: Citing History Sources from Artifacts to Cyberspace*. Also in 1997, the Board for Certification of Genealogists (BCG) developed the Genealogical Proof Standard (GPS) which remains a requirement for certification. Why does almost every family history application support GEDCOM but not the GPS and standardized source citations?

If we were to take what we have learned over the past two decades of family history software development and start over from scratch with a new application, what would it look like? I think it would include these 10 ideas: Genealogical Proof Standard, organize, transcribe, source citation, stress-free imports, link and layer, source provenance, visualize, remember, and share.

# 1. Genealogical Proof Standard

The Genealogical Proof Standard has been around for more than 10 years, but is seldom mentioned outside the requirements for becoming a Certified Genealogist. This 5 step process is intended to systematically get us as close to the truth as possible when conducting genealogy research:

1. We conduct a reasonably exhaustive search in reliable sources for all information that is or may be pertinent to the identity, relationship, event, or situation in question;
2. We collect and include in our compilation a complete, accurate citation to the source or sources of each item of information we use;
3. We analyze and correlate the collected information to assess its quality as evidence;
4. We resolve any conflicts caused by items of evidence that contradict each other or are contrary to a proposed (hypothetical) solution to the question; and
5. We arrive at a soundly reasoned, coherently written conclusion.

This Genealogy Research Process map is a way to visualize the concepts from the BCG and Elizabeth Shown Mills:

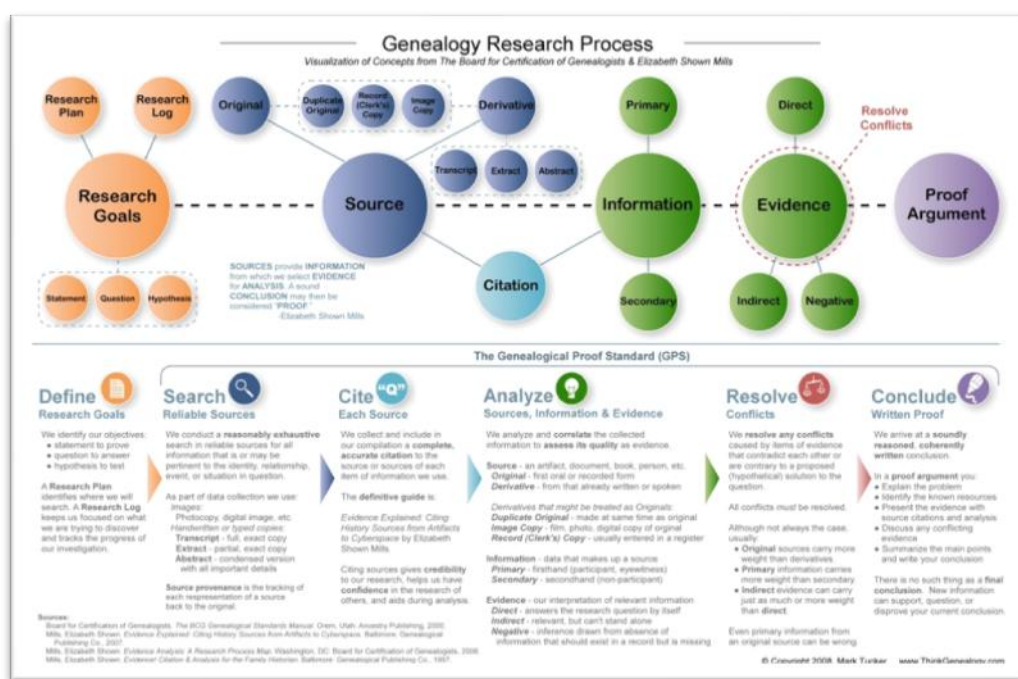


Figure 1 - Genealogy Research Process Map

We should incorporate this workflow into every family history application.

## GPS – Step 1

The first step of the GPS is to conduct a reasonably exhaustive search. Consider what can be seen in Family Tree Maker 2008's Web Search feature that allows you to search Ancestry and add information to your family tree:

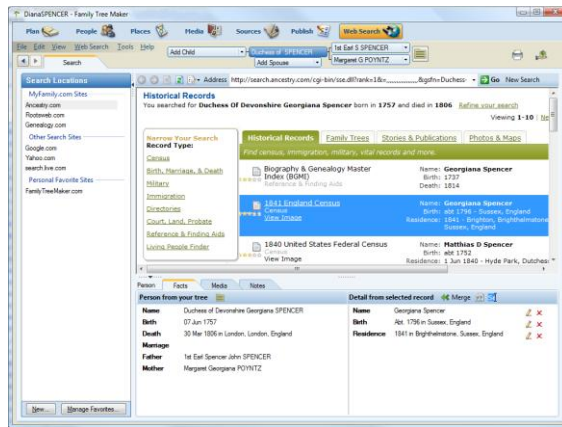


Figure 2 - Family Tree Maker 2008: Web Search

Imagine this scope expanded to search World Vital Records, Footnote, and others.

## GPS – Step 2

With our search conducted and sources found and added to our application, they can be organized and transcribed. Next we need to provide an accurate source citation. Organization, transcription, and source citation will be discussed in the next sections.

## GPS – Step 3

Sources must first be classified as original or derivative with special cases (from *Evidence Explained*) for duplicate original, image copy, and official record copy. Information then needs to be identified in the sources and classified as primary or secondary. We must then correlate this information with the questions we want to answer or the problems we want solved. When information is applied to problem statements it becomes evidence. Evidence needs to be classified as direct or indirect with special handling of negative evidence.

## GPS – Step 4

We then turn to resolution of conflicting evidence. This is where the source, information, and evidence classification tasks help to provide weight or credibility to one statement over another.

The classification of source, information, and evidence is very powerful in its ability to help prove identity, relationships, life events, and biographical details. This is why the Genealogical Proof Standard is known as a credibility standard.

## GPS – Step 5

But we fall short unless we write down the conclusion. We present our case in a *proof argument*. It is important to remember that there is no such thing as a final conclusion. As more information is discovered, classified, analyzed, and correlated our previous conclusion can be supported, called into question, or proved incorrect.

Imagine if our genealogy software guided us through the steps of the GPS – giving hints and examples along the way.

## 2. Organize

One of the challenges of doing family history research is organizing all the photos, documents, or other sources that come our way. Many of these sources are in paper form, but an increasing number are digital. Regardless of the way you organize your sources (surname, document type, etc.); there are advantages to organizing sources digitally. One of the advantages is being able to quickly find your sources. Keywords you add to the documents as well as the source text can be searched. You can also generate reports for use in your research.

The Clooz software is an electronic filing cabinet that helps you organize your documents and link them to people. It uses input templates based on the work of Elizabeth Shown Mills.

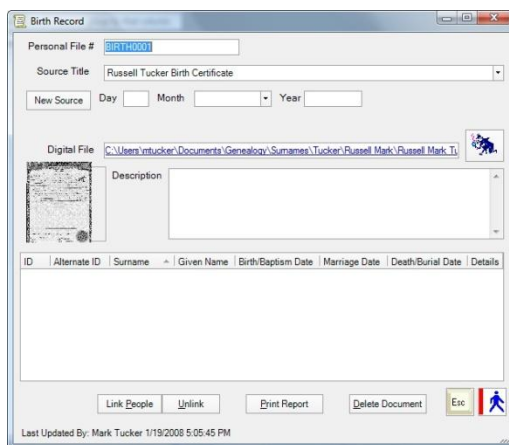


Figure 3 - Clooz 2.1: Birth Record Entry

Photoshop Elements is not a genealogy application, but I like how some of its organization features can be used in the design of genealogy software. All images can be added to the Organizer and then grouped in albums. Related images can be stacked together. Images can be tagged and then quickly found using search filters.

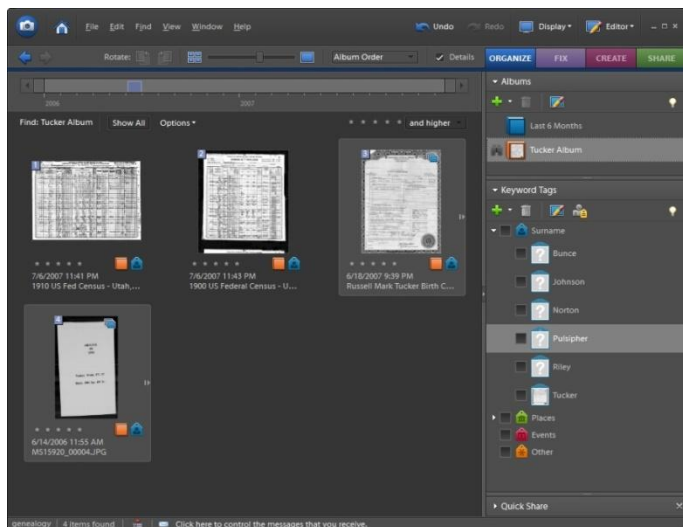


Figure 4 - Photoshop Elements Organizer showing Genealogy Sources

You can view photos by album, date taken, import batch, file system location, and location on a map. For a genealogy application you should be able to group documents by type (birth, marriage, death, census, wills, etc.) as well as by surname or family.



Figure 5 - Photoshop Elements Organizer: Display menu

The timeline is helpful to visualize the number of photos by date, narrow the range of photos to view, and quickly navigate to photos by month.



Figure 6 - Photoshop Elements Organizer: Thumbnail Timeline

A similar timeline is available when viewing photos by Import Batch.



Figure 7 - Photoshop Elements Organizer: Import Batch Timeline

For this to better support genealogy sources, the timeline would need to show the date of the source document, not the date of the image file.

### 3. Transcribe, Extract, and Abstract

Some digital sources will be in a format that is readily searchable, while others must be transcribed, extracted, or abstracted. Some images contain typed text so OCR software can make a first pass at conversion then we can make finishing corrections. Other document types require us to do the transcription ourselves. A simple word processor with only those features needed to support transcription would be very helpful. It would contain help and guides so that all genealogists would do correct transcriptions without needing to take a course.

Another idea is to allow highlighting on the image that would bring up a transcription box to do a quick extract of important parts of the document:

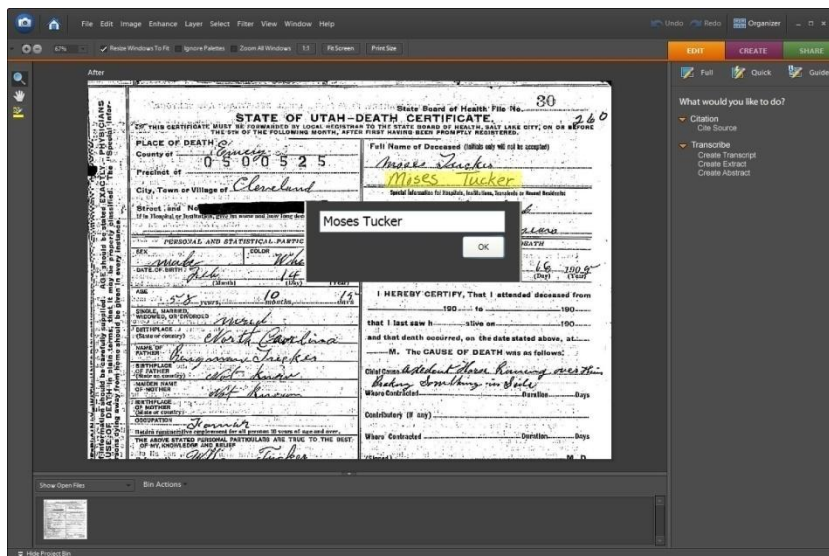


Figure 8 - Document Highlight and Quick Extract

This highlight feature goes beyond what is found in FamilySearch Indexing that allows you to highlight a portion of the document that is being extracted:

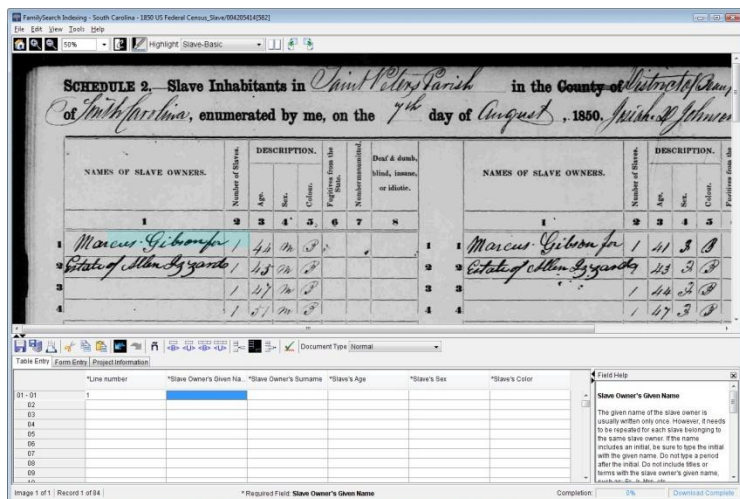


Figure 9 - Highlighting in FamilySearch Indexing

It is more than an annotation feature, although the source citation could be a layer over the document so when it is printed the citation is printed as well.

This feature is more of what is found in Footnote.com where you can annotate a portion of the document as a person, place, date, or text:

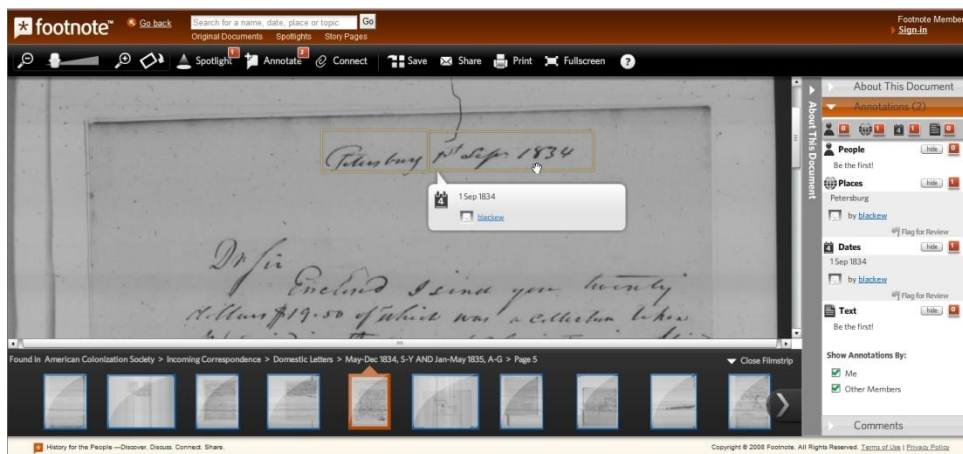


Figure 10 - Annotations in Footnote

The difference between the proposed design and what is currently available is that besides the annotations accompanying the document image, it is used as a primary way to enter data into your database. Data entry would follow the rules for creating an extract, transcript, and abstract. This enhanced feature also allows you to indicate if each piece of information in the document is primary or secondary following the GPS.



## 4. Source Citation

Source citation is vital to genealogy research, and it should be easy to do correctly. The new standard guide to citation is Elizabeth Shown Mills's book, *Evidence Explained: Citing History Sources from Artifacts to Cyberspace*. The book defines citation models for many source types with the information needed for the citation as well as the format. These citation models should be part of the software so that we can browse or search by keyword to find the correct model for the source. The first application to implement this is Legacy Family Tree, version 7.

Online databases need to provide better source citations. Not all databases provide citations and when they do they are in a text format that must be copied and pasted into your genealogy program. In all cases, they do not follow the standardized citation models defined by Mills.

These first two examples not only show the text-based representation of online source citations, but also inconsistent formatting for the same citation. The first is from FamilySearch.org:



Figure 11 - FamilySearch.org 1880 Census Citation

While this one is from Ancestry.com:

**Source Citation:** Year: 1880; Census Place: Laurel, Ashe, North Carolina; Roll: T9\_952; Family History Film: 1254952; Page: 607.2000; Enumeration District: 22; Image: 0218.

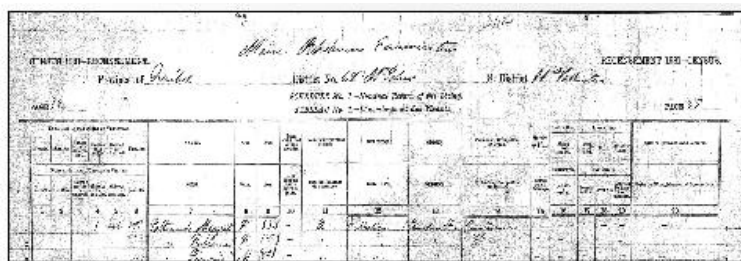
Figure 12 - Ancestry.com 1880 Census Citation

The above two examples don't follow the citation format found in *Evidence Explained*:

1880 U.S. census, Ashe County, North Carolina, population schedule, Laurel Township, p. 14B, dwelling 117, family 129, Moses Tucker; digital image, *Ancestry.com* (<http://www.ancestry.com> : accessed 19 January 2008); citing Family History Library microfilm 1,254,952.



One of the best non-commercial sites that demonstrate online citations is by Stephen J. Danko. Here is an example of a census citation from his site:



**1881 Canadian Census Record for Margrite Patenaude and Family**

SOURCE: 1881 Census of Québec, Canada, St. Jean County, St. Valentin, page 39, lines 1-3, Margrite Pattenaude and family; Library and Archives Canada, Ottawa; FHL microfilm 1,375,835.

**Figure 13 - Census Citation from Stephen J. Danko's site**

When you click on a document to download the image, the link should instead point to citation information in a standard XML format. Then the genealogy software could 1) read the information, 2) add the image to the application, and 3) add the proper source citation. This is the same principle that makes RSS work for blogs. The citation models defined in *Evidence Explained* should be used to define the XML for the citation:

First (Full) Reference Note				
CENSUS ID		JURISDICTION	SCHEDULE	
1. 1850 U.S. census, Marion County, Iowa, population schedule,				
CIVIL DIVISION	PAGE ID	HOUSEHOLD ID	PERSON(S) ...	
Lake Prairie, p. 290 (stamped), dwelling 151, family 156, Virgil W. and				
... OF INTEREST	ITEM TYPE OR FORMAT	WEBSITE TITLE	URL (DIGITAL LOCATION)	
Wyatt B. Earp; digital image, <i>Ancestry.com</i>			(http://www.ancestry.com : ac-	
... DATE	CREDIT LINE (SOURCE OF THIS SOURCE)			
cessed 16 January 2006); citing NARA microfilm publication M432, roll 187.				

**Figure 14 - Census QuickCheck Model from "Evidence Explained" page 240**

Here is one representation of the citation model in XML:

```
<CitationModel>
  <Category>Census Records</Category>
  <SubCategory>Digital Images Online Commerical Site</SubCategory>
  <Fields>
    <CensusId>1850 U.S. census</CensusId>
    <Jurisdiction>Marion County, Iowa</Jurisdiction>
    <Schedule>population schedule</Schedule>
    <CivilDivision>Lake Prarie</CivilDivision>
    <PageId>p. 290 (stamped)</PageId>
    <HouseholdId>dwelling 151, family 156</HouseholdId>
    <PersonsOfInterest>Virgil W. and Wyatt B. Earp</PersonsOfInterest>
    <ItemTypeOrFormat>digital image</ItemTypeOrFormat>
    <WebsiteTitle>Ancestry.com</WebsiteTitle>
    <Url>http://www.ancestry.com</Url>
    <Date>accessed 16 January 2006</Date>
    <CreditLine>citing NARA microfilm publication M432, roll 187</CreditLine>
  </Fields>
</CitationModel>
```

```
</Fields>
</CitationModel>
```

The Citation Style Language can be used in conjunction with this XML to define the citation format.

There already exists a tool that captures online citations, stores them, and then outputs them as formatted citations. It is a Firefox plug-in called Zotero and it could be used in conjunction with genealogy software.

An alternative approach is to download the file and have it automatically load in the genealogy application in a similar manner that Quicken or Microsoft Money is able to add transactions from a financial site:



Figure 15 - Financial Site Download

What could be easier for a genealogist than every time a document image is downloaded from an online database, the source was automatically cited?

## 5. Stress-free Imports

When we get a GEDCOM file from someone, how many of us refuse to import it into our main database? For various reasons we don't want to merge their data and have it get mixed up with ours.

What if importing didn't mean we had to merge? What if their information resided in our database but remained separate? Imagine an import history list that shows information such as when an import was done and where the information came from. With a single click, all the information from an import could be hidden or shown throughout the application.

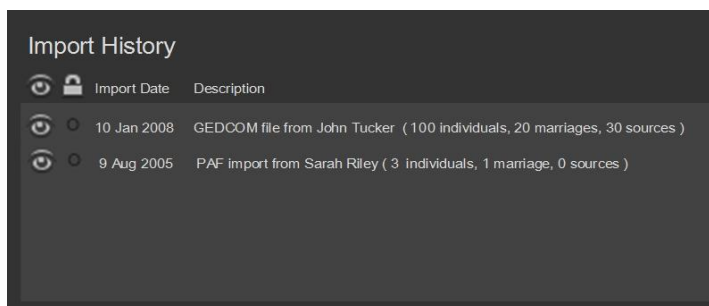


Figure 16 - Design of Import History screen

This is similar to how many graphics programs use layers that can be hidden or locked. If desired, the import could be deleted without destroying your information with it. This would require the next item, link and layer.

## 6. Link and Layer

Most of us have heard the term match and merge. This usually comes into play after we have imported some information into our database. We find two individuals in our database and determine that they are the same person and then we merge their information together. Once the merge is done, there is no going back to the way it was.

To explain the concept of link and layer, we will return to Photoshop Elements. This application allows modification of an image and one of its powerful features is layers.

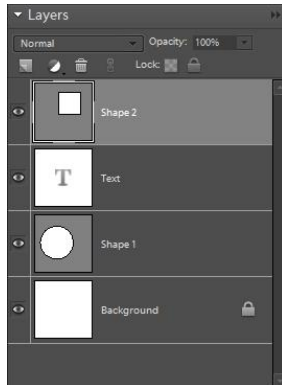


Figure 17 - Photoshop Elements Layers

Think of layers as transparent sheets that can be stacked. The image is a composite of all the layers. Work can be done on one layer without affecting other layers. Layers can be hidden to see the effect without them present. A layer can also be locked to prevent accidental editing.

What if instead of merging information, it was linked and each link was shown as a source layer. Take for example the footnoted Family Group Sheet from page 103 of *Evidence!* by Elizabeth Shown Mills:






Family Group Sheet			
HUSBAND 'S FULL NAME	Louis Solastie Rachal		
	DATE	CITY/COUNTY	STATE/COUNTRY
BIRTH	c. May 1797 <sup>1</sup> c. 1792 <sup>2</sup>	Cane River <sup>1</sup> Natchitoches Parish <sup>2</sup>	Louisiana <sup>1,2</sup>
BAPTISM	18 Oct 1799 <sup>1</sup>		
MARRIAGE	— — 1817 <sup>2</sup>	Natchitoches Parish <sup>2</sup>	La. <sup>2</sup>
DEATH	— Dec 1835 <sup>3</sup>	Cane River <sup>3</sup>	La. <sup>3</sup>
SOURCES:			
 			
 1 Elizabeth Shown Mills, <i>Natchitoches, 1729-1803: [Translated] Abstracts of the Catholic Church Records</i> (New Orleans: Polyanthos, 1976 ), no 2923.			
 2 Register 5, no. 1817: 10, St François Church, Natchitoches [now Immaculate Conception].			
 3 Succession 429 ( 1835 ), Louis Solastie Rachal, Clerk of Court 's Office, Natchitoches Parish Courthouse.			

Figure 18 – Design of Link and Layer with all sources shown

Each of the three sources has some information that we have determined are about the same person. Our database would have three different records (one for each source) each containing only the information from the respective source. Each person record would be linked to each other with a “self” relationship.

Notice that there are multiple birth dates and locations from two different sources (Source #1 and Source #2). Also notice what other information comes from Source #2. Since each source is listed as a separate “layer” then the information can be shown together or individual sources can be hidden. When you hide Source #2, then any date or location from that source is hidden:

**Family Group Sheet**

HUSBAND 'S FULL NAME	Louis Solastie Rachal		
	DATE	CITY/COUNTY	STATE/COUNTRY
BIRTH	c. May 1797 <sup>1</sup>	Cane River <sup>1</sup>	Louisiana <sup>1</sup>
BAPTISM	18 Oct 1799 <sup>1</sup>		
MARRIAGE			
DEATH	— Dec 1835 <sup>3</sup>	Cane River <sup>3</sup>	La. <sup>3</sup>

**SOURCES:**

☒   
☐

1 Elizabeth Shown Mills, *Natchitoches, 1729-1803: [Translated] Abstracts of the Catholic Church Records* (New Orleans: Polyanthos, 1976 ), no 2923.

2 Register 5, no. 1817: 10, St. François Church, Natchitoches [now Immaculate Conception].

3 Succession 429 (1835), Louis Solastie Rachal, Clerk of Court 's Office, Natchitoches Parish Courthouse.

**Figure 19 – Design of Link and Layer with Source #2 hidden**

When you see this information in your genealogy software, you see the composite of the information. You can choose to hide a layer or look at it all. If you later find that the connection was incorrect, you can break the link and the rest of the information is unaffected.

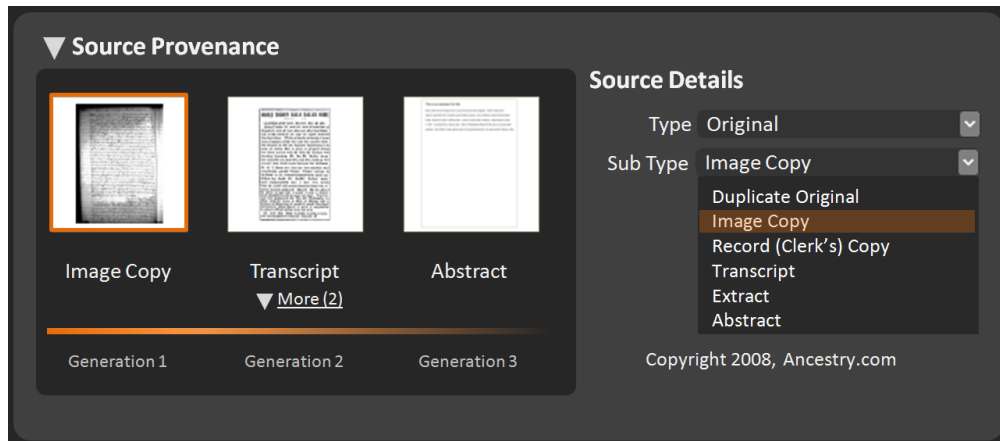
Hiding all information for an import (discussed previously) is equivalent to hiding all layers associated with the import.

The lock feature could be used to prevent accidentally modifying information.

## 7. Source Provenance

According to the GPS, when a source is first analyzed it should be classified as original or derivative. When downloading a document from an online database, the metadata could include this classification as original or derivative. All sources should include the source provenance all the way back to the original. Who would be in a better position to know that than the content owner? This identification would then only have to be done once correctly and it would save each researcher from doing that work themselves with the possibility of mistakes.

Here is one design for the source provenance feature:



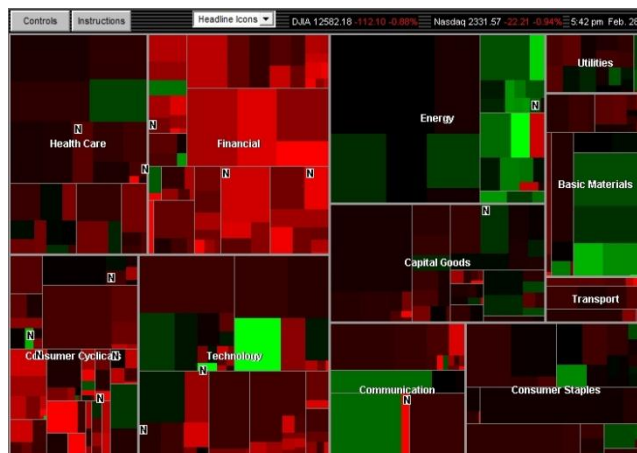
### Figure 20 - Source Provenance Design

This example shows the provenance with the source that is closest to the original on the left followed by its derivatives. In this example, the Image Copy of the source is treated as an original and it has three derivatives with only the Transcript showing. The Abstract was derived from the Transcript so it is listed next. Because the original is likely not in a digital format, then the Image Copy is one generation from the original and the Abstract is three generations from the original.

## 8. Visualize

Genealogy software contains more and more information which can be better understood through visualizations. Following are some examples of visualizations that would help the family historian understand and evaluate their research.

One type of visualization is the treemap as can be seen from the SmartMoney site:



**Figure 21 - Treemap Visualization from SmartMoney.com**

A treemap is great when you want to be able to get a big picture of how things relate and then drill down into more detail. One application of a treemap to genealogy software could be the visual representation of all the surnames in your database. It could also be used to show people in your database first by location and then by surname.

You can quickly see the relative number of people sharing a surname and filter the person list by using a Surname Cloud:



Figure 22 - Family.Show Surname Cloud

Or you conduct a search to return all female ancestors and then you want to see a histogram of their age at death:

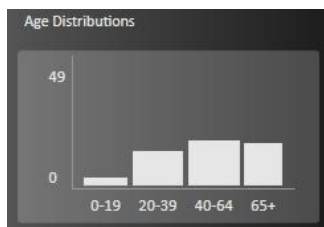


Figure 23 - Family.Show Age Distributions

There are many other visualizations possible to show related sources, relationships, and locations for a specific ancestor.

## 9. Remember

Genealogy software should remember so we don't have to. It should remember whenever you add or change information. It should keep track of your research goals. It should keep a Research Log of all the sources accessed and when. For certain types of resources like web sites; it should automatically add a To Do List reminder to check for changes in 6 months. It should remember what you were doing the last time you did research on a specific person.

The time spent on family history is very precious. As family historians, we don't want to waste any time trying to remember what we did last. A dashboard could provide quick links to the last 10 items worked on and other key information. This is similar to the Outlook Today dashboard in Microsoft Outlook.



Sometimes our research gets put on hold for a few weeks or even many years. We should be able to pick right back up where we left off and have the software remember all the important things so we have the needed context.

## 10. Share

Our software should support our comfort level for sharing information. Some information we want to keep private to our local database, some we want to share with a select group, and some with the world. Maybe we want to share with those researching the same people or surnames. Or we can share with others researching the same location or grouped together for a specific project.

The main format for sharing information today is GEDCOM 5.5 which was released in 1996. It allows for sharing individuals, families, names, dates, places, notes, sources, and repositories. Not widely implemented are media files: video, audio, images, pdf, and text documents. GEDCOM is in need of a major revision with a likely move to XML. Think of word processing applications for the PC. We started with a format that accepted only text (txt), and then we could add formatting with the Rich Text Format (rtf). As document contents become advanced Word Document (doc) was released and most recently (docx). Today's word processing applications can support multiple formats. This should be possible with our family history software.

As genealogy social networking sites like FamilyPursuit, FamilyLink, Geni, and GeneTree become more prevalent, we will want to share our information. It shouldn't matter that I want to use a desktop application and you want to use a web application, we should still be able to connect and share data. But currently each of these online applications is its own island. Each wants to own the list of family and friends. There needs to be a way for us to connect these together and say this account on FamilyPursuit is me as well as this account on GeneTree and that I want to share my data with this person running a desktop application.

## Conclusion

There are many innovations that can be made to make genealogy software better. Design is a key element and it all starts with ideas. Whether the ideas are for incremental changes or 10x "jumping curves" type innovations, we as a genealogy community must work together.

Alan Cooper in the book *About Face 3: The Essentials of Interaction Design*, shows an evolution of the software development process that finally leads to design at the beginning of the process and more involvement by the people that will use the software:

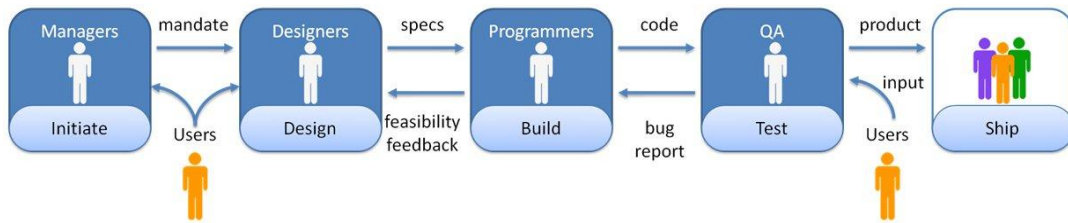


Figure 24 - Goal-Directed Design Process

The genealogy community includes genealogists, developers, and designers that can work together, to share and explore the future of genealogy software. From this “think tank” can emerge the best ideas that will improve our experience with the software and help us advance our genealogy research. Soon these features can be as common in our family history applications as GEDCOM is today.

## Bibliography

Board for Certification of Genealogists. *The BCG Genealogical Standards Manual*. Orem, Utah: Ancestry Publishing, 2000.

Cooper, Alan; Robert Reimann; and David Cronin. *About Face 3: The Essentials of Interaction Design*. Indianapolis, IN: Wiley Publishing, Inc., 2007.

Mills, Elizabeth Shown. *Evidence Analysis: A Research Process Map*. Washington, DC: Board for Certification of Genealogists, 2006.

Mills, Elizabeth Shown. *Evidence! Citation & Analysis for the Family Historian*. Baltimore: Genealogical Publishing Co., 1997.

Mills, Elizabeth Shown. *Evidence Explained: Citing History Sources from Artifacts to Cyberspace*. Baltimore: Genealogical Publishing Co., 2007.

## Websites

Family Tree Maker

<http://www.familytreemaker.com/>

World Vital Records

<http://www.worldvitalrecords.com/>

Footnote.com

<http://www.footnote.com/>

Clooz 2.0

<http://www.clooz.com/>

Photoshop Elements

<http://www.adobe.com/products/photoshopelwin/>

FamilySearch Indexing

<http://www.familysearchindexing.org/>

Legacy Family Tree

<http://www.legacyfamilytree.com/>

FamilySearch.org

<http://www.familysearch.org>

Ancestry.com

<http://www.ancestry.com>

StephenDanko

<http://www.StephenDanko.com>

Citation Style Language

<http://xbiblio.sourceforge.net/csl>

Zotero

<http://www.zotero.org/>

SmartMoney.com

<http://www.smartmoney.com>

Family.Show

<http://www.vertigo.com/familyshow.aspx>

FamilyPursuit

<http://www.familypursuit.com/>

FamilyLink

<http://www.familylink.com>

Geni

<http://www.geni.com>

GeneTree

<http://www.genetree.com>

Genealogy Research Process Map

<http://www.thinkgenealogy.com/2008/02/24/genealogy-research-map>