

## A Template for Evaluating Evidence

*Genealogical Computing* 24 (April–June 2004)

[extracted from unpaginated archived copy, *Ancestry.com*

(<http://www.ancestry.com/learn/library/article.aspx?article=9270> : 16 October 2011)

Surety levels. Confidence numbers. Primary versus secondary classes. For years, genealogical data management software has offered these and other schemes by which family history researchers might (theoretically) evaluate the reliability of the data they enter. Unfortunately, all suffer from two fatal flaws: they are totally subjective; and none of them permit researchers to evaluate the radically different characteristics of a single source.

Data management programs need a meaningful and objective way for users to identify the strengths and weaknesses within their data. A method does exist—a simple one—made possible by dynamic changes in evidence analysis that have occurred in the genealogy field since the mid-1990s.

### **Subjectivity**

Any two genealogists can look at the same piece of data and reach widely varying conclusions as to its reliability. Experience is a major factor, but no factor weighs in more heavily than the sheer subjectivity of the approaches that have been taken. As with the stars assigned by opinions posted at Amazon.com, what one user of a source considers a 5, another may deem a 3. When data is exchanged, confidence or surety levels are meaningless. Each recipient of the data must personally locate and appraise the source before he or she is able to make even a preliminary judgment as to the reliability of the information.

### **Primary vs. Secondary**

The field of logic defines “false choice” as a forced choice between two things falsely presented as an either/or scenario. Here lies an even more serious problem with the evaluation of genealogical evidence: the false choice of primary versus secondary, applied indiscriminately to sources, information, and evidence. The terms primary and secondary are time honored in some other fields—particularly history, journalism, and law. By the standards and needs of those fields, the terms may still be adequate. For genealogy, they are not.

To use history as a comparison: Historians typically focus on the “big picture.” Their objective is to understand society; they pursue that objective by studying events and causes rather than obscure, individual lives. Because they typically work with humanity in the aggregate and use individual cases as “anecdotes” to illustrate their points, if they make a mistake on a small particular here or there, it seldom affects the broad conclusions presented in their “big picture.” Genealogists, however, study history in microcosm. We focus on the individual. If we make a mistake on a genealogical detail it can have immense consequences for all further research we conduct on that person and

that line. Consequently, genealogists have to be far more precise in measuring the reliability of minute pieces of evidence.

The impossibility of forcing all evidence into one of two neat pigeonholes—primary or secondary (firsthand or secondhand)—is easily demonstrated by a death certificate. That document typically asks for death date, place, and cause. For these pieces of data, a death certificate should provide firsthand information, as supplied by the attending physician. However, death certificates also call for birth dates and places, names of parents, and birthplaces of parents. Unless both parents are alive, present, and serving as the source of the information (a rare situation), then a death certificate cannot be deemed a “primary source” for that information.

Similarly, a deed may recite a chain of title. The person supplying the information may be a source of “primary” information for his own acquisition or sale of the property at the same time that he is a source of “secondary” information for the chain of title from the original patentee. Those two traditional classes of evidence—primary and secondary—may even coexist in the same sentence. If our software asks us to classify the deed itself as a primary source or a secondary source, then we are forced to make a false choice.

### **The Solution: The Analysis Tree**

Over the past several years, a new schemata has been developed for the handling of genealogical evidence—the Analysis Tree, as depicted in figure 1. Whatever the research problem, whatever the type of source, whatever the type of information, we can measure our findings against this tree in a precise and totally objective fashion.

The Analysis Tree has essentially three branches: the source is the package or the container that holds the information; the information is the factual details we extract from the source; and the evidence is what we mentally draw from that information.

- **Sources.** Sources can be people, books, microfilm, journals or magazines, CDs, databases, or websites. They can be artifacts of any type from tombstones to tea towels on which a female forebear embroidered her name and dates. By whatever name a source is called, the quality of a source is essentially determined by asking one question: Is it an original or a derivative? As a rule, original sources are more reliable than derivatives—although all rules have exceptions.

- **Information.** The details or data that we extract from a source is our information. It is evaluated by asking ourselves: Did the informant or creator of this detail have firsthand (primary) or secondhand (secondary) information? Diarists, witnesses, or letter writers, for example, who offer details about experiences they personally lived through, would provide primary information. Those who relate tradition or report details secondhand would provide secondary information. As a rule, primary informants are more reliable than secondary informants—although all rules have exceptions.

- **Evidence.** All information we find may not be relevant to our problem, even if it deals with our individual. If our problem is the identification of John's wife's maiden name,

then information stating that John voted in the contested election of 1876 would not likely be evidence for the problem we have defined. A marriage record, on the other hand, typically would be evidence. Once we decide that a piece of information is evidence for a problem, then we measure that evidence by asking: Is this direct evidence or indirect evidence?

For example, a church marriage registration that says John married Mary Smith, the daughter of Samuel and Susan Smith, supplies evidence that directly answers the research question we defined. On the other hand, if the marriage record were an ancient, original bond and the bride's maiden surname were torn away, but one Samuel Smith served our John as a bondsman, then the evidence we draw from this source and its information—i.e., Mary may have been kin to Samuel Smith—would be indirect evidence usable toward a solution of our problem. As a rule, direct evidence is preferred to indirect evidence, although indirect evidence, carefully constructed, could indeed outweigh direct evidence drawn from a false claim.

### **Applying the Tree to Software**

This schemata provides an objective, standardized means by which all researchers, regardless of experience, can measure the quality of the information they record in their data management software. At least, it could if genealogical software would provide a simple “Data Quality” screen on which users can tick off options. (see figure 2) The screen should be attached not to the source but to the individual piece of data that is being referenced.

The result would be three simple tickmarks, three totally objective choices, each carrying a standard meaning. Considering that standard meanings are a fundamental principle for the exchange of data, this system of measuring the quality of data would seem a no-brainer.

### **Beyond the Tree**

For newer genealogists who lack the experience needed to make valid judgments using a 1,2,3,4,5 scoring system, the Analysis Tree eliminates agonizing decisions and misjudgments. Experienced researchers will use the Analysis Tree as a framework for a more precise scrutiny of their data. With experience, researchers learn that even information from firsthand informants can be wrong if the informant was biased or had faulty recall. They learn that a transcribed marriage register, carefully copied—or even a “compiled genealogy,” carefully prepared—can provide a reliable marriage date, whereas an actual witness to the marriage may “misremember” the date.

To facilitate this kind of analysis, citation templates should continue to offer software users the ability to enter a free-form “source description.” Along with the usual details that citation manuals outline for various types of research materials, researchers also need to record a critical description of that source. For example, when using documents and registers, we add a note if ink is faded to the point of dubious legibility or penmanship is scrawled. When drawing data from family artifacts, we need to record the provenance of that artifact. When using film, we might note not only legibility problems but also

whether part of the text is bound so tightly in the book that we cannot be sure we have all data copied correctly. When using a published book, we note whether the author has cited his or her source for each assertion. Because a source is the package that contains our information, and damaged packaging can adulterate the contents, it is important to describe the condition of that packaging. This type of information is correctly placed with the source template itself.

Concurrently, we also need a separate screen, attached to the specific data each time we extract a piece of information from the source. Here, we need to identify the three basic qualities on which credibility rests: 1) whether the data's source is an original or a derivative, 2) whether the specific data is from a primary or secondary informant, and 3) whether that data represents direct or indirect evidence of the date, name, place, event, or circumstance we are trying to prove.

### **Basic Evidence Terms**

*Citation* A statement in which we supply details to identify our source. It should include not only standard citation details (outlined in citation manuals) but also a critical source description.

*Evidence* The information we choose as relevant to the question we need to answer. Evidence is either direct or indirect, depending upon whether a) it directly answers our research question or b) it is relevant but cannot answer the question without being combined with other evidence.

*Information* The data that a source provides. Information may be either primary or secondary (i.e., firsthand or secondhand). Any given source can contain both types of information.

*Proof* The evidence and reasoning (analysis) that support our conclusion.

*Proof Argument* The written explanation in which we detail the evidence and reasoning that support our conclusion.

*Record Recorded information.* A single source can contain many records or only one. The source of a record can be of any type—e.g., an original document, a published book, a taped interview, a tombstone, etc. It may or may not be reliable.

*Source* Anything that offers information—a document, register, publication, artifact, website, database, person, etc. Sources are either original or derivative.

- **Original source.** One that contributes written, oral, or visual information not derived from a prior written record or oral communication. Official copies (i.e., verbatim transcriptions entered into church or governmental registers by persons charged by law with recording the information accurately) may be treated as an original if the actual “loose sheet” original is no longer available. Microfilmed images and digitized images

may be treated as an original when produced by an official or otherwise reputable agency.

- **Derivative source.** One that contributes information copied, transcribed, abstracted, summarized, or repeated from information in a previously existing source. Photocopies and digital images made and distributed by individuals do not carry the same weight as originals because of the possibility of tampering. Certificates (birth, marriage, death, etc.) issued by church and governmental agencies are not considered originals because copying errors are more frequent and data may be intentionally omitted to fit the form of the certificate.

Comparative examples: 1) A manuscript church register is usually an original; a set of abstracts or translations made from the register is a derivative; 2) A DNA lab report (a document) is an original; an article written about DNA test results would be derivative.

*Source Description* A critical description of a source, noting any flaws that affect the reliability of that source. This detail should be added, free-form, to every source citation.

*Verification* A situation in which an assertion is supported by direct and credible evidence found in an independently created source. A mere repetition of an assertion, in an equally weak source, does not constitute verification. Rather, the term “verification” implies that the supporting evidence is of sufficient quality to credibly stand alone.

**Elizabeth Shown Mills, CG, CGL, FASG**, a past president of the American Society of Genealogists and the Board for Certification of Genealogists, recently retired from a 16-year post as editor of the *NGS Quarterly*. Her best-known publications are *Evidence! Citation & Analysis for the Family Historian* and *Professional Genealogy: A Manual for Researchers, Writers, Editors, Lecturers, and Librarians*.