

[Snapshot of the page 11 Apr 2011](#) Created by Adrian
[Snapshot 15 May 2011](#)

Introduction

I have been concerned for some time that there is a gap in our analyses between the genealogical research process, the evidence & conclusion model and my aspirations for a "scientifically" robust process and write-up. This page is an attempt to fill in that gap and define more specifically what we need - in my *personal* opinion - to be documenting and therefore what I would like to see in my ideal genealogy research software. Ultimately though, the focus of the page is less on the process, and more on driving out what data items are involved in that process, so I can be more confident about what's needed in the BetterGEDCOM Data Model.

As with all my process descriptions, I usually don't mention whether or not software will be used at any point, because a process is about "what" is done, not "how". Where I do mention software, it's because avoiding doing so makes the words harder to read.

It may very well be that this page does no more than document what you, the reader, felt was obvious. If so fine, but I have satisfied myself that I have filled in the gap.

This process researches a specific event, attribute or relationship concerning something or someone. It does not attempt to set an overall strategic direction. As a result, someone writing a family history will go through this process many times. (*Note - I have not yet tested this process in my mind against large-scale family reconstruction, but suspect the same steps appear - just more often.*)

Inspiration for this process description has been taken from a variety of sources, credited in the text but especially from Mark Tucker's Genealogy Research Process diagram.

1. Set a focussed goal

Set and record a focussed goal (e.g. "Who were the parents of X?") (*c.f. Tom Jones, "Inferential Genealogy" course handout, 2010, Family Search*)

Input

- Current conclusions about people, things, etc

Output

- Focussed goal – what things or people are to be investigated? What relationships, events or attributes do we want to know about them?

2. Create or revise research plan

It is likely that you need to take several steps to reach the goal. So, split the necessary work into portions, each of which contributes a step towards the overall goal. Each portion will have its own specific objective, which is lower-level / more detailed / more specific than the overall focussed goal. Unless the plan is deliberately intended to be only a partial plan (see below), the last step should pull everything together and provide a solution to the overall focussed goal.

From the steps, with their own objectives, create a “reasonably exhaustive” research plan describing what to search for *and* how to analyse it. The research plan needs to be broad enough in time, space and people to trap potentially useful information (c.f. Tom Jones, also the Genealogical Proof Standard.) Do not be afraid to include speculative items, e.g. "Anything in Chester Quarter Sessions records for the 1820s?".

This plan starts as an initial plan - it is highly likely that it will be necessary to loop back here and create a revised plan later on in the light of information found - "No plan survives first contact" ([Field-Marshal Helmuth von Moltke the Elder](#)). Indeed, if you are uncertain of the direction your research may take, the initial plan may be only a partial plan, with the rest being defined only in the light of the first set of discoveries.

Input

- Focussed goal (from previous step).
- Current conclusions about people, things, etc
- Useful data about types of sources to look for when looking for evidence relating to specific attribute / event values (e.g. “parents’ names can be found on Scots death certificates”) (probably this data is held elsewhere in books, magazines, on-line catalogues and research notes from archives and record offices, for instance.)
- Useful data about those types of sources e.g. “pre-1837 marriages in England were always in Church of England parish churches” or “post-1837 marriage certificates in England were created by register offices and / or churches”(probably this data is held elsewhere)
- Where to look for those potential sources (e.g. list of where Cheshire parish registers can be found) (probably this data is held elsewhere)

Output

- List of work portions, each with specific objectives and description of what is intended to be done. This describes "what" is intended and "how". For instance, "What? "Find X's marriage". "How? Look for all marriages in Nantwich, Acton and Wistaston between 1820 and 1840 with groom's name = X. For each of the couples, are they found in the subsequent censuses? Does the age in the census match X?")
- Location of potentially *relevant* sources ("where") needed to fulfil the objectives above, forming

an initial search log (e.g. "Microfilms of parish registers (PRs) for Nantwich, Acton and Wistaston, at Chester Archives". This log will then be completed in the next steps to summarise what has been found.) (*The list of objectives and the initial log, taken together, make up a search plan*)

- Assumptions made for each work portion – “As both X and Y live close to where they were born, according to the censuses, it is assumed that they were married reasonably close to there.” Also “It is assumed that neither of N and W were Quaker or Jewish” – these are the two exceptions to the rule about marriage in CofE church.

3. Carry out research

For each of the work-portions in the research-plan, carry out the research according to the current plan. (*When searching for paper-based records, this step takes place in Archives, Record Offices, etc. When searching internet based records, this takes place at a computer terminal and the division between this and subsequent steps might tend to disappear*). (There are personal decisions to be made about how much to record for sources that are close to the search criteria but do not match - if it's been a long journey and you won't be back for a while, it might be tempting to record "close" sources for long-term storage somehow in case they turn out to be for a relative.). Important - a search does not find a record - make a note of that as this will save you looking for it in the same place again. Also, the lack of a record where it might reasonably be expected to be, could be significant.

Input

- List of specific objectives, each with a statement of “how” I intend to fulfil each objective (from previous step)
- Location of potentially *relevant* sources needed to fulfil the objectives above, forming an initial search log (from previous step)
- Documented assumptions (from previous step)

Output

- Contents of the researched sources, documented in such a way as to be understandable (e.g. a series of marriage transcripts for people satisfying the search criteria) and with enough data to enable accurate citations, record provenance, etc.
- Updated search log saying for each source, what's been searched, what was missing, etc.

3.5 Understand the Records

For each of the work-portions, check your understanding of the records that have been judged to have useful information. Understanding why a record was created will help interpret the information in it (*For instance, does the grant of probate say 'Personal Effects' and / or 'Real Estate' - do you understand the difference?*). (c.f. Tom Jones, "Inferential Genealogy")

4. Select & Analyse the Evidence

For each of the work-portions, assemble the research from this work-portion. Assess the quality of the source material. Using your research plan, select the evidence (i.e. the information that's relevant to the objective of this work portion). This evidence can come from this work-portion, previous portions *and* the evidence already in your database. Look for any patterns, matches, differences, etc. that might be meaningful. *(These are not meant to be sequential steps but can take place in parallel)*

Can you demonstrate ("prove") that the person referred to in the evidence is the one that the objective needs? Analyse the evidence to see if you can answer the questions posed by the objective for this work-portion.

What are the conclusions for this work-portion? Is there any conflicting evidence? *(e.g. this looks like him but it's the wrong father)* Any partial progress? *(e.g. These are the marriages matching our couple but there is more than 1 match, so we cannot yet tell which is their marriage)*. Any evidence that contradicts any hypothesis? *(e.g. Implication of complete search is that they were not married after all)*

If there are conclusions, record the analysis and conclusions for this work-portion in some form of proof summary or proof argument, referring back as necessary to previous conclusions.

Input

- List of specific objectives, each with a statement of "how" I intend to fulfil each objective (from previous steps)
- Contents of the researched sources, documented in such a way as to be understandable and with enough data to enable accurate citations, record provenance, etc. (from previous step).
- Updated search log saying for each source, what's been searched, what was missing, etc (from previous step)
- An understanding of the nature of the researched sources (probably held elsewhere)

Output

- Evidence relevant to the objective(s), including evidence of identity
- Proof (summary or argument) giving results of analysis and conclusions – if any (See BCG for [samples](#)).
- Conflicting evidence – if any

5. Has the Objective been met for this Work-Portion?

If the objective for this work-portion has not been met, *or* if there is conflicting evidence that cannot be resolved, return to create new search plan with revised specific objectives *(Some conflicts can be accepted if they can be resolved, i.e. explained away in a plausible manner - e.g. it was 50 years after the marriage and the son giving the information was born long after that marriage.)*

5.5 Record Conclusions

For each specific objective, if it *has* been met with *no* unresolved conflicting evidence, enter any conclusions into the genealogy application, either

- merging these conclusions into existing people, events, objects, etc, deleting or replacing conclusions that are no longer accepted (i.e. use the "conclusion-only" model) or
- creating new people, events, objects containing the new conclusions plus old conclusions that are still accepted (leaving the old detail there but marked up as "superseded") (i.e. use the "evidence conclusion model")

6. Go onto next work portion in research plan

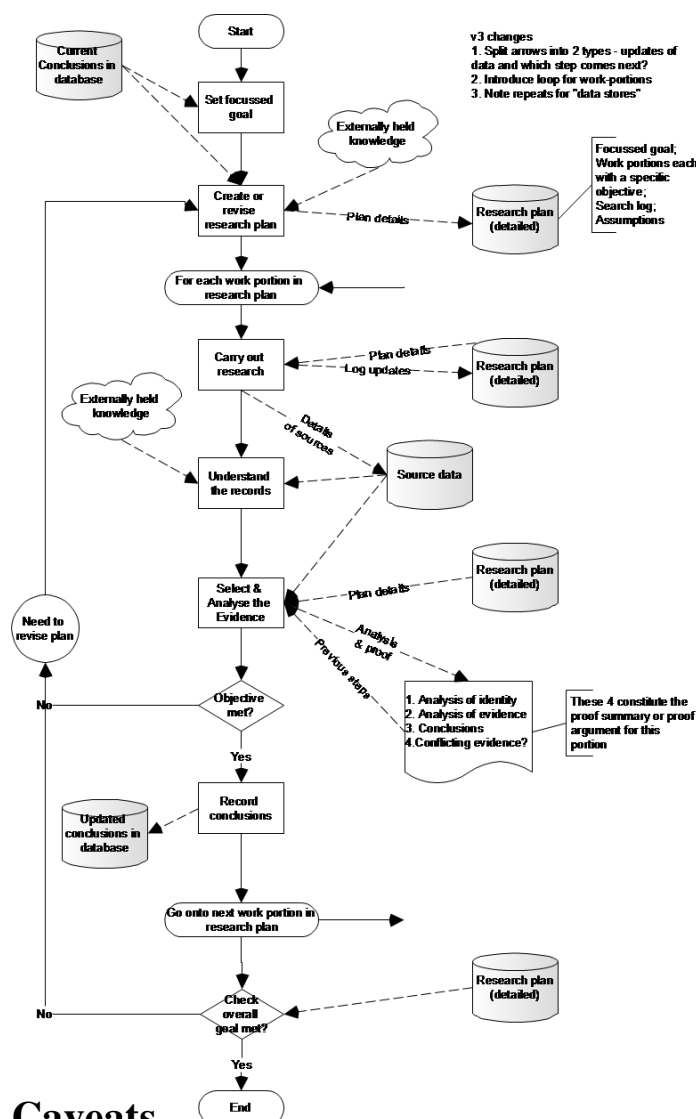
Go onto the next work portion in the research plan.

7. Check overall goal has been met

If all the objectives from all the work portions have been completed and there is **no** unresolved conflicting evidence, double check to see if the overall focussed goal has been met (*remember, in this way of working, the last portion is meant to provide the answer to that focussed goal*) and that all the conclusions have been entered into the genealogy application, including the final conclusions. If it has, then this research process is complete.

If there is any unresolved conflicting evidence or if the overall focussed goal has not been met, return to create new search plan with revised specific objectives.

Diagrammatically



Caveats

I talk of "proof" and "final conclusions". It is unlikely that proof in a complex genealogical study can reach the standard of proof required in a criminal case ("beyond a reasonable doubt") - the [Genealogical Proof Standard](#) exists to provide criteria to judge the standard of proof obtained. Nor is any conclusion really final as the appearance of previously unsuspected information may throw everything into suspicion.

Application Software

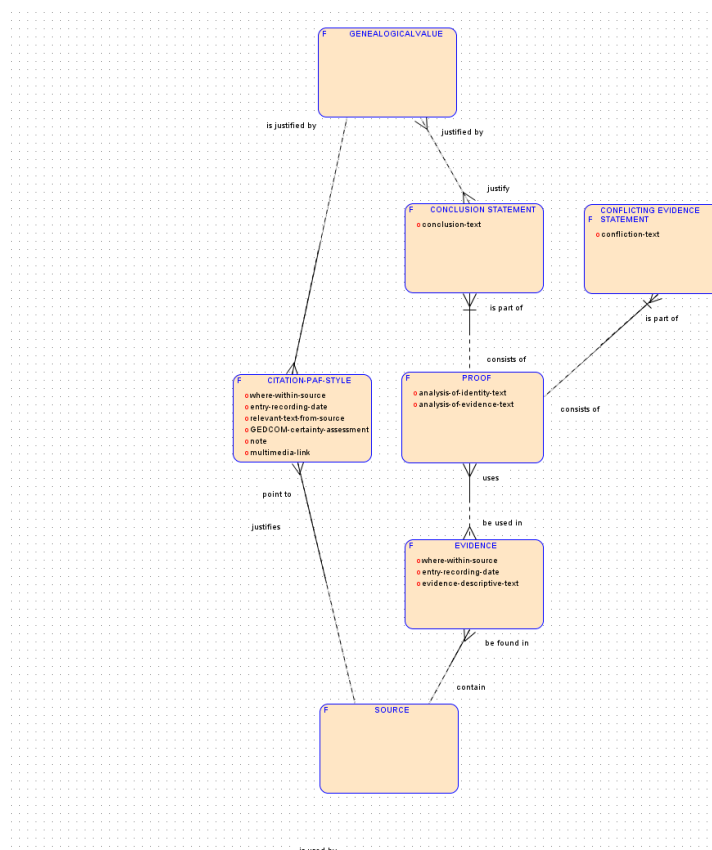
The crux of the matter is this - what do I want to see in an application? And therefore in the BetterGEDCOM Data Model? The answer is - *everything* that's recorded as an input or an output above. Actually, there are exceptions denoted above by the phrase "probably this data is held elsewhere" since otherwise we'd end up dumping all the text books into our applications. But after these first steps, I'd really like all that lot to go into my application.

Data Analysis

First cut list of entities - excluding those already clearly covered by GEDCOM - i.e. persons, families, etc. This list is somewhat descriptive, rather than specific.

- Focussed goal – *what things or people are to be investigated? What relationships, events or attributes do we want to know about them?*
- Research Plan -
 - List of work portions, each with specific objectives and description of what is intended to be done.
 - Initial search log containing location(s) of potentially relevant sources ("*where*") needed to fulfil the objectives above. *This log is later updated completed to summarise what has been or what **cannot** be found where.*
 - Assumptions made.
- Contents of the researched sources, documented in such a way as to be understandable, and with enough data to enable accurate citations, record provenance, etc (*as per GEDCOM Source-entity but with extra attributes compared to current GEDCOM?*)
- Proof argument or summary for each work portion made up of
 - Evidence used in the work portion, including evidence of identity
 - Results of analysis
 - Conclusions if any
 - Conflicting evidence if any
- Updated values of a person / group / place etc recorded in database for each work portion

The indented bullets are intended to imply a probable relationship - e.g. the higher level bullet consists of the lower level ones.



First cut of data model shown above. Notes:

- Multiple cardinality on relationships is shown by crows-feet (a) because that's what I use and (b) because the (free) modelling tool does it like that.
- Entity "Genealogical Value" is meant to represent any attribute of any person, family, group, place, ship, artefact, etc. entity. Or any event or relationship involving same.
- Entity "Citation PAF style" is meant to refer to the citation structure as encoded in GEDCOM, PAF, etc, etc.
- I have only resolved one many-to-many relationship - that involving "Genealogical Value" and "Source", which is resolved by the "Citation PAF style".
- "Citation PAF style" remains on the model for 2 reasons - (i) compatibility and (ii) because I feel that it could otherwise be impossible to see which bit of a source contributed to the value of the "Genealogical Value". In other words, I see both this and proof / conclusion in use at the same time.
- Conclusion and Evidence statements are just envisaged here to be free format text. Rightly or wrongly.
- A conclusion statement may justify several values - e.g. "John Doe was born to Richard Doe on 31/1/99" justifies a parent's name and a birth date.
- The text in the Proof, Conflicting Evidence and Conclusion entities will need to be substantial in length.
- A proof gives rise to (potentially) many conclusions. Or even none if it's a disproof(?).
- A conclusion comes from only 1 proof. (No reason that several proofs can't come to an *equivalent* conclusion).

- A proof must arise from *some* evidence.
- A piece of evidence comes from only one source. But a source contains (perhaps) many bits of evidence.

Comments?