20.05.2011 G.Thorud

Comments on the page "Research process, Evidence and GPS" of 15.05.2011

A. General comments

This document is written with close to no knowledge about formally defined processes published in the US, there are others that know those much better than me. In most cases I am not writing comments about data against the clause data analysis, but in the process steps – that is where the data are actually described.

Please read the document twice, I have not spent time trying to create a logical sequence.

Detail: Each major step in the page could be identified as a step in its title.

Priority: Priority should be given to definition of the output data that need to go into persons, events, citations, sources – i.e. those things that could affect the rest of the total BG model.

It might be helpful to look at www.lineascope.com – but I am not saying "do it that way". (Yes I know, I should tell you if I find anything interesting in there, but it is not easy to see all the structure from the videos.)

Definitions

If the document is to form a basis for further work it should define many terms used in the document, at least those that will be reflected as data structures in a data model that I assume will be the result of the documents. We cannot have data structures without definitions. Examples alone are not a good way to write definitions.

Also miss a definition of:

(Re)Search log – this is the only working document mentioned. Is this a document containing only the info explicitly specified as entered into the log? Or will it contain all the notes made by the researcher throughout the whole process? (I assume the log will be in a computer program or on paper) It should be defined initially. Should it be called research log rather than search? Also, I am not sure if a log is a good term, to me a log is something recording something happening and it is written as a linear structure, sequentially/chronologically, which I don't expect something in a program to look like. Anyway, the Log may not translate into a record, so it may not matter, but an application might have such a view of the records.

Administration in the Requirements Catalog

The Rec Cat contains a summary of most features and data handled by some current programs and will have to be compared and probably merged with the data model resulting from this document. We cannot neglect the experience of the current programs. But it seems like your Goal fits with the Objective in the Rec Cat and your Objective fits with Task, so that is a good start.

There is another orthogonal dimension to what is described in this document and that is the "project management type of info" e.g. hours used, status, priority and where/when to do something which are probably relevant to many of the records related to a process. This info is correctly kept out of this document (except where), but it needs to go into the end product.

B. Re. Introduction

It would be useful to state if the document is based on one or more process defined elsewhere, incl. define GPS. One problem is that processes essentially doing the same thing can be structured in many ways, so what is implemented in BG may need to have some flexibility if possible.

Re. last paragraph. The type of problem that is solved in this document seems to be a "relatively small problem", i.e. someone writing a Family genealogy would have to go through this process hundreds and probably thousands of times. They will need a structure into which they can organize these "small problems" into a project, and professional researchers will have several projects. A large project working on a genealogy for a village or parish would probably follow other steps, e.g. transcribing a number of sources as an initial step.

C. Goal

Re. Output. It should be possible to link to persons and events (and more main records) but it may be more difficult to reference relationships unless we have records for them – we could.

D. Search Plan

Re. Input

- Useful data about what sources (and the two next bullets): Should it be "what source TYPES"? I interpret this to be held in an (online) archive/library catalog or similar (or your brain), or a feature in some programs trying to guide you to sources (e.g. Gensmarts in Legacy or a separate program supporting Rootsmagic, TMG,FTM and more) – I am not sure how successful those features are. Research Guidance (with Gensmarts) in Legacy is something like a skeleton of a simple research process – it could be extended – have a look. Genbox can record e.g. the types of info in sources (event type), time period, jurisdiction, responsible agency and is likely to extend on its research guidance features. In other words, the info need not be external to the program and you could envisage user organizations creating files with this type of info, but the best way would be if archives etc. could codify the info types – but I am not sure how realistic this is and it is a question if they would encode according to event types in BG – most likely not, so it is more likely users would do it in a way useful to programs. As I expect jurisdiction and time period to be covered by citation elements, you could get a long way by attaching event types and a probability of finding that type of info in the source. (The above is not the focus of your document, but since looked at these programs to see if there were something relevant to your doc, I wanted to make note of it.)

- Useful data about sources – to me it looks like the example here is a combination of the two previous steps – what is this info used for? Can this be defined more precisely?

Re. Output

- Objectives – to me this seems like what in the Rec Cat (and in several programs is called a Task/Target/To-do. Besides – what is the difference between an objective and a goal? The use of Objective is confusing to me., the terms are too similar. Tasks are more general than just search for info. If an objective is not a task, what is the relation between an objective and a lookup in a single source? Can you lookup in several sources?
  The words "what" and "how" should be expanded to full sentences
- At some stage you will need to identify the source. That could be done is in this step. If you have that source registered, which is often the case, you should be able to link to it.
- Also, you might have vague formulations about the info you want to find, and you want to ask an archivist when you get to the archive – so when you are in the archive you want to find those vague formulations possibly related to that archive.
- A consequence of these three comments above is that we need to look at objectives/tasks in the context of the other entities in BG. How does the concepts relate to a single source, and later in the document a single citation (and other structures where data is recorded) – one-one/one-many.
- Documented assumptions – should read "Document"? If not, I read this as something that has already been documented – where? (Also, this step is then not only about a search plan as stated initially.)

E. Research

You want to record that you did **not** find a source for the info (see previous step), or did not find a particular source you are looking for in a particular archive, or did not find the info in a source.

I think we have enough other discussions about how to record the source data so will not go into that.

F. Check your understanding – Assemble research

A lot could be said about how to do this, but I don't think it will affect the data structure.

There is no place to record text info about the quality of the source and the info in the source evaluation – in text – or is that in Evidence?

A problem is the term "Evidence" it has many variants, text extracts, multimedia, source summary, codified – if you ask (as in Data Analysis) where this should go in a data model (records) or a report, you will get different answers depending on the variant. But it seems to me that we are here talking about yet another variant. Or is it the Summary, or …?

G. Proof

For my reference a link to a posting by gene about Proof summary and Proof argument:
http://bettergedcom.wikispaces.com/message/view/Research+Process%2C+Evidence+%26+GPS/37632782#37702356

For the record: I get the understanding that proof summary or argument are results of analysis and conclusions - evidence and conflicting evidence is not in summary or argument. Or?

It seems also that the intention is to have goal and objective records containing evidence (?, cf last paragraph in F), analysis, conclusion and conflicting evidence? Or may we just link to fields holding some of these?

Do any or all of the 4 (Evidence, analysis ..) above output in published reports (not drafts with the researchers notes) or do we just have to duplicate info? If they output, an important question is, where do we record (or automatically import/copy) the proof summaries, proof arguments, evidence and conflicting evidence **for output**

- Separate document – possibly linked to via multimedia – possibly an “appendix”
- Citations (actually you have to look at the citation elements that can hold such info, there is “reasoning” and “summary” in there and the relevant ones should be identified. Gene – Help!)
- The “main” (climbing) note of a person, place, group
- In notes for events
- (Gene uses a special “Tag” (event))
- Or do we output the fields in the goal and objective records?

And where will they end up in reports – we know where the elements in this bulleted list end up (except the last) – so if we map evidence, analysis into these we will know where it ends up in reports? Will “where” depend on how much text – probably – as has been discussed. Do we merge various fields into consecutive paragraphs in reports?

Maybe a table with the possible locations for each would help?

Also, most likely different researchers would choose different places, e.g. some in citations and some in the “main note” of a e.g. a person record (could perhaps be configured in a program or let the user choose in each case – or must user copy the info into the relevant places?).

How can we minimize duplicate entry of info?

How can we separate between info recorded for the researchers “internal” use (which could be printed in draft reports) and info intended for publication.

If we copy the information for output in reports into other structures, we want to link from a goal/objective/task to the data structures holding the recorded output, although these links would not produce output (or may some of them?) – they are for the researchers convenience.

Just an idea: Could you have a keyword/code/? (with a record id) that could be entered anywhere in the main note of a person that would import a data field (eg. analysis, conclusion) from a goal or objective record?

If you have collected, and linked to a lot of records derived from sources that turn out not to be relevant, do you want to classify some as relevant or irrelevant – or degree of relevance?

H. Data analysis

The progress against … info and the Proof … info seems to be the same. What is the difference other than one type being for an objective and one for the goal? Are we again duplicating info – will users enter the same info twice –are they likely to enter the info against the goal only? Also, the info will be very fragmented making it difficult to see all the info at the same time, but that can in theory be solved by views in the application.
Is there any difference between the two, goal and objective info, wrt to output in normal published reports (not in special reports for internal use by the researcher)?

(Are we sure that we have covered all the types of info that people want to record, i.e. are there other research process descriptions that would structure this differently?)

As has been described in the discussion about Admin01, it is vital to have a (large) general field in each of the records to hold various info (most likely for the researchers use only). One use is as a container for info converted from other structures, and another is the things we have not thought about.

NB! Also, a third is to use it to record the info in an unstructured way that would be printed in a report – using dedicated fields for the Evidence, Results, Conclusion and conflicts invites the user to enter info in a way that could require output of these field labels in reports and that will not look nice. NB!