

Citations in BetterGEDCOM – Some high level considerations

Some thoughts for discussion.

The current trend

It looks like several of the major genealogy programs in the US have, or will be, implementing support for Evidence Explained, although with differing interpretations (and possibly different source type subsets?). It is likely that this will result in a requirement to implement features in BG to support recording of a large and complex set of citation elements. It should be a relatively easy task to provide the basic technical mechanisms, but it may be more difficult to standardize the set of elements. The technical solution must transfer an open set of citation elements, templates and possibly more. There are however other complicating issues that will be discussed below.

Evidence Explained and other style guides

Evidence Explained covers the USA, and to some extent the UK. It is designed to help the user record citation elements for a particular source, using element types specific to a (few) source type(s), and thus have a huge number of citation element types – one list ([Yates](#)) has almost 600 types, but I have not verified this so the real number may be smaller. EE also defines where these elements should be included in citations (e.g. footnotes and elements in bibliographies) and how the citation should be formatted – thus describing templates for a large number of source types.

A major problem with EE is that it has not attempted to generalize the element types, as some other similar style guides have. EE has taken the easy way out, the difficult aspect of creating a style guide is to make it simple, and yet powerful enough to perform its main task, which is to identify the text (or whatever) referenced.

It will be necessary to define similar style guides for other countries, although they could be simpler, with more general elements. The set of citation elements could overlap with EE, but there will most likely be elements that are not common. There are also other style guides used in the USA (I don't know to what extent they are used in genealogy), and at least one program has implemented citation templates from other guides in addition to EE (RM). New style guides, and extensions to existing ones, will appear in the future.

Citation elements using different style guides (US or other countries) could be exchanged (possibly in the same BG file), and could end up in the same report. Depending on the differences between style guides, this may produce different looking citations in the same report.

It is interesting to note that when the elements are output in a citation, very little (or none) of the classification info appear, so the classification seems to be most important as a guidance to the person that records the elements. If it was possible to define general elements that would end up in the same position in citations, it would not matter if the specificity of the element is lost. The very

specific classification of an element type is not important, many users receiving a BG file will not look at it, they will just print it in a footnote.

Import of citation data from various databases

There are currently several solutions for collection of citation info on the internet. One solution is called [Zotero](#), which is an extension to Firefox, and is based on the “Citation Style Language”. It is used to download source data and the associated citation data from various network databases. Many more such solutions are likely to appear in the future. The same functionality can be built into genealogy programs, or these programs could import data from programs/services such as Zotero if citation elements are exported from them. It would then be possible to import citation data from databases operated by archives, libraries and document databases more or less specializing in genealogy. Such downloads are likely to improve the quality of genealogy data by increasing the number of citations.

Most of the databases where the citation data comes from use a normalized set of data element types that are much smaller than in EE. There are attempts to standardize the element types, and there is a considerable number of data elements that are common to the standards, but it could take many, many years before we see something close to a 100% harmonization(if ever). I consider it very unlikely that general libraries and archives will adopt EE for their databases, it would be a tremendous task to do the conversion. And, what about Familysearch and Ancestry, what can users download from them now and in the future– with which citation elements? If EE shall have any chance to survive in the long term in the US, it will require support from FS and/or Ancestry, and it will require that general public libraries and archives are NOT able to fund digitization of their collections so they can become competitors.

It is however clear that if genealogy programs want’s to benefit from such downloading, they must use a smaller set of more general data elements than in EE, and this set must be able to evolve over time.

A medium sized set of elements

There is a question if it is possible to group and standardize a medium sized set of generalized element types so that it would be possible to convert data from one style guide (template) to another. The set of elements will be larger than in Gedcom and considerably smaller than in EE. (Zotero has 50-60+ elements and extensions have been proposed, e.g. for archives. Evidence! has 20+ elements and TMG has about 30 “source [element] groups”) A medium sized set for BG could end up with 50-80 elements. Some will consider this too many. Since such a set will evolve over time, there must be a way to evolve it without changing the base BG standard.

Backwards compatibility

BG will also have a backwards compatibility problem, it will take a long time before the full features of BG will be in general use. It will be a complex task to try to map the large number of citation elements EE to the fields in Gedcom, and in most cases it will be impossible, as demonstrated by how programs export EE based citations today. Transfer of data to programs supporting only the fields in Gedcom will not work without loss of data. There may be a need to define a set of elements that is a little bit larger than the one in Gedcom.

Language

When templates contain fixed words, there will be problems if the templates use different language. Zotero may have solved this using a small standardized set of words, which can be translated by applications. If this is not a solution it will be necessary to translate all templates in a style, and there will be problems with user defined templates.

Conversion

One might envisage functions in genealogy (or other) programs that more or less automatically converts data between citations based on different sets of citation elements, and one might also envisage that the conversion is assisted by rules that could be downloaded (in a standard format) from databases or transported in BG. (Based on a source type, one might even envisage conversion from a generalized set of elements to the set in EE (?) or to a simple Gedcom style). When there are more than 3 styles, the least number of conversion rule sets is needed when conversion is based on a harmonized set of elements, rather than converting directly from one style to another – but the latter approach may be more accurate. Conversion may be considered a too complex function, compared to the severity of the problem.

Summary of implications for BetterGEDCOM

- BG must be able to handle several citation styles, some possibly as complex as EE, and these will evolve over time.
- Agreement on how to implement EE is needed.
- A medium sized set of generalized citation elements will simplify exchange and conversion between various services and citation styles. Also this set will evolve over time. Templates should be defined for this set.
- Programs supporting a large set must be required to fill in the fields in current Gedcom, maybe with a few extensions such as source type (and perhaps identifiers for access databases containing source citation element?), perhaps elements that the importer could discard without major consequences.
- User defined citation guides, citation elements, source types and templates could use the same technical solutions as the other style guides, but there may be a need for recording of explanatory info or other info defined in guides. Programs could give the user control over how these elements are mapped to “current Gedcom like” elements, if not, there will be problems. There will also be language problems.

The next step

Creation of a set of styles and a generalized set of elements, possibly accompanied by citation templates for the generalized set, possibly for a basic general set of source types, will be a major effort. A first step in the BetterGEDCOM project should therefore focus on the creation of general data structures that will facilitate an open set of styles, including a generalized one. It may also consider principles for creation of a generalized style (it cannot simply be a concentrate of EE), and possibly have special data structures that make the creation/use easier? Such a work should preferably be based on some practical cases. Although it will most likely be too much work to create a complete set of generalized elements based on EE, one could spend some time on a subset of it, in

order to identify possible principles, problems and technical solutions. The “Citation Style Language” and possibly other standards (e.g. MARCxx) or ISO 690 could be checked for inspiration. Evidence! and TMG contains a reasonable, but probably too small, set of elements – larger than Gedcom – that should be looked at.

Other possible requirements to a harmonized set of citation elements

- There will be a need to have a few elements (title, author, “jurisdiction”?) that are common for all source types that can be used to find a source stored in by a program.
- There might also be a need for some common fields (not necessarily citation elements) that can be used for grouping of source types, making selection of the source type easier.
- There may be a need to have elements for sorting in bibliographies.
- Could a common set of elements be useful in detecting duplicate sources and merge them?
- It should be checked if there may be any language issues, eg. fixed words used in citations.