

An architecture for sources, reference notes and bibliographies

1 Contents

2	Introduction.....	2
3	Definitions	3
4	Style Guides.....	5
5	Source types	7
6	Source Type Sets	8
7	Selection of source types, Generalization of source types and Modularization of citation elements.....	9
7.1	Selection of source types	9
7.2	Generalization of Source Types and Citation Element Modules.....	10
8	Citation elements	12
8.1	General Citation Elements.....	12
8.2	Extensions to the set of General Citation Elements.....	13
8.3	Standard Citation Elements.....	13
8.4	Specialized Citation Elements	14
8.5	Definition of Citation Elements	14
8.6	Other Citation Element issues.....	15
9	Citation Element Sets	15
10	Citation Element Data Types	15
11	Citation Elements for Source Types	16
11.1	Source Type independent Citation Elements.....	16
11.2	Source type dependent Citation Elements.....	17
12	Templates.....	18
13	Template Sets.....	21
14	Repositories.....	21
15	Translations (localization).....	22
16	Data structures in BG	22
16.1	Value structures – those holding the data for a particular citation in a BG file.....	23

17	Conversion templates.....	23
18	Publication/distribution	24
19	Citation Style Language and Zotero.....	24
20	Program operation	26
21	Various related documents:	26
22	Do we want to do this work – How - Prioritization of work items.....	26
22.1	Prioritization.....	27
23	Appendix 1. CSL definition of the top level pieces in a citation and bibliography	27

2 Introduction

The following is a first attempt to start collecting technical elements that **could** go into a solution for sources and citations. It has been written in brainstorming mode, as a basis for discussion. I have not tried to remove bad ideas, but rather to collect everything. Another purpose of the document is to try to understand the numerous aspects of a solution and try to understand some of the features of existing methods and solutions. There is a lot of work remaining. Several balancing acts may have to be performed wrt flexibility vs. complexity.

Unfortunately, due to various circumstances beyond my control, this document does not have the quality that I had hoped. There are unfinished parts, inconsistencies, especially in terminology, maybe also in semantics, and probably issues discussed twice, but I feel I have to publish a first version now.

This version does not reflect the discussions on the wiki during the last 3-4 weeks.

Note that the term citation is – at the moment – used with the meaning of EE’s reference note. I have retained the term Citation Element as used on the BetterGEDCOM wiki. Hope that is not too confusing.

The goal of the work could be to specify a proposal for

- a) a solution for exchange of information about sources and reference notes that will allow the “importer” of a BG file to print citation data coming from different BG files or his/her genealogy program, using a style guide determined by the “importer” producing a report (to the extent possible), that would work internationally (incl. Evidence Explained for the US) and could support several style guides.
- b) a solution for exchange of definitions of Source Types, Citation Elements and Templates and possibly other definitions needed.

- c) a solution that can download data from source meta data in databases (this is currently not discussed in detail in the document)

The foundation, and the most difficult work item, would be a set of GENERAL citation elements that are international and which “corresponds” to source meta data in repository databases and other databases.

In addition a technical solution is needed for transfer of data (probably not a big problem), and one or more sets of source types could be developed based on the general elements.

This document is based on conclusions in the document “[Geir's high level considerations on citations in BetterGEDCOM.pdf](#)” published on the BetterGEDCOM wiki, and thus, goes further than would be required to simply implement a solution for support of only the book Evidence Explained (EE) by Mills.

The document is also based on the scenario described in “[From repository meta data to BetterGEDCOM and reports – A future scenario](#)” which tries to describe **possible** elements in a future solution.

And, finally, it is based a lot of work that that has been put into solutions for sources and citations in current programs, and current style guides. See also [BetterGEDCOM support for Evidence Explained...](#)

Some (incl. the author of this document) are likely to comment that the solution is very complex compared to Gedcom. However, the reality, in terms of implementation in programs, is that it is not much more complex than what can be found in the leading genealogy programs today.

It might be necessary to read this document twice in order to understand everything. Some definitions can be found on the wiki. I have not yet had a chance to check the document for consistency and have yet not checked if there are any additional requirements that may be needed to handle sources for my own country – so more work is required.

Suggestions for improved terminology are welcome.

There may be additional useful ideas in existing genealogy programs.

An important clause is “Do we want to do this work – How - Prioritization of work items”.

3 Definitions

This list is not complete. (Not in alphabetic order.)

Reference note (from Evidence Explained) – A footnote, endnote or an Inline note identifying a source, where in source, an extract/summary of the cited text and an argument. Inline notes usually contains only a reference to the source, and possibly where in source. Footnotes and endnotes may also just contain an arbitrary comment or detailed explanation etc, not referencing a source. (The Chicago manual of style uses just “note”, but that has a different meaning in BG. --- Check if a Bibliography entry is a Chicago “note”)

Bibliography – A list of the referenced sources and their repositories.

Citation Element – this term may later be changed to Reference element - SEE BG DEFS

Citation Style Language (CSL) - An XML based language used to specify how Reference Notes and Bibliographies should be rendered according to some style guide. Many styles (e.g. several for the Chicago Manual of Style) have been specified using CSL. Provided that a specified set of Citation Elements are used, a user can select an arbitrary(?) style for the rendering, and also the (spoken) language to be used. CSL is used by Zotero and other programs, see later in this document. For those interested in more detail about CSL, see later in this document and read the CSL specification, or my summary of it– you can get a good picture of the issues it handles, even if you are not a technical expert.

Style Guide Rule Set (Term invented by the author, there is most likely some better term) – The rules (with exceptions) that describe how Reference notes and Bibliographies should look like, as specified by a Style Guide. This set of rules is thus a subset of the content in the Style Guide, as most existing Style Guides also gives rules about things not related to Reference notes and bibliographies. Most (all?) of these rules are independent of the types of documents that are cited.

Source Type – Examples of source types are Church records, Books, Journals, Census records and many more. Although source types are in general referred to as e.g. a Census Record, Source Types may be specific for e.g. a country and time period, e.g. a Census for England, Norway or the US are different source types, and different Source types may even be defined for specific years. Many other criteria that would be used to distinguish between source types exists.

Source Type Definition – Specifies the criteria that distinguishes a Source Type from others, and specifies the citation elements that may be recorded for a Source Type, and specifies more – TBD.

Source Type Set – A collection of Source Type Definitions for e.g. the source types used in one or more countries, or in some cases a repository, or those defined by a Style Guide book or some other context. Note however that a Source Type Set is in principle independent of Style Guide Rule Sets, so the fact that Source Types Sets may appear in Style Guide books together with a Style Guide Rule Set, is just because they are in the same binder.

Source Type Class – Examples are books, census, church record, passenger list, map, military records etc. There may be many Source Types in each class, each identified by certain properties such as author type, medium, type of geographic area covered etc.

Template – A formal program readable specification of some aspects of how a Reference Note or Bibliography entry should be rendered for a specific Source Type, according to a specific Style Guide Rule Set.

Template Set – A collection of Templates for the Source Types in a Source Type Set, based on a specific Style Guide Rule Set. There may in theory be several Template Sets per Source Type Set, each for different Style Guide Rule Sets.

Citation Element Module – A collection of one or more Citation Elements that are included in a Reference Note or Bibliography when the Source Type have a specific characteristic, e.g. published

on a certain media, or being a republication, or a transcription etc. (The usability of this concept is still to be proven.)

Template Module – A definition of how a Citation Element Module shall be rendered in a Reference Note or Bibliography. Could be used to generate part of a template, or could be an “addition” to a “core” template for the source type.

4 Style Guides

There seems to be no precise definition of what a style guide is, and the topics of current style guides varies somewhat. Examples of style guides are The Chicago manual of style and Evidence Explained, but there are others (Some current styles guides are listed on the wiki). Both these guides describe how Reference notes and Bibliographies should be written for a number of source types, but also provide more general guidelines for writing – and EE also touches genealogy research guidelines. For the purpose of this document I have limited the scope of a Style Guide to how Reference notes and Bibliographies should be rendered, and called this subset of the guide a Style Guide Rule Set (any better suggestions?) – although they may not be clearly listed as general rules in a Style Guide.

Style guides seems to be self-contained (e.g. they define their own citation elements, source types and citation elements per source type) and are in general not compatible in terms of these three types of content. There are also several versions of the guides, which indicates that the guides may also change in the future.

Observation: However, CSL seems to have been able to define citation elements and source types that are independent of a large number of style guides, but leaves it to the specifications based on CSL to describe how each element should be rendered in a Reference note based on a specific style. CSL, Zotero and MS Word 2010 and most likely some/all specifications do not specify the elements to be used for a source type – they may let the user choose a source type, but lets him choose the elements to use, and then formats them according to a style. Styles are in principle independent of the individual source types. This should also be possible in genealogy.

Thus, I have made two assumptions that may or may not be realistic:

- a) All style guides supported by BG must be able to use a general, reasonably sized, set of citation elements
- b) Source type names and definitions, and the set of citation elements that may be present for a source type, will be independent of style guides – the guide thus only controls the layout of reference Notes/Bibliographies as specified by its “Style Guide Rule Set”.¹

If this is not realistic, general citation elements and elements per source type must be defined per style. In any case, it is unlikely that one Style Guide will contain all Source Type Sets and corresponding Template Sets covering all countries or source types. The advantage of having the

¹ In theory, if e.g. two Style Guides use different General Elements for what is essentially the same or similar type of information, this will not work – so the challenge will be to select General elements that prevents this. It could also be that it would be necessary to adjust rules and content of style guides to fit with a more general approach. The only way to find out is to start on the work with an assumption that it can be made to work, then find the problems, and then solve them.

elements independent of style is that the user could select one style for a report independent of the origin of citation elements.

BG could support an open ended set of style guides.

Observation: It may be that most genealogists, in many countries, would be satisfied by a style based on The Chicago manual of style (as EE is). Nevertheless, I have chosen to keep the concept of a “Style guide definition” as below, so that there can be several, just to see if there is any need for it. My personal opinion is that things can be improved, for example by reducing repetition of information when the same source is cited many times or when many sources are stored in the same repository, saving paper and making the material less boring.

It is assumed that various organizations may develop BG-Style guides. BG could define requirements to such guides.

It may be that the work discovers things that current Style Guides does not handle well, perhaps in the area of digital or digitalized sources. We may have to fix things.

The following is a check list of what could be specified for a Style Guide, but it is unclear to me if there is a need for a Style guide entity/specification in BG, so the following is retained just to make sure we have considered every aspect when we say that a program supports a style guide.

- a) Style guide name
- b) Style guide issuer info (name, contact info, URL)
- c) Style guide version
- d) Publication date
- e) Style guide UUID (unique per name and version)
- f) Style guide language (language in Reference notes/Bibliographies, each set of templates for a style guide could have one language, or templates can be language independent - TBD)
- g) Description of the style guide (per version)
- h) URI for download of guide definition file
- i) Style guide documentation (for download (not in BG file))
- j) The source types it applies to, and specification of the information related to each source type (incl. citation elements). This should be replaced by references to separate Source Type Sets.
- k) ((The (Extended) general citation elements specific to the style (in addition to standard elements, and “General Citation Elements” maintained in a specification outside the standard).)) --- not the right place for this
- l) Citation Template Sets for each Source Type Set, (each set has an assigned language TBDiscussed – it might be possible to define language independent templates). This should be replaced by references to separate Template Sets.
- m) Possibly various parameters controlling the appearance of various bits and pieces in templates – across all templates in a set.
- n) Possibly various additional rules for formatting that applies independent of templates.
- o) Translation (localization) information for various languages/countries. Can also be style independent as in CSL.

- p) Identification of other specifications that the style is based on (not discussed in detail), possibly by modifying/extending an existing style.
- q) More?

m) and n) would most likely be defined by reference to an existing style guide, and it would be useful to identify the relevant parts of such guides. There is a need to check if there are language/culture issues.

It could be that a structure's sole purpose could be to identify a style guide, i.e. not provide any rules or other content found in a guide.

Each style guide (and other structures in this specification) should have its own "namespace" or e.g. source types, in order to avoid name clashes.

Could there be copyright issues related to publishing the essence of EE in a standard style guide?

(Genealogy programs could define their own styles (tailored to how they currently work), and transfer definitions thereof, or just citation templates, and require/expect the importing program to use it – that is not a desirable situation unless it can be ensured that the citation printed in a report adheres to the same style that the importing program's user want to use, and even then there might be other complications with varying definitions of citation elements and source types.)

5 Source types

For each source type, the following could be specified

- a) Source type ID (ID within Source Type Set)
- b) Source type name
- c) Source type name aliases (a source type may be known by several names).
- d) Source type short name??? (for selection lists, check if any program uses this)
- e) Source class (e.g. census, maybe subclasses, see also f), for selection)
- f) Source type selection criteria (e.g. media type, list TBD, used by the user to select the source type, see below.).
- g) Source type short description (typically max 100-200 characters, for use in the field before anything is entered, see RM/Legacy)
- h) Source type long description
- i) Ordered list of source related citation elements, see separate clause
- j) Ordered list of "where in source" related citation elements, see separate clause
- k) Is subordinate to source type (if multilevel source types, may depend on e), see below)
- l) Reference – reference to where in style specification, e.g. clause/chapter/page (cf. RM)

Multi-level source types should be considered. Used for e.g. an article in a journal, also most archives use a tree structure – so you could store info about archive entities higher than the single source (cf. Genbox). Elements from all levels must map into a single citation template. There will most likely be restrictions on which Citation Elements can occur at which level. Could map from a program internal multilevel structure to a single level structure in BG, but preservation of the multiple levels could be useful when transferring between one users programs (or one might envisage source meta data

catalogues exchanged in BG files). Should two levels be output separately or be merged in biographies? Probably merged.

Users could in theory receive, and end up with storing, differing definitions of the same source type. This might require special functionality when merging sources. But several source types for the same source may be rare.

(Source types may be generic or very specialized. Is there any benefit in defining more general, rather than specialized, types?)

Why have source type definitions in programs: Two situations 1) entering source info 2) receiving citation data in a BG-file. 1) Let the user enter appropriate citation elements 2) May not be required if citation templates for the source type is available. Could give the user information about an imported source type – a description – ((but this could also be carried by a source record (rather than a source type def) although most likely creating duplicates)).

Why exchange definitions of source types: Users can exchange information about source types that is not initially supported by their program, possibly source types relevant to a foreign country. It could also make it a bit easier for vendors to implement large sets of source types, and handle updates.

Alternative solution: Vendors do not need this exchange to make things work, and could let their users define source types, but interchange between different programs MAY (TBD) require the same source definition to be entered into or being supported by the other program. Publication of a standard for source types that users and/or vendors would need could be done outside a BG-file, and will then not require recording in BG.

Consequence of if this or alternative solutions are not implemented:

Benefits of this solution compared to alternative: Users and vendors don't have to type in source definitions. If a definition is not built into the program, users must have a document describing the source type.

Dependencies: Definitions of citation elements

6 Source Type Sets

A Source Type Set describes a collection of source types, that appear in some context, typically a country or maybe an archive /repository. It specifies citation elements applicable to the source types and other info relevant to the types. It is independent of Citation Styles. And more ...

It is assumed that various organizations may develop Source Type Sets, and at least one Template Set for it. Special cases are those defined by users.

There may be several translations of Source Type Set definitions, and the Source Types within them.

For each Source Type Set, the following could be specified:

- ID (unique per version)
- Name/Title

- Issuer
- Issuer URL
- Date
- Context (Country, archive etc.)
- Description
- Version (each version need an ID)
- Language
- Translator
- Original specification ref (if translated)
- (One or more styles that it may be used with)
- Source type selection criteria
- List of Source type defs
- Special Citation Element Set, see separate clause
- Extended General Citation Elements Set (if any) --- check terminology

7 Selection of source types, Generalization of source types and Modularization of citation elements

The current approach in programs and Evidence Explained seems to be to define new source types for the same source type when it is republished, published on a new medium, transcribed, etc. This has undesirable implications:

1. A large number of source types must be defined
2. An even larger number of templates must be defined
3. It becomes more complicated for the user to choose from a long list of source types.

7.1 Selection of source types

The most important issue is the last one. If it will be possible to transfer definitions of source types and corresponding templates, it must be possible for the receiving program to let its user select source types in an efficient and user friendly way. One program which with a reasonably good functionality for source type selection is Legacy, which has classified the sources into 55 groups (which I have called Source Type Classes), and after having selected one, the user has to make up to three choices (one being medium) in order to select the source type. The other two choices depend on the Source Type Class.

There seems to be two alternatives ways to implement a stepwise selection mechanism

- A multi-dimensional matrix, where a dimension could be Source Type Class, medium etc, each resulting in a choice.
- A “decision tree” where a choice presented to the user depends on previous choices made, the choices being similar to the dimensions. This alternative seems to be the more powerful and user friendly.

It is a question if it is possible to define a set of choices (and alternatives) that could be used across Source Type Sets, or if each set has to define its own – probably the latter is necessary, but this will

be less user friendly. One possibility is that after going through the choices, a small list of Source Types will be presented to the user.

In order to determine this, a list of choices (**Source type selection criteria**) must be compiled:

- Country(ies)
- Source Type Class
- Medium
- Author type
- Original/Transcript/Index
- More ..

It must be possible to exchange definitions of Source type selection criteria in a BG file.

Is it possible to “merge” (some or all) criteria from several Source Type Sets in the user interface so the user can select between source types in several STSs.

(There could perhaps be Source Type sub-classes?)

Several choices might end up with the same source type.

A structure must be defined to transfer the selection criteria, but the overall mechanism must be decided first.

Source Type must in many cases also be selected if meta data is downloaded from an internet database.

7.2 Generalization of Source Types and Citation Element Modules

I have not checked, but I assume that most other sciences have managed with a much smaller number of source types than what can be found in EE, and the reason why EE have more types is not only because there are more original source types – it is also because it lists a lot of variants depending on the source type selection criteria. Some of the principles/rules for citation are built into the selection criteria. (It may even be a question if EE has described all the possible combinations, most likely not (Yates has 170 source types, 510 templates, there are about 400 in EE - about 1200 templates, TMG has 18 EE templates for a Book.)

One might envisage that the number of source types could be reduced substantially if the source type selection criteria were carefully chosen, and having a base set of citation elements for a somewhat generic source type. The generic source type could then be specialized for e.g. a specific media, by adding a number of elements needed to cite a source published on that media, and by recording the type of media together with the other elements. I have invented the term “Citation Element Module” to hold the citation elements to be added for e.g. a specific medium. Each value of a choice used to select a source type, see previous clause, could have a corresponding Citation Element Module with elements to be added, and a corresponding Template module. Thus one might envisage a much smaller number (than EE) of generic source types, independent of Source type selection criteria. (Also, one might perhaps envisage general source types being independent of style guides, although the citation elements could or could not vary from guide to guide.)

Citation Element Modules are inspired by Citation Style Language (CSL) where each citation is built up of high level modules (called macros, having if/else capability) which shows that it is possible to have a limited set of source types as long as the approach is structured². It might not be easy to understand, but the top level modules (macros) for a Chicago manual style is shown in Appendix 1. Some macro-like, but simpler, functionality can be found in the languages used by some genealogy programs to define templates.

A simple example would be a generic source type “Book” (having elements title and author) that when published on the media type “Internet” would also have a Citation Element Module type “media=internet” added, containing the elements URL and retrieval date in a Citation Element module.

Also, consider sources that are copies (or transcript, original, index) of another source, the generic source would be those for the source of the source and additional elements would be added depending on the media/publisher of the copy (the latter being building blocks of citations elements independent of the original source type).

There may be other source type selection criteria than media or copies (e.g. one or multiple authors, country or local level, container – the book/journal an article is contained in, republication, published/not published, copy repository&/original repository), where different values gives rise to a lot of different criteria specific source types. Such criteria could be found by comparing the various variants of citations for e.g. church records in Evidence Explained. See also the source type selection functionality in Legacy, (and maybe also the spreadsheet published by [Yates](#)). Other Citation Element Modules could be found by looking at the specifications of CSL based styles in Zotero.

Citation Element Modules could be used in two ways

- 1) As a tool or set of rules to be used when developing Source Types and corresponding templates. This would most likely not have technical consequences of its own, so the number of source types would not be reduced, and more work must be put into standardizing the source types. (However, one might envisage a macro/include/ preprocessor tool that would do much of the style definition work automatically, based on defined modules and selection criteria.)
- 2) Citation Element Modules and Template Modules could be exchanged as separate pieces of data in BG files, so that the importing program producing a citation in a report would put together the relevant generic source and the modules.

Alternative 2) was my initial approach, but it seems a solution as in 1) with many source types, a good source selection mechanism and a powerful template language would be simpler to implement in programs, and be more flexible wrt changes – I hope I am right (and looking at existing programs, I am probably not the first one to come to this conclusion.) It could also mean that any work on Citation Element Modules might be done “later”?

² It seems that the people working on CSL have a much better understanding of the generic building blocks in citations than is currently evident in specifications of how citations should look like for genealogy sources.

(It might not only be the set of citation elements that could depend on such rules, but also the formatting of parts of a citation? E.g. in EE (2.22) (based on Chicago) the title of published work should be in italics, otherwise no italics but in quotation marks unless it is a title you created because the source has no title.)

Could also group elements that are all always present in source types, if one of the elements in the group are present. The presence of a CE module could in general be made dependent on the presence or value of a CE.

8 Citation elements

Citation elements can be more or less general or specific. General elements apply to many Source Types and specific elements apply to one, or a few, Source Types. Evidence Explained have few general Citation Elements, and many specific ones (500+) since they are defined in the context of each Source Type (or a few Source Types) – this makes it very easy for the user to find out how to record citation elements for a source. Most other style guides have considerably fewer Citation Elements. CSL, which handles many style guides, has approximately 50 elements (called variables) that are independent of the styles. Evidence! has about 20. Most databases, in libraries or archives, containing meta data about sources use a relatively small set of elements.

Three types of CEs, and their relations, are define below

1. General Citation Elements
2. Standard Citation Elements
3. Specific Citation Elements

8.1 General Citation Elements

If all style guides were to define specific citation elements for a large number of Source Types, worldwide, there would be a huge number of elements – many thousand. It would be very difficult to map elements from one style to those of another style, and also difficult to exchange data with other solutions for citations or databases holding source meta data.

It is therefore desirable to develop a set of General Citation Elements for BetterGEDCOM, that are independent of Style Guides, Source Types and cultures. Since CSL has about 50 elements, it might be possible to limit this set to about 80 elements (this is just a guesstimate!).

General elements should as far as possible be independent of country or culture, e.g. geographical/administrative entities must be generic – most likely just a hierarchy of names. Geographic/cultural dependencies could be handled by Specific CEs – if they can be ignored by the importing application (see below).

It is necessary to define a plan for how general elements can be developed, so we don't embark on work that we cannot finish.

Development of general citation elements used internationally will require a lot of work with obtaining information about "citation elements" used in archives and libraries in various countries.

The possible existence of standards for archives should be investigated. Do we have the necessary expertise – well, anyway we have to start somewhere, and assume that the expertise will join in at some stage. The work will most likely also require a lot of compromises.

The “Citation Style Language” and possibly other standards (e.g. MARCxx) or ISO 690 could be checked for inspiration on how to create generic elements. Evidence! and TMG contains a reasonable, but probably too small, set of elements – larger than Gedcom – that should be looked at. TMG also has 34? General elements. GeneJ has made an initial attempt to generalize the elements in EE, see <http://bettergedcom.wikispaces.com/Citation+Elements>

Principles for design of BG generic elements should be considered. What are the factors that determine if a set of specific elements can be mapped to a generic element? Are there any other reasons for having specific elements than making it easier for the user (in the context of the style type)? What are the (negative) consequences of having too specific elements? What are the problems if an element becomes too general? Does it matter too much how precisely elements are defined when transferred in BG, as long as they end up in the correct place in a citation?³ Defining general citation elements will be a balancing act, which may require some compromises.

General CEs could be defined in a document separate from a standard so it can be updated more frequently.

8.2 Extensions to the set of General Citation Elements

It cannot be ruled out that some source types, somewhere in the world, will require definition of Citation Elements that are not contained in the current set of General Citation Elements. There must be a way to transfer definitions of such elements, and that is probably best done in the context of a Source Type Set ((or Style Guide?)). The number of such definitions should be minimized. Future revisions of the set of General Citation Elements may include these elements. User defined elements is another example where extensions should be allowed. It may be necessary define “Standard GCEs” for those defined by a BG document.

8.3 Standard Citation Elements

BetterGEDCOM must be backwards compatible with Gedcom, and cannot (initially) expect all programs to handle all features described in this document.

A few elements (Gedcom 5.5 elements + perhaps a few new ones?) could be specified by the standard (Standard Citation Elements).

Candidates for new standard elements are: URI for source meta database, source type, ISBN/ISSN. These extensions should be droppable (not be required to import)????

I have not worked much on how Standard CEs could be generated from GCEs or CEs specific to sources. One possibility is some sort of template, but it may perhaps be possible to design a general solution based on general CEs only, or even simpler, the Standard CEs could be just a subset of the GCEs.

³ One test that should be easy to perform programmatically is if a general citation element is made so general that it would end up twice in a citation. This may be easy to test in a database, based on a mapping of Yates 580 EE citation elements to a set of general elements to be tested.

It might be necessary to always record Standard Elements in the BG-file (possibly causing duplication of info), or a GCE could be accompanied by a value indicating that a CE is also a Standard CE. A template solution, if that's the solution, must be implemented in the exporting program, not the importing.

The simplest solution might be to just use the Gedcom 5.5 elements for what they were intended for, perhaps fix the definitions to be more precise, and align them with general elements – and not allow concatenation of a lot of things into the fields, as many programs currently do.

It may be that the use of Standard Citation Elements, only, may not transfer all the info entered about a source and citation in an advanced program, but that is the prize of being backwards compatible. Something is better than nothing.

8.4 Specialized Citation Elements

In order to bridge the gap between the large number of specific Citation Elements in existing implementations based on Evidence Explained and a set of General Citation Elements, Specialized Citation Elements may be defined, based on General Citation Elements - a Specialized element is a subtype of a General one.

Specialized elements have the same Data Type as the General one it is based on, and its definition refers to the General element. The Specialized CE may have its own Long and Short Descriptions that overrules those of the GCE. Other parts of the GCE definition that are inherited, or others may be overruled are TBD.

Specialized CEs must occur in the same place as the corresponding GCE in Reference Notes and Bibliographies. Templates are based on the corresponding GCE, not the Specialized element.

When a Specialized CE is transferred, it is always accompanied by the ID (and possibly name?) of the corresponding GCE. If an importing program has no knowledge of the SCE, it may ignore the SCE and use its knowledge of the corresponding GCE instead.

Specialized CEs for EE will require agreement among program vendors on how to interpret/implement EE.

It will be useful to maintain a list of Specialized CEs used for a GCE in various STSs, in order to ensure consistent use of Specialized CEs and also utilization of the GCEs.

8.5 Definition of Citation Elements

For each citation element, the following is specified:

- a) Citation Element ID
- b) General Citation Element Reference (Only for Specific Citation Elements)
- c) Name (used as prompt in the user interface)
- d) Acronym (may be needed in templates, uniqueness in some context necessary)
- e) Short description (RM and Legacy displays this in the field where data is to be entered)
- f) Long description
- g) Data type (only for General Citation Elements)
- h) Choices, only for GCEs (if data type is choice, see data types, for e.g. author type)

- i) Can have short value (see “Citation Style Language” --- more later)
- j) For use in “source” or “where in source” (??needed?? Most likely yes)
- k) (Reference to “where” in defining document, e.g. Style Guide/EE)

Translations of Citation element definitions

8.6 Other Citation Element issues

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 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?

BG should develop a mapping table with GCEs and SCEs for EE, assuming one definition of EE.

9 Citation Element Sets

Citation Element Sets is a way to record a collection of Citation Elements, either the collection of General Citation Elements or the Specific Citation Elements defined by a Source Type Set.

For each Citation Element Set, the following is specified:

- a) ID
- b) Name/Title
- c) Issuer
- d) Issuer URL
- e) Date
- f) Description
- g) Version (each version need an ID)
- h) Language
- i) Translator
- j) Reference to original version (if translated)
- k) A number of Citation Element Definitions

There may be several translations of a set.

10 Citation Element Data Types

RootsMagic has data types for citation elements. Each element must be assigned one of the following data types:

- a) text
- b) date (for validation of dates on input, control format on output, extract year), what about periods, is calendar an issue?)
- c) name (for extraction of surname only on output, reverse name, problems with international rules for reversing and locating surname in a name, CSL has a structure for names that would handle e.g. where to insert commas in reversed names using different types of prefixes)
- d) place (reverse name parts, extract short name, link place to place record?)

Other possible data types (not in RM):

- e) URI/URLs, E-mail – clickable in electronic documents. (Possibly linked to a BG-structure for URLs) Word 2010 and maybe earlier versions support links in footnotes.
- f) single choice – accompanied by a list of possible values – values are text (the possible values are specified in the CE definition) (one use might be author type)

Programs must implement defined rules and functions that are specified for a data type.

Some of the functionality provided by data types could alternatively be solved by splitting a citation element into several text elements (e.g. day, month, year or given name, surname). Legacy has chosen this approach for names. It may be more flexible for e.g. names since it does not hardwire handling of names in a program (as a data type does) but will occupy more screen space (more elements). Some functions related to data types cannot be handled by splitting into fields.

CSL should be consulted for other possible data types. (Some ideas for types are Integer, Identifier (id type, value – for e.g. ISBN/ISSN/URN rather than having separate elements for each possible identifier scheme (i.e. not separate CEs for ISBN and ISSN, used for e.g. database lookup), Sequence (hierarchy of values, printed/entered with some delimiter, for hierarchies in archives and geographical areas – with an indicator of ascending/descending order when recorded) but these may not be needed?)

Data types must be independent of culture, cf. Name and Date.

((The allowed character set for a data type is probably not an issue. Also for names of source types etc.))

It may be necessary to reserve characters for special purposes, such as separators in lists, or characters to isolate specific parts of names.

11 Citation Elements for Source Types

11.1 Source Type independent Citation Elements

For each citation there are some elements that are independent of source type (they can appear for all source types, but they are not listed per source type):

- a) Reasoning/rationale/analysis/conclusion (written by the researcher)
- b) Abstract/summary
- c) Extraction (text transcribed from the source)
- d) Free text (for reference notes that do not contain references to sources)
- e) Repository info (several elements, not discussed in this version of the document)
- f) No-print (This is not really a citation element. If the citation is not intended for printing, but only to link data about persons etc. to evidence record(s) – see my working document on Evidence&Conclusion.)

The variants in a) have been discussed in the Research process discussion on the wiki. Some claim that this is not according to Evidence Explained or Chicago, but that is not correct in my view. Most programs support notes and extracts/abstracts for the source and the citation – and allow output of

them in Reference notes. I am personally not too concerned with any limitations that style guides may impose, rather I would include what I see as useful. An alternative for naming such elements could be to call them prefix and suffix, describing where the value would appear in a reference note, or use even more generic names so that templates would control where they appear. (An issue is their location in a Reference note referring to several sources.)

Citation templates are likely to place restrictions on where a), b) and c) can occur in a Reference note – I have seen examples from GeneJ where this type of information is located at almost arbitrary locations within a reference note, relative to the Source and “Where in source” elements. The Chicago manual places comments/reasoning (other stuff than what identifies the source(s) and where in source) at the end of the reference note, while some programs can also place such text in the beginning. (In order to handle such reference notes, the note would have to be an arbitrary string of text entered by the user, where you would use field codes (enclosed in square brackets in many programs) to tell where the Source, “Where in source” and repository related elements are to appear (or possibly with one field code that indicates all three as a unit, as defined by some template). Such an approach will have problems being translated or presented in an arbitrary style.

How realistic is translation of reference notes if a large part of it is free text. Is there a need to be able to record the source type independent CEs in several languages?

Location of these elements in Reference Notes referring to several sources is an issue.

11.2 Source type dependent Citation Elements

In addition there are two ordered lists of citation elements specified for each source type,

- 1) those related to the source and
- 2) those related to “where in source”.

For each element in these two lists, the following is specified in a Source Type Set:

- a) Citation Element ID (General or Specific)
- b) Is short version (see “Citation Style Language” - needs more work)
- c) Max number of occurrences (repeats) of this element in sequence (e.g. several authors)
See below.
- d) Valid choices (see CE Data Types, may not be desirable per source type, may not be needed.)
- e) (?Default value? – just an idea)
- f) Short and long descriptions (what is best, to have these in Specialized CEs or here?)

It might be possible/necessary to define applicable Citation Element Modules (see above) which could be imported as a group into the lists of citation elements for each source type. (The reason will be to harmonize the layout across source types.) Thus, the list of citation elements above would also be defined for Citation Element Modules.

Where there may be several occurrences of a citation element (e.g. several authors), an issue is if this should be transferred as several separate data fields, or a list of values separated a separator character and space in one field. Choice of separator might be an issue, see CE data types

12 Templates

Templates are used to produce the output of Reference Notes and Bibliographies, based on one citation style, source type and the values of citation elements – and more - rules. Several genealogy programs currently have built in templates and some allows user defined templates, and others have them more or less hard wired in the program without using templates.

Several Citation templates may be defined for each source type in a template set (the terms used below are most likely not correct):

- a) Inline citation (e.g. [Hanson 1978]) (Possibly with where in source)
- b) First citation
- c) “Secondary”/Short citation – following the first citation to the source (and where in source?)
- d) “Referring” citation (term invented by the author, referring to a bibliography for the source info, in order to reduce the amount of text by not repeating the source info in every citation).⁴ Various alternatives should be considered. One possible variant is produced by Gramps, a combination of endnotes and bibliography where each bibliography entry is numbered and has sub items a) b) etc. containing “where in source” so the superscript reference is e.g. 1a. The purpose is to save paper. Some use the author date format in a foot/end note to refer to the Bibliography, and adds where in source in the foot/endnote – but it might necessary to also use other things than author since our sources do not always have author (need to check EE on how this is solved in e.g. Bibliography sorting). Some authors use an acronym for the Source title in the Reference note, that refers to a Bibliography sorted on the acronym (Geir’s ref: NST).
- e) Bibliography entry
- f) Templates for production of Standard Citation Elements (see separate clause above). This template type may not be needed.

Regarding f): There could be one special template that specifies how the elements for a source type shall be mapped into Standard CEs. This type of template is called a “conversion rule” in the scenario document referenced in the beginning of this document. All programs must export these Gedcom fields (but they could perhaps be encoded as ordinary elements in order to avoid duplication of information??). This type of template may be a group of templates, one for each Standard CE, or one template could produce a string with keywords and data that can be easily split into the Standard CEs – there are several ways to do it. As mentioned elsewhere, it might not be necessary with a temple.

Citation templates are contained in Template Sets.

Citation Templates are based on General CEs only, not on specialized CEs.

Citation templates could have the following info defined:

- 1) Template ID (within set)
- 2) Source type ID (that the template applies to)

⁴ In populations that are stationary (i.e. migration is not common), e.g. Norway pre 1900 (except emigration to America), the same sources such as church records, censuses, probates and land records will be cited MANY times in a genealogy report.

- 3) Type (Inline, First footnote/citation, etc. see a)-f) above)
- 4) A sequence of “citation element references” (see lists of citations per source type), “translated terms” (standard words (e.g. “Page”, “Vol” that can be translated, see CSL below)), characters (letters, punctuation) – and more
- 5) Conditional structures - Text/element output may depend on presence of element and perhaps also element value, possibly with alternative (e.g. if no value or a specific value or any value).
- 6) Template modules (building blocks, analogous to other modules described for lists of elements per source type, see also macros in CSL). The same module could apply to several citation types, e.g. in footnotes and bibliography entries. (Consider the more powerful CSL macros if needed.)

Other features that must be considered (far from complete):

- I. Special handling of repeatable elements?
- II. Formating - bold, italics etc. (font?) (possibly conditional?)
- III. Data type dependent functions (e.g. extract year from date, see data types)
- IV. Separate Name spaces for elements, and also for translatable terms
- V. Escape sequence for output of otherwise reserved characters
- VI. There may be a need to identify elements (or parts of) for sorting in bibliographies, e.g. author name.
- VII. Localization of special characters (e.g. punctuation, list separators)
- VIII. See “Source type independent Citation elements” above
- IX. Ibid
- X. Capitalization
- XI. CR/LF/TAB
- XII. Quotes (varying styles?)
- XIII. Hanging indent (should be defined by the program, not the template)
- XIV. Comments (for debugging, not output – see Genbox)
- XV. Number style (4 vs. IV)
- XVI. Multiple sources in one citation, sharing reasoning.
- XVII. Two? Variants of “source of the source” – a copy/transcript or referenced by another author (cf. EE 2.21)
- XVIII. Hierarchies in archives ...provenance/jurisdiction ... from higher to lower or reversed.
- XIX. More ...

A number of other template related issues can be found in CSL – more work needed. The languages used by genealogy programs for template definitions should also be looked at in more detail. Also consider any options that programs might have for controlling the layout/appearance of a citation in a report – other than selecting between footnotes and endnotes. Do current programs support all “features” in EE wrt how a citation or bibliography should look like?

The nitty-gritty details of some of the layout (and perhaps not so important) features, e.g. “et. al.”, *ibid?*, sorting in Bibliographies, Hanging indent, are probably best left to programs to sort out themselves, and some is preferably controlled by the user when generating a report. This just has to be sorted out wrt what might need support in templates.

See discussion of complexity in the template mechanisms versus a large number of templates in the clause about Zotero below.

Some aspects of the definition of templates could be too complex to do for ordinary users. Some features could be restricted, that is, not to be used in user defined templates ??? But the languages used by current programs are also not simple to use. Vendors will not like too much complexity.

Transfer of preformatted complete citations, as some programs export today (RM), are not considered a good solution since they cannot be translated to a different style or language without a lot of manual intervention, thus producing inconsistent reports.

Template modules could be implemented by letting a template include another template, it might ease production of templates, and reduce file the size – but can also be done by preprocessing before the template is published. “Include” keyword (cf. programming languages) might be useful for user defined template.

Genbox has a “Global template” used to specify where Source type independent elements should go if not contained in the source type specific template – useful?.

RootsMagic can override templates with free text, but this will be a problem if data is exchanged across languages.

((A question is if it might be possible to do without templates, and just have source type independent rules for where GENERAL citations elements should end up in citations, (if they are present). Most likely not?))

Why have citation template definitions in programs: Allows the user to select the output style, provided that citation elements from a different style can be used. Allows the users to define their own styles. Allows output in a different language than was used when the citation was entered, but abstract/reasoning would have to be translated.

Why exchange definitions of citation templates: Users can exchange templates for source types that are not initially supported by their program, possibly source types relevant to a foreign country. Templates can be exchanged together with a Source Type Set, or only those relevant to the data in the BG file.

Alternative solution: Vendors do not need this exchange to make things work, and could let their users define templates, but interchange between different programs MAY (TBD) require templates for the same source type to be entered into or being supported by the other program. Publication of a standard for templates for source types that users and/or vendors could use, would be needed, but would not require recording in BG. Alternative: It may perhaps be possible to design source types (or a few groups/classes of source types) independent citation templates using CSL, which would only look at the citation elements present –based on general rules for where the various citation elements are supposed to appear in citations. This is the approach in CSL. Alternative: Transfer the citation as it should print (cf. RM) – if a citation element is to be changed, the citation must also be changed. The user cannot change the style, and the citation will be language dependent.

Consequences if this or alternative solution is not implemented:

Benefits of this solution compared to alternative: Users and vendors don't have to type in template definitions. If a definition is not built into the program, users can download it or get it in a BG-file.

Dependencies: Definitions of citation elements, definitions of source types

13 Template Sets

Template Set – A collection of Templates for the Source Types in a Source Type Set, based on a specific Style Guide Rule Set. There may in theory be several Template Sets per Source Type Set, each for different Style Guide Rule Sets.

A template Set will most likely be defined at the same time as a Source Type Set, but it would be possible for various organizations to develop Template Sets based on Source Type Sets developed by others. Special cases are those defined by users.

Templates may also be translated into various languages, there should be one template set per language. Alternatively the concept of localization in CSL could be used – this would probably make it easier for users because they will not have to locate and obtain the relevant version for their language, and there would be no need to create translations of Template Sets (see also Translation below).

For each Template Set, the following could be specified:

- a) UUID
- b) Issuer
- c) Issuer URL (and/or directly to (related) Template set)
- d) Date
- e) Description
- f) Version
- g) Language (language used in the rendered Reference note)
- h) The Source Type set it applies to
- i) The style guide it is based on
- j) List of Template defs
- k) --- possibly more, see style guides

14 Repositories

Needs more work ...

One interesting thing from EE 2.33: A website is not a repository...hm. What about a websites with e.g. books published only on that site?

EE 2.31 Copy/original repository. EE 2.33: Info for recovery of a broken link.

Repository name acronym (Geir's ref: NST)

Original repository vs. Copy repository (cf. EE 2.31) – and which of these to include in a citation.

Some styles allow repository only in Bibliographies, while some authors place them in Reference Notes (Geir's ref: NST).

A source can be located in several repositories.

15 Translations (localization)

This needs more work ... please skip.

CSL has style independent localization files (Locales).

Citation elements could be assigned numeric identifiers, within a context (style?).

Translation tables for translation of the number into a name in a specific language (and country) could be defined in BG or web. The same could apply to "terms" (cf. CSL) used in templates. Also possible for some generic source types – if any. To represent styles 100% correctly it will be necessary to localize date formats, special characters etc. (see CSL).

16 Data structures in BG

This needs more work

Data structures must be defined for

- a) Repositories
- b) Sources (incl. element type, element value, ?Is short value?)
- c) Where in source (incl. element type, element value)
- d) Reasoning, Summary, Extract (could have several per where in source, summary could be grouped with Reasoning)
- e) Citation Element
- f) Templates
- g) Template Sets
- h) (Citation style or reference to paper/web)
- i) Source Type Sets
- j) Source Types
- k) Translations (localization) ?
- l) More

There can be several (style guides/) Source Type Sets and Template sets per BG file, possibly with a default identified in the header. If elements in the header are referenced by numeric IDs, there could be optional short names ala Gedcom Tags that will make the file more readable for humans.

For backwards compatibility, all programs must export the Standard Citation Elements (unless they are a subset of the General Citation Elements).

It must be possible to transfer Source Type Sets and Template sets separately (no person info in the file, see Independent Record Collections in the BG Requirements catalogue), (same with source+repository info).

The records above will contain more info than is discussed in this document.

There is a need to look into efficient coding, e.g. transferring default info in the file header, and be able to escape the default inside the file. What should be placed in the file header rather than being repeated many times in the file.

Must source type definitions, element defs and templates be included in all BG files, or can they be downloaded? Files for archiving should contain everything actually used in the file. On the other hand it may create a lot of overhead if all this information is transferred in every file, but it could also be that it will not be too much info – considering that most current systems are able to handle large video files. Another factor is time to import/export files.

Several of the structures in this document contain info about issuer, version, translator, description etc. These could be defined in a common structure.

16.1 Value structures – those holding the data for a particular citation in a BG file

This need more workplease skip

General structure for citation elements

- Citation element ID
- General citation element ID
- Consider name for readability
- Value
- ? Data type (if other than text, here or in some definition) ?

Reference note

- ??? Should allow inclusion of repository info FOR ONE OF THE SOURCES– or not – in the citation??? (What do I mean here?)

17 Conversion templates

This is a possibility – not more than that at the moment.

From the document “Geir’s high level considerations ...”: 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.

Zotero and MS Word 2010 allows choice of citation style independent of the data stored for a citation (have not tested if it works with all styles). I assume this is possible because the citation elements most likely are the same across styles (they are at least in Zotero, cf. “variables” in CSL. This indicates that it might be possible to do conversions.

Could be used to translate citation elements from one style to another, or to a generic set of elements. (Going from a generic set to a more specific one would be possible if qualifiers are present, but probably program specific.) See templates for conversion to Standard Citation Elements, see above. Conversion rules are likely to be language specific. Rules could define decision points where user intervention is required.

See also “[From repository meta data to BetterGEDCOM and reports – A future scenario](#)”

18 Publication/distribution

(Style guides,) Source Type Sets, Template sets and other definitions can be defined on paper, web (structured or unstructured) and/or in a BG specified file format, inside a BG file (possibly in the header) or outside (it may require a specific file type) possibly downloadable from the web. This needs further work

There is a need to look into how the various structures should be packed together, or how they can be found/obtained. E.g. how to find Template Sets for a Source Type Set. (This is a relatively simple task compared to the creation of a general Citation elements.)

19 Citation Style Language and Zotero

Zotero’s Citation Style Language to be inspected for more ideas.

The CSL 1.0 standard specification: <http://citationstyles.org/downloads/specification.html>

Schema: <https://bitbucket.org/bdarcus/csl-schema/downloads>

CSL blog: <http://citationstyles.org/>

Zotero: www.zotero.org

Zotero on the BG wiki:

<http://bettergedcom.wikispaces.com/message/view/Source+Entity/32067062#32072620>

Zotero variables <https://bitbucket.org/bdarcus/csl-schema/src/855dcc00cba7/csl-variables.rnc>

Example style specifications: All current styles: <http://www.zotero.org/styles>

Locales (language adaptations): <https://github.com/citation-style-language/locales>

Geir's summary of CSL:

<http://bettergedcom.wikispaces.com/file/view/Citation+Style+language+Summary.pdf/225661842/Citation%20Style%20language%20Summary.pdf>

There are at least the following important aspects of Zotero and CSL

- The ability to download citation elements from an internet service
- The ability to specify citation styles by means of Citation Style Language (CSL)
- It handles Language/Country specific adaptations (possibly independent of style)

A drastic approach would be to use CSL and CSL program code as the core, and adopt it to our purpose. An open source Java (and later other languages) program generates citations and bibliography text based on CSL. CSL would solve a lot of detailed problems. However, CSL and the programs would have to be extended to support additional citation types (secondary/short and other citation types not supported), additional citation elements and element qualifiers. **It might however be possible to solve many of these problems by having language specific templates for very specific source types** – which I don't think CSL have by default??? – but it will be more work translating templates and specifying more templates.

A possibility is to define a format, separate from a BG file (but maybe with some common data structures), that could be used for encoding of source/citation (meta) data from web services (the format made available by the service) since Zotero etc. does not support all the citation elements we WISH to use in our style guides, but the question is if the databases of archives and libraries support our fields – probably not – it might be that the elements of these institutions can be handled by Zotero. (Mark Tucker has a demo video of download of such info on his blog, see my scenario document.). The meta data download formats supported by Zotero should be investigated.

It should be noted that Zotero does not have definitions of the citation elements for each source type, rather the style definition file (which has program like structures, e.g. if-elseif-else) checks the source type and produces different output depending on the type. See for example the Chicago manual of style's style definition – link above. The differences between the sources are not many, and there are many sources that are treated the same. It indicates that there, wrt what is output in a citation it may not be a need for many source types – but this could be different in genealogy.

CSL gives the style creator the ability to control many “advanced features” in how citations and bibliographies are formatted (e.g. “suppression” of author name in bibliographies when several sources with the same author appear in sequence, “max number of authors before “et.al” is used, handling of ibid and more). These features are not user controllable by means of templates in genealogy programs, and I have not checked if such features are hardwired into the programs. (There may be more important such advanced features that could be identified by reading the CSL summary.)

Note that CSL has a structure for person names with elements that may have interest in person names in BG.

Observation: In many cases the citation elements that can be downloaded from an internet repository, will not be specific enough for a complete citation – unless the source is transcribed, it will often not contain the details of the specific information that is relevant for the citation. But it will ease the job of creating the citation.

I have not looked into the solutions in Zotero for creating shared databases of citations. Given the diversity of genealogy sources – especially in an international context – I am not convinced about the feasibility of this in the short term. In an “all digital” world a citation could be replaced by a link to such a database, or the link could refer to an entry in a database operated by a repository (as currently provided by e.g. Norway’s National Archives – but that does not generate a citation, only source info in human readable format).

20 Program operation

How should a program handle import of data referring to unknown definitions of source types, templates etc?

Further rules for programs

21 Various related documents:

Specifications related to citation of documents in archives (from Ancestry Insider)

http://www.unk.edu/uploadedFiles/academics/library/gov_doc/about/Citing%20Government%20Documents%20-%20CMOS%202003.pdf

<http://www.archivists.org/glossary/index.asp>

22 Do we want to do this work – How - Prioritization of work items

We need to spend some time determining if the solution described in this document is realistic and how to do the work.

- a) has anyone attempted anything similar
- b) are there alternative solutions
- c) is this overkill
- d) who are interested in contributing to the work
- e) are we willing to work on something that is more advanced than just making implementations of EE interwork – which might be a simpler task?
- f) do we have the necessary expertise (also internationally) – do we know what meta data are in existing (internal) databases that could be made available on the net?
- g) what do we know about any existing standards for archives (libraries are maybe more advanced wrt standards?– e.g. MARC)
- h) what are the benefits of each feature above
- i) what MUST be in a standard, i.e. the minimum.

- j) would program vendors be interested once a (draft) specification is ready – what are the consequences for them?
- k) would database operators / internet repositories be interested to implement it – what are the consequences for them?
- l) are we satisfied with making a specification even if it may not be implemented?
- m) how do we resolve conflicts?
- n) do we work openly? – will open discussions going in all directions kill the interest from any newcomers – do we just publish intermediate results?

We should write a “project plan” before starting.

22.1 Prioritization

The minimum implementation may be the structures in a file holding the data for citations, sources, and “lookups”. What the priorities beyond the minimum (top priority first)

- i. The development of general citation elements is a critical component.
- ii. Citation element data types – templates depend on them
- iii. Transfer of templates – identified by source type – does not require transfer of source types
- iv. Source type definitions

Additional issues

1. What are the consequences for users in countries outside the US, will they be able to develop specifications that users can load into their programs
2. What are the consequences for interworking depending on what we choose to include
3. What can be handled by program vendors without a standard solution
4. What are users (or user groups) able to handle wrt complexity
5. Are there things in this we do not need

23 Appendix 1.

CSL definition of the top level pieces in a citation and bibliography

This is for Chicago Manual of style “Full reference note with bibliography”. The full definition is here

<http://www.zotero.org/styles/chicago-fullnote-bibliography/dev>

```
<citation et-al-min="4" et-al-use-first="1" et-al-subsequent-min="4" et-al-subsequent-use-first="1"
disambiguate-add-names="true">
```

```
  <layout prefix="" suffix="." delimiter="; ">
```

```
    <choose>
```

```
      <if position="ibid-with-locator">
```

```
        <group delimiter=", ">
```

```
          <text term="ibid" text-case="capitalize-first"/>
```

```
          <text macro="point-locators-subsequent"/>
```

```
        </group>
```

```
      </if>
```

```

<else-if position="ibid">
    <text term="ibid" text-case="capitalize-first"/>
</else-if>
<else-if position="subsequent">
    <group delimiter=", ">
        <text macro="contributors-short"/>
        <text macro="title-short"/>
        <text macro="point-locators-subsequent"/>
    </group>
</else-if>
<else>
    <group delimiter=", ">
        <text macro="contributors-note"/>
        <text macro="title-note"/>
        <text macro="description-note"/>
        <text macro="secondary-contributors-note"/>
        <text macro="container-title-note"/>
        <text macro="container-contributors-note"/>
    </group>
    <text macro="locators-note"/>
    <text macro="collection-title" prefix=", "/>
    <text macro="issue-note"/>
    <text macro="locators-newspaper" prefix=", "/>
    <text macro="point-locators"/>
    <text macro="access-note" prefix=", "/>
</else>
</choose>
</layout>
</citation>
<bibliography hanging-indent="true" et-al-min="11" et-al-use-first="7" subsequent-author-
substitute="———" entry-spacing="0">
    <sort>
        <key macro="sort-key"/>
        <key variable="issued"/>
    </sort>
    <layout suffix=". ">
        <group delimiter=". ">
            <text macro="contributors"/>
            <text macro="title"/>
            <text macro="description"/>
            <text macro="secondary-contributors"/>
            <group delimiter=", ">
                <text macro="container-title"/>
                <text macro="container-contributors"/>
                <text macro="locators-chapter"/>
            </group>
        </group>
    </layout>
</bibliography>

```

```
        </group>
    </group>
    <text macro="locators"/>
    <text macro="collection-title" prefix=". "/>
    <text macro="issue"/>
    <text macro="locators-newspaper" prefix=", "/>
    <text macro="locators-journal"/>
    <text macro="access" prefix=". "/>
</layout>
</bibliography>
```

-----THE END