

OpenOffice Bibliographic Project Discussion and Issues

2003/04/15

Contents

1. Document Style Options
 2. Revert to declared style
 3. Add symbols to the automatic footnote numbering scheme.
 4. Provide Support for Footnote Citation Styles by Adapting the Field Editor for the Bibliography Tables.
 5. Modify the 'Bibliographic Entry>Edit' Panel to support appropriate field entry options for the citation type selected.
 6. Defining citations and references different treatment may be needed for first and subsequent use of the citation.
 7. There needs to be a mechanism for transferring the internal Document Bibliographic entries to and from the database to allow import and export from OpenOffice.
 8. Some of the Bibliographic Database fields need to be increased in length.
 9. Should we persist with the DBASE Bibliographic Database ?
 10. It would be nice to have search capability to Internet Bibliographic Databases
 11. Can we Utilise other OpenSource applications ?
 12. Issue: Internal Representation of Bibliographic Data
 13. Develop style definitions
-

A printer friendly PDF version of this page is available [detailed.pdf \(63Kb\)](#)

1. Add a selection option for 'Document Style' to cover whole document.

We need to add user selection for a 'Document Style' to apply to a whole document. I see two options for doing this -

1. Adding a new tab-panel 'Document Styles' to File>Properties.

or

2. Adding a new tab-panel 'Document Styles' to Format>Page. (Does Page format only apply to the current section or the whole document ?)

This tab-panel would contain a 'Document Style' selection list which include the supported styles eg APA, ASA, Chicago, MLA, German DIN, French ISO, etc. {need to collect full list.}. Also we need provision for the users to create their own styles. (See the list of BibTeX supported styles [BibTeX supported styles.](#))

1.1. Document Style Options

Other options in the tab panel could be such things as -

- Style Data import / export; (Endnote ^(TM) and Refman ^(TM) provide style definitions to the general public which could be imported if their import format was decoded)
- 'Strictly Enforce Style' (Y/N), This would prevent the user modifying the document style aspects defined by the Style Manual. This could make the OO word processor easier to use as many of the functions would be grayed-out. The user would not be so bewildered with choice. The user could at any time turn off 'Strictly Enforce Style', and have access to all settings.

There would also need to be provision for options which are provided by the selected style that the user can choose. Eg. The Chicago Style allows selection of either in-text, footnote or endnotes, or a dual footnote comments endnote citation scheme. Perhaps a 'Style Option' tab would handle this.

The information associated with Document Style would include the following methods for Bibliographic citation -

- in-text
- footnotes
- endnotes
- endnotes and footnotes. (requires symbols detailed in item 2 the field formats for Bibliographic citation and Bibliographic Tables (Reference Lists and Bibliographic Lists).
- special editing for Bibliographic Tables eg repeated author names indicated by a three-em dash eg. (they should be a continuous line not dashes)-

Charles Dickens, *A Tale of Two Cities*. London: Penguin Books, 2000.
———. *Nicholas Nickleby*. London: Penguin Classics, 1956.
———. *Oliver Twist*. New York, Random House, 1965.

1.2. Other Style considerations

Page formatting constraints (the MLA Manual specifies page margins (need to check details)
There may be particular requirements for Section, Chapter and Heading styles.

[Top of Page](#)

1.3. Conversion code

When a new 'Document Style' was selected there would need to be some format conversion process that would be applied to an existing document. This would change the citation style from the current to the new, the trickiest ones to do would be 'author-date in-text' to footnote or endnote or visa versa.

An Issue: In order to allow a user change citation methods it will be necessary to consider how we treat the footnote entry. It seems to me there are two options. Utilise the standard footnote entry code, or to have a special citation footnote.

2. Revert to declared style

Is there a 'revert to declared style' function in OpenOffice ? In MS Word there is a function that resets text to the declared format. It removes all user made minor adjustments with fonts, margins etc. This facility is probably required to make the above work and very useful in it own right.

3. And symbols to the automatic footnote numbering scheme.

The Chicago Manual of Style stipulates that when Endnotes and Footnotes are both used the Endnotes are consecutively numbered (1,2,3 ...) and for the Footnotes symbols are to be used. The series they suggest are -

*** , † , ‡ , § , || , # .**

As more symbols are needed they are doubled and trebled -

**** , †† , ‡‡ , §§ , || || , ## : *** , ††† , ‡‡‡ , §§§ , || || || , ###**

The proposal is to add this set of symbols to the automatic numbering options for footnotes.

The processing of the footnotes will need to be adjusted for symbols, as the symbol sequence is restarted at the initial symbol (*), for each new page they are use on.

4. Provide Support for Footnote Citation Styles by Adapting the Field Editor for the Bibliography Tables.

The current Insert>Indexes and Tables>'Bibliographic Entry' mechanism [[screen image](#)] which is use to select fields and format Tables [including the bibliographic table] should be extended to enable defining footnote citation fields. Currently Bibliographic Entry has only two options available - selection from bibliographic database or document content. Then there is the selection box for the 'short name'; which is the reference string to be inserted with the text, eg. [CharlesE1978]. This mechanism needs to be extended to the footnote or endnote method. These methods require this information in a more complex format in a footnote or endnote; for example³⁴,

34. T.M. Charles-Edwards,"Honour and status in Some Irish and Welsh Prose Tales.",*Eriu*, xxxvi, 1978.

The Bibliographic Entry selection box should be changed from 'short name' to 'citation type' - Book, Article, Journal Article, Collection etc. A new button 'Customise Citation Styles' would bring up a new field selection and formating panel very similar (maybe identical) to the table format editor. Note that the Citation Style and style options will have been set in the new File>Properties tab-panel 'Document Styles' mentioned above at section 1. Also the 'Customise Citation Styles' button would be inoperative and greyed out if 'Strictly Enforce Style' is set to 'Yes', so as not mess up the style settings.

Note: when the citation formats are defined and provided for the various Document Style Conventions the average user will not have to use this. All the fields will be predefined.

[Top of Page](#)

5. Modify the ‘Bibliographic Entry>Edit’ Panel to support appropriate field entry options for the citation type selected.

Currently the Insert>Indexes and Tables>Bibliographic Entry>Edit>New Panel [[screen image](#)] offers a selection box that allows the user to selection citation type from a pick list eg Book, Article, Manual, Incollection etc. And is presented with a selection of 29 fields in which to place their data. The preferred action would be that the selection of a citation type would indicate which fields are Required, Optional, or Ignored for that citation type. See a screen-print of sixpack as an example of such an [editor](#) .

The list of citation types and fields that are used in BibTeX are given below and should be considered as a minimum requirement.

article - An article from a journal or magazine. Required fields: author, title, journal, year. Optional fields: volume, number, pages , month, note.

Book - A book with an explicit publisher. Required fields: author or editor, title, publisher , year. Optional fields: volume or number , series, address, edition, month , note .

Booklet - A work that is printed and bound, but without a named publisher or sponsoring institution. Required field: title. Optional fields: author, howpublished , address, month, year, note.

Conference - The same as INPROCEEDINGS , included for *Scribe* compatibility.

Inbook - A part of a book, which may be a chapter (or section or whatever) and/or a range of pages. Required fields: author or editor, title, chapter and/or pages, publisher, year . Optional fields: volume or number, series, type , address, edition, month, note.

Incollection - A part of a book having its own title. Required fields: author, title, booktitle, publisher, year. Optional fields: editor, volume or number, series, type, chapter , pages, address, edition, month , note.

Inproceedings - An article in a conference proceedings. Required fields: author, title, booktitle, year. Optional fields: editor, volume or number, series, pages,address, month , organization, publisher, note .

Manual - Technical documentation. Required field: title. Optional fields: author, organization, address, edition, month, year, note.

Mastersthesis - A Master’s thesis. Required fields: author, title, school, year . Optional fields: type, address, month, note .

Misc - Use this type when nothing else fits. Required fields: none. Optional fields: author, title, howpublished, month, year, note.

Phdthesis - A PhD thesis. Required fields: author, title, school, year . Optional fields: type, address, month, note.

Proceedings - The proceedings of a conference. Required fields: title, year. Optional fields: editor, volume or number, series, address , month, organization, publisher, note.

Techreport - A report published by a school or other institution, usually numbered within a series. Required fields: author, title, institution, year. Optional fields: type,

number, address , month, note.

Unpublished - A document having an author and title, but not formally published. Required fields: author, title, note. Optional fields: month, year.

[The full text with field definitions and usage hints.](#)

[Top of Page](#)

6. Defining citations and references different treatment maybe needed for first and subsequent use of the citation.

There is a tedious and fault prone aspects of the footnote / endnote citation method - the maintenance the Initial and Subsequent citations in the correct order as one edits the text. In the mad rush to complete the paper as the deadline approaches, a piece of text is moved and the Initial Citation reference now comes after the Subsequent reference, and the examiner gives you a red mark.

I do not know enough about the internals of OpenOffice to suggest how this would be implemented. Two options would be to include a check after bibliographic entry or to check and adjust when the update fields function is activated. This would require the citation field to have an identifier which signified that a citation's initial and subsequent appearances were related to the same reference in order to check if the current Initial citation is still the initial one after some text editing. That is not using string matching but reference identifiers.

This also requires that with style that require it all of the citation formats will have two versions; the full initial one and the shorter subsequent one.

Example. An Initial Reference -

'Oisín and Patrick' in *Irish Myths and Legends*, Lady Gregory. (London: Running Press, 1989), 412-421.

A Subsequent Reference -

'Oisín and Patrick' in *Irish Myths and Legends*, 450.

or just -

'Oisín and Patrick', 25.

6.1. Some Implementation Issues

In order to allow a user change citation methods it will be necessary to consider how we treat the footnote / endnote entry. It seems to me there are two options. Utilise the standard footnote / endnote entry code, or to have a special citation footnote.

I will give some examples. A document has in-text author-date citations. This is what is currently implemented. eg.

The story of Finn (Gregory1989) is central to our argument./

We now change the citation setting to footnote method and it deletes the In-text field, creates a footnote and places the citation field into the footnote.

The story of Finn¹ is central to our argument.

1. Lady Gregory, *Irish Myths and Legends*, (London: Running Press, 1989)

This seems simple enough. But what happens if we are starting with the footnote / endnote style, and change to the in-text style. Will it be just as simple a matter to find all the bibliographic entries, go to the footnote / endnote reference, delete the footnote / endnote and insert a new in-text author-date citations? What do we do if the user has added extra text into citation footnote / endnote ? -

1. Lady Gregory, *Irish Myths and Legends*, (London: Running Press, 1989) Note that this was originally published in 1903.

Do we then leave the footnote / endnote with just the added text? And deleted if there was no added text ?

The story of Finn (Gregory1989)¹ is central to our argument. eg.

1. Note that this was originally published in 1903.

As more symbols are needed they are doubled and trebled - The proposal is to add this set of symbols to the automatic numbering options for footnotes.

Would it help to have a special field for footnote / endnote or citations, and if we did how do we deal with added comments? A user writable text area with the citation field ?

[Top of Page](#)

6.2 The inset citation dialog will need to support types of citations within a citation style.

This will be required in order to support the following type of citation system -

If I quote a document(book/journal/article etc.) without a page it looks like

"This method is very reliable (AUTHOR YEAR)"

f.e. "This method is very reliable (BASLER 2003)

or

"But AUTHOR (YEAR) showed that..."

f.e. "But WILSON(2002) showed that this method is not very reliable."

If a page or some pages are quoted (direct or indirect) it must look like

"'This method is very reliable' (AUTHOR YEAR:12)"

f.e. "'This method is very reliable' (BASLER 2003:12)"

or

"'This method is very reliable' (BASLER 2003:12 ff.)"

or

"AUTHOR (YEAR:12) stated 'This method is very reliable.'"

f.e. "BASLER (2003:12) stated 'This method is very reliable.'"

or

"BASLER (2003:12 f.) concludes that this method is very reliable.'"

Some comments on that:

1. The AUTHOR(s) must be in small capitals (as every person everywhere in the text).
2. One author: "AUTHOR 2003" Two authors: "AUTHOR1 & AUTHOR2 2003" Three or more: "AUTHOR1 ET AL. 2003"

If there are more than one publication of an author in the same year it must look like:

"BASLER 2003a" and "BASLER 2003b" a.s.o.

In the bibliography the above example would look like

BASLER, M. (YEAR): Book title. City1 et al.

or

BASLER, M. & D. WILSON (YEAR): Book title. City1 et al. or BASLER, M. D., WILSON, A.

NONAME & B. NONAME (YEAR): Book title. City1 et al.

Note that in the bibliography index ALL authors including their initials must be stated, in the text citations only the first, followed by "ET AL."

[Top of Page](#)

7. There needs to be a mechanism for transferring the internal Document Bibliographic entries to and from the database to allow import and export from OpenOffice.

Currently if a users creates Bibliographic entries with the Insert>Indexes and tables>Bibliographic Entry>New function, these entries can not be easily exported from the document.

As the first priority there needs to be an export and import of bibliographic data in the BibTeX format. As this is the most common format and the one most other bibliographic tools will utilise.

8. Some of the Bibliographic Database fields need to be increased in length.

Some users have run into the field length limit of the varchar datatype (or of the Input form). At least Title and Authors, and possibly some others should be made longvarchar. Currently if the user changes the field definitions in the database the bibliographic functions - Bibliographic Entry, and Bibliographic Table cease to work on the altered table. Some journal articles have hundreds of authors !

9. Should we persist with the DBASE Bibliographic Database?

Or, to replace it with storage in BibTeX format and utilise other OpenSource (eg. sixpack or pybliographer) work (code or design) for a BibTeX editor and browser. We should also check the OpenOffice Database project for their view and plans.

10. It would be nice to have search capability to Internet Bibliographic Database

P.S from an Institute of Biochemistry requested the ability to connect and query on-line databases (like Medline), or at least import records from saved queries. He wrote - "PubMed for example would allow your browser to save the queries in several formats that are easy to import."

11. Can we Utilise other OpenSource applications ?

Decide if we want to provide support for or utilise OpenSource bibliographic tools such as Latex/BibTeX, Sixpack, bp, or Pybliographer to work with OpenOffice (they could provide significant parts of the bibliographic functions).

The Pybliographer developers are considering major enhancements to the application and have produced a planing document which has a very good discussion of bibliographic and user client program issues. It is well worth a look at.

-<http://canvas.gnome.org:65348/pybliographer/design/todo.pdf>

There several relevant and interesting projects -

BibTeXML. BibTeX as XML markup. A bibliography DTD and schema for XML that expresses the content model in BibTeX -- the bibliographic system for use with LaTeX, which is widely adopted by the scientific community. <http://bibtexml.sourceforge.net/>

The BibX is to be the XML equivalent to RIS and BibTeX, but to significantly improve on them.

RefDB will support BibX (but not BibTeXML, though there may well be XSLT files to convert back-and-forth). This support will take time to implement. Nevertheless, this doesn't preclude working with the system now. BibX will simply provide a new import/export format, and also a richer data model than RefDB currently has (which is based on RIS).

There also the UniMark Project which relates to the use of SGML/XML in bibliographic data management based upon MARC formats (used in Library catalogs)
<http://www.oasis-open.org/cover/marc.html>

It would not be difficult to provide interaction with sixpack and or pybliographer (at least on Linux). If a Unix style pipe for text input was developed for OpenOffice, this would allow the other application to directly insert citations eg. [wilson2002] into the text. Sixpack can produce a bibliographic reference table of the cited works in html which could be imported into OpenOffice Writer.

A brief description of these, and other related projects, and links, are to be found in the [Software Pages](#).

12. Issue: Internal Representation of Bibliographic Data

We may wish to utilise work already done on XML representation of Bibliographic data. Refer to BibTeXML and BibX at item 11 above.

Also see BiblioML - XML for UNIMARC Bibliographic Records @ <http://www.oasis-open.org/cover/biblioML.html> MARC is a different way of representing bib data, mostly tied to libraries, and there are a variety of attempts to XML-ify it.

[Top of Page](#)

13. Develop style definition

Provide for collections of bibliographic field definitions that are selected via a select document style option. Most of the styles are already defined in the latex/BibTeX, but this is not a simple format to extract information from them.

13.1 BibTeX and conversion tools and General Information

bp is a Perl library that is designed to:

- Let you quickly make tools to access bibliographies
- Let you quickly make tools to access bibliographies
- Let you access multiple bibliography formats transparently
- Let you convert between formats

- Let you convert between character sets

Developer - 'My first goal when designing the [bp] package was only the first -- I had written a number of tools that accessed my BibTeX bibliographies, and I saw that I was reusing a lot of code. So I decided to make a generic package to access BibTeX bibliographies. About a year later I decided that it would be even better if the package could read multiple formats, and convert between them. The result is bp. Parts of the package are still missing (namely documentation, automatic format recognition, and a good set of utilities), but everything needed for a working system exists'. Home Page <http://www.ecst.csuchico.edu/~jacobsd/bib/bp/index.html>

Field List for the Canonical Bibliography Format

<http://www.ecst.csuchico.edu/~jacobsd/bib/bp/CanonicalFields.html>

[List of BibTeX document types and field definitions, with some hints on use.](#)

[List of BibTeX supported citations style .](#)

For a document which explains how to hack the BibTeX style macros see- [Macros](#) .

13.1.a Cl-BibTeX A BibTeX system in Common Lisp

Developer - 'This is a replacement for the BibTeX program in Common Lisp. The aim is to enable the user to format bibliographic entries using Common Lisp programs, rather than using the stack language of BibTeX style files. A powerful tool is needed for dealing with citations in scientific documents. BibTeX is good for formatting bibliographies, but customizing the format is a pain because it requires writing/changing a program in BAFL (BibTeX Anonymous Forth-Like Language -- Drew McDermott in comp.lang.lisp)'; <http://www.nongnu.org/cl-bibtex/>

The rationale for the program is that LISP is more comprehensible and hence more easy to modify and develop new styles than the original BibTeX language.

By [David N. Wilson](#)