

Request For Comments: New Bibliographic System for Libre Office

Lucas de Sousa Brito
Artificial Intelligence Expert
<lucas75@gmail.com>

FEV/2024

1

INTRODUCTION

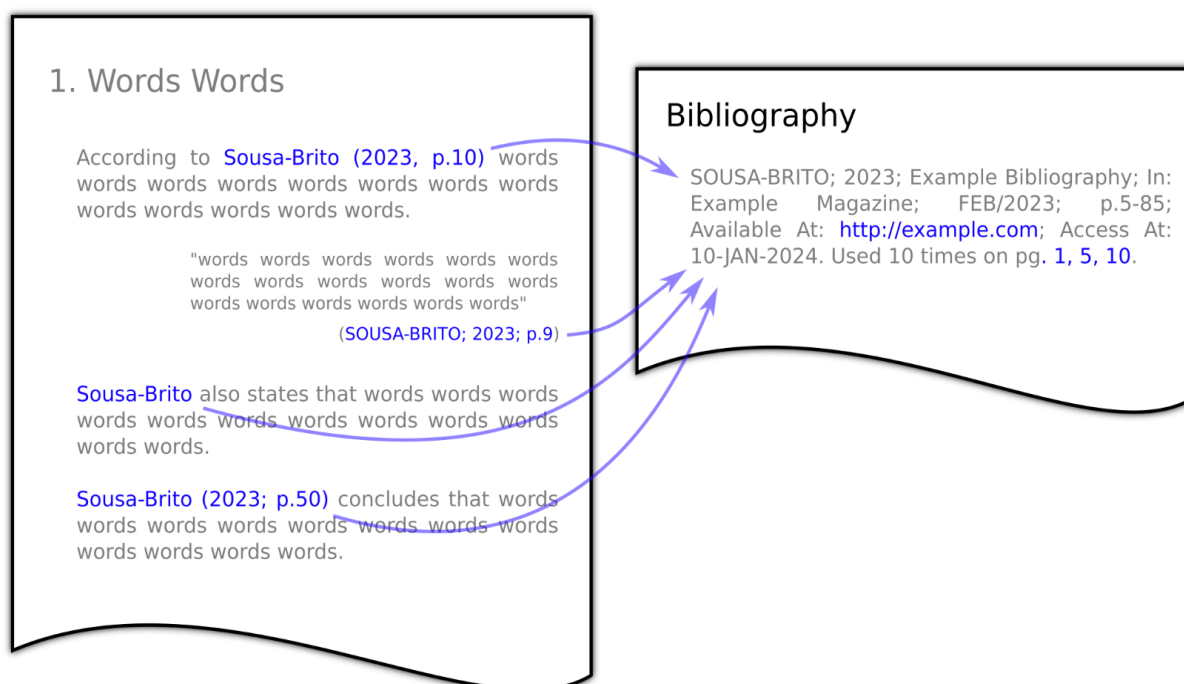
This document contains a initial feature proposal for improving LibreOffice bibliographic management. This document doesn't pretend to be a final spec, is only to start the conversation.

This covers, more or less, the Enhancements Requests [157064](#), [121945](#), [153350](#), [73275](#) and [92572](#) from LibreOffice Bugzilla and a few other ideas.

2 DESIRABLE FEATURES

2.1 Support to Author-Year bibliographic styles is desirable

The current bibliographic system is not designed to support “author-year” bibliographic styles such as APA and Brazilian NBR. In these systems we can have multiple references with different texts pointing to the same bibliography entry like the example bellow:

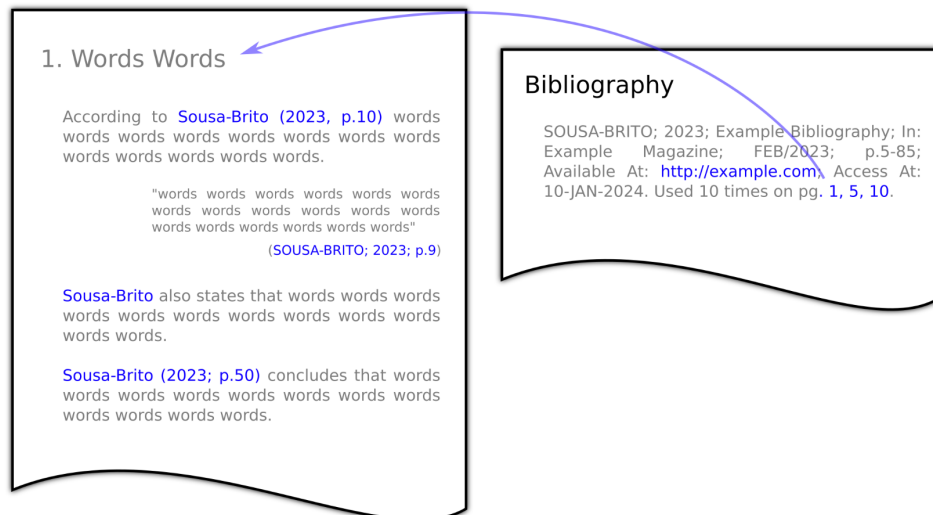


The Enhancement Requests [157064](#), [153350](#) and, more or less, asks for support to Author-Year styles.

2.2

Support for back references is desirable

Some Latex bibliographic styles generate back references which is a cool feature to review long documents.



2.3

Support for hyperlink styles is desirable

A different color for hyperlinks is important to make easier to the reader to recognize the reference is an active element.

2.4

Better support for non-paged media, laws and regulation is desirable

Not every reference will point to a page within a document. On non-paged media, laws and regulations the references will usually be to section numbers within the referenced document.

2.5

Better support for external bibliographic databases is desirable

The Enhancement Requests [121958](#) and [121945](#) suggests to follow the example and/or improve integration with Zotero.

Zotero, EndNote and Mendeley provide advanced bibliographic databases with thousands of bibliographic styles and improve integration to all of them could give the user more choice.

2.6

Better presentation of bibliography in the dialog combos

The Enhancement Request [73275](#) talks about the difficult of finding the correct citation on the dialog combo and suggests a wider presentation in the combos with author name, year and publication instead of the short name.

2.7

Bibliographic support on Presentations and Drawings is also desirable

The Enhancement Request [92572](#) asks for this. Although this proposal is about LibreOffice Writer, there is no reason this can't be implemented in LibreOffice Draw and Impress.

3

DESIGN CONSIDERATIONS

3.1

At its core, the bibliography is just an index not unlike the table of contents

This is relevant because one of the LibreOffice “superpowers” is its cross reference capabilities and this can be leveraged to make a generic index class that is suited for the bibliographic needs.

3.2

Index management and bibliographic formatting can be easily separated

This is the key concept on this proposal that will make the bibliographic system more powerful and support both internal and external bibliographic databases the same way.

Bibliographic databases will have the responsibility of formatting the Reference Label and the Bibliographic Index Entry while LibreOffice internally will handle the index itself.

3.3

Bibliographic database and bibliographic style can't be easily separated

The bibliographic style formatter will define which fields of the bibliographic record must be present. To allow separation between database and formatting we would need:

- A fixed bibliographic data model; or

This means the bibliographic data model will be slow to evolve because will depend on committee decisions and will hinder creativity in this space.

- A protocol to allow the bibliographic database and bibliographic style to communicate and pass along the information of required fields.

Plugins systems have the inconvenience that even in minor version of the main software, many plugins stop working and depends on the developers to update to the new plugin API.

Although this would be a nice thing to have, the complexity would probably negate the benefits of separating the database and formatting functions.

3.4

An advanced internal bibliographic database may be overkill and unnecessary

A bibliographic database capable of formatting the references according to several bibliographic standards is a lot of work and may be unnecessary:

- First because Zotero, EndNote and Mendeley already provide bibliographic database services with support to thousands of bibliographic styles;
- Second because Zotero is open source;
- Third because most users will want a cloud based bibliographic database anyways;
- Forth because with an easier and dependable API, several new services may arise.

Even if LibreOffice Community decide to develop an internal advanced bibliographic database of its own, the proposal in this document is a nice first step.

Bear with me. It will become obvious in a few minutes. As we will show ahead, for basic use an advanced bibliographic database and formatting is unnecessary.

4 THE PROPOSAL

4.1 Separate the index from formatting through an HTTP API

The core idea is to have a new index class on LibreOffice that will handle the links, backrefs and index and left the bibliographic style to be handled for by the external databases through an HTTP API.

API	Parameters	Results
List Available Styles	Text Filter (opt)	List of styles
List Available References	Style Id (req) Text Filter (opt)	List of References
		List of Placement Options
		Err: Style Not Supported
Format Index Entry	Style Id (req) Entry Id (req) DuplKeySeq (opt) Extra Parameters=(opt)	Formatted Entry
		Biblio Data (opt)
		DuplKey (req)
		SortKey (opt)
		Err: Style Not Supported
		Err: Entry Not Found
Format Reference Label	Style Id (req) Entries=[] For each entry: { Entry Id= (req) DuplKeySeq=(req) Extra Parameter=(opt) }	Formatted Label
		Err: Style Not Supported
		Err: Entry Not Found
		Err: Multiple Source not Supported

4.2

The formatting will happen in the “Add/Edit Entry” dialog

The sketch bellow illustrates that the formatting of the **Index Entry** and the **Reference Label** can be done in the Add/Edit Index Entry dialog. The document will only have to store the formatted ones and manage the hyperlinks and backrefs.

1. Words Words

According to [Sousa-Brito \(2023, p.10\)](#) words words words words words words words words words words words words.

"words words words words words words words words words words words words words words words words"

(SOUSA-BRITO; 2023; p.9)

Bibliography

SOUSA-BRITO; 2023; Example Bibliography; In: Example Magazine; FEB/2023; p.5-85; Available At: <http://example.com>; Access At: 10-JAN-2024. Used 10 times on pg. 1, 5, 10.

Add/Edit Index Entry

Index: TheBibliography ▼ Placement: Online ▼ Reference Label: Sousa-Brito (2023; p.10) [B I N]

Source: Sousa-Brito; 2023; Example Bibliography; Example Magazine (@Example.com) ▼ Extra parameters: p.10

Index Entry: [B I N Link] SOUSA-BRITO; 2023; Example Bibliography; In: Example Magazine; FEB/2023; p. 5-85; Available At: <http://example.com>; Access At: 10-JAN-2024. Used 12 times in this document

☒ Automatic Formatting

Add Source... OK Cancel

The inner workings of this screen is:

- When the user begins to types something on the Source the dialog call the “List Available References” function from the API of every configured bibliographic source;
- When the user selects a Source Key, the dialog will call the “Format Index Entry” and “Format Reference Label” from the specified bibliographic source;

- If successful enable the Extra Parameters and mark the Reference Label and Index Entry as automatic formatting;
- When the user changes the Extra Parameters, the dialog will call the "Format Reference Label" again;
- If the user disables the automatic formatting, the dialog will no longer update those fields automatically;
- If the user reenable the automatic formatting, the dialog will call the "Format Index Entry" and "Format Reference Label" functions of the API accordingly.
- We will discuss the "Placement", "Extra Parameters" and duplicate resolution ahead.

4.3 Manual Entries

This illustration bellow describes how the bibliographic index can work when not connected to a bibliographic source. The user will provide a **Reference Label** and the **Index Entry** manually:

1. Words Words

According to [Sousa-Brito \(2023, p.10\)](#) words words words words words words words words words words words.

"words words words words words words words words words words words words words words words"

([SOUSA-BRITO; 2023; p.9](#))

Bibliography

SOUSA-BRITO; 2023; Example Bibliography; In: Example Magazine; FEB/2023; p.5-85; Available At: <http://example.com>; Access At: 10-JAN-2024. Used 10 times on pg. **1, 5, 10.**

Add/Edit Index Entry

Index: TheBibliography ▼ Placement: Manual ▼ Reference Label: Sousa-Brito (2023; p.10) [B I N]

Source: Sousa-Brito; 2023; Example Bibliography; Example Magazine (@ThisDocument) ▼ Extra parameters: p.10

Index Entry: [B I N Link] SOUSA-BRITO; 2023; Example Bibliography; In: Example Magazine; FEB/2023; p. 5-85; Available At: <http://example.com>; Access At: 10-JAN-2024. Used 12 times in this document

☐ Automatic Formatting

Add Source... OK Cancel

This will be useful for:

- beginners;
- when a bibliographic style doesn't have an automatic formatting available;
- when there are government/company rules doesn't allow the use of a cloud based bibliographic database to index sensitive documents.

4.4

The Placement Option

The **Placement** options control the formatting style of the **Reference Label**. Example of placement options usually found in LaTeX bibliographic system includes:

- **Online:** For citations such as “Author (Year; Param) said that”;
- **Author Name:** For citations such as “Author also said that”;
- **Parenthesized:** For citations such as “(AUTHOR, Year, Param)”;
- **Year:** For citations such as “In Year, he derived that”

4.5

The Extra Parameters

Annotate “where” the content is on a large document is important. This mean the page numbers in books, section numbers in Laws, documents on non-paged media, time range on videos and audios, keyword in encyclopedias and dictionaries etc.

4.6

Multiple Sources

LaTeX and Zotero are examples of bibliographic sources that supports citation for multiple sources at the same time. This is necessary because when a citation come from multiple sources, the **Reference Label** must be combined.

Without support to multiple sources on the same entry, the user would have to cite twice and the reference label would look like “(SOUSA-BRITO, 2013)(DOE, 2003)” which is not acceptable for most bibliographic styles.

The diagram illustrates the process of adding index entries to a document. It shows a text document with a section titled "1. Words Words" containing placeholder text and a citation "(SOUSA-BRITO; 2013; p.9)". A blue arrow points from this citation to a "Bibliography" section, which lists two sources: "SOUSA-BRITO; 2023; Example Bibliography; In: Example Magazine; FEB/2023; p.5-85; Available At: <http://example.com>; Access At: 10-JAN-2024. Used 10 times on pg. 1, 5, 10." and "DOE, JOHN; 2003. Another Example Bibliography; In: Example Encyclopedia; Available At: <http://example.com>; Access At: 10-JAN-2024. Used 5 times on pg. 1, 5, 10." A blue arrow then points from the bibliography to a software interface titled "Add/Edit Index Entry".

The "Add/Edit Index Entry" interface has a header bar with "Add/Edit Index Entry" and a "BIN" button. It contains two main sections for adding entries. The first section has fields for "Index" (TheBibliography), "Placement" (Online), and "Reference Label" (SOUSA-BRITO; 2023; p.10; DOE; 2023; p.50). It also has a "Source" field (Sousa-Brito; 2023; Example Bibliography; Example Magazine (@Example.com)) and an "Extra parameters" field (p.10). The "Index Entry" field contains the formatted citation: "SOUSA-BRITO; 2023; Example Bibliography; In: Example Magazine; FEB/2023; p. 5-85; Available At: <http://example.com>; Access At: 10-JAN-2024." The "Used 12 times in this document" status is shown. The "Automatic Formatting" checkbox is checked. The second section has fields for "Source" (Doe J; 2003; Example Bibliography; Example Magazine (@Example.com)) and "Extra parameters" (p.50). The "Index Entry" field contains the formatted citation: "DOE, JOHN; 2003. Another Example Bibliography; In: Example Encyclopedia; Available At: <http://example.com>; Access At: 10-JAN-2024." The "Used 5 times in this document" status is shown. The "Automatic Formatting" checkbox is checked. At the bottom, there are buttons for "Add Source...", "OK", and "Cancel".

This is also why a fallback with manual formatting is so important. There are a lot of quirks in bibliographic styles and is doubtful the bibliographic source will handle them all.

4.7

Duplicate Resolution

Author Year styles can have duplicated keys when the same author has more than one document within a year. Usually this is resolved with a suffix in the year ex: "Sousa-Brito (2023B)".

For that reason the APIs have to return a "DuplKey" and when the document already have another entry with the same "DuplKey", LibreOffice must assign a Duplicated Sequence Number and refresh the Reference Label and Index Entry Text. That way the bibliographic database will resolve the differences.

4.8

The database selection should be part of the document

URL	Short Name	Style
https://example.com/users/lucas	@Example.com	APA v7
https://127.55.22.33/mydb	@MyDb	APA v7

The bibliographic database configuration should be done at the document level specifying the URL to the databases. This is important because replacing the source database from a document may render it inconsistent because different databases can use the

same name for different references. I also expect the environment to have a "default bibliographic source" configuration, so that new documents inherit them.

When the user select the style, this dialog will access the "List Available Styles" function from the API to select the desired style.

4.9

Other configurations

- Order:
 - Order of appearance in the document
 - Alphabetical order
- Placement
 - As index
 - As footnotes
- Backrefs
 - Page Numbers
 - Section Number 1st level
 - Section Number 2nd Level

4.10 Hyperlink style

For hyperlink styles the solution is simple. Is enough to create character styles like "Biblio Reference Label" and "Biblio Backref" with default style of hyperlinks (underlined blue).

4.11 The backrefs

The backrefs will require a configuration of words for "Used in pages ..." or "Used in page ..." (plural and singular).

4.12 Document Model Considerations

The examples bellow is more or less what the XML within the document needs to be:

Example reference in the text

```
<NewIndexMark
  src-index="index001"
  src-entry-id="1234"
  src-placement="Online"
  src-extra-parameters="p.10">
SOUSA-BRITO (2023; p.10)
</NewIndexMark>
```


Example of the corresponding index

```
<NewIndex name="index001">
  <sources>
    <source url="https://example.com/users/lucas75/academic"
      style="APA v7"/>
    <source url="https://127.55.22.33/mydb"
      style="APA v7"/>
  </sources>
  <NewIndexEntry entry-id="1234"
    src-url="https://example.com/users/lucas75/academic"
    src-style="APA v7"
    src-dupl-key="Sousa-Brito (2023)"
    src-dupl-seq="0"
    src-entry-id="SousaBrito2023"
    src-extra-paramenters=""
  >
    SOUSA-BRITO; 2023; Example Bibliography;
    In: Example Magazine; FEB/2023; p.5-85;
    Available At:
    <a href="http://example.com">http://example.com</a>;
    Access At: 10-JAN-2024.
  </NewIndexEntry>
</NewIndex>
```

Example of the index

```
<NewIndexRender
  placement="asindex" order="alpha" backrefs="page_number"
/>
```

5

FINAL CONSIDERATIONS

It is important to consider that this documents is superficial and is meant to start the conversation. Is doesn't pretend to be a finished solution and will require some work to be completed.

Of course, the full functionality depends on Zotero, EndNote and Mendeley to support LibreOffice API, but is reasonable to think they will because maintain an API is easier than maintain an extension.

This seems to be a big change for LibreOffice, but it may be worth to modernize LibreOffice bibliographic handling.