

# OVERVIEW OF LIBREOFFICE 7.0 INTEROPERABILITY FIXES



openSUSE + LibreOffice Virtual Conference, 17.10.2020



# Who is speaking?

## Vasily Melenchuk

- CIB software GmbH employee
- The Document Foundation member

## Michael Stahl

- Joined the Sun Hamburg OpenOffice.org Writer team in 2007
- employed by CIB labs GmbH

# Interoperability and OOXML formats

- OOXML formats (DOCX, XLSX, PPTX) are popular (obvious!)
- Still in wide use older binary formats and RTF
- LO should import/export them without losing quality

# Working on interop bugs

- 1) Minimize/reduce testcase
- 2) ??????????
- 3) Profit!

# Working on interop bugs

- 1) Minimize/reduce testcase
- 2) ??????????
- 3) Profit! Verify original testcase is fixed
- 4) Profit, probably

# Reducing testcases (steps)

10 Make a file backup

20 Remove something looking unnecessary

30 Verify file is still okay in MS Word

40 Verify file is still broken in Writer

50 GOTO 10

# Reducing testcases (details)

- Preferably should be done on XML level, not in Writer or Word
- Need to edit unformatted XML inside zip-archive
  - easier for vim/emacs users
  - sad for others mostly

# XML pretty-printing

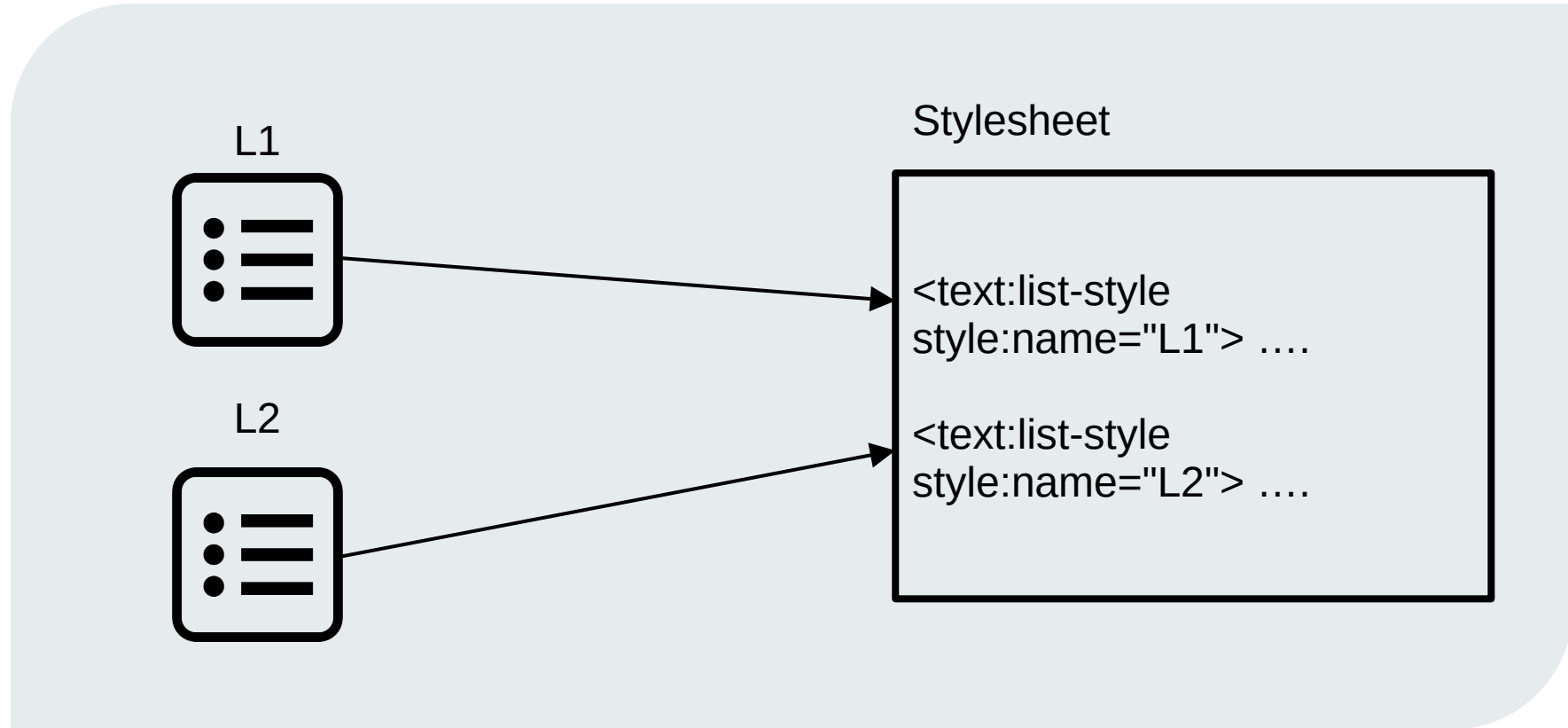
- In Options->Advanced->Open Expert Configuration find and enable "Pretty Printing"
  - Works only for ODF files
  - Only on save
- "xmllint --format"
- online XML formatters



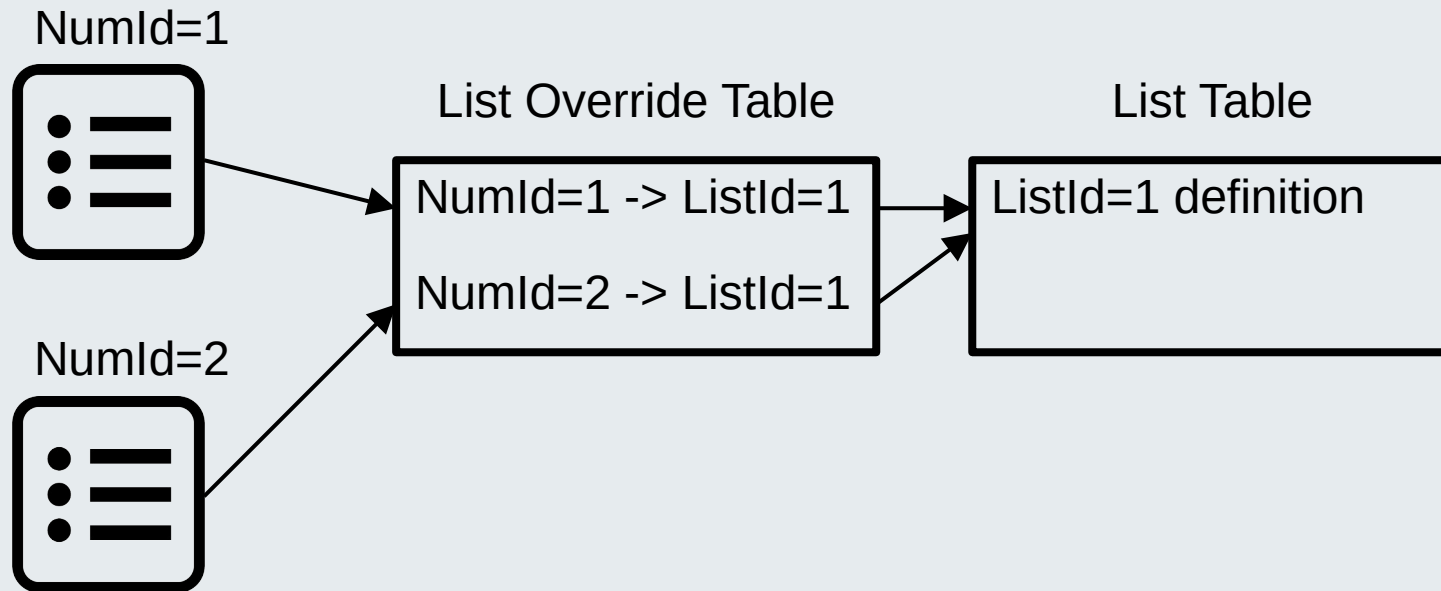
# XML pretty-printing (my solution)

- Primitive script:
  - unpack document package to temp folder
  - xmllint them all
  - pack files back
- <https://github.com/bzzzil/doctools>

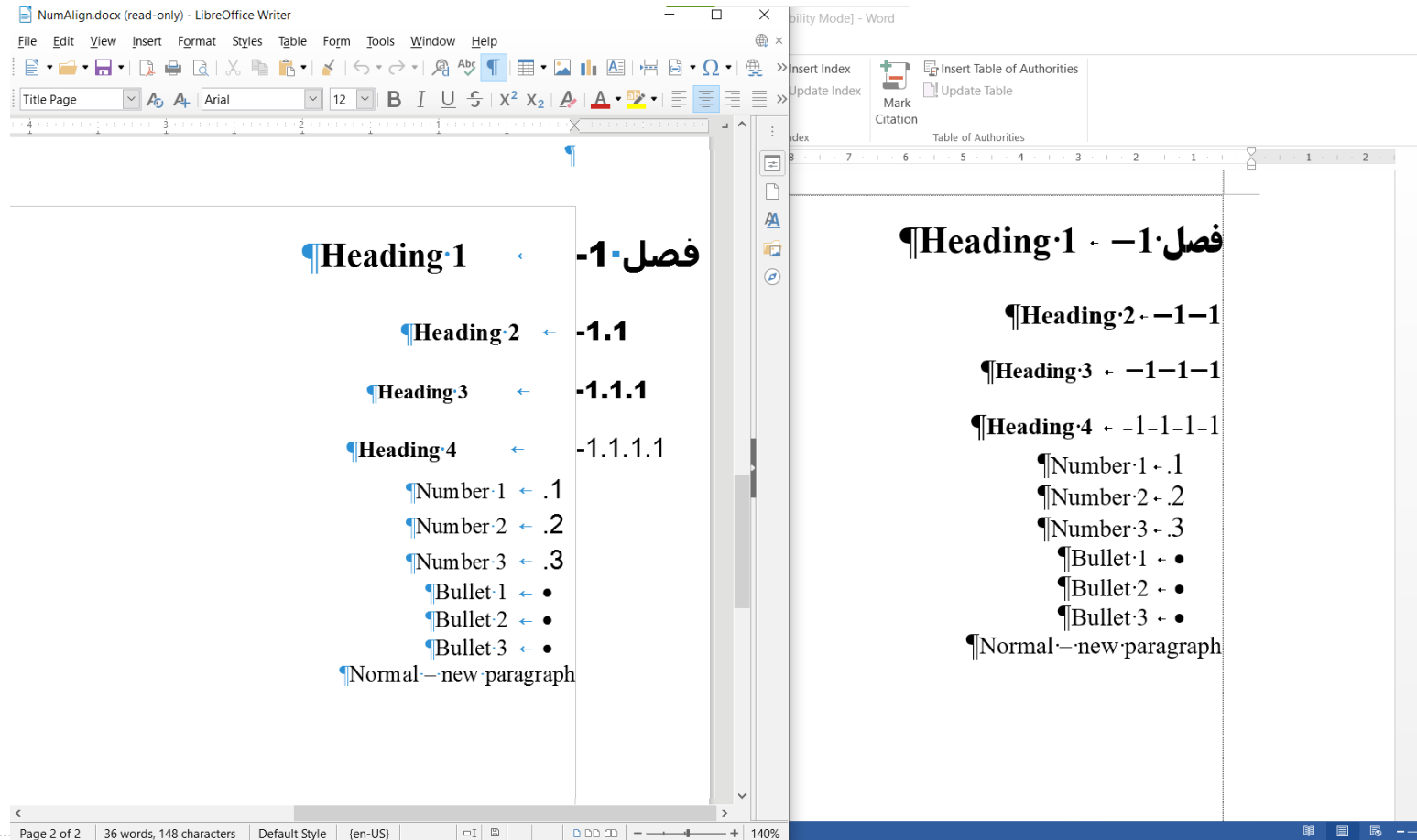
# Lists in ODF



# Lists in DOCX



# tdf#116883: problem

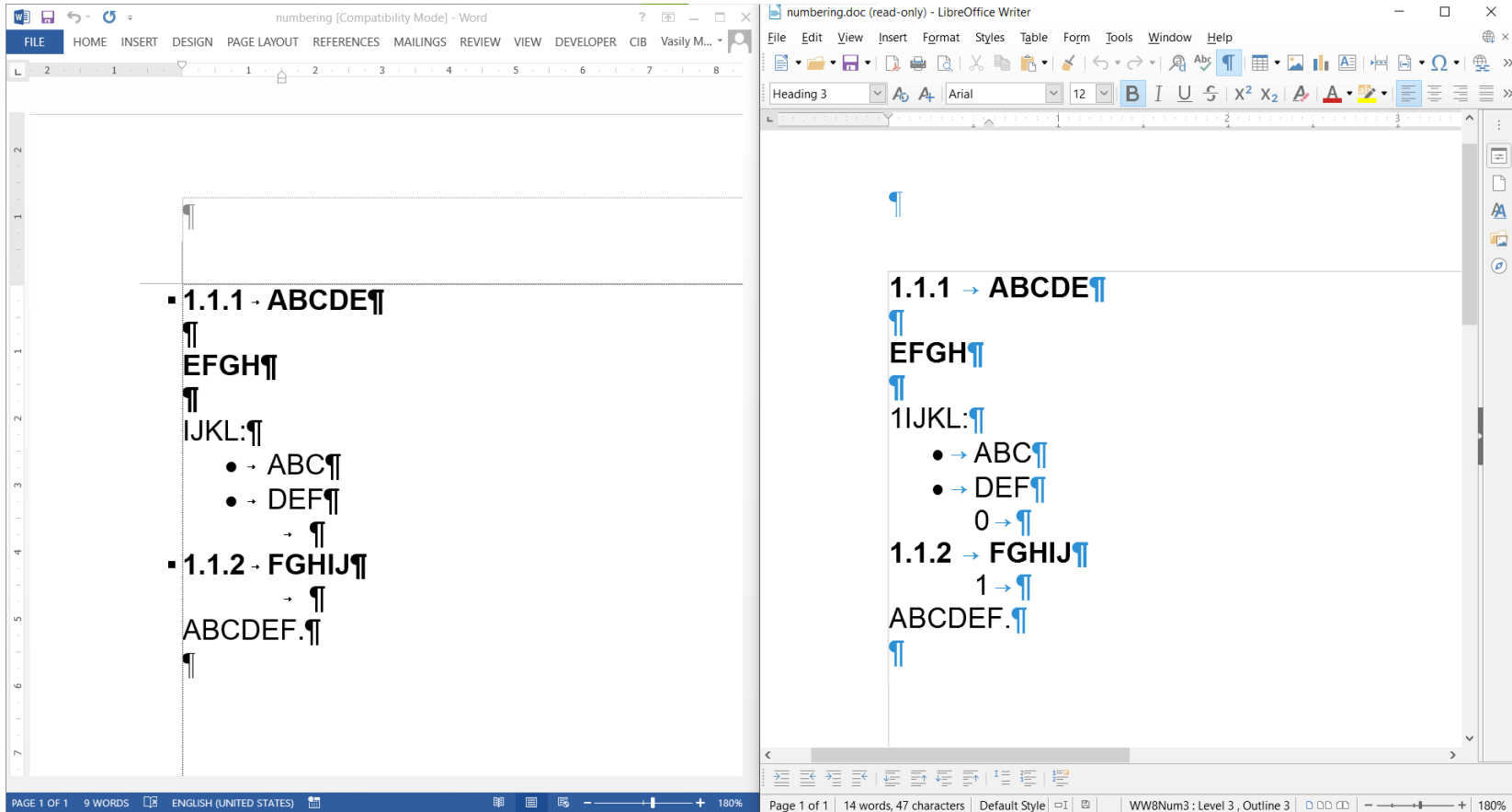


The image shows a side-by-side comparison of a document's table of contents and its actual content in LibreOffice Writer. The left window displays the table of contents for 'NumAlign.docx', listing levels from 'Heading 1' to 'Normal' with corresponding Arabic numerals (1, 1.1, 1.1.1, 1.1.1.1, .1, .2, .3, •, •, •, •). The right window shows the document content, where the same structure is rendered with Arabic numerals and symbols (فصل 1, 1.1, 1.1.1, 1.1.1.1, .1, .2, .3, •, •, •, •, •). The status bar at the bottom indicates 'Page 2 of 2', '36 words, 148 characters', and '140%' zoom.

# tdf#116883: format string

- Word: has list level format string "%1-%2-%3-" but it could be even "**Hello%1!---%2World!%3 :)**"
- Writer: only prefix & suffix and hardcoded dot (".") as a separator
- Right now LO uses internally format string
- Not in UI
- prefix/suffix as a fallback an for ODF
- ODF stadard need to be extended?

# tdf#120394: problem



The image shows two side-by-side screenshots of document editors. The left window is Microsoft Word in Compatibility Mode, showing a document with a list structure. The right window is LibreOffice Writer, showing the same document with a different list structure. The problem is that the numbering in LibreOffice Writer does not match the intended structure in Word.

**Microsoft Word (Left):**

- 1.1.1 - ABCDE¶
- EF GH¶
- IJKL:¶
- -> ABC¶
- -> DEF¶
- > ¶
- 1.1.2 - FGHIJ¶
- > ¶
- ABCDEF.¶

**LibreOffice Writer (Right):**

- 1.1.1 → ABCDE¶
- EF GH¶
- 1 IJKL:¶
- -> ABC¶
- -> DEF¶
- 0 -> ¶
- 1.1.2 → FGHIJ¶
- 1 -> ¶
- ABCDEF.¶

- Lists can refer each other not only via override table, but also with `w:styleLink`

# DOCX lists support: resume

- in total 16 commits with "list" in first commit line (i'm too lazy to verify commits)
- many of them own regressions :(
- it is not one step forward two step back
- thanks to Xisco Fauli for assistance & patience :)



# Paragraph Mark Formatting

- Word can apply formatting to paragraph mark
  - is also applied to list label
- Writer: there is no paragraph mark
- Status quo: was implemented as a 0 length text attribute at end of the paragraph
  - imported from DOCX paragraph mark→exported to DOCX again
  - a hack that couldn't handle all cases, e.g., if the format of the last character happened to be the same as the paragraph mark

# Paragraph Mark Formatting Improvement

- added paragraph item RES\_PARATR\_LIST\_AUTOFORMAT
- paint as list label
- round-trip DOCX
- hard formatting in 6.4.0
- character style too in 6.4.7
- TODO: RTF, ODF extension, UI, paint as pilcrow too

# Example

tdf131801.docx - Word

DATEI START EINFÜGE ENTWURF SEITENLA VERWEISE SENDUN ÜBERPRÜ ANS

Calibri (Textkörper) 11

F K U abc x<sub>2</sub> x<sup>2</sup>

Einfügen

Zwischenablage

Schriftart Absatz Format

1. → Style-mark ¶
2. → *Style-mark-and-text* ¶
3. → No-style ¶
4. → *Style-text* ¶
5. → Style+hard-16pt-mark ¶
6. → *Style+hard-16pt-mark-and-text* ¶
7. → No-style ¶
8. → *Style+hard-16pt-text* ¶

Unbenannt 1 - LibreOffice Writer

Datei Bearbeiten Ansicht Einfügen Format Vorlagen Tabelle Formular Ex

List Paragraph Calibri 16 F K U

1. → Style-mark ¶
2. → *Style-mark-and-text* ¶
3. → No-style ¶
4. → *Style-text* ¶
5. → Style+hard-16pt-mark ¶
6. → *Style+hard-16pt-mark-and-text* ¶
7. → No-style ¶
8. → *Style+hard-16pt-text* ¶

Untitled 1 - LibreOfficeDev Writer 7.1 [2a58902f0eecff8d68ad2669270524b99675c39c]

File Edit View Insert Format Styles Table Form Tools Window Help

List Paragraph Calibri 11 pt B

1. → Style-mark ¶
2. → *Style-mark-and-text* ¶
3. → No-style ¶
4. → *Style-text* ¶
5. → Style+hard-16pt-mark ¶
6. → *Style+hard-16pt-mark-and-text* ¶
7. → No-style ¶
8. → *Style+hard-16pt-text* ¶

# CIB

IDEEN MIT SYSTEM



**OUR PRODUCTS:**

<https://libreoffice.cib.de>