Collabora Productivity

OOXML support maintenance

By Tamás Zolnai

Software Engineer at Collabora Productivity
tamas.zolnai@collabora.com

Introduction

Why OOXML compatibility is an issue?

- Migration or mixed environment

- Replacement of MSO

What will I speak about?

- Areas of comp. issues

- Improvements done in last year

- Future possibilities

Collabora Productivity

Form controls

Form control example
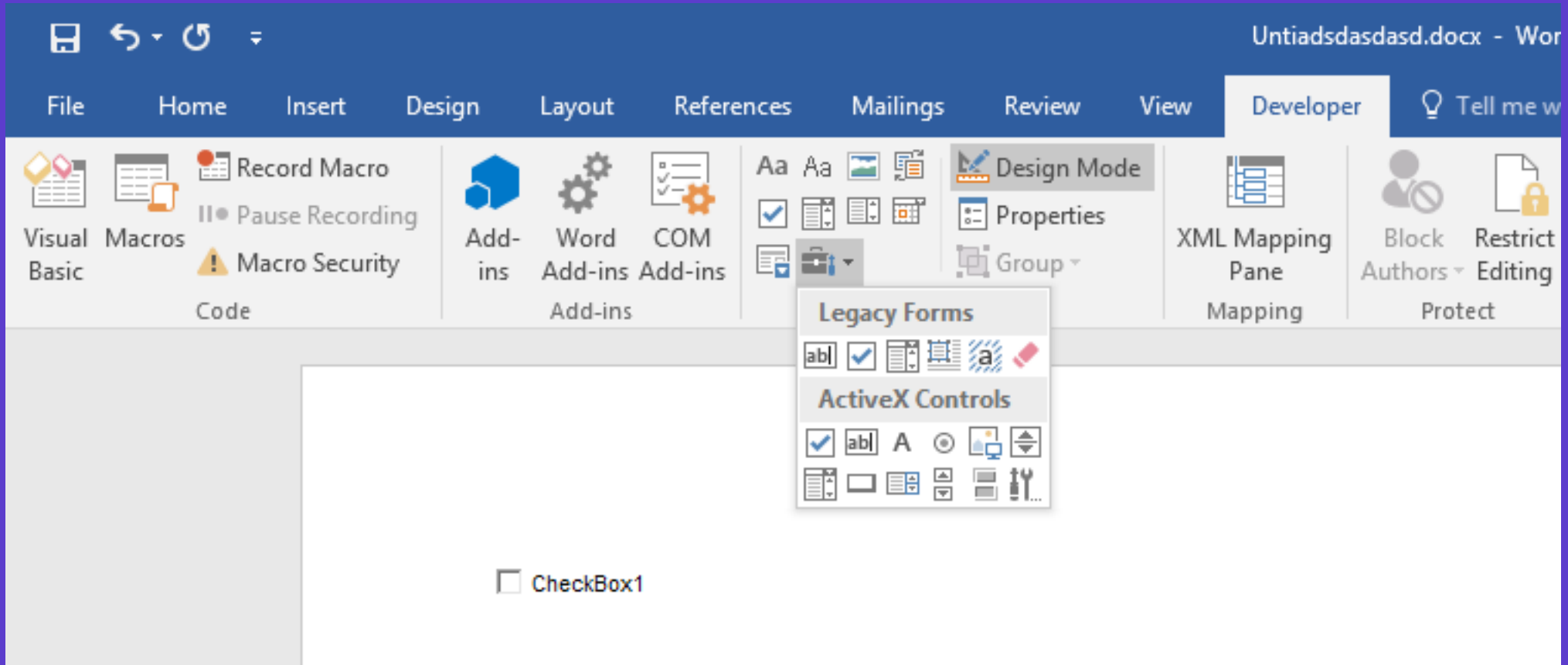
Form controls

LibreOffice

- Lots of different control types: option button, radio button, pushbutton, combobox, listbox, picture button, etc

- Floating objects: can be anchored as text or as a shape object

MS Office

- Three kind of controls: legacy, ActiveX and content

- Content control: new controls, control types: plain text, image, checkbox, combobox, listbox, date picker

- ActiveX controls: floating, more visual properties (background color, triplestate, wordwrap, image, alignment)

MSO control types

What's done about form controls?

ActiveX controls

- Now imported from DOCX files (tdf#91384)

  - Before only XSLX and PPTX was supported

- LO form controls are exported as MSO ActiveX controls (tdf#50097)

  - ActiveX controls are matching the best to LO form controls

    - Floating objects, similar additional properties, similar types
    - One serious limitation is the missing preset items for list controls

  - Except date picker, which has a better matching with content control

Implementation details

Position, size and anchor

- Specified in /word/content.xml, defined as a VML shape

- A bit different XML tags when the control is a floating object (pict or object context)

Control properties

- Binary file inside DOCX, /word/activeX/activeX1.bin

- /word/activeX/activeX1.xml contains some options how to read that binary

- This binary data seems to be picked up from the DOC binary format

Collabora Productivity

HSC South Eastern Health and Social Care Trust

Future possibilities

Check specific controls whether ActiveX the best match

- Maybe list controls (combobox and listbox) would be better to save as content control (because of preset items)
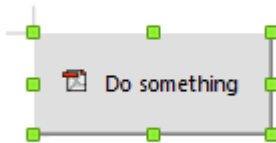
Future possibilities

Small issues in ActiveX control support (easy hacks)

- ActiveX control locked / read only property is not imported (tdf#111417)

- Additional picture of button control is not saved to DOCX

- Text color of ActiveX controls are not saved to DOCX

Vertical text in a table

Vertical text in MSO

Vertical text in a table

MS Office

- Vertical text for most of the container objects (tables, shapes, text boxes)
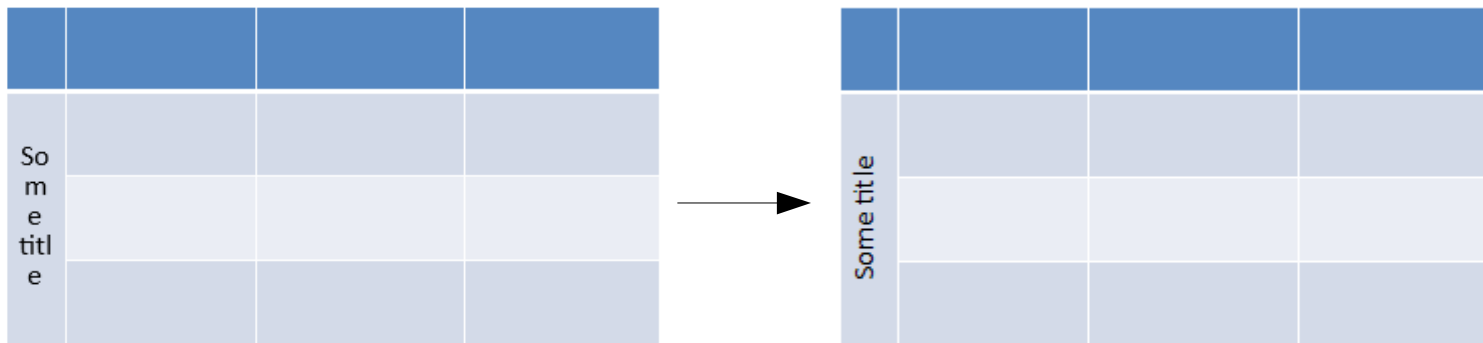
- Two supported rotation angle (90° and 270°)

LibreOffice

- Has some vertical text support, but this is related to different language scripts (right-to-left, top-to-bottom, etc)

- Needed to have an actual rotation property which is independent of other font script properties.

Collabora Productivity

SUSE
We adapt. You succeed.

What's done about vertical text?

Implemented text rotation for Impress tables

- Now text rotation is imported form PPTX files (tdf#100926)

- The same two rotation angles are supported

- By now it's only available via UNO API, can't set on the GUI

Implementation details

Different levels of text rotation implementation

- Low level code was already there (Calc has text rotation in cells)

- Medium level was added during implementation.

  - Editengine now handles TopToBottom and BottomToTop text.

  - Both during editing and in non-active mode.

- Heigher level

  - PPTX import filter now handles the vertical properties (vert, vert270)

  - Calc table cells has this property already, so we did not need to extend API or ODF.

Future possibilities

GUI option for vertical text in Impress tables

- tdf#111633: Missing GUI option for text rotation inside Impress table

Extend vertical text support for other objects

- Writer tables
  - They have a different implementation (sw core)
  - ODF property is already there
- Shapes, text boxes, text frames

Collabora Productivity

SUSE
We adapt. You succeed.

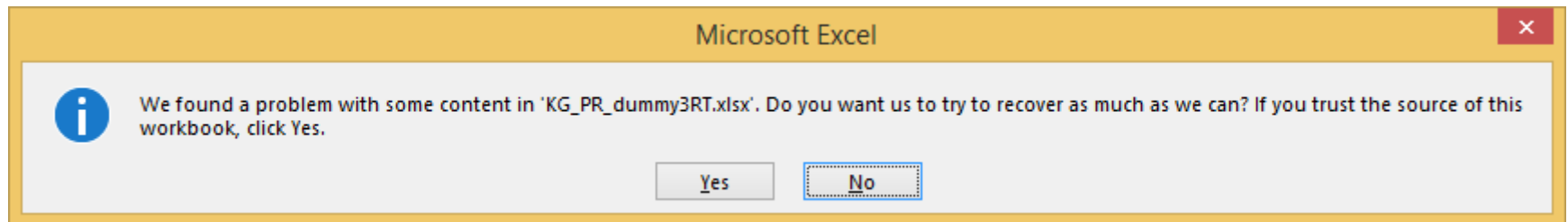# Pivot tables

A pivot table example

Pivot table

In general

- Special table to organize and aggregate data

- Complex feature supported both by LO and MSO

Issues

- Lots of issues related to MSO compatibility

- Most serious type: corruption and data loss after RT in Calc

Microsoft Excel

We found a problem with some content in 'KG_PR_dummy3RT.xlsx'. Do you want us to try to recover as much as we can? If you trust the source of this workbook, click Yes.

Yes   No

What's done about pivot tables?

Performance improvement (tdf#102694)

- Pivot cache behind the pivot table

- Groups are shared between tables

- Specific test document import time 20 min → 0,5 min

Larger and smaller fixes (still working on that)

- Pivot table's layout was wrongly exported (tdf#112946)

- Corrupted XLSX was generated by LO (tdf#112893)

- Filtering of different fields were not preserved (tdf#107711)

Collabora Productivity

HSC South Eastern Health and Social Care Trust

SUSE
We adapt. You succeed.

Filtering of a pivot field

Implementation details

MSO pivot table

- Different table layout modes: outline, tabular and compact.

  - LO supports outline and tabular form, but not the compact mode

- Pivot table definition and pivot cache

  - Pivot table: fields, layout, etc

  - Cache contains the shared items, groups, etc.

  - Redundancy in pivot cache definition for robustness (repair dialog)

Collabora Productivity

HSC South Eastern Health and Social Care Trust

Future possibilities

Support compact mode in LO

- It needs to have a different pivot table output, with some additional buttons

- LO already support outline and tabular mode, so might be easy to add a new mode (e.g. ODF extension)

Improve performance of XLSX import

- UNO API usage seems slow

- Also pivot cache handling might has issues

Collabora Productivity

HSC South Eastern Health and Social Care Trust

Outline mode vs compact mode

Thank you!

By Tamás Zolnai

tamas.zolnai@collabora.com