



Collabora Productivity

A year in LibreOffice's PDF support

By Miklos Vajna

Senior Software Engineer at Collabora Productivity

2017-10-13



About Miklos

- From Hungary
 - More blurb: <http://vmiklos.hu/>
- Google Summer of Code 2010/2011
 - Rewrite of the Writer RTF import/export
- Writer developer since 2012
- Contractor at Collabora since 2013



Thanks

- Collabora is an open source consulting company
 - What we do and share with the community has to be paid by someone
- Sponsors of the work presented here are:
 - Dutch Ministry of Defense in cooperation with Nou&Off
 - Professional Media Group nv

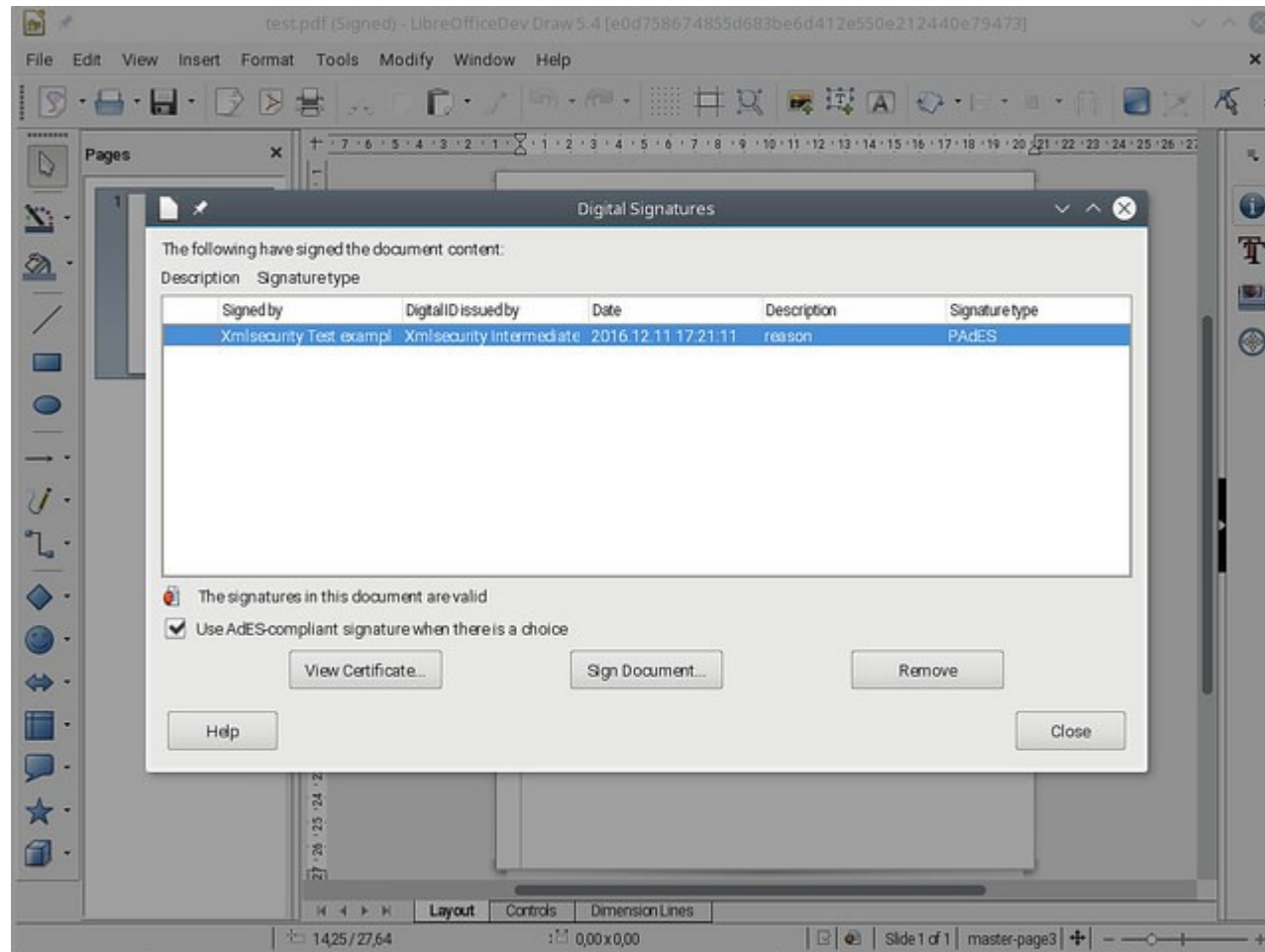


New PDF features from the past year



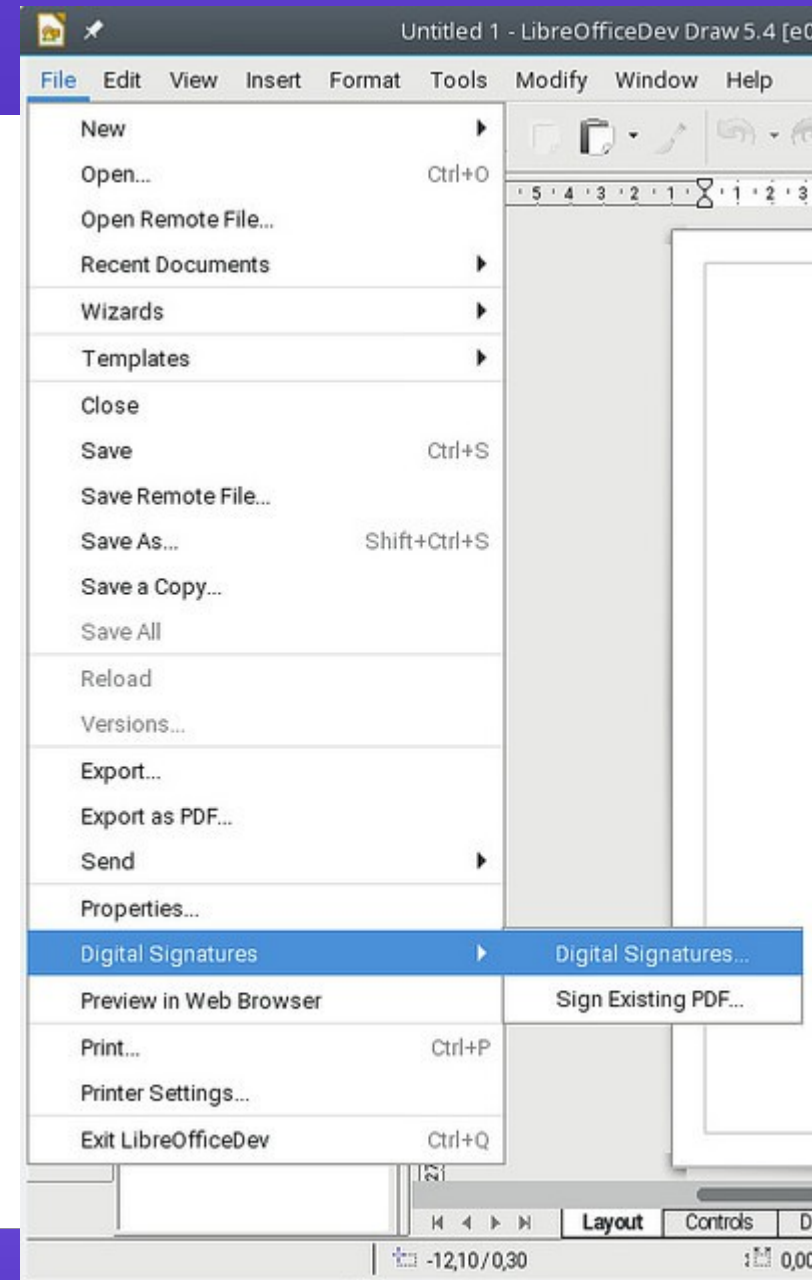
PDF signature verification

- Open already signed PDFs
- Verify their signatures
- May be multiple signatures
- Own tokenizer
 - sdext/boost, poppler, pdfium found suboptimal for this purpose



Signing of an existing PDF

- Signing as part of PDF export was already supported
- Here: incremental updates
- Use-case:
 - Multiple signatures
 - Signing PDF produced outside LO
 - Signed PDF 1.5+ documents
 - We produce 1.4 currently



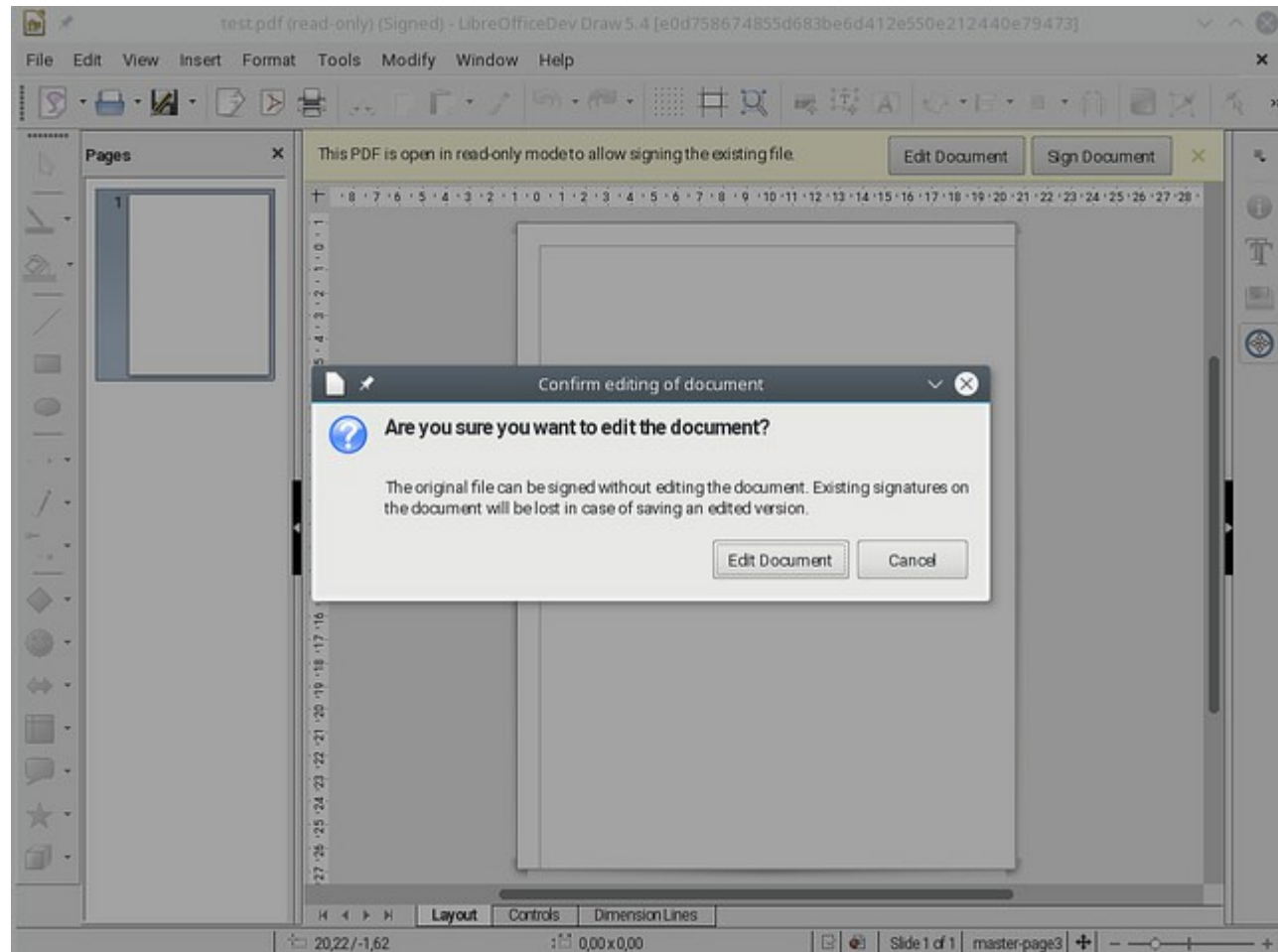
PDF signing: SHA1 → SHA256

- PDF signature verification:
 - Checking if the hash matches
 - Validating the signing certificate
- SHA1 is relevant for the first step
- SHA1 is considered to be weak today
- ODF/OOXML signing already used SHA256
 - PDF signing is now up to date with them



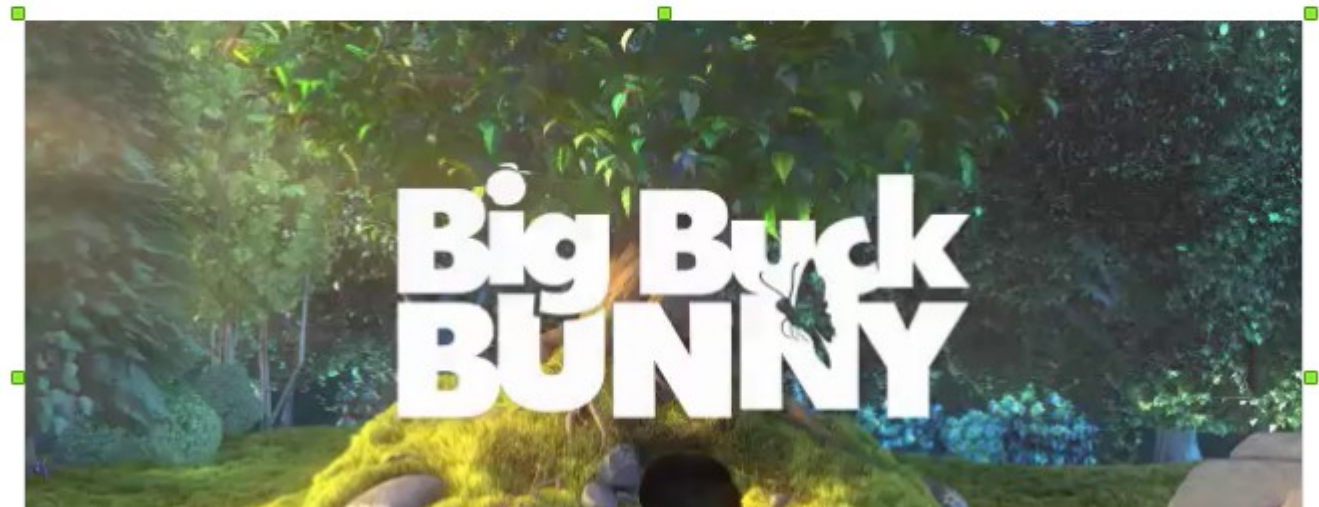
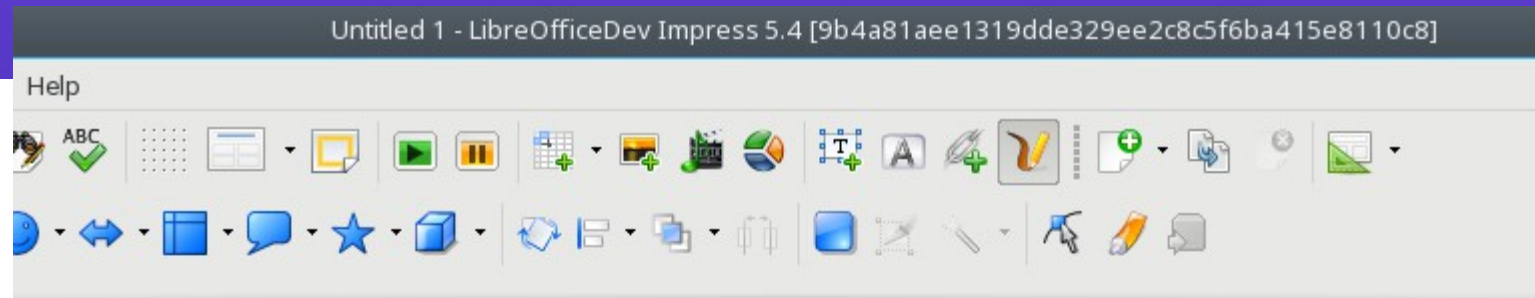
PAdES support

- A set of additional restrictions over normal PDF signatures
- Brings the possibility, so that the signature is legally binding
- Signs the certificate (necessary, as there can be multiple certificates for the same private key)



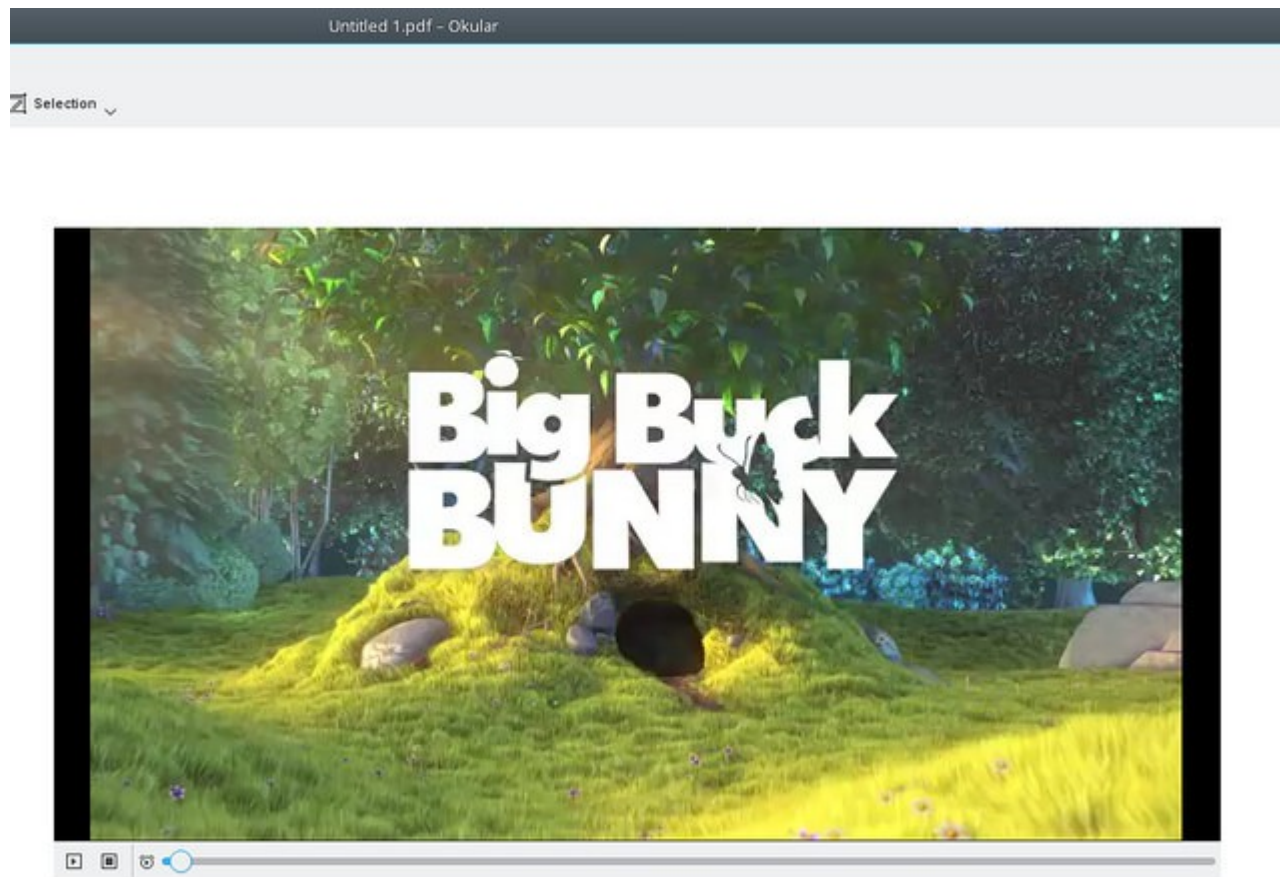
PDF export of linked videos

- Export of media shapes to PDF
- Actual video is a URL
- Snapshot image by avmedia
- Free of flash – not something Acrobat writes (but it can read it)



PDF export of embedded videos

- Embedding case: video in PDF can be viewed offline
- LO still just transfers the byte array



PDF export of text fill color

- Relevant for Impress/Draw, Writer already created a separate rectangle for this purpose
- Initial version, then one that handles rotation
- pdfium API
 - For test purposes



hello



hello



pdfium to render PDF images

- Old way: import via poppler, an external process and ODF into Draw, then copy the Draw page as a metafile
- New way: render into a bitmap by pdfium
- Better rendering:
 - e.g. embedded fonts
 - Quality of Foxit
 - Now part of Chrome

Embedded fonts test

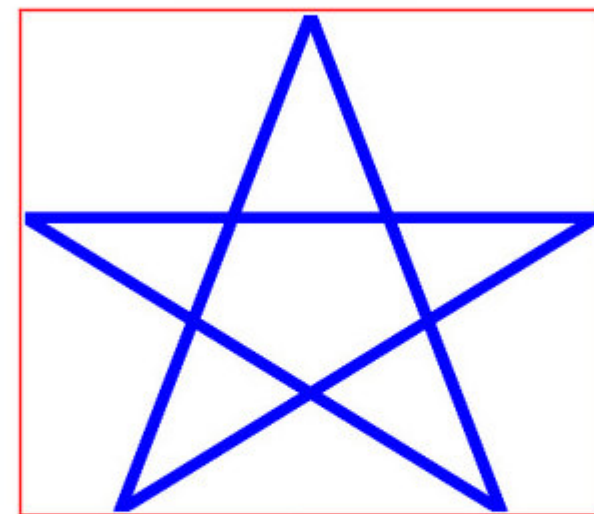
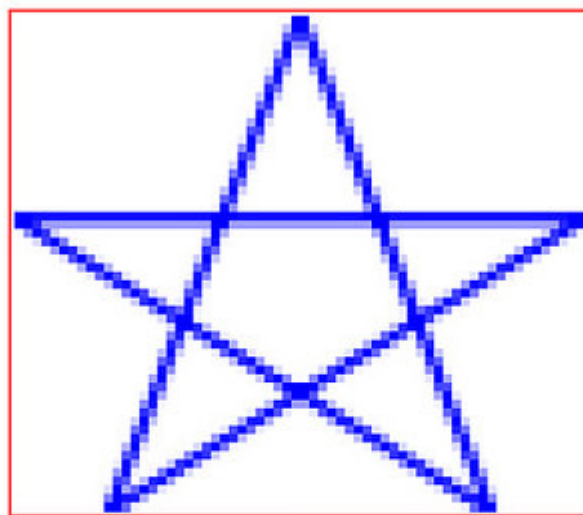
Libre - Callibri	Libre - Eroadway
Libre - Cambria	Libre - Ariel Black
Libre - Bauhaus	LIBRE - CASTELLAR
<i>Libre - Blackadder</i>	Libre - Berlin Sans FB
<i>Libre - Vladimir Script</i>	Libre - Gill Sans MT
¶ © ∞ X ¶	Libre - Elephant
<i>Libre - Jokerman</i>	Libre - Impact
<i>Libre - Lucida Cal</i>	<i>Libre - Magneto</i>
Libre - Old English T.	Libre - Paybill
Libre - Revié	LIBRE - STENCIL
Libre - Sketchflo	<i>Libre - Vinetti</i>

Embedded fonts test

Libre - Callibri	Libre - Eroadway
Libre - Cambria	Libre - Ariel Black
Libre - Bauhaus	LIBRE - CASTELLAR
<i>Libre - Blackadder</i>	Libre - Berlin Sans FB
<i>Libre - Vladimir Script</i>	Libre - Gill Sans MT
¶ © ∞ X ¶	Libre - Elephant
<i>Libre - Jokerman</i>	Libre - Impact
<i>Libre - Lucida Cal</i>	<i>Libre - Magneto</i>
Libre - Old English T.	Libre - Paybill
Libre - Revié	LIBRE - STENCIL
Libre - Sketchflow	<i>Libre - Vinetti</i>

Roundtrip PDF images to PDF: reference XObjects

- Problem: pdffium renders to a bitmap
- Export back to PDF contains this bitmap
- Idea: use the reference XObject markup
 - Can wrap a page from an existing PDF as an image



Roundtrip PDF images to PDF: form XObjects

- Problem: form XObject markup is ~only supported by Acrobat
- Solution: use form XObjects, which can refer to an existing PDF object
 - Much more work, all references has to be recursively copied over from the original file
 - References are unique identifiers, so all references have to be also rewritten
 - At the end works nicely, supported ~everywhere



Roundtrip PDF images to PDF: form XObjects, down-conversion

- Additional problem: we write PDF 1.4, what if the PDF image is 1.5+?
- Turns out that the problematic markup has equivalent in PDF 1.4, just less optimal (no way to compress, etc.)
- Solution: use pdium to down-convert 1.5+ to 1.4, and then feed that into the form XObject embedder



PDF export from Writer: the magic “subtract flys” option

- Writer compatibility option: paint order not only depends on z-order, but also on anchoring hierarchy
- Requires to not paint the full background in one go
 - rounding errors, unexpected white lines
- Not enabled for new documents, but users still suffer
- Fixed a number of rounding errors in the PDF export
 - Also there is now UI to disable the legacy behavior if you don't depend on it



asdf



How are these
implemented?



Code pointers: PDF signature handling

- xmlsecurity has the doc signing bits:
 - `xmlsecurity/source/helper/pdfsignaturehelper.cxx`
 - `xmlsecurity/source/pdfio/pdfdocument.cxx`
- Shared “sign a byte array” code:
 - `svl/source/crypto/`
- PDF tokenizer:
 - `vcl/source/filter/ipdf/pdfdocument.cxx`
 - Used for PDF image roundtrip and signing

Code pointers: pdfium

- PDF image import filter:
 - `vcl/source/filter/ipdf/pdfread.cxx`
- PDF image roundtrip, export code:
 - `vcl/source/gdi/pdfwriter_impl.cxx`
 - `PDFWriterImpl::writeReferenceXObject()`
 - `PDFWriterImpl::copyExternalResources()`
 - This is the recursive function, handling the object graph

Code pointers: PDF export & testcases

- PDF export shared bits:
 - vcl/source/gdi/pdf*
 - The PDF export is an output device you can draw on at the end
- Application-specific bits, like link handling:
 - sw/source/core/text/EnhancedPDFExportHelper.cxx
 - sd/source/ui/unoidl/unomodel.cxx
 - ImplPDF*() functions
- Testsuite: CppunitTest_vcl_pdfexport
 - Parses the result with pdfium & asserts with its API

Summary

- PDF support in LibreOffice improved significantly in the past year:
 - PDF signature handling
 - pdium integration
 - PDF image roundtrip
 - Various PDF export / testing improvements
- Thanks for the sponsors and for listening! :-)
- Slides: <https://vmiklos.hu/odp>

