

Automatic Documents, packed with content and signed

Michael Meeks

CEO, Collabora Productivity

michael.meeks@collabora.com

2025-02-01



Collabora
Online



FOSDEM



Automatic Documents: Extract / Transform API



Why it is useful

- We can get or set data without actually opening the document.
(Extract/Transform)
- Transform: Scripts can generate pre-filled documents.
(form with personal user info already filled)
- Extract: Scripts can search specific contents in multiple documents
- Extract + Transform: Scripts can gather data, and generate statistic, charts from it.
- Many more use cases...



Usage - new cool/ end-points:

HTTP POST to
/cool/extract-document-structure
/cool/transform-document-structure

Extract: curl -k -F "data=@**test.docx**" -F "filter=contentcontrol"
https://localhost:9980/cool/extract-document-structure > **out.json**

Transform: curl -v -k -F "data=@**test.docx**" -F "format=docx" -F
"transform=\$(cat **transform.JSON**)"
https://localhost:9980/cool/transform-document-structure > **out.docx`**



Actually supported capabilities:

Writer:

- **Content Controls**
overwrite text / contents
checkbox, date-picker ...
- **Charts**
Add/delete data/description
set data with matrix / vector
- **Document Properties**
overwrite standard properties
add/remove custom properties

Impress:

- **Slides**
Insert / delete / reorder slides
change layout / master slide
overwrite texts

Probably there will be much more...



Extract

Document properties

Write data to JSON file

Properties of "temp2"

General Description Custom Properties Font Statist

Title:	Title TxT
Subject:	Subject TxT
Keywords:	
Contributor:	Contributor1 TxT, Contributor2 TXT
Coverage:	Coverage TxT
Identifier:	Identifier TxT
Publisher:	Publisher TxT, Publisher2 TXT
Relation:	Relation TxT, Relation2 TXT
Rights:	Rights TxT
Source:	Source TxT
Type:	Type TxT
Comments:	Description TxT



```
"DocStructure": {  
  "DocPops": {  
    "Author": "Author TxT",  
    "Generator": "Generator TxT",  
    "CreationDate": "2024-01-21T14:45:00",  
    "Title": "Title TxT",  
    "Subject": "Subject TxT",  
    "Description": "Description TxT",  
    "Keywords": [ ],  
    "Language": "en-GB",  
    "ModifiedBy": "ModifiedBy TxT",  
    "ModificationDate": "2024-05-23T10:05:50.159530766",  
    "PrintedBy": "PrintedBy TxT",  
    "PrintDate": "0000-00-00T00:00:00",  
    "TemplateName": "TemplateName TxT",  
    "TemplateURL": "TemplateURL TxT",  
    "TemplateDate": "0000-00-00T00:00:00",  
    "AutoloadURL": "",  
    "AutoloadSecs": 0,  
    "DefaultTarget": "DefaultTarget TxT",  
    "DocumentStatistics": {  
      "PageCount": 300,  
      "TableCount": 60,  
      "ImageCount": 10,  
      "ObjectCount": 0,  
      "ParagraphCount": 2880,  
      "WordCount": 78680,  
      "CharacterCount": 485920,  
      "NonWhitespaceCharacterCount": 411520  
    },  
    "EditingCycles": 12,  
    "EditingDuration": 12345,  
    "Contributor": [ "Contributor1 TxT", "Contributor2 TXT"],  
    "Coverage": "Coverage TxT",  
    "Identifier": "Identifier TxT",  
    "Publisher": [ "Publisher TxT", "Publisher2 TXT"],  
    "Relation": [ "Relation TxT", "Relation2 TXT"],  
    "Rights": "Rights TxT",  
    "Source": "Source TxT",  
    "Type": "Type TxT",  
    "UserDefinedProperties": {  
      "NewPropName Bool": { "type": "boolean", "value": true},  
      "NewPropName Numb": { "type": "long", "value": 1245},  
      "NewPropName Str": { "type": "boolean", "value": false},  
      "NewPropName float": { "type": "float", "value": 12.45}  
    }  
  }  
}
```



Document Properties Transform

Properties of "docStructureChartExampleOrig

General Description Custom Properties Font

Title:

Subject:

Keywords:

Contributor:

Coverage:

Identifier:

Publisher:

Relation:

Rights:

Source:

Type:

Comments:

Properties Font Statisti

Title TxT

Subject TxT

Contributor1 TxT, Contributor2 TXT

Coverage TxT

Identifier TxT

Publisher TxT, Publisher2 TXT

Relation TxT, Relation2 TXT

Rights TxT

Source TxT

Type TxT

Description TxT

type:

Comments:

```
"Transforms": {  
  "DocumentProperties": {  
    "Title": "Title TxT",  
    "Subject": "Subject TxT",  
    "Keywords": [ ],  
    "Contributor": ["Contributor1 TxT", "Contributor2 TXT"],  
    "Coverage": "Coverage TxT",  
    "Identifier": "Identifier TxT",  
    "Publisher": ["Publisher TxT", "Publisher2 TXT"],  
    "Relation": ["Relation TxT", "Relation2 TXT"],  
    "Rights": "Rights TxT",  
    "Source": "Source TxT",  
    "Type": "Type TxT",  
    "Description": "Description TxT",  
  }  
}
```



Content Control Transform

Original Document

Plain text: This is plain text
Rich text: This *is* a rich text
Check Box:
Combo Box: C1
Drop Down List: first
Date: Oct 7, 2024

Transform

```
{  
  "Transforms": {  
    "ContentControls.ByIndex.0": {  
      "content":""  
    },  
    "ContentControls.ByTag.Rich tag": {  
      "content":"New text, losing formats",  
      "alias":"RichTag"  
    },  
    "ContentControls.ByAlias.Date title": {  
      "content":"May 14, 2024",  
      "date":"2024-05-14"  
    },  
    "ContentControlsById.1086976183": {  
      "checked":true  
    },  
    "ContentControls.ByAlias.Drop title": {  
      "content":""  
    },  
    "ContentControls.ByAlias.Combo title": {  
      "content":"C3"  
    }  
  }  
}
```

Result Document

Plain text: Click here to enter text
Rich text: New text, losing formats
Check Box:
Combo Box: C3
Drop Down List: Choose an item
Date: May 14, 2024



Chart Extract

Write data to JSON file:

```
{  
  "DocStructure": {  
    "Charts.ByEmbedIndex.0": {  
      "name": "Object1",  
      "title": "Paid leave days",  
      "subtitle": "Subtitle2",  
      "RowDescriptions": [ "James", "Mary", "Patricia", "David"],  
      "ColumnDescriptions": [ "2022", "2023"],  
      "DataValues": [  
        "Row.0": [ "22", "24"],  
        "Row.1": [ "18", "16"],  
        "Row.2": [ "32", "32"],  
        "Row.3": [ "25", "23"]  
      ]  
    }  
  }  
}
```

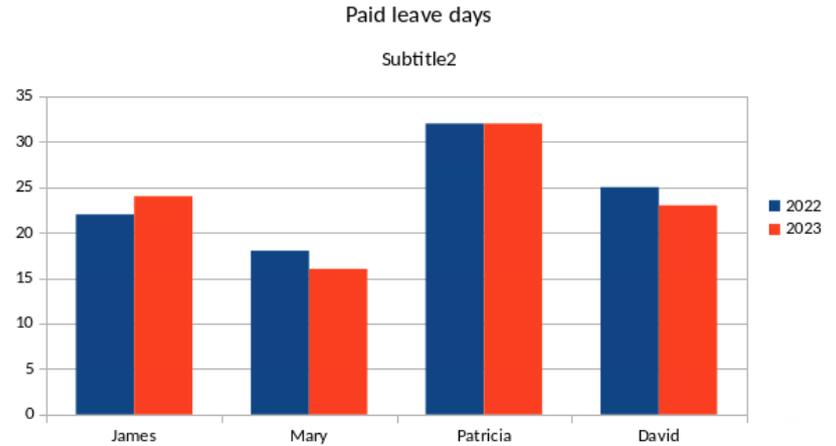
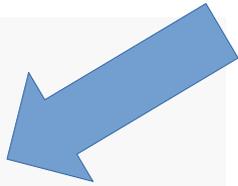
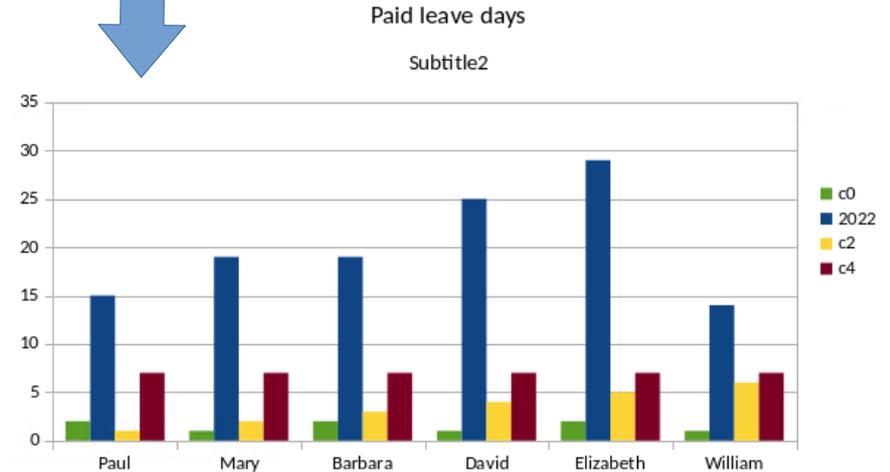




Chart Transform

Commands are in JSON file:

```
{  
  "Transforms": {  
    "Charts.ByEmbedIndex.0": {  
      "modifyrow.1": [ 19, 15 ],  
      "datayx.3.1": 37,  
      "deleterow.0": "",  
      "deleterow.1": "",  
      "insertrow.0": [ 15, 17 ],  
      "setrowdesc.0": "Paul",  
      "insertrow.2": [ 19, 22 ],  
      "setrowdesc.2": "Barbara",  
      "insertrow.4": [ 29, 27 ],  
      "setrowdesc.4": "Elizabeth",  
      "insertrow.5": [ 14, 26 ],  
      "setrowdesc.5": "William",  
      "insertcolumn.1": [ 1, 2, 3, 4, 5, 6 ],  
      "insertcolumn.0": [ 2, 1, 2, 1, 2, 1 ],  
      "insertcolumn.4": [ 7, 7, 7, 7, 7, 7 ],  
      "setcolumndesc.4": "c4",  
      "setcolumndesc.0": "c0",  
      "setcolumndesc.2": "c2"  
    },  
    "Charts.BySubTitle.Subtitle2": {  
      "deletecolumn.3": ""  
    }  
  }  
}
```

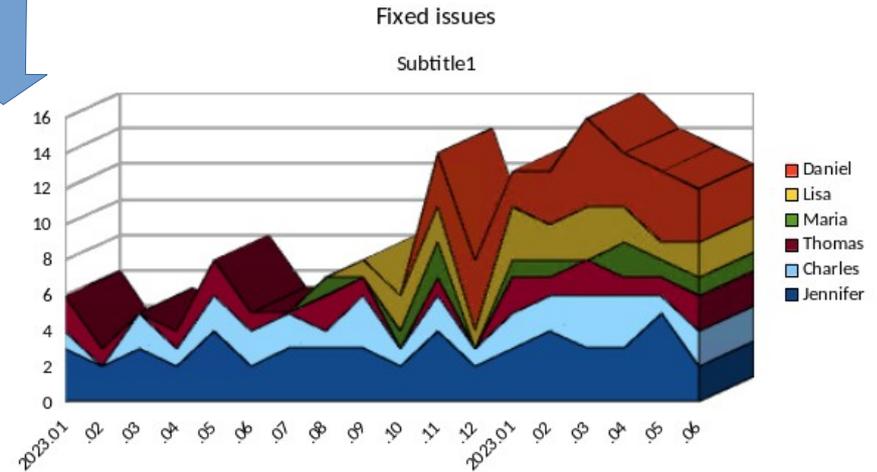
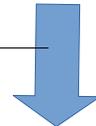




Multiple ways...

Set with matrix/vector:

```
"Charts.ByTitle.Fixed issues": {  
  "data": [ [ 3,1,2 ],  
            [ 2,0,1 ],  
            [ 3,2,0 ],  
            [ 2,1,1 ],  
            [ 4,2,2 ],  
            [ 2,2,1 ],  
            [ 3,2,0,0 ],  
            [ 3,1,2,1 ],  
            [ 3,3,1,0,1 ],  
            [ 2,1,0,1,2 ],  
            [ 4,2,1,2,2,3 ],  
            [ 2,1,0,0,1,4 ],  
            [ 3,2,2,1,3,2 ],  
            [ 4,2,1,1,2,3 ],  
            [ 3,3,2,0,3,5 ],  
            [ 3,3,1,2,2,3 ],  
            [ 5,1,1,1,1,4 ],  
            [ 2,2,2,1,2,3 ] ],  
  "setrowdesc": [ "2023.01", ".02", ".03", ".04", ".05", ".06",  
                  ".07", ".08", ".09", ".10", ".11", ".12",  
                  "2023.01", ".02", ".03", ".04", ".05", ".06" ],  
  "setcolumndesc": [ "Jennifer", "Charles", "Thomas",  
                     "Maria", "Lisa", "Daniel" ]  
}
```





Slides Transform

1 Collabora Online
Powerful Online Collaboration

2 Collabora Online
Distinctives

3 Friendly Open Source Project

4 Performance - faster & simpler

5 With thanks to our Partners,
Customers & Community!

```
{  
  "Transforms": {  
    "SlideCommands": [  
      {"JumpToSlideByName": "Slide 3"},  
      {"MoveSlide": 0},  
      {"RenameSlide": "Slide3-Renamed"},  
      {"DeleteSlide": 2},  
      {"JumpToSlide": 2},  
      {"DeleteSlide": ""},  
      {"JumpToSlide": 1},  
      {"DuplicateSlide": ""},  
      {"RenameSlide": "Slide1-Duplicated"},  
      {"InsertMasterSlide": 1},  
      {"RenameSlide": "SlideInserted-1"},  
      {"ChangeLayout": 18},  
      {"JumpToSlide": "last"},  
      {"InsertMasterSlideByName": "Topic Separator white"},  
      {"RenameSlide": "SlideInserted-Name"},  
      {"ChangeLayoutByName": "AUTOLAYOUT_TITLE_2CONTENT"},  
      {"SetText.0": "first"},  
      {"SetText.1": "second"},  
      {"SetText.2": "third"},  
      {"DuplicateSlide": 1},  
      {"MoveSlide.2": 6}  
    ]  
  }  
}
```

1 Friendly Open Source Project

2 Collabora Online

3 Collabora Online

4

5 With thanks to our Partners,
Customers & Community!

6 Title
first
second
third

7 Collabora Online



Documentation

For more details, explanations of actually usable commands, samples check the documentation:

https://sdk.collaboraonline.com/docs/extract_transform_api.html

Or check Unittests in the sourceCode search for:

`.uno:TransformDocumentStructure`

`.uno:ExtractDocumentStructure`

Electronic signing



Motivation

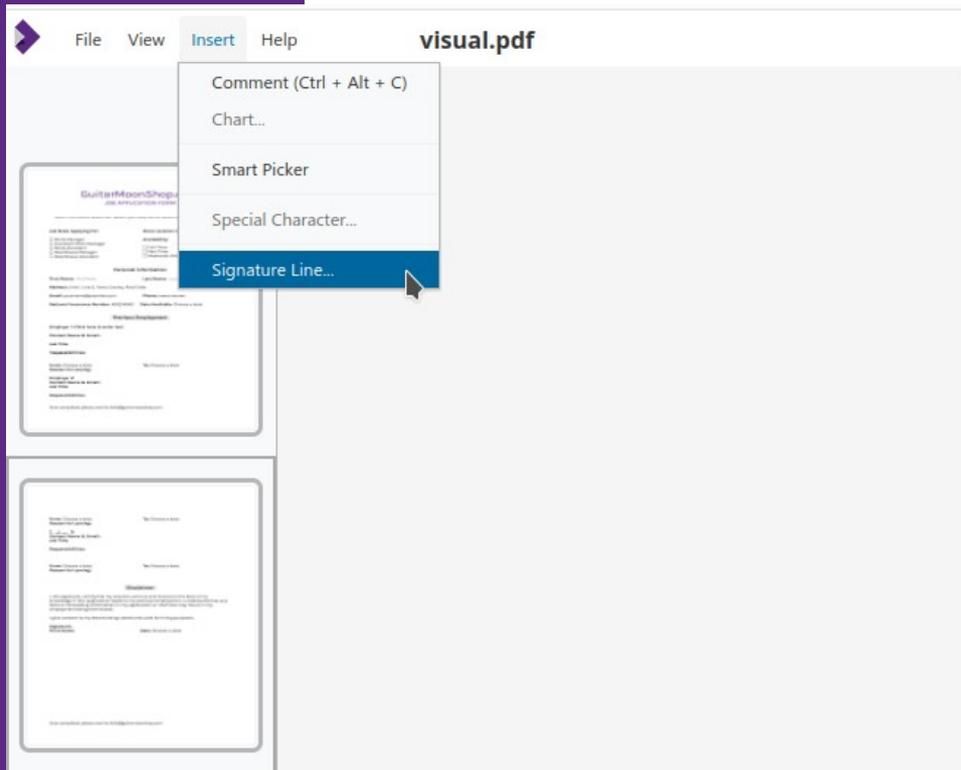
Qualified electronic signatures: can be considered as a digital equivalent to handwritten signatures

- Create proper electronic signatures in COOL, based on your personal ID / passport
- Not share your document with a 3rd party, just the hash of the PDF
- Integrate with Nextcloud, so just need to install the Nextcloud AIO image, nothing else – SDK documentation also makes this possible for any other COOL integration, too
- Potentially combine with Secure View
- Work with visual signing to place a signature widget at a custom position



Step 1: insert a signature widget

Open a PDF in COOL and select Insert → Signature line



Employer 3:
Contact Name & Email:
Job Title:

Responsibilities:

From: Choose a date
Reason for Leaving:

To: Choose a date

Disclaimer

I, the Applicant, certify that my answers are true and honest to the best of my knowledge. If this application leads to my eventual employment, I understand that any false or misleading information in my application or interview may result in my employment being terminated.

I give consent to my details being stored and used for hiring purposes.

Signature:
Print Name:

Date: Choose a date



Step 2: position the signature widget

Place it similar to your handwritten signature or DocuSign

Disclaimer

I, the Applicant, certify that my answers are true and honest to the best of my knowledge. If this application leads to my eventual employment, I understand that any false or misleading information in my application or interview may result in my employment being terminated.

I give consent to my details being stored and used for hiring purposes.

Signature: Digitally signed by: Miklos Vajna
Print Name: Miklos Vajna
Date: 01/28/2025

Date: Choose a date

The signature line can now be moved or resized as needed. [Finish electronic signing](#)



Step 3: specify country & provider

Select your country and pick one available provider there, e.g. Evrotrust works in EU countries & in the UK, etc.

Digitally signed by:
Iiklos Vajna
Date: 01/28/2025

Date: Choose a date

INSERT ELECTRONIC SIGNATURE

Country:
Estonia

Provider:
Smart-ID

OK



Step 4: sign the document

The document hash is sent to eID Easy, which wraps many providers

The screenshot shows a PDF viewer interface with a document titled "visual.pdf". The document content includes a form with fields for "Employer 3:", "Contact Name & Email:", and "Job Title:". A signature overlay is visible on the document. In the foreground, a browser window titled "eID Easy - High quality eSignatures" is open, displaying a login form for SMART-ID. The form includes a "Personal identity code" field with the value "30303039914" and a "SIGN" button. A "Cancel" button is also visible below the form. The browser address bar shows the URL: "https://test.eideasy.com/single-method-signature?client_id=2IaeiZXbcKzIP1KvjZH9gnty2IJKM8Lg&doc_id=6dQB1Q6jf3KXlxGZ".



Step 5: view the created signature

File → Digital signatures can also show these later

The screenshot shows a PDF viewer interface with a menu bar (File, View, Insert, Help) and a document titled 'visual.pdf'. The document content includes a 'Disclaimer' section and a 'DIGITAL SIGNATURES' dialog box. The dialog box displays a table of signatures and a 'View Certificate' button.

From: Choose a date
To: Choose a date
Reason for Leaving:

Disclaimer

I, the Applicant, certify that my answers are true and honest to the best of my knowledge. If this application leads to my eventual employment, I understand that I may result in my employment for the purposes of the company.

DIGITAL SIGNATURES

The following have signed the document content:

Signed by	Digital ID	Issued by	Date	Description	Signature type
TESTNUMBER,OK	TEST	of EID-SK	2016 01/28/2025 14:49:03		PaDES

The signatures in this document are valid

View Certificate Remove

Help Close

Once completed, please send to hello@guitarmoonshop.com



A real signature, verified in Acrobat

This time using a real personal ID and the Evrotrust provider

- A Qualified Electronic Signature according to EU regulations

Signed and all signatures are valid. **Signature Panel**

1 SIGNATURES

	TITLE
My company	Customer
Signature: Digitally signed by: Miklos Vajna Date: 01/28/2025	Signature:

Signatures

- ✓ Rev. 1: Signed by MIKLOS VAJNA
 - Signature is valid:
 - Source of Trust obtained from European Union Trusted Lists (EUTL)
 - This is a Qualified Electronic Signature according to EU Regulation
 - Document has not been modified since this signature was applied
 - Signer's identity is valid
 - The signature includes an embedded timestamp.
 - Signature is not LTV enabled and will expire after 2026/05/21 00:00:00
- > Signature Details
 - Last Checked: 2025.01.28 15:50:17 +01'00'
 - Field: Signature1 on page 1
 - [Click to view this version](#)



How is esign support implemented?

Collabora Online integrates with eID Easy to support many providers

- The document hash is extracted
- The hash is sent to eID Easy
- The hash is signed in a popup, on eID Easy servers or with a redirect (similar to online payments)
- The signed hash gets downloaded
- The downloaded hash is serialized to the PDF file as an appended signature



Summary

Available in Collabora Online 24.04 + LibreOffice technology 25.8:

- Automatic documents provides
 - Extract information about documents automatically
 - Transform / fill in documents automatically
- Electronic signing:
 - Qualified electronic signatures: can be considered as a digital equivalent to handwritten signatures
 - Signature according to EU regulations