XWIKI
YOUR INFORMATION ORGANIZED

Importing data in an Open Source Wiki

FOSDEM'21 Online

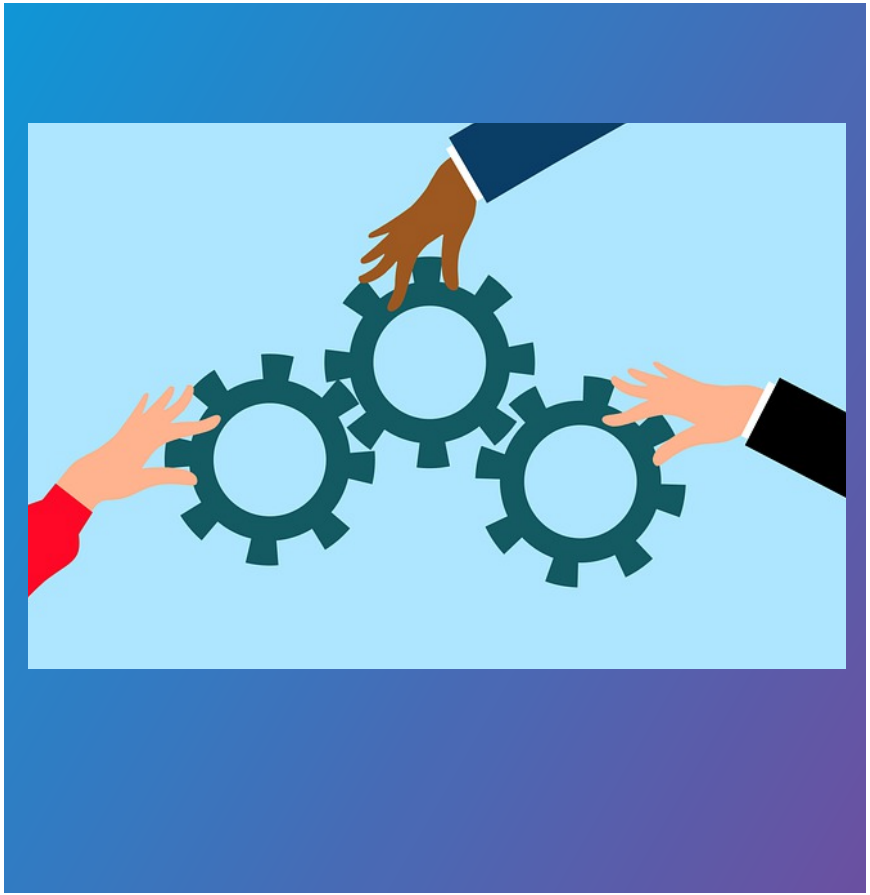
by Ludovic Dubost, XWiki SAS

Who am I ?

Ludovic Dubost

Creator of XWiki & Founder of XWiki SAS

- **XWiki & XWiki SAS**
- **CryptPad**
- **All Open Source**
- **16 years of Open Source business**



Why use a Wiki and XWiki ?

Improved Knowledge sharing and organization of information

Information is easier to create and organize than in Office files (Microsoft Sharepoint)

XWiki is a prime Open Source alternative to Atlassian Confluence

public domain image by **j4p4n**

Why importing ?

Even when users want to go for Open Source, the transition process is a key difficulty

A Wiki works much better with data in it, to help users understand the value of the Wiki concept

Types of Imports

Sources of data are multiple

- **Office files**
- **A database**
- **Another wiki (Confluence, MediaWiki, GitHub wiki,)**
- **HTML Source**

Available Methods

Manual / Coding

Tools

- **Manual Wiki syntax conversion & Office file import**
- **Script using APIs of XWiki**
- **XWiki Batch Import including Office batch import**
- **Filter streams & Confluence, MediaWiki, DokuWiki, GitHub Wiki Filters**

The Manual Way

Wiki syntax conversion
Office File Importing

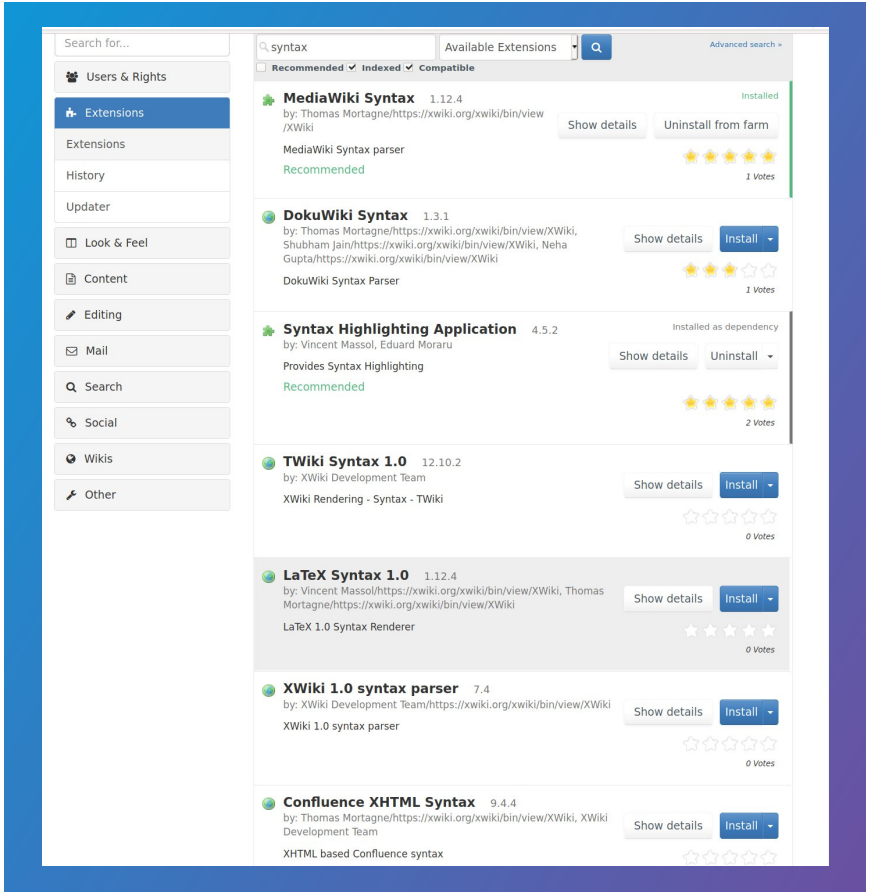


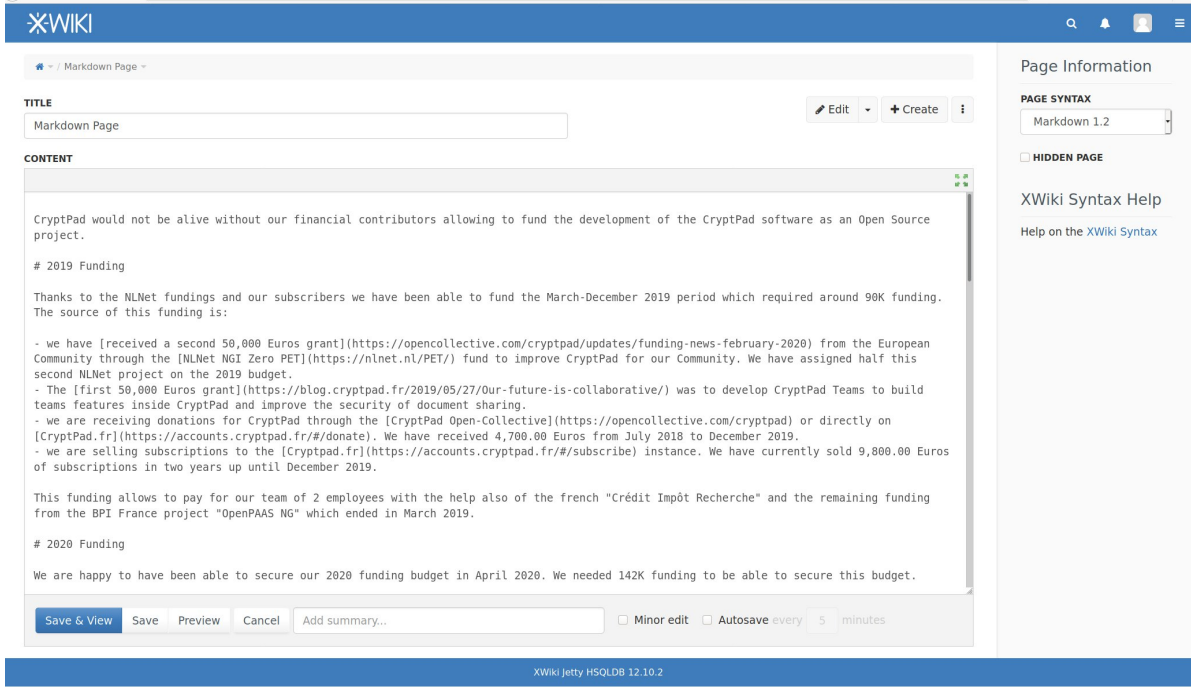
image from Tobie Langel

Manual Tools

Wiki & HTML Syntax Importing

1. Install and/or activate additional syntaxes in XWiki

Manual Tools: Wiki & HTML Syntax Importing

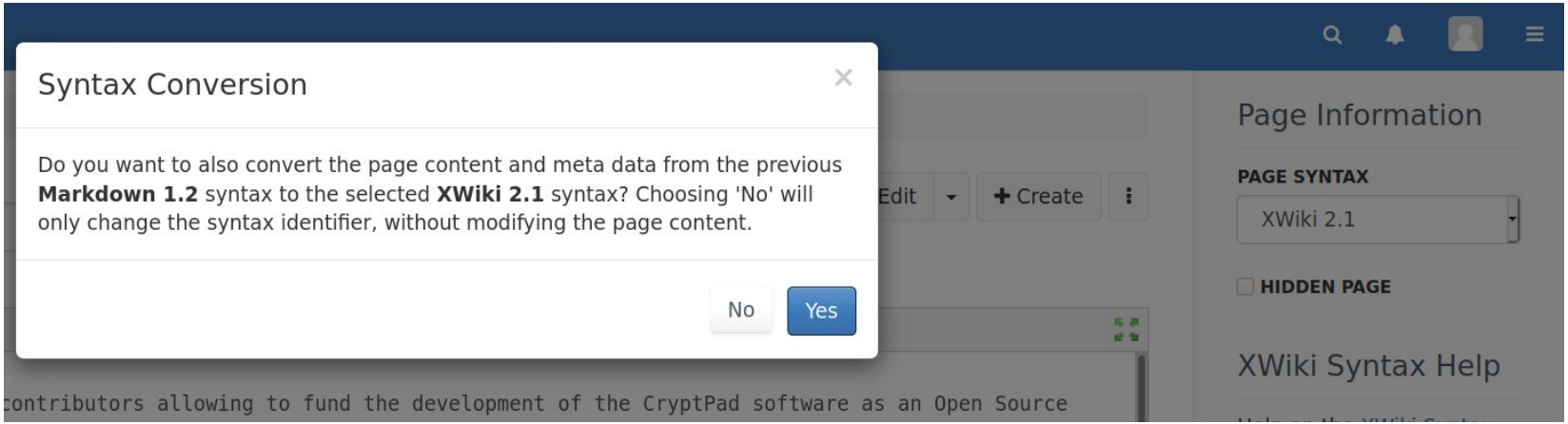


2. Copy paste HTML or Wiki Syntax from an outside source

3. Create XWiki document using the source syntax

4. Copy paste the content in it

Manual Tools: Wiki & HTML Syntax Importing



5. Edit, change to XWiki syntax and accept the conversion

Manual Tools: Wiki & HTML Syntax Importing

Imports / Markdown Page

Markdown Page

Edit | Create

Last modified by Administrator on 2021/01/16 14:57

CryptPad would not be alive without our financial contributors allowing to fund the development of the CryptPad software as an Open Source project.

2019 Funding

Thanks to the NLNet fundings and our subscribers we have been able to fund the March-December 2019 period which required around 90K funding. The source of this funding is:

- we have received a second 50,000 Euros grant from the European Community through the NLNet NGI Zero PET fund to improve CryptPad for our Community. We have assigned half this second NLNet project on the 2019 budget.
- The first 50,000 Euros grant was to develop CryptPad Teams to build teams features inside CryptPad and improve the security of document sharing.
- we are receiving donations for CryptPad through the CryptPad Open-Collective or directly on CryptPad.fr. We have received 4,700.00 Euros from July 2018 to December 2019.
- we are selling subscriptions to the Cryptpad.fr instance. We have currently sold 9,800.00 Euros of subscriptions in two years up until December 2019.

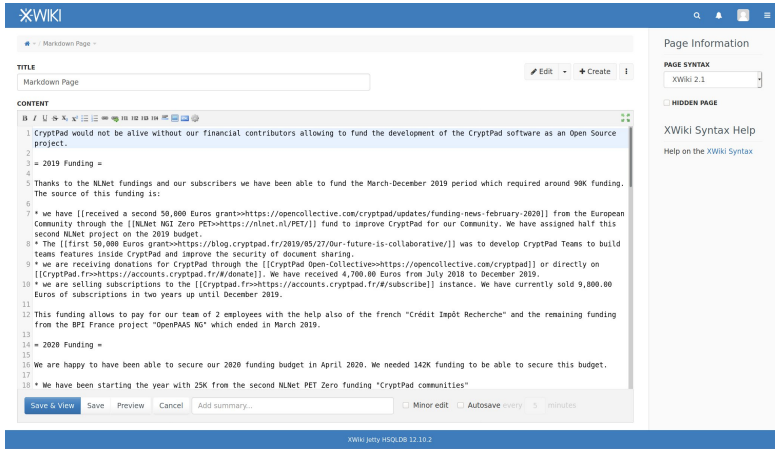
This funding allows to pay for our team of 2 employees with the help also of the french "Crédit Impôt Recherche" and the remaining funding from the BPI France project "OpenPAAS NG" which ended in March 2019.

2020 Funding

We are happy to have been able to secure our 2020 funding budget in April 2020. We needed 142K funding to be able to secure this budget.

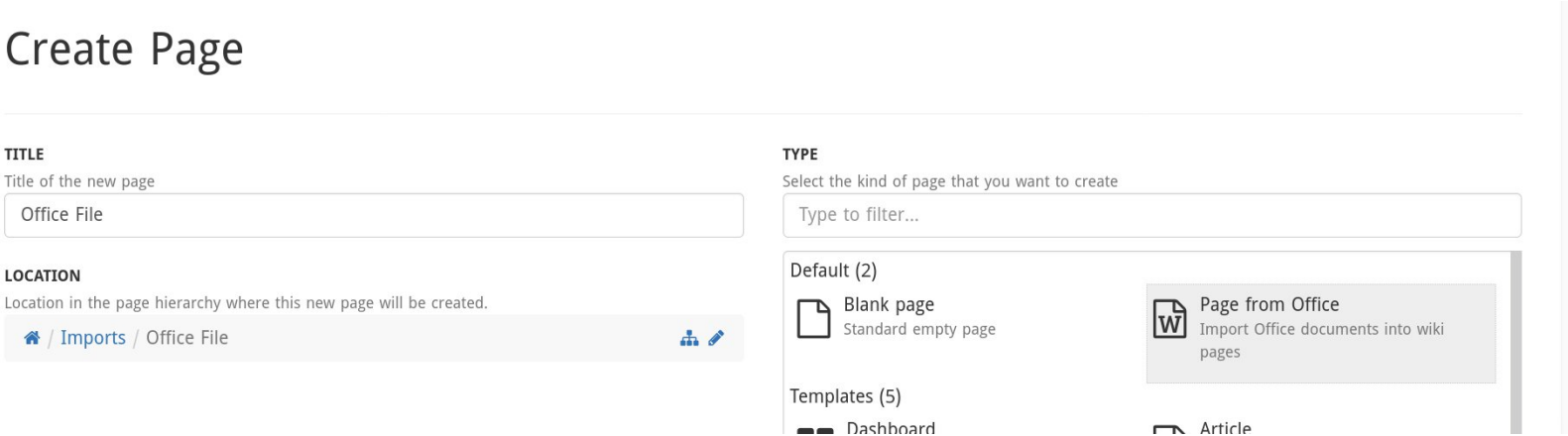
- We have been starting the year with 25K from the second NLNet PET Zero funding "CryptPad communities"
- We have been happy also to have our project "CryptPad SMC" (Secure Mobile Communication) been accepted by NGI Trust with a 75K funding for this project
- We have also received a funding of 9,1K from the MOSS program from the Mozilla Foundation after our pitch at FOSDEM
- Finally we are now expecting over 30K of revenue from subscriptions to CryptPad.fr and from donation on OpenCollective, which is higher that what we initially expected.
- A third NLNet PET Zero project "CryptPad Dialogue" has also been accepted for 50k funding, from which 3k will be assigned in 2020.

This brings our total funding for 2020 to 142KEuros



HTML can also be copy-pasted in the Wysiwyg editor

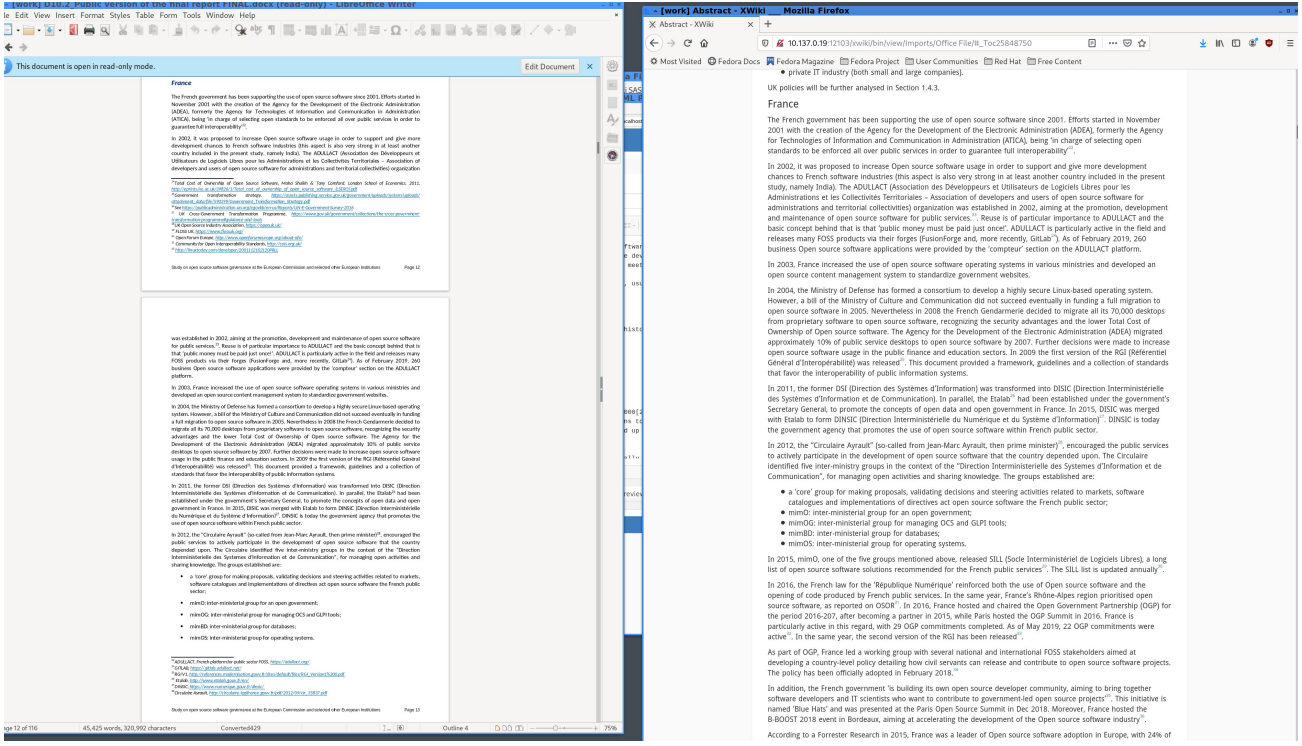
Manual Tools: Office Import



Uses LibreOffice server to convert to HTML then to XWiki syntax. Possibility to filter styles & split document in multiple pages.
Office Import is also possible from the Wysiwyg editor.

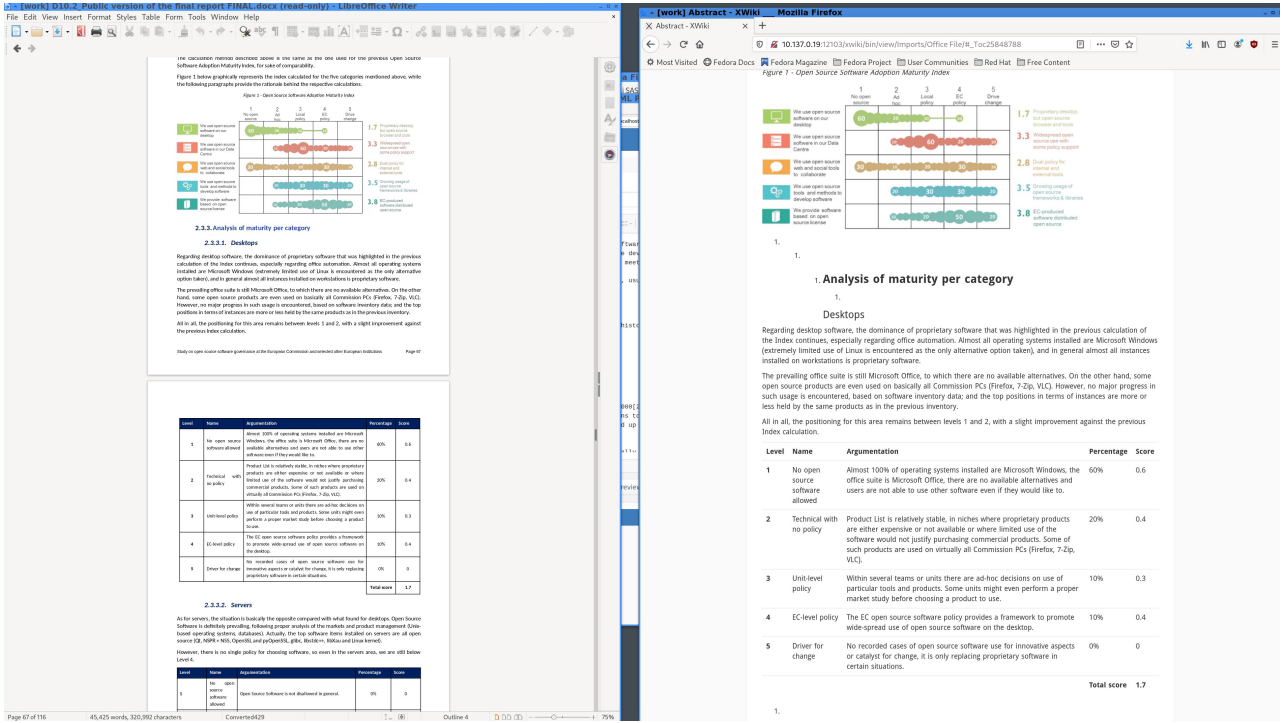
See:
https://extensions.xwiki.org/xwiki/bin/view/Extension/Office%20Importer%20Application

Manual Tools: Office Import



Uses LibreOffice server to convert to HTML then to XWiki syntax.
Possibility to filter styles & split document in multiple pages.

Manual Tools: Office Import



Imports images, tables, footnotes and others elements.

The developer way

Script using APIs

The Developer way

- Many APIs availables to convert HTML or other syntaxes to XWiki
- Java available through Groovy
- Job Macro allows to run long tasks
- Allows to customize fully the import process
- Examples: HCL Connections import for Lenovo, HTML imports with custom cleanup

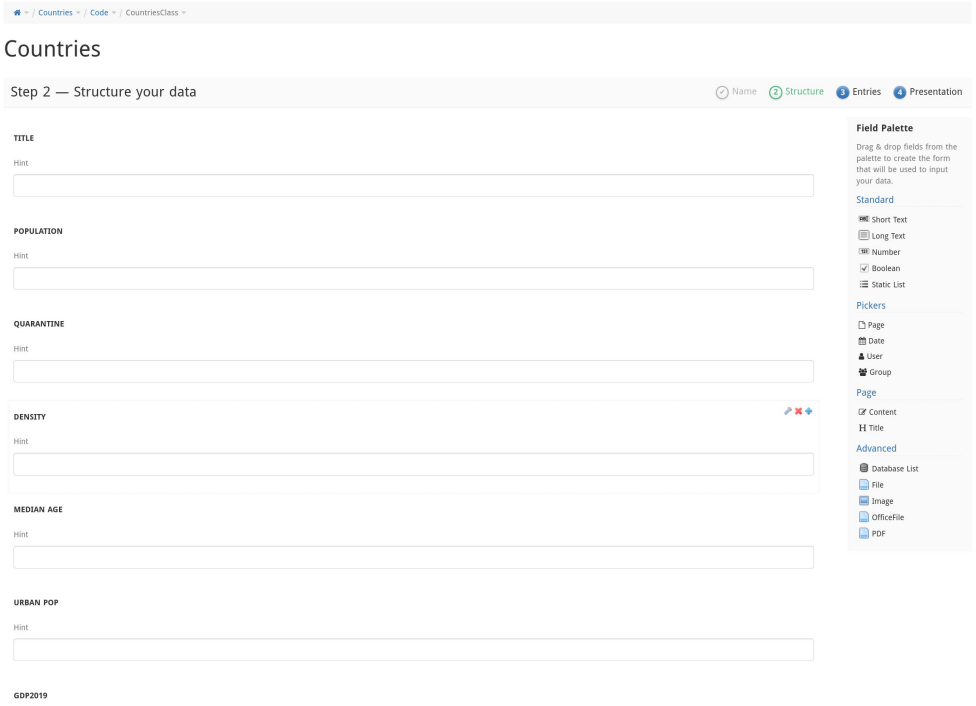
Batch Importing

Importing databases and automate file uploading and conversion in XWiki

Batch Import: General Principles

- Prepare a data structure in XWiki or use an existing one (Client list, Project list, Countries, Users)
- Prepare an Excel file of the data to import (from a database, by exporting data)
- List attached files in a column and upload the files to the server as a zip file
- Batch import to configure a mapping of the spreadsheet
- Preview and run the final import

Batch Import: create an App

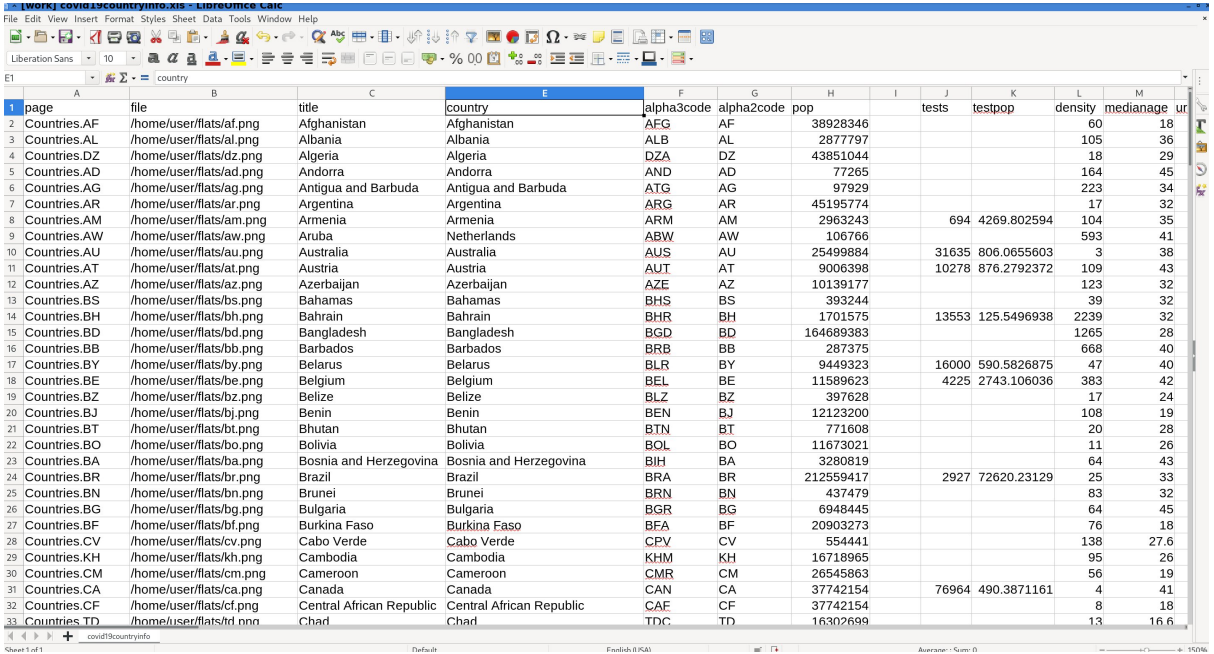


AppWithinMinutes is a powerful tool to create XWiki data structure

Your data structure will receive the imported data

It's a great feature of XWiki allowing to have better navigation in information

Batch Import: prepare the spreadsheet/csv



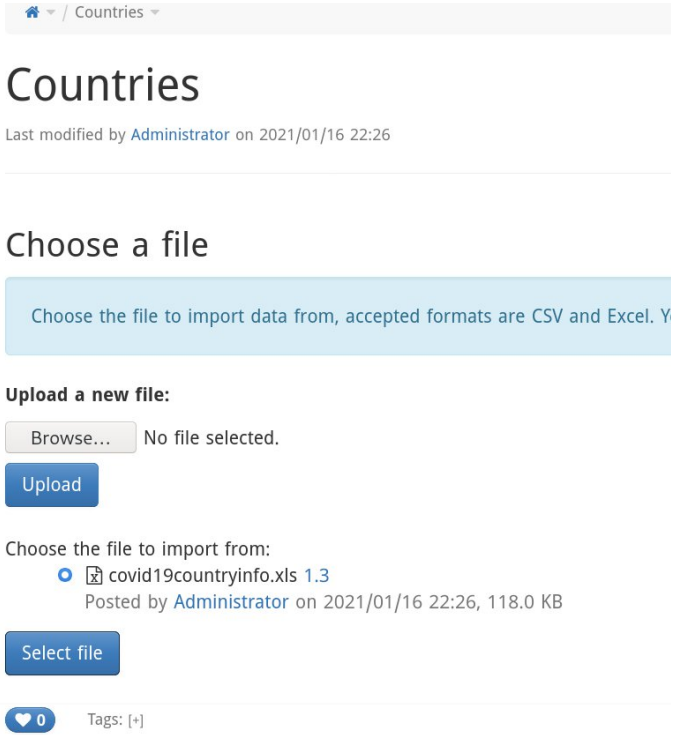
page	file	title	country	alpha3code	alpha2code	pop	tests	testpop	density	medianage	ur
Countries.AF	/home/user/flats/af.png	Afghanistan	Afghanistan	AFG	AF	38928346			60	18	
Countries.AL	/home/user/flats/al.png	Albania	Albania	ALB	AL	2877797			105	36	
Countries.DZ	/home/user/flats/dz.png	Algeria	Algeria	DZA	DZ	43851044			18	29	
Countries.AD	/home/user/flats/ad.png	Andorra	Andorra	AND	AD	77265			164	45	
Countries.AG	/home/user/flats/ag.png	Antigua and Barbuda	Antigua and Barbuda	ATG	AG	97929			223	34	
Countries.AR	/home/user/flats/ar.png	Argentina	Argentina	ARG	AR	45195774			17	32	
Countries.AM	/home/user/flats/am.png	Armenia	Armenia	ARM	AM	2963243	694	4269.802594	104	35	
Countries.AW	/home/user/flats/aw.png	Aruba	Netherlands	ABW	AW	106766			593	41	
Countries.AU	/home/user/flats/au.png	Australia	Australia	AUS	AU	25499884	31635	806.0655603	3	38	
Countries.AT	/home/user/flats/at.png	Austria	Austria	AUT	AT	9006398	10278	876.2792372	109	43	
Countries.AZ	/home/user/flats/az.png	Azerbaijan	Azerbaijan	AZE	AZ	10139177			123	32	
Countries.BS	/home/user/flats/bs.png	Bahamas	Bahamas	BHS	BS	393244			39	32	
Countries.BH	/home/user/flats/bh.png	Bahrain	Bahrain	BHR	BH	1701575	13553	125.5496938	2239	32	
Countries.BD	/home/user/flats/bd.png	Bangladesh	Bangladesh	BGD	BD	164689383			1265	28	
Countries.BB	/home/user/flats/bb.png	Barbados	Barbados	BRB	BB	287375			668	40	
Countries.BY	/home/user/flats/by.png	Belarus	Belarus	BLR	BY	9449323	16000	590.5826875	47	40	
Countries.BE	/home/user/flats/be.png	Belgium	Belgium	BEL	BE	11589623	4225	2743.106036	383	42	
Countries.BZ	/home/user/flats/bz.png	Belize	Belize	BLZ	BZ	397628			17	24	
Countries.BJ	/home/user/flats/bj.png	Benin	Benin	BEN	BJ	12123200			108	19	
Countries.BT	/home/user/flats/bt.png	Bhutan	Bhutan	BTN	BT	771608			20	28	
Countries.BO	/home/user/flats/bo.png	Bolivia	Bolivia	BOL	BO	11673021			11	26	
Countries.BA	/home/user/flats/ba.png	Bosnia and Herzegovina	Bosnia and Herzegovina	BIH	BA	3280819			64	43	
Countries.BR	/home/user/flats/br.png	Brazil	Brazil	BRA	BR	212559417	2927	72620.23129	25	33	
Countries.BN	/home/user/flats/bn.png	Brunei	Brunei	BRN	BN	437479			83	32	
Countries.BG	/home/user/flats/bg.png	Bulgaria	Bulgaria	BGR	BG	6948445			64	45	
Countries.BF	/home/user/flats/bf.png	Burkina Faso	Burkina Faso	BFA	BF	20903273			76	18	
Countries.CV	/home/user/flats/cv.png	Cabo Verde	Cabo Verde	CPV	CV	554441			138	27.6	
Countries.KH	/home/user/flats/kh.png	Cambodia	Cambodia	KHM	KH	16718965			95	26	
Countries.CM	/home/user/flats/cm.png	Cameroon	Cameroon	CMR	CM	26545863			56	19	
Countries.CA	/home/user/flats/ca.png	Canada	Canada	CAN	CA	37742154	76964	490.3871161	4	41	
Countries.CF	/home/user/flats/cf.png	Central African Republic	Central African Republic	CAF	CF	37742154			8	18	
Countries.TD	/home/user/flats/td.png	Chad	Chad	TDC	TD	16302699			13	16.6	

Add a page name, space and title column to prepare where the data will go in XWiki

A column will refer to the attached files (images, word file)

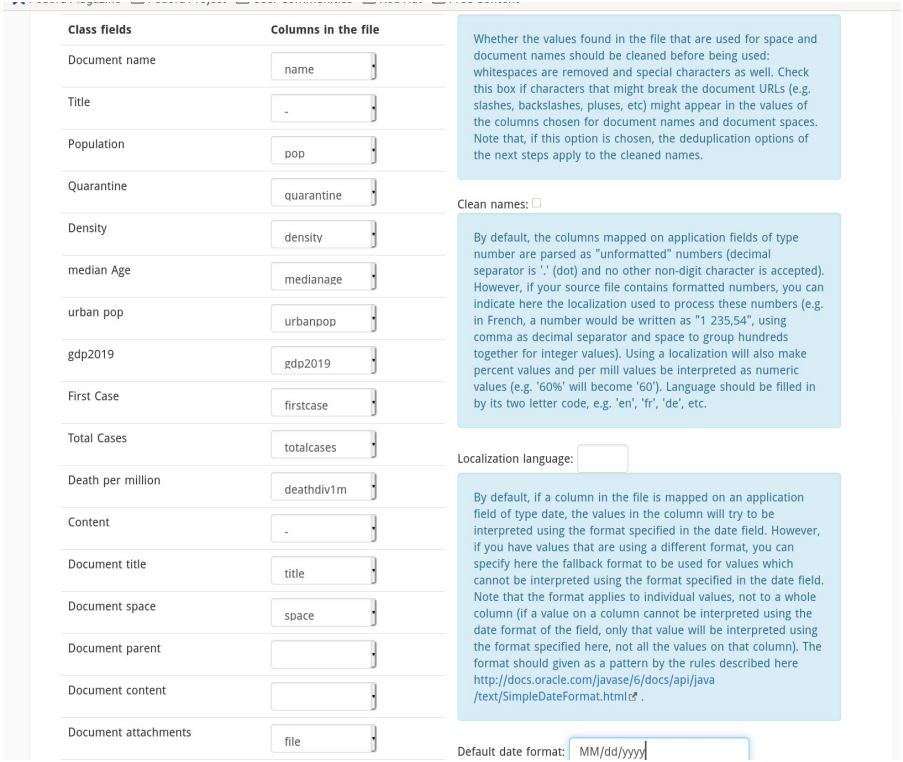
Clean up format of numbers and dates if necessary

Batch Import: upload the file



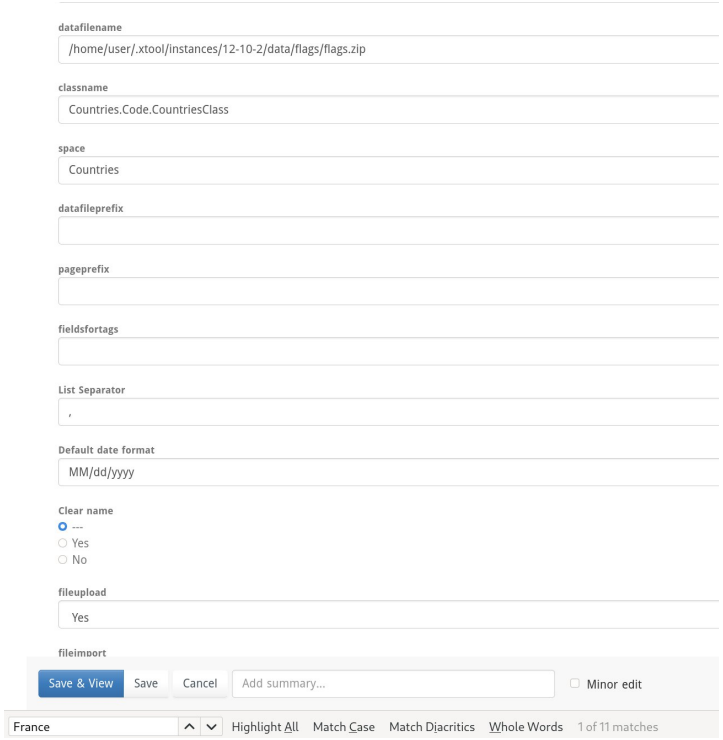
Both CSV and Excel are supported. The Excel plugin needs a configuration in xwiki.cfg and a restart

Batch Import: configure the mapping



Each column in the mapping can be mapped to a column in the App Within Minutes application

Batch Import: file attachment settings



Some hidden settings are necessary to perform file attachment or file conversion (from word to XWiki)

The files to be uploaded need to be transfered on the server as a zip file. The XWiki servers needs to be able to access it

Batch Import: preview imports

Countries

Edit | Create

Last modified by Administrator on 2021/01/16 22:42

Preview of data mapping and parsing

Following is a preview of the first 10 lines of the source file, with the values for the fields as they will be interpreted by the import system. A warning icon means that there is an error on processing that value and **the whole row will be skipped on import**, an empty cell means that there is no value in the input file for the column associated to that field.
You can go back and forth between mapping and preview until you get the result you want.
Note that the errors presented here refer to only the 10 first rows, if you need a complete verification of the file, advance to the import step and perform a simulation.

	Document full name	Population	Quarantine	Density	median Age	urban pop	gdp2019	First Case	Total Cases	Death per million	Document title	Document space	Document attachments
1	Countries.AF	38928346		60.0	18.0	25.0	18734	02/11/2021	273	0.2	Afghanistan	Countries	af.png
2	Countries.AL	2877797		105.0	36.0	63.0	15418	03/07/2020	304	6.0	Albania	Countries	al.png
3	Countries.DZ	43851044		18.0	29.0	73.0	172781	02/12/2021	986	2.0	Algeria	Countries	dz.png
4	Countries.AD	77265		164.0	45.0	88.0	91527	03/01/2020	439	207.0	Andorra	Countries	ad.png
5	Countries.AG	97929		223.0	34.0	26.0	1688	03/12/2020	9		Antigua and Barbuda	Countries	ag.png
6	Countries.AR	45195774	03/08/2021	17.0	32.0	93.0	445469	03/02/2020	1265	0.9	Argentina	Countries	ar.png
7	Countries.AM	2963243		104.0	35.0	63.0	13444	02/05/2022	736	2.0	Armenia	Countries	am.png
8	Countries.AW	106766		593.0	41.0	44.0	2903	03/12/2020	60		Aruba	Countries	aw.png
9	Countries.AU	25499884		3.0	38.0	86.0	1376255	01/12/2021	5350	1.0	Australia	Countries	au.png

Previewing allows to check if the data is converted properly.

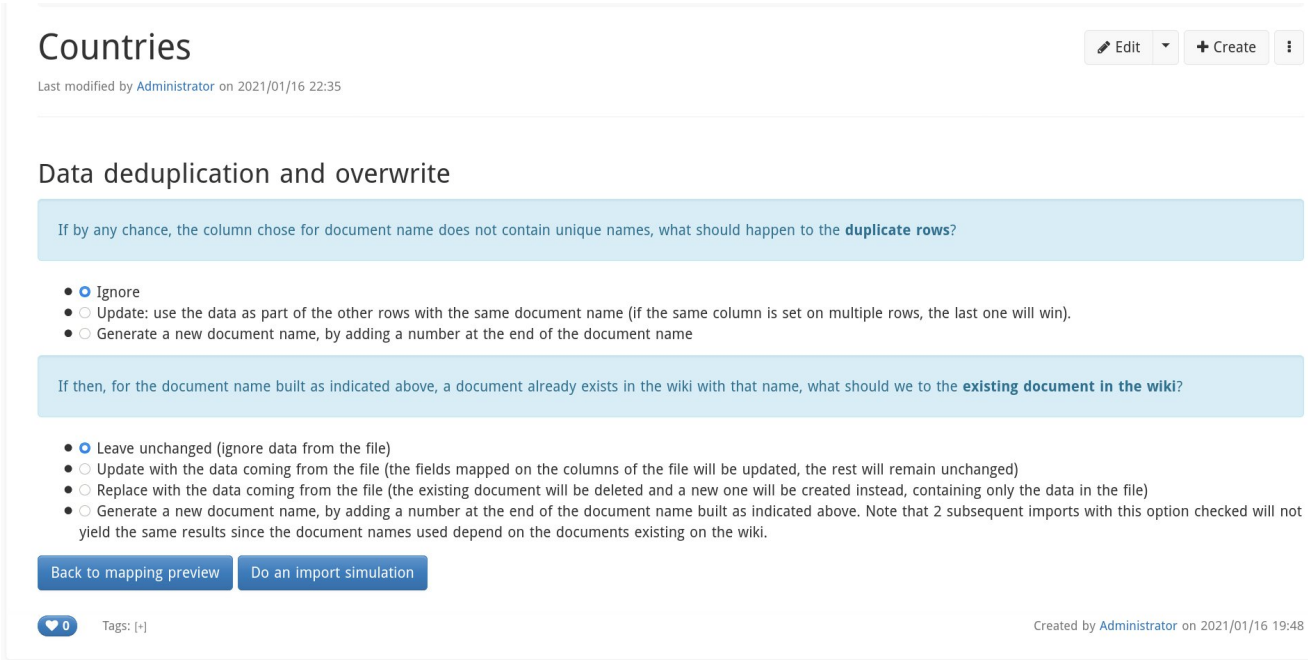
In particular check the document full name

Date conversions are possible.

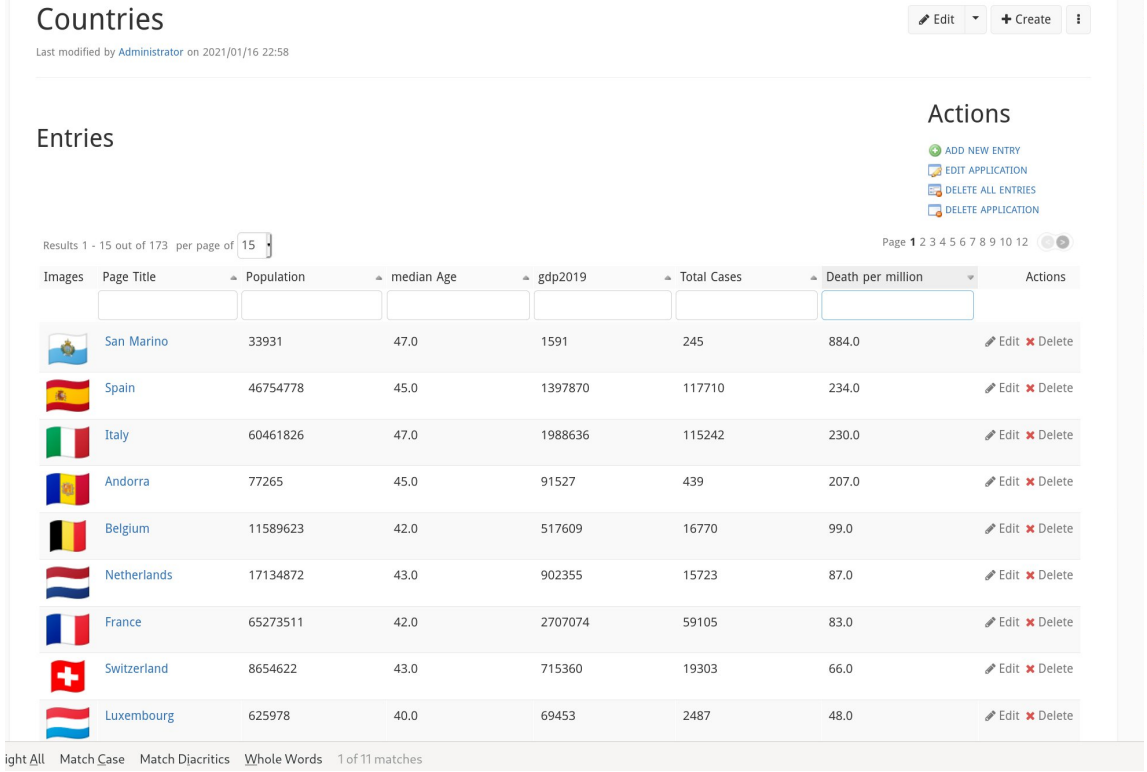
You can verify the attached file column is properly read

XWIKI

Batch Import: deduplication settings



Batch Import: enjoy the results



Countries

Edit | + Create

Last modified by Administrator on 2021/01/16 22:58

Entries

Actions

ADD NEW ENTRY
EDIT APPLICATION
DELETE ALL ENTRIES
DELETE APPLICATION

Results 1 - 15 out of 173 per page of 15

Page 1 2 3 4 5 6 7 8 9 10 12

Images	Page Title	Population	median Age	gdp2019	Total Cases	Death per million	Actions
	San Marino	33931	47.0	1591	245	884.0	Edit Delete
	Spain	46754778	45.0	1397870	117710	234.0	Edit Delete
	Italy	60461826	47.0	1988636	115242	230.0	Edit Delete
	Andorra	77265	45.0	91527	439	207.0	Edit Delete
	Belgium	11589623	42.0	517609	16770	99.0	Edit Delete
	Netherlands	17134872	43.0	902355	15723	87.0	Edit Delete
	France	65273511	42.0	2707074	59105	83.0	Edit Delete
	Switzerland	8654622	43.0	715360	19303	66.0	Edit Delete
	Luxembourg	625978	40.0	69453	2487	48.0	Edit Delete

ight All Match Case Match Diacritics Whole Words 1 of 11 matches

Batch Import: additional possibilities

- Convert files from Word to XWiki
- Convert files from Office formats to PDF
- Write custom processes at each stage of the import process (reading CVS, reading each line, after line import, etc..)
- Convert HTML or many Wiki syntax in fields to XWiki syntax
- Work on the spreadsheet to create meta-data prior to XWiki import (useful for Sharepoint or DMS)

See https://extensions.xwiki.org/xwiki/bin/view/Extension/Batch+Import+Application

XWiki Filter Streams

Confluence, MediaWiki
GitHub Wiki, DokuWiki

XWiki Filter Streams & Syntax Conversion

- A framework to import/export from to and from XWiki
- Uses XWiki's syntax conversion capabilities
- Also can import/export content in XWiki format (useful for migration from old XWiki instances)

See https://extensions.xwiki.org/xwiki/bin/view/Extension/Filter%20Module

Supported Wikis

- **Confluence**
- **MediaWiki**
- **DokuWiki**
- **GitHub Wiki**

Warning:

No import is perfect. There is always work to do. It can be because of custom features or macros, absolute links or changes in recent versions

A word about Confluence

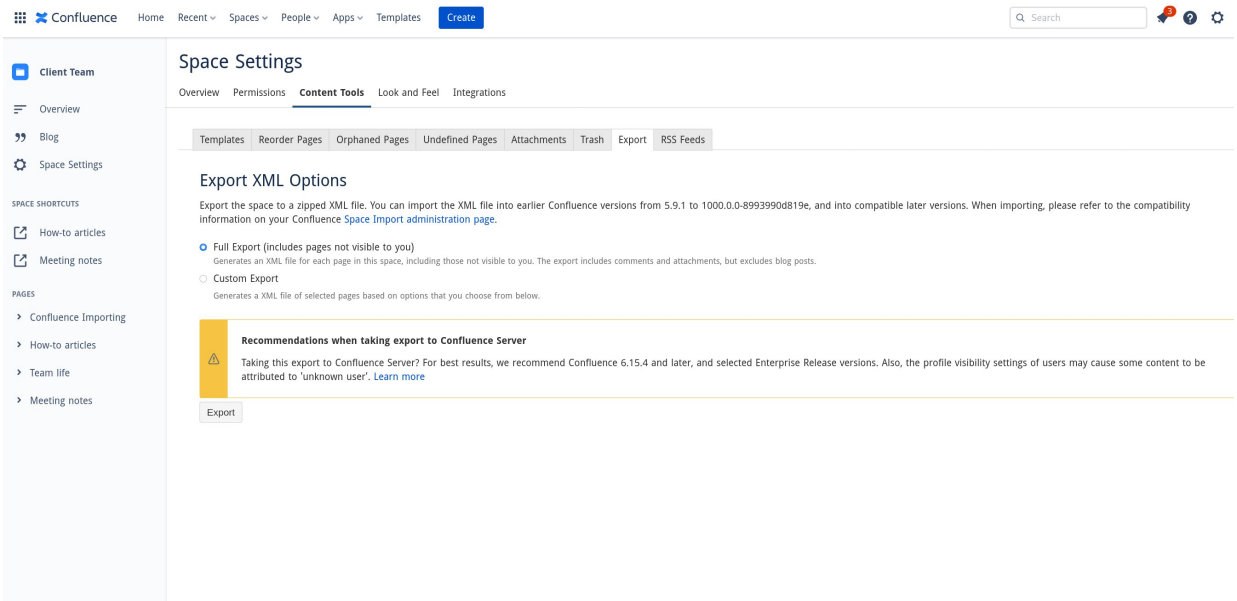
- Wiki used by many developers, including FOSS fans
- Major pricing changes by Atlassian over time
  - Regular pricing increase on Cloud and on-Premise
  - Major shift away from onPremise (only Data Center offer)
- XWiki receiving more calls for Confluence migration

Confluence Import process

- Export from Confluence as XML
- Run Confluence Filter
- Run Nested Migrator
- List & adapt macros or replace them

See https://extensions.xwiki.org/xwiki/bin/view/Extension/Confluence/XML/

Confluence Import: export Confluence XML

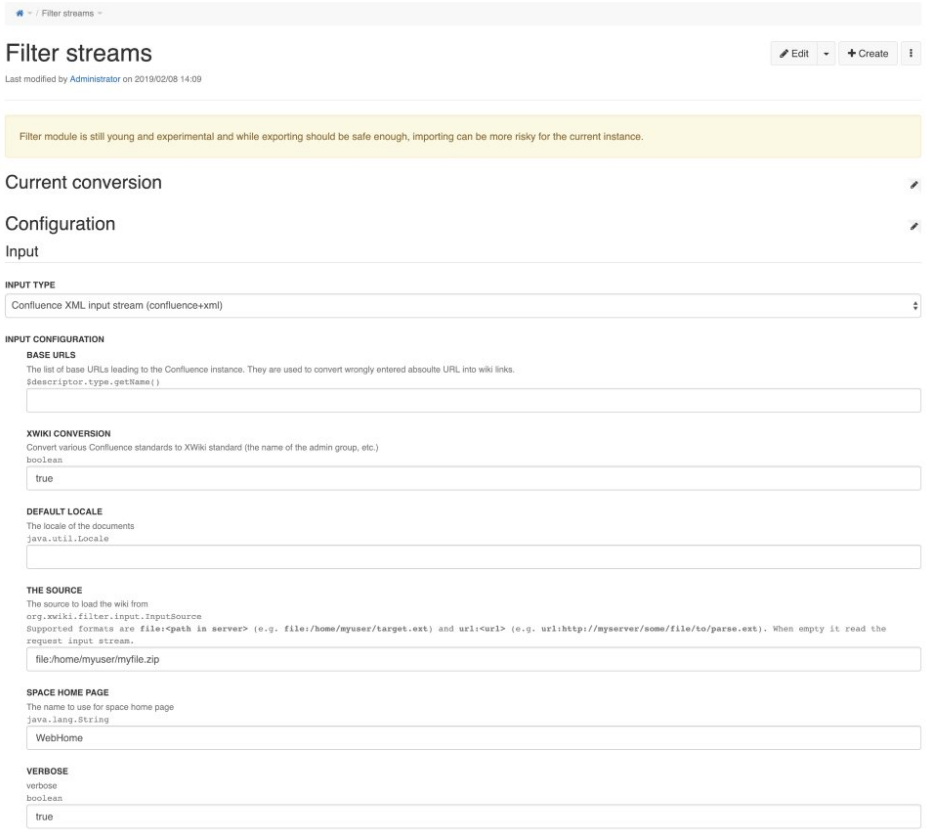


Exporting is available in Space Settings / Content Tools / Export

Choose XML output. It's possible to export all or only selected pages

Note that not being an admin might lead to a partial export

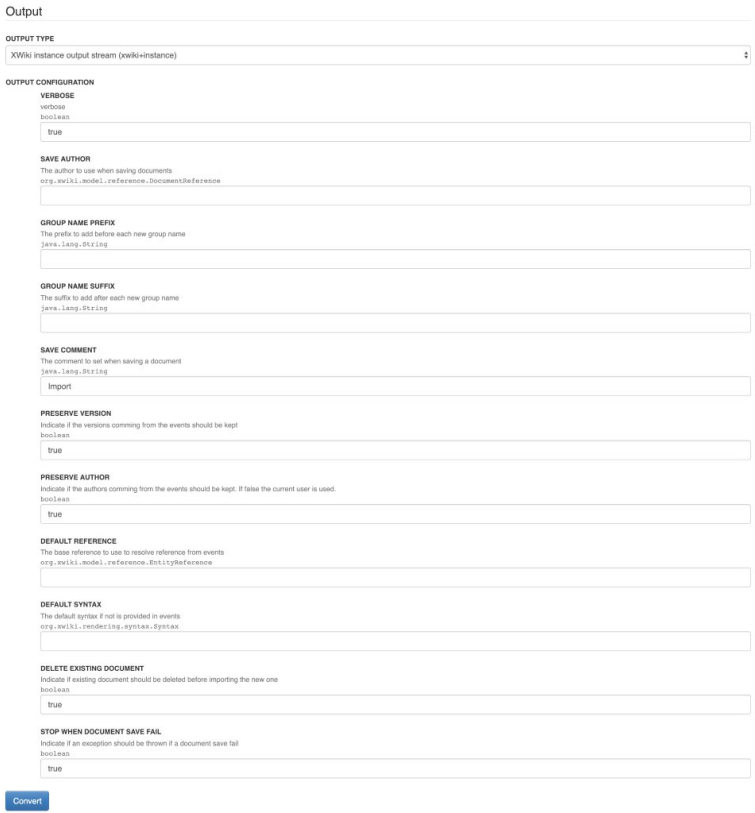
Confluence Import: configure filter export



Choose confluence+xml

The Confluence export zip file should be uploaded on the server in an area available

Confluence Import: configure filter import



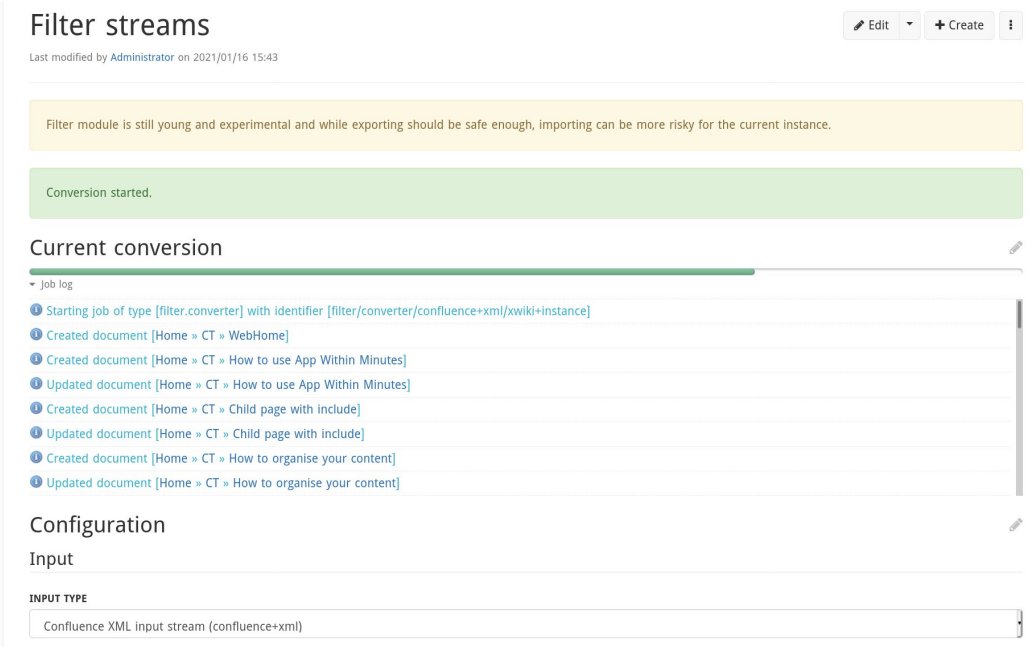
Choose xwiki+instance

With or without versions

Run import

The result should show up in a space named with the Confluence space Key all pages under that space as XWiki "terminal pages"

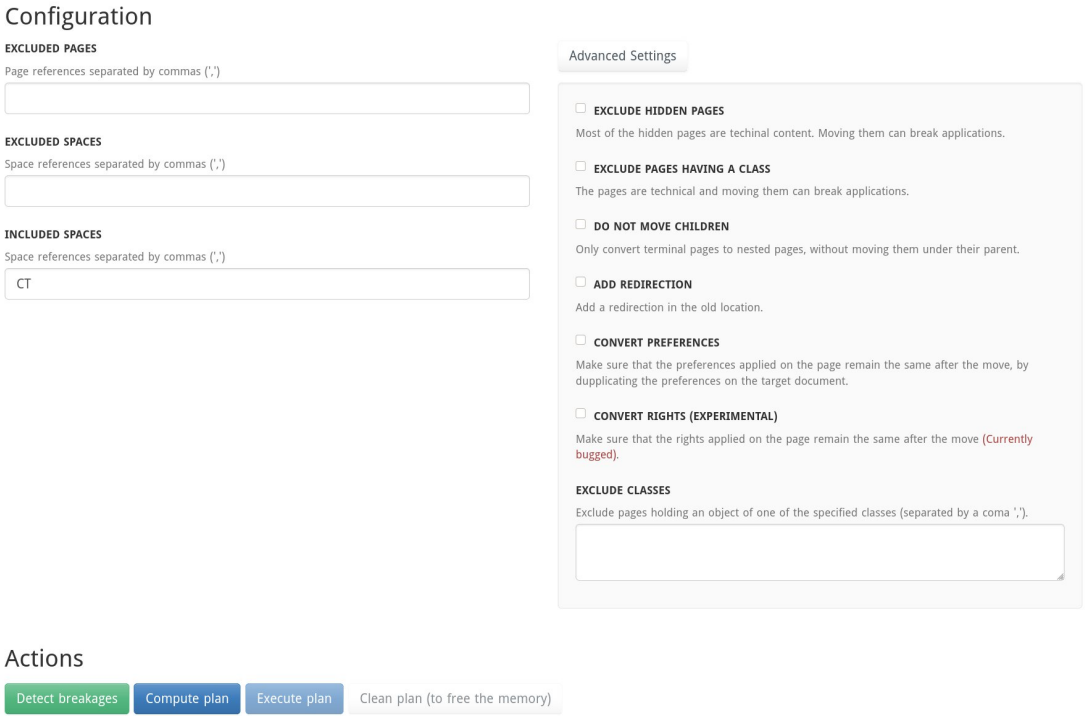
Confluence Import: follow the progress page by page



You can see each page import. Errors should not stop the process

If there are massive errors, check disk space. Unfinished import runs can lead data behind in tmp directories

Confluence Import: Nested Pages Migrator



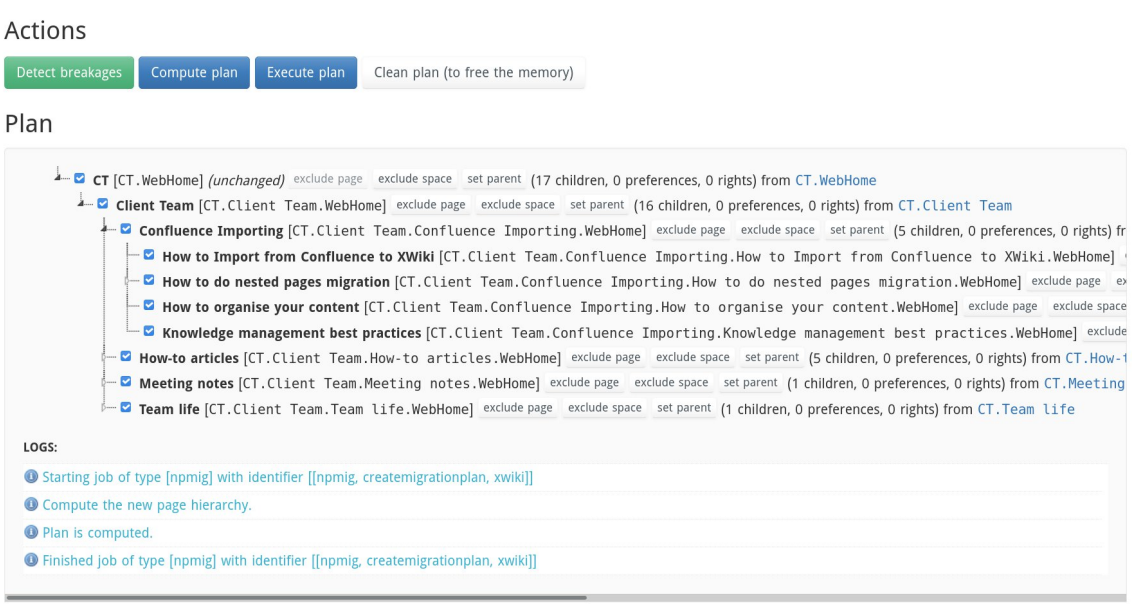
Choose space to convert to nested pages (usually the Confluence space key)

Uncheck all advanced settings

Run "compute plan"

See https://extensions.xwiki.org/xwiki/bin/view/Extension/Nested%20Pages%20Migrator%20Application

Confluence Import: check compute plan result

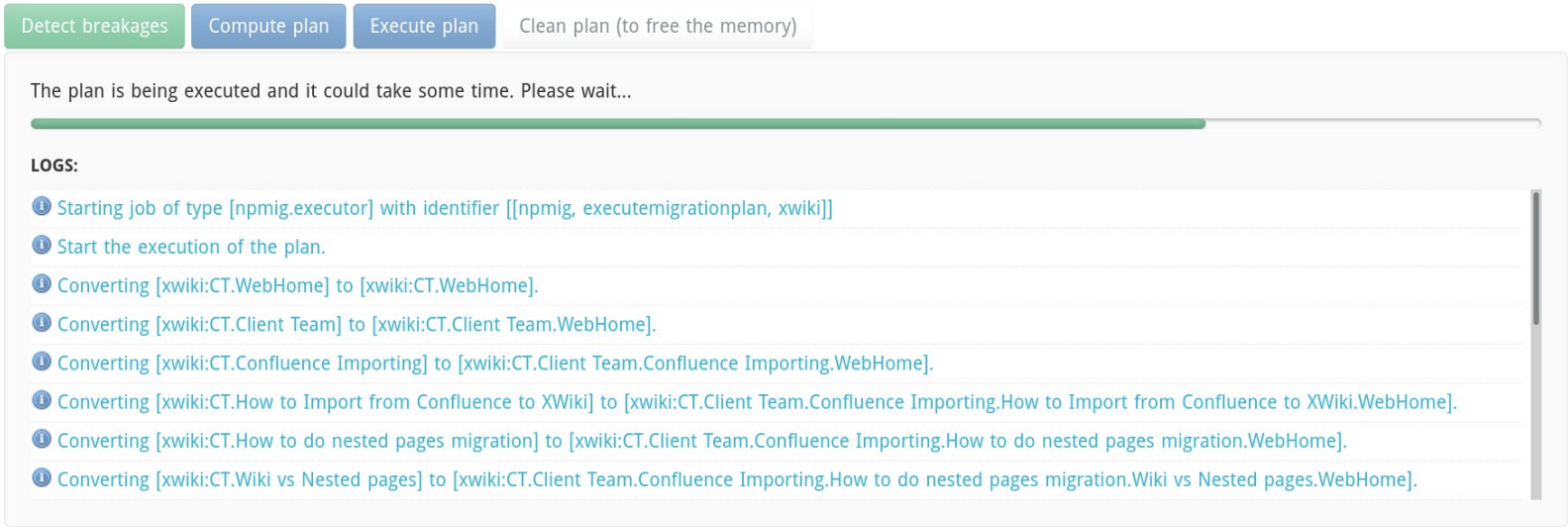


The tool lists how the pages will be organized

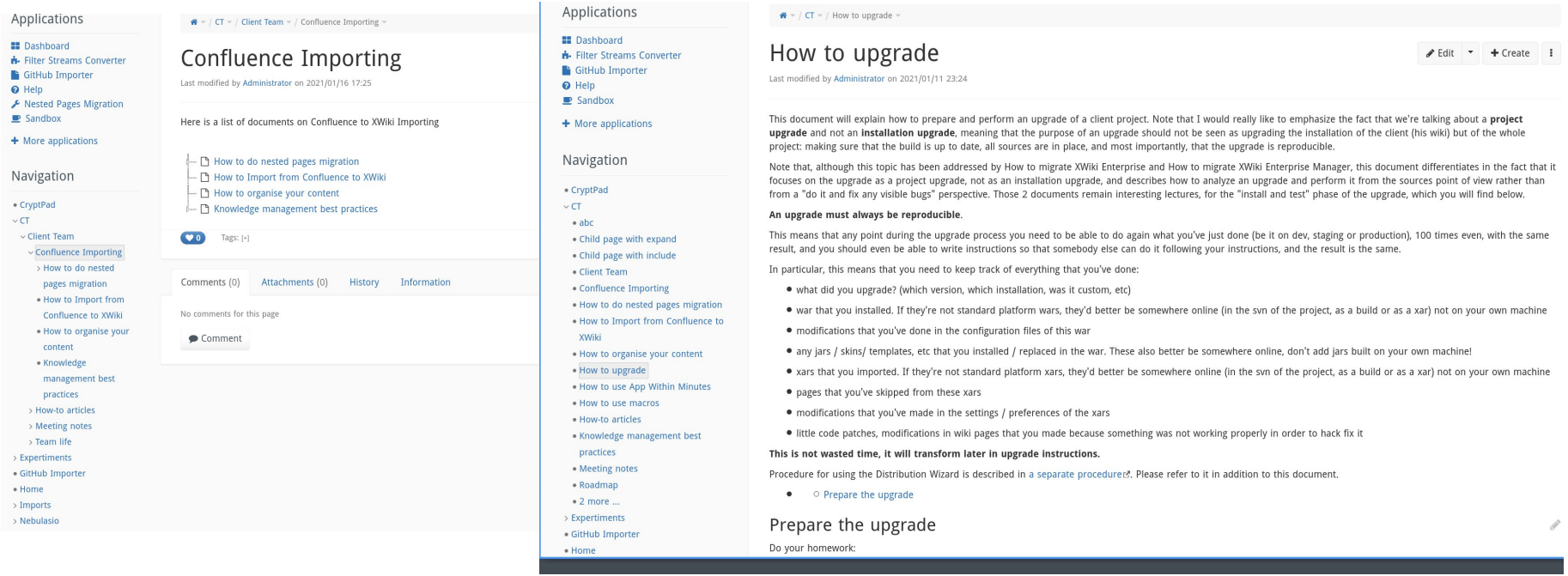
Check the result which should be similar to the Confluence hierarchy

If errors are listed of "too long page names", XWiki has a full page name limit of 255 characters. Page Name need to be changed but not the title. Find the intermediary page with a too long name

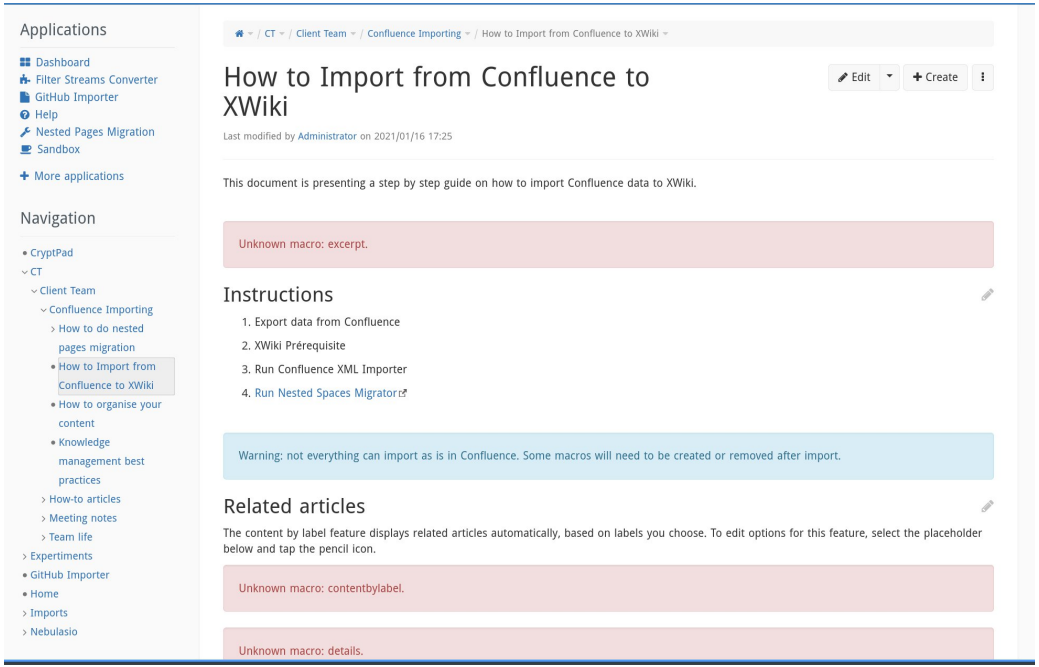
Confluence Import: execute plan



Confluence Import: enjoy the results



Confluence Import: what about Macros ?



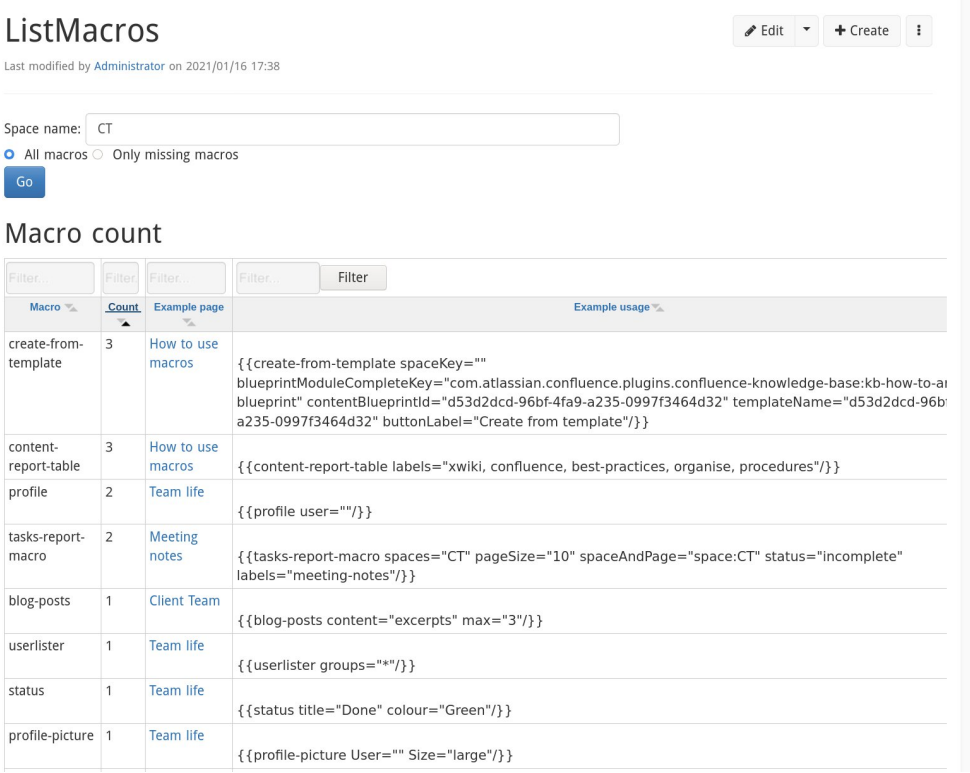
What if you see red boxes ?

Confluence has a lot of macros, many are not supported

Some macros can be bridged to XWiki macros (need Wiki Macro development)

We are working on it. Some macros will be paying.

Confluence Import: what about Macros ?



ListMacros

Edit | + Create

Last modified by Administrator on 2021/01/16 17:38

Space name: CT

All macros Only missing macros

Go

Macro count

Filter

Macro	Count	Example page	Example usage
create-from-template	3	How to use macros	{{create-from-template spaceKey="" blueprintModuleCompleteKey="com.atlassian.confluence.plugins.confluence-knowledge-base:kb-how-to-ar blueprint" contentBlueprintId="d53d2dcd-96bf-4fa9-a235-0997f3464d32" templateName="d53d2dcd-96b a235-0997f3464d32" buttonLabel="Create from template"/}}
content-report-table	3	How to use macros	{{content-report-table labels="xwiki, confluence, best-practices, organise, procedures"/}}
profile	2	Team life	{{profile user=""/}}
tasks-report-macro	2	Meeting notes	{{tasks-report-macro spaces="CT" pageSize="10" spaceAndPage="space:CT" status="incomplete" labels="meeting-notes"/}}
blog-posts	1	Client Team	{{blog-posts content="excerpts" max="3"/}}
userlister	1	Team life	{{userlister groups="*"/}}
status	1	Team life	{{status title="Done" colour="Green"/}}
profile-picture	1	Team life	{{profile-picture User="" Size="large"/}}

We are building a tool to list which macros are being used

Contact us if you want to check which are the macros used in your instance, and help us decide which macros are more important

XWIKI

Confluence: a work in progress

- Import more content
  - PR by contributor for comments / labels
- Support more macros of Confluence
- Maintain importer for future Confluence evolution (Cloud & On-premise). Confluence export format is not officially documented

If your company wants to migrate, sponsoring the work is possible

Conclusion

Conclusion

- Many tools available for importing content in XWiki
- Many of these tools have been build during service missions from XWiki SAS for clients. Thanks for the support.
- Thanks to the contributors who are reporting issues and also wrote code
- These tools will improve and in particular the Confluence one, with more users coming to XWiki

THANK YOU

Talk build with Open Source:

- **Operating System: QubesOS / Fedora**
- **Presentation software: CryptPad OnlyOffice Slides (alpha)**
- **Screenshots: Shotwell**
- **Recording: Open Broadcasting Software**
- **XWiki and XWiki Extensions !**

XWIKI
YOUR INFORMATION ORGANIZED