



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**

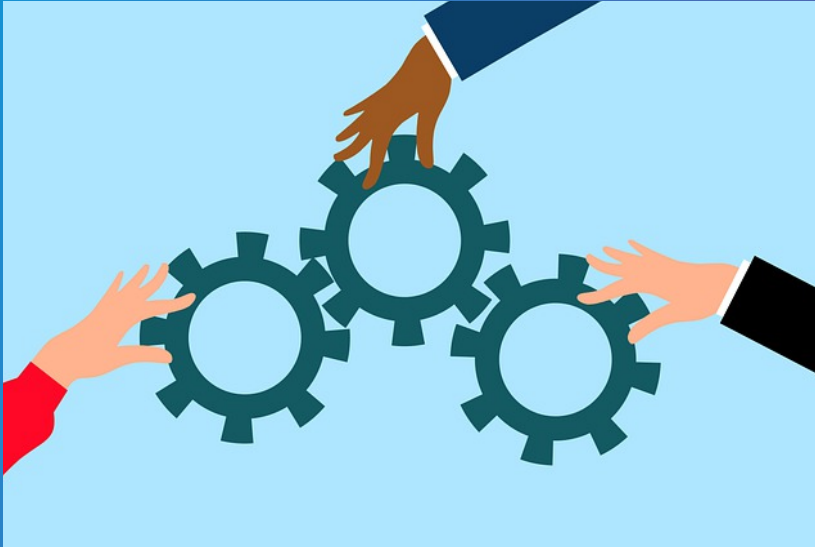
 **16 years of Open  
Source business**

# 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

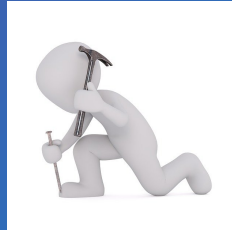
# Types of Imports

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**

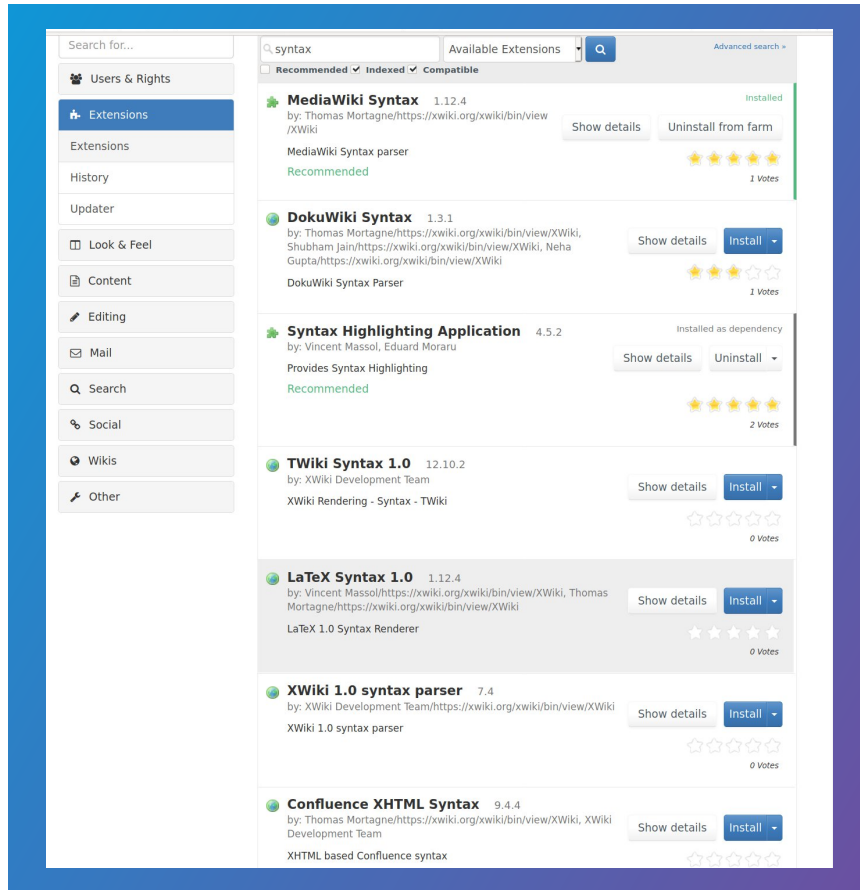




# Manual Tools

## Wiki & HTML Syntax Importing

### 1. Install and/or activate additional syntaxes in XWiki



The screenshot displays the XWiki extension marketplace interface. On the left is a navigation sidebar with categories like 'Users & Rights', 'Extensions', 'Look & Feel', 'Content', 'Editing', 'Mail', 'Search', 'Social', 'Wikis', and 'Other'. The main area shows a search for 'syntax' with filters for 'Recommended', 'Indexed', and 'Compatible'. A list of extensions is shown, including:

- MediaWiki Syntax** 1.12.4 (Installed): by Thomas Mortagne. Includes 'MediaWiki Syntax parser' (Recommended, 1 Vote).
- DokuWiki Syntax** 1.3.1: by Thomas Mortagne, Shubham Jain, and Neha Gupta. Includes 'DokuWiki Syntax Parser' (1 Vote).
- Syntax Highlighting Application** 4.5.2 (Installed as dependency): by Vincent Massol and Eduard Moraru. Provides 'Syntax Highlighting' (Recommended, 2 Votes).
- TWiki Syntax 1.0** 12.10.2: by XWiki Development Team. Includes 'XWiki Rendering - Syntax - TWiki' (0 Votes).
- LaTeX Syntax 1.0** 1.12.4: by Vincent Massol, Thomas Mortagne, and Thomas Mortagne. Includes 'LaTeX 1.0 Syntax Renderer' (0 Votes).
- XWiki 1.0 syntax parser** 7.4: by XWiki Development Team. Includes 'XWiki 1.0 syntax parser' (0 Votes).
- Confluence XHTML Syntax** 9.4.4: by Thomas Mortagne and XWiki Development Team. Includes 'XHTML based Confluence syntax' (0 Votes).

image from Tobie Langel

# Manual Tools: Wiki & HTML Syntax Importing

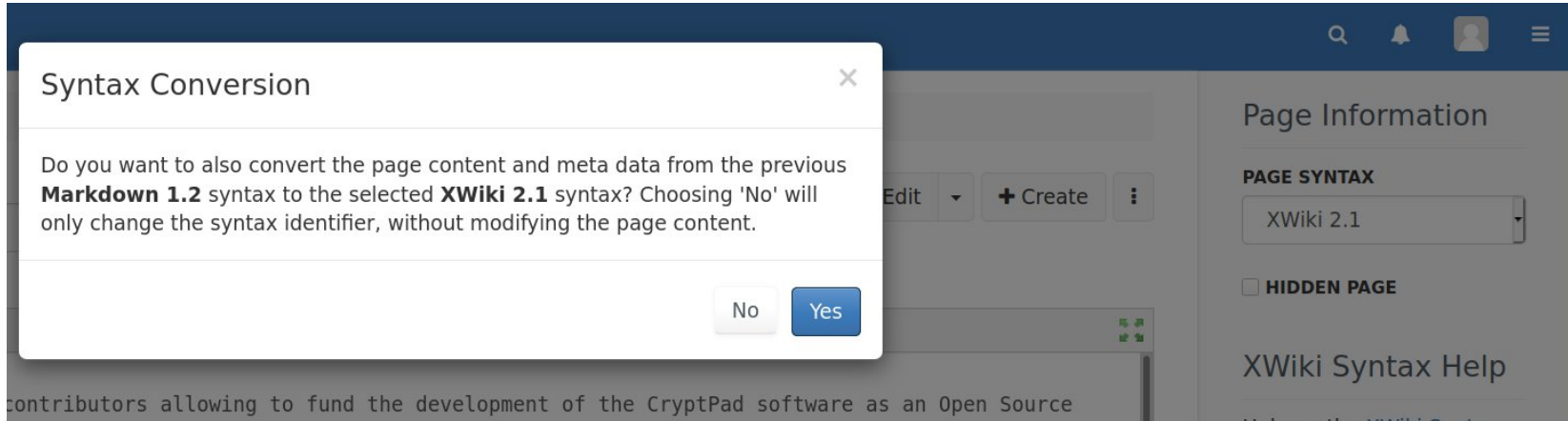
The screenshot shows the XWiki editor interface. At the top left is the XWiki logo. Below it, the page title is 'Markdown Page'. The main content area contains text about CryptPad funding, including a section for '2019 Funding' and '2020 Funding'. The right sidebar shows 'Page Information' with 'PAGE SYNTAX' set to 'Markdown 1.2' and 'HIDDEN PAGE' checked. Below that are links for 'XWiki Syntax Help' and 'Help on the XWiki Syntax'. At the bottom of the editor, there are buttons for 'Save & View', 'Save', 'Preview', and 'Cancel', along with a field for 'Add summary...' and checkboxes for 'Minor edit' and 'Autosave every 5 minutes'. The footer of the page indicates 'XWiki Jetty HSOLDB 12.10.2'.

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

Home / Imports / Markdown Page

## Markdown Page

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 than 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

The screenshot shows the XWiki editor interface. The page title is "Markdown Page". The content area displays a list of funding sources for CryptPad in 2019 and 2020. The interface includes a top navigation bar with the XWiki logo, a right sidebar with "Page Information" and "Page Syntax" (set to XWiki 2.1), and a bottom status bar showing "XWiki Jetty HSQldb 13.1.0.2".

HTML can also be copy-pasted in the Wysiwyg editor

# Manual Tools: Office Import

## Create Page

### TITLE

Title of the new page

### LOCATION

Location in the page hierarchy where this new page will be created.


 / Imports / Office File

### TYPE

Select the kind of page that you want to create

#### Default (2)

 **Blank page**  
Standard empty page

 **Page from Office**  
Import Office documents into wiki pages

#### Templates (5)

 **Dashboard**

 **Article**

**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

The screenshot displays a web browser window with two panes. The left pane shows a document in read-only mode, with a table of contents and a list of links. The right pane shows the document content in XWiki syntax, including a table of contents and a list of links.

**Table of Contents (Left Pane):**

- France
- was established in 2002, aiming at the promotion, development and maintenance of open source software for public services. "Reuse of particular experiences to ADULLACT and the basic concept behind that is that public money must be paid just for cost." ADULLACT particularly values the field and releases many FOSS products via their forge (Bistrouffage and, more recently, GitLab). As of February 2019, 300 business Open source software applications were provided by the "compete" section on the ADULLACT platform.
- In 2003, France increased the use of open source software operating systems in various ministries and developed an open source content management system to standardize government websites.
- In 2004, the Ministry of Defense has formed a consortium to develop a highly secure Linux-based operating system. However, a bill of the Ministry of Culture and Communication did not succeed eventually in funding a full migration to open source software in 2005. Nevertheless in 2008 the French Government decided to migrate all its 70,000 desktops from proprietary software to open source software, recognizing the security advantages and the lower Total Cost of Ownership of Open source software. The Agency for the Development of the Electronic Administration (ANSA) migrated approximately 10% of public service desktops to open source software by 2007. Further decisions were made to increase open source software usage in the public finance and education sectors. In 2009 the first version of the RG3 (Deferred General of Interoperability) was released. This document provided a framework, guidelines and a collection of standards that favor the interoperability of public information systems.
- In 2011, the former DISC (Direction des Systèmes d'Information) was transformed into DISIC (Direction Interministérielle des Systèmes d'Information). In parallel, the "Etats" had been established under the government's Secretary of State, to promote the concepts of open data and open government in France. In 2015, DISIC was merged with Etalab to form DINISIC (Direction Interministérielle du Numérique et du Système d'Information). DINISIC is today the government agency that promotes the use of open source software within French public sector.
- In 2012, the "Circulaire Aynaud" (so called from Jean-Marie Aynaud, then prime minister), encouraged the public services to actively participate in the development of open source software that the country depended upon. The Circulaire identified five inter-ministry groups in the context of the "Direction Interministérielle des Systèmes d'Information et de Communication", for managing open activities and sharing knowledge. The groups established are:
  - core group for making proposals, validating decisions and steering activities related to markets, software catalogues and implementations of directives act open source software the French public sector;
  - minMI: inter-ministerial group for an open government;
  - minMDS: inter-ministerial group for managing ODS and OLPs tools;
  - minMBI: inter-ministerial group for databases;
  - minMOS: inter-ministerial group for operating systems.
- In 2015, minMI, one of the five groups mentioned above, released SILL (Socle Interministériel de Logiciels Libres), a long list of open source software solutions recommended for the French public services. "The SILL list is updated annually".
- In 2016, the French law for the "République Numérique" reinforced both the use of Open source software and the opening of code produced by French public services. In the same year, France's Rhône-Alpes region prioritised open source software, as reported on OSGE. In 2016, France hosted and trained the Open Government Partnership (OGP) for the period 2016-2017, after becoming a partner in 2015, while Paris hosted the OGP Summit in 2016. France is particularly active in this regard, with 29 OGP commitments completed. As of May 2019, 22 OGP Commitments were active". In the same year, the second version of the RG3 has been released".
- As part of OGP, France led a working group with several national and international FOSS stakeholders aimed at developing a country-level policy detailing how civil servants can release and contribute to open source software projects. The policy has been officially adopted in February 2018."
- In addition, the French government is building its own open source developer community, aiming to bring together software developers and IT scientists who want to contribute to government-led open source projects. This initiative is named "Blue Hat" and was presented at the Paris Open Source Summit in Dec 2018. Moreover, France hosted the B-BOOST 2018 event in Bordeaux, aiming at accelerating the development of the Open source software industry".
- According to a Forrester Research in 2015, France was a leader of Open source software adoption in Europe, with 24% of

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

# Manual Tools: Office Import

The screenshot shows two windows side-by-side. The left window is LibreOffice Writer, displaying a document with a table of contents and a section titled '2.3.3.1 Analysis of maturity per category'. The right window is Mozilla Firefox, displaying a webpage with a similar table and a chart. The browser address bar shows '10.137.0.19:12103/bwwiki/bin/view/Imports/OfficeFile/#\_Toc25848788'.

**Table 1: Analysis of maturity per category (from LibreOffice document)**

Level	Name	Argumentation	Percentage	Score	
1	No open source software allowed	Almost 100% of operating systems installed are Microsoft Windows, the office suite is Microsoft Office, there are no available alternatives and users are not able to use other software even if they would like to.	60%	0.6	
2	Technical with no policy	Product List is relatively stable, in niches where proprietary products are either expensive or not available or where limited use of the software would not justify purchasing commercial products. Some of such products are used on virtually all Commission PCs (Firefox, 7-Zip, VLC).	20%	0.4	
3	Uni-level policy	Within several teams or units there are ad hoc decisions on use of particular tools and products. Some units might even perform a proper market study before choosing a product to use.	10%	0.3	
4	EC-level policy	The EC open source software policy provides a framework to promote wide-spread use of open source software on the desktop.	10%	0.4	
5	Driver for change	No recorded cases of open source software for innovative aspects or catalyst for change, it is only replacing proprietary software in certain situations.	0%	0	
				<b>Total score</b>	<b>1.7</b>

**Table 2: Analysis of maturity per category (from Firefox browser)**

Level	Name	Argumentation	Percentage	Score	
1	No open source software allowed	Almost 100% of operating systems installed are Microsoft Windows, the office suite is Microsoft Office, there are no available alternatives and users are not able to use other software even if they would like to.	60%	0.6	
2	Technical with no policy	Product List is relatively stable, in niches where proprietary products are either expensive or not available or where limited use of the software would not justify purchasing commercial products. Some of such products are used on virtually all Commission PCs (Firefox, 7-Zip, VLC).	20%	0.4	
3	Uni-level policy	Within several teams or units there are ad hoc decisions on use of particular tools and products. Some units might even perform a proper market study before choosing a product to use.	10%	0.3	
4	EC-level policy	The EC open source software policy provides a framework to promote wide-spread use of open source software on the desktop.	10%	0.4	
5	Driver for change	No recorded cases of open source software for innovative aspects or catalyst for change, it is only replacing proprietary software in certain situations.	0%	0	
				<b>Total score</b>	<b>1.7</b>

**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

🏠 / Countries - / Code - / CountriesClass -

## Countries

Step 2 — Structure your data

Name Structure Entries Presentation

**TITLE**

Hint

**POPULATION**

Hint

**QUARANTINE**

Hint

**DENSITY**

Hint

**MEDIAN AGE**

Hint

**URBAN POP**

Hint

GDP2019

**Field Palette**

Drag & drop fields from the palette to create the form that will be used to input your data.

**Standard**

- Short Text
- Long Text
- Number
- Boolean
- Static List

**Pickers**

- Page
- Date
- User
- Group

**Page**

- Content
- H Title

**Advanced**

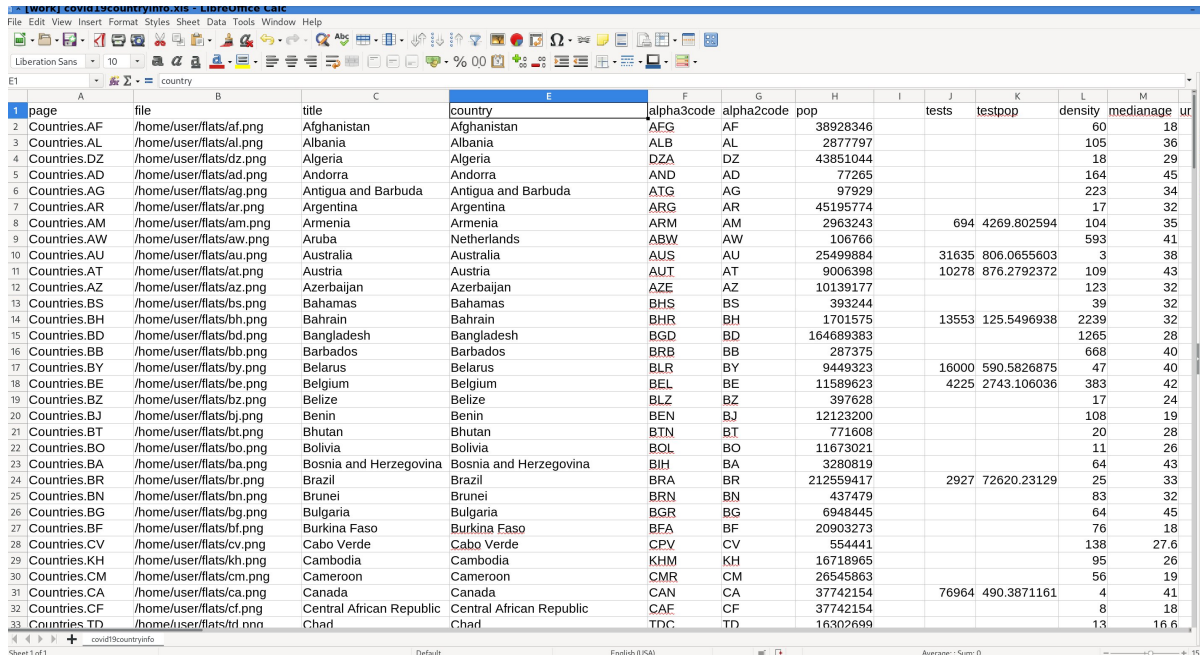
- Database List
- File
- Image
- Officefile
- PDF

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

🏠 / Countries ▾

## Countries

Last modified by [Administrator](#) on 2021/01/16 22:26

### Choose a file

Choose the file to import data from, accepted formats are CSV and Excel. You

#### Upload a new file:

No file selected.

#### Choose the file to import from:

covid19countryinfo.xls 1.3

Posted by [Administrator](#) on 2021/01/16 22:26, 118.0 KB

Tags: [+]

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

# Batch Import: configure the mapping

Class fields	Columns in the file
Document name	<input type="text" value="name"/>
Title	<input type="text" value="-"/>
Population	<input type="text" value="pop"/>
Quarantine	<input type="text" value="quarantine"/>
Density	<input type="text" value="density"/>
median Age	<input type="text" value="medianage"/>
urban pop	<input type="text" value="urbanpop"/>
gdp2019	<input type="text" value="gdp2019"/>
First Case	<input type="text" value="firstcase"/>
Total Cases	<input type="text" value="totalcases"/>
Death per million	<input type="text" value="deathdiv1m"/>
Content	<input type="text" value="-"/>
Document title	<input type="text" value="title"/>
Document space	<input type="text" value="space"/>
Document parent	<input type="text" value=""/>
Document content	<input type="text" value=""/>
Document attachments	<input type="text" value="file"/>

Whether the values found in the file that are used for space and document names should be cleaned before being used: whitespaces are removed and special characters as well. Check this box if characters that might break the document URLs (e.g. slashes, backslashes, pluses, etc) might appear in the values of the columns chosen for document names and document spaces. Note that, if this option is chosen, the deduplication options of the next steps apply to the cleaned names.

Clean names:

By default, the columns mapped on application fields of type number are parsed as "unformatted" numbers (decimal separator is '.' (dot) and no other non-digit character is accepted). However, if your source file contains formatted numbers, you can indicate here the localization used to process these numbers (e.g. in French, a number would be written as "1 235,54", using comma as decimal separator and space to group hundreds together for integer values). Using a localization will also make percent values and per mill values be interpreted as numeric values (e.g. '60%' will become '60'). Language should be filled in by its two letter code, e.g. 'en', 'fr', 'de', etc.

Localization language:

By default, if a column in the file is mapped on an application field of type date, the values in the column will try to be interpreted using the format specified in the date field. However, if you have values that are using a different format, you can specify here the fallback format to be used for values which cannot be interpreted using the format specified in the date field. Note that the format applies to individual values, not to a whole column (if a value on a column cannot be interpreted using the date format of the field, only that value will be interpreted using the format specified here, not all the values on that column). The format should given as a pattern by the rules described here <http://docs.oracle.com/javase/6/docs/api/java/text/SimpleDateFormat.html>.

Default date format:

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

# Batch Import: file attachment settings

The screenshot shows a web form for configuring file attachment settings during a batch import. The form includes several input fields and a radio button group:

- datafilename:** /home/user/.xtool/instances/12-10-2/data/flags/flags.zip
- classname:** Countries.Code.CountriesClass
- space:** Countries
- datafileprefix:** (empty)
- pageprefix:** (empty)
- fieldsfortags:** (empty)
- List Separator:** ,
- Default date format:** MM/dd/yyyy
- Clear name:** Radio buttons for --- (selected), Yes, and No.
- fileupload:** Yes
- fileimport:** (empty)

At the bottom, there is a control bar with buttons for "Save & View", "Save", "Cancel", and "Add summary...". A "Minor edit" checkbox is also present. Below the control bar, a search bar shows "France" and a dropdown menu with options: "Highlight All", "Match Case", "Match Djacritics", "Whole Words", and "1 of 11 matches".

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

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



# Batch Import: preview imports

## Countries

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

 Edit  + Create 

### 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

# Batch Import: deduplication settings

## Countries

[Edit](#) [+ Create](#) [⋮](#)

Last modified by Administrator on 2021/01/16 22:35

### Data deduplication and overwrite

If by any chance, the column chose for document name does not contain unique names, what should happen to the **duplicate rows**?

- Ignore
- Update: use the data as part of the other rows with the same document name (if the same column is set on multiple rows, the last one will win).
- Generate a new document name, by adding a number at the end of the document name

If then, for the document name built as indicated above, a document already exists in the wiki with that name, what should we do to the **existing document in the wiki**?

- Leave unchanged (ignore data from the file)
- Update with the data coming from the file (the fields mapped on the columns of the file will be updated, the rest will remain unchanged)
- Replace with the data coming from the file (the existing document will be deleted and a new one will be created instead, containing only the data in the file)
- Generate a new document name, by adding a number at the end of the document name built as indicated above. Note that 2 subsequent imports with this option checked will not yield the same results since the document names used depend on the documents existing on the wiki.

[Back to mapping preview](#)

[Do an import simulation](#)

[♥ 0](#) Tags: [-]

Created by Administrator on 2021/01/16 19:48

# Batch Import: enjoy the results

## Countries

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










[Edit](#) [+ Create](#) [⋮](#)

### Entries

Results 1 - 15 out of 173 per page of  Page [1](#) [2](#) [3](#) [4](#) [5](#) [6](#) [7](#) [8](#) [9](#) [10](#) [12](#)

### Actions

- [ADD NEW ENTRY](#)
- [EDIT APPLICATION](#)
- [DELETE ALL ENTRIES](#)
- [DELETE APPLICATION](#)

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

[Right 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

The screenshot shows the Confluence interface. At the top, there's a navigation bar with 'Confluence', 'Home', 'Recent', 'Spaces', 'People', 'Apps', 'Templates', and a 'Create' button. A search bar is on the right. The left sidebar shows 'Client Team' and 'Space Settings'. The main content area is titled 'Space Settings' and has tabs for 'Overview', 'Permissions', 'Content Tools', 'Look and Feel', and 'Integrations'. Under 'Content Tools', there are sub-tabs: 'Templates', 'Reorder Pages', 'Orphaned Pages', 'Undefined Pages', 'Attachments', 'Trash', 'Export', and 'RSS Feeds'. The 'Export' tab is active, showing 'Export XML Options'. It explains that you can export to a zipped XML file for various Confluence versions. Two options are listed: 'Full Export (includes pages not visible to you)' and 'Custom Export'. A yellow warning box contains 'Recommendations when taking export to Confluence Server', advising to use Confluence 6.15.4 or later and to check user visibility settings. An 'Export' button is at the bottom.

# Confluence Import: configure filter export

Filter streams

Last modified by Administrator on 2019/02/08 14:09

Filter module is still young and experimental and while exporting should be safe enough, importing can be more risky for the current instance.

Current conversion

Configuration

Input

INPUT TYPE

Confluence XML input stream (confluence+xml)

INPUT CONFIGURATION

**BASE URLS**  
The list of base URLs leading to the Confluence instance. They are used to convert wrongly entered absolute URL into wiki links.  
sdescriptor.type.getName()

**XWIKI CONVERSION**  
Convert various Confluence standards to XWiki standard (the name of the admin group, etc.)  
boolean  
true

**DEFAULT LOCALE**  
The locale of the documents  
java.util.Locale

**THE SOURCE**  
The source to load the wiki from  
org.xwiki.filter.input.InputSource  
Supported formats are file:<path in server> (e.g. file:/home/myuser/target.ext) and url:<url> (e.g. url:http://myserver/some/file/to/parse.ext). When empty it read the request input stream.  
file:/home/myuser/myfile.zip

**SPACE HOME PAGE**  
The name to use for space home page  
java.lang.String  
WebHome

**VERBOSE**  
verbose  
boolean  
true

Choose confluence+xml

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

# Confluence Import: configure filter import

Output

OUTPUT TYPE

XWiki instance output stream (xwiki+instance)

OUTPUT CONFIGURATION

**VERBOSE**  
Whether  
boolean  
 true

**SAVE AUTHOR**  
The author to use when saving documents  
org.xwiki.model.reference.DocumentReference

**GROUP NAME PREFIX**  
The prefix to add before each new group name  
java.lang.String

**GROUP NAME SUFFIX**  
The suffix to add after each new group name  
java.lang.String

**SAVE COMMENT**  
The comment to set when saving a document  
java.lang.String

**PRESERVE VERSION**  
Indicates if the versions coming from the events should be kept  
boolean  
 true

**PRESERVE AUTHOR**  
Indicates if the authors coming from the events should be kept. If false the current user is used.  
boolean  
 true

**DEFAULT REFERENCE**  
The base reference to use to resolve reference from events  
org.xwiki.model.reference.EntityReference

**DEFAULT SYNTAX**  
The default syntax if not is provided in events  
org.xwiki.rendering.syntax.Syntax

**DELETE EXISTING DOCUMENT**  
Indicates if existing document should be deleted before importing the new one  
boolean  
 true

**STOP WHEN DOCUMENT SAVE FAIL**  
Indicates if an exception should be thrown if a document save fail  
boolean  
 true

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

The screenshot shows the 'Filter streams' configuration page in Confluence. At the top right, there are 'Edit', '+ Create', and a help icon. Below the title, a note states: 'Filter module is still young and experimental and while exporting should be safe enough, importing can be more risky for the current instance.' A green bar indicates 'Conversion started.' Below this is a 'Current conversion' section with a progress bar and a 'Job log' table. The job log contains the following entries:

Job log
Starting job of type [filter.converter] with identifier [filter/converter/confluence+xml/xwiki+instance]
Created document [Home » CT » WebHome]
Created document [Home » CT » How to use App Within Minutes]
Updated document [Home » CT » How to use App Within Minutes]
Created document [Home » CT » Child page with include]
Updated document [Home » CT » Child page with include]
Created document [Home » CT » How to organise your content]
Updated document [Home » CT » How to organise your content]

Below the job log is the 'Configuration' section, which includes an 'Input' field with the value 'Confluence XML input stream (confluence+xml)'.

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

## Configuration

### EXCLUDED PAGES

Page references separated by commas (,)

### EXCLUDED SPACES

Space references separated by commas (,)

### INCLUDED SPACES

Space references separated by commas (,)

## Advanced Settings

### EXCLUDE HIDDEN PAGES

Most of the hidden pages are technical content. Moving them can break applications.

### EXCLUDE PAGES HAVING A CLASS

The pages are technical and moving them can break applications.

### DO NOT MOVE CHILDREN

Only convert terminal pages to nested pages, without moving them under their parent.

### ADD REDIRECTION

Add a redirection in the old location.

### CONVERT PREFERENCES

Make sure that the preferences applied on the page remain the same after the move, by duplicating the preferences on the target document.

### CONVERT RIGHTS (EXPERIMENTAL)

Make sure that the rights applied on the page remain the same after the move (Currently bugged).

### EXCLUDE CLASSES

Exclude pages holding an object of one of the specified classes (separated by a coma ',').

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

Uncheck all advanced settings

Run "compute plan"

## Actions

Detect breakages

Compute plan

Execute plan

Clean plan (to free the memory)

# Confluence Import: check compute plan result

## Actions

Detect breakages Compute plan Execute plan Clean plan (to free the memory)

## Plan

LOGS:

- Starting job of type [npmig] with identifier [[npmig, createmigrationplan, xwiki]]
- Compute the new page hierarchy.
- Plan is computed.
- Finished job of type [npmig] with identifier [[npmig, createmigrationplan, xwiki]]

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

Detect breakages

Compute plan

Execute plan

Clean plan (to free the memory)

The plan is being executed and it could take some time. Please wait...

## LOGS:

- Starting job of type [npmig.executor] with identifier [[npmig, executemigrationplan, xwiki]]
- Start the execution of the plan.
- Converting [xwiki:CT.WebHome] to [xwiki:CT.WebHome].
- Converting [xwiki:CT.Client Team] to [xwiki:CT.Client Team.WebHome].
- Converting [xwiki:CT.Confluence Importing] to [xwiki:CT.Client Team.Confluence Importing.WebHome].
- Converting [xwiki:CT.How to Import from Confluence to XWiki] to [xwiki:CT.Client Team.Confluence Importing.How to Import from Confluence to XWiki.WebHome].
- Converting [xwiki:CT.How to do nested pages migration] to [xwiki:CT.Client Team.Confluence Importing.How to do nested pages migration.WebHome].
- Converting [xwiki:CT.Wiki vs Nested pages] to [xwiki:CT.Client Team.Confluence Importing.How to do nested pages migration.Wiki vs Nested pages.WebHome].



# Confluence Import: enjoy the results

This screenshot shows the left sidebar of the Confluence 'Confluence Importing' page. It includes sections for 'Applications' (Dashboard, Filter Streams Converter, GitHub Importer, Help, Nested Pages Migration, Sandbox, More applications) and 'Navigation' (CryptPad, CT, Client Team, Confluence Importing, How to do nested pages migration, How to Import from Confluence to XWiki, How to organise your content, Knowledge management best practices, How-to articles, Meeting notes, Team life, Experiments, GitHub Importer, Home, Imports, Nebulasio).

This screenshot shows the main content area of the 'Confluence Importing' page. It features a breadcrumb trail, a title, a modification date, a list of documents, a 'Comments (0)' section, and a 'Comment' button.

This screenshot shows the left sidebar of the 'How to upgrade' page. It includes sections for 'Applications' (Dashboard, Filter Streams Converter, GitHub Importer, Help, Sandbox, More applications) and 'Navigation' (CryptPad, CT, abc, Child page with expand, Child page with include, Client Team, Confluence Importing, How to do nested pages migration, How to Import from Confluence to XWiki, How to organise your content, How to use macros, How-to articles, Knowledge management best practices, Meeting notes, Roadmap, 2 more ..., Experiments, GitHub Importer, Home).

This screenshot shows the main content area of the 'How to upgrade' page. It includes a breadcrumb trail, a title, a modification date, an 'Edit' button, and a 'Create' button. The main text explains the concept of a project upgrade and provides instructions on how to prepare for it.

## How to upgrade

This document will explain how to prepare and perform an upgrade of a client project. Note that I would really like to emphasize the fact that we're talking about a **project upgrade** and not an **installation upgrade**, meaning that the purpose of an upgrade should not be seen as upgrading the installation of the client (his wiki) but of the whole project: making sure that the build is up to date, all sources are in place, and most importantly, that the upgrade is reproducible.

Note that, although this topic has been addressed by How to migrate XWiki Enterprise and How to migrate XWiki Enterprise Manager, this document differentiates in the fact that it focuses on the upgrade as a project upgrade, not as an installation upgrade, and describes how to analyze an upgrade and perform it from the sources point of view rather than from a "do it and fix any visible bugs" perspective. Those 2 documents remain interesting lectures, for the "install and test" phase of the upgrade, which you will find below.

**An upgrade must always be reproducible.**

This means that any point during the upgrade process you need to be able to do again what you've just done (be it on dev, staging or production), 100 times even, with the same result, and you should even be able to write instructions so that somebody else can do it following your instructions, and the result is the same.

In particular, this means that you need to keep track of everything that you've done:

- what did you upgrade? (which version, which installation, was it custom, etc)
- war that you installed. If they're not standard platform wars, they'd better be somewhere online (in the svn of the project, as a build or as a xar) not on your own machine
- modifications that you've done in the configuration files of this war
- any jars / skins/ templates, etc that you installed / replaced in the war. These also better be somewhere online, don't add jars built on your own machine!
- xars that you imported. If they're not standard platform xars, they'd better be somewhere online (in the svn of the project, as a build or as a xar) not on your own machine
- pages that you've skipped from these xars
- modifications that you've made in the settings / preferences of the xars
- little code patches, modifications in wiki pages that you made because something was not working properly in order to hack fix it

**This is not wasted time, it will transform later in upgrade instructions.**

Procedure for using the Distribution Wizard is described in a [separate procedure](#). Please refer to it in addition to this document.

- [Prepare the upgrade](#)

## Prepare the upgrade

Do your homework:

# Confluence Import: what about Macros ?

The screenshot shows a Confluence page titled "How to Import from Confluence to XWiki". The page is part of a "Client Team" workspace under "Confluence Importing". The main content area contains a heading "How to Import from Confluence to XWiki" with "Edit" and "Create" buttons. Below the heading, a paragraph states: "This document is presenting a step by step guide on how to import Confluence data to XWiki." There are three red error boxes: one at the top containing "Unknown macro: excerpt.", one below the "Instructions" section containing "Unknown macro: contentbylabel.", and one at the bottom containing "Unknown macro: details." The "Instructions" section lists four steps: 1. Export data from Confluence, 2. XWiki Prerequisite, 3. Run Confluence XML Importer, and 4. Run Nested Spaces Migrator. A blue warning box states: "Warning: not everything can import as is in Confluence. Some macros will need to be created or removed after import." The left sidebar shows navigation menus for "Applications" and "Navigation".

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

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

Edit Create

Space name: CT

All macros  Only missing macros

Go

### Macro count

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

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

---

# 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 !**

