



Using Excel to prepare XML files for Europe PMC's External Links Service

> EMBL-EBI Wellcome Trust Genome Campus Hinxton Cambridge CB10 1SD UK www.ebi.ac.uk

http://www.ebi.ac.uk/support

4th April 2014, Doc Version 1.3

Table of Contents

Introduction	3
Populating the spreadsheet with required data	3
Mapping the data to the XML Schema	5
Exporting links to an XML file	7
Preparing a provider details XML file	7

Introduction

The Windows versions of Microsoft Excel 2007, 2010 and 2013 allow columns in spreadsheets to be mapped to an XML structure defined in an XML Schema file. This document provides instructions for how to use this feature to generate the files that are necessary to participate in the External Links Service.

Since the first version of this document, the ability to identify records in Europe PMC by their DOIs has been added. However, because Excel does not fully support the W3C XML Schema recommendation, it is not possible to generate such files using Excel. What follows assumes you have used the Europe PMC web site or web services to obtain a list of IDs of the records to which you wish to add links.

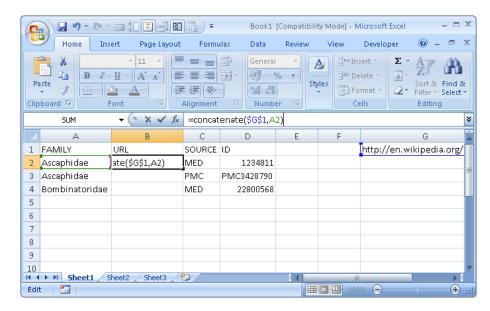
Populating the spreadsheet with required data

The following procedure to generate an XML links file covers one potential example of URL formatting. However depending on the required URL structure, other formatting may be needed. Consider someone who wishes to provide links from articles that mention families of frogs to an external resource with more details about those specific families (for the purposes of this demonstration, the links will be to Wikipedia pages.) The raw data might look something like the following:

	А	В	С	
1	FAMILY	SOURCE	ID	
2	Ascaphidae	MED	1234811	
З	Ascaphidae	PMC	PMC3428790	
4	Bombinatoridae	MED	22800568	
5				

The first task is to generate URLs to the resources. To do this:

- 1. Enter the base URL in cell F1. In this example we will use: <u>http://en.wikipedia.org/wiki/</u>
- 2. Right-click on the B-column and select Insert.
- 3. Give the new B-column a heading, say URL.
- 4. In the cell B2, enter this formula: =concatenate(\$G\$1,A2)



5. Right-click on cell B2 and select **Copy**.

6. Select all the cells in the B-column for which there are values in the A-column, right-click and select **Paste**.

) - (21 -		-	Book1 [Compati	bility Mode	e] - Microsoft E	ixcel	- = x
	Hom	e Ins	ert Page Lay	yout Formulas	Data Reviev	v View	/ Develop	er 🕜 –	⊐ X
	iste ▼ ✓	Calibri BI Calibri		E E E E E E E	General • • • • • • • • • • • • • •			Σ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓	Select *
	B3		- (9	fx =CONCATENA	TE(\$G\$1,A3)				*
	Δ			В		С	D	E	-
1	FAMILY		URL			SOURCE	ID		
2	Ascaphid	lae	http://en.wik	ipedia.org/wiki/As	caphidae	MED	12348	11	_
3	Ascaphid	ae	http://en.wik	(ipedia.org/wiki/As	caphidae	PMC	PMC34287	90	
4	Bombina	toridae	http://en.wik	ipedia.org/wiki/Bc	mbinatoridae	MED	228005	68	
5									
6									
7									
8									
9									
10	▶ ► Sh	eet1 / S	iheet2 / Sheet3	/*]	14				
			ress ENTER or ch	A	Count: 2	₩ 🛛 🖽	100% 🕞 —	0	- 🕂 ,::

One further piece of information is needed before export to XML can begin: the provider ID to associate with each link (and under whose heading the links will be displayed on the Europe PMC web site.) For most contributors, a single provider ID is likely to be sufficient, but it is possible to maintain multiple profiles (e.g. one for each research group within an organisation, or for each text mining method.) To add this:

- 1. Right-click on the B-column and select Insert.
- 2. Give the new B-column a heading, say *PROVIDER ID*.
- 3. In the cell B2, enter the provider ID. This should be an integer given to you when you applied.
- 4. Right-click on cell B2 and select **Copy**.
- 5. Select all the cells in the B-column for which there are values in the A-column, right-click and select **Paste**.

		II (I I I I I I I I I I I I I I I I I I	₹	Book1 [Compat	ibility Mo	ode] - Microso	ift Excel	- = X
C	Home Ins	ert Page Layout	Formulas	Data Review	w Vi	ew Deve	loper	🔞 🗕 📼 🗙
	Calibri B Z Baste V board S	• 11 • <u>U</u> • A A A Sont S		General • • • • • • • • • • • • • •	Styles	Gells	i	Sort & Find & Filter + Select + Editing
	B3	▼ () f _x	1000					×
	А	В		С			D	E
1	FAMILY	PROFILE ID	URL				SOURCE	ID
2	Ascaphidae	1000	http://en.wikip	edia.org/wiki,	/Ascapł	nidae	MED	12348
3	Ascaphidae	1000	http://en.wikip	edia.org/wiki,	/Ascaph	nidae	PMC	PMC342875
4	Bombinatoridae	1000	http://en.wikip	edia.org/wiki/	/Bombi	natoridae	MED	228005
5			Ê.					
6								
7								
8								
9								
10								
		heet2 Sheet3)	
Sele	ect destination and p	ress ENTE Avera	ge: 1000 Count:	2 Sum: 2000		🔟 100% 🧲		. +

Mapping the data to the XML Schema

All the information required to export to XML is now present in the spreadsheet. Next, load the schema file that will define the structure of the XML for Excel, as follows:

- 1. Click on the **Office Button** (the large round in the top-left of the window in Excel 2007) or the **File** tab (in the ribbon in Excel 2010/2013.)
- 2. Select **Excel Options**. In the **Popular** section of the options dialog, select the **Show Developer tab in the Ribbon** checkbox, if it is not already checked. In some versions of Excel, the same checkbox is in a section of the option called **Customize** Ribbon. Click **OK**.
- 3. Select the **Developer** tab which will now be included in the ribbon.
- 4. In this tab there is a section labelled **XML**. Click the **Source** icon within that section. A panel entitled **XML Source** will be displayed on the left of the main Excel window.

6	n) 🖬 🤊 - (° -		🗐 🖛 🛛 Book1 [Comp	atibility Mode] - Micro	soft Excel			_ = X
P	Home Ins	ert Page Layout	Formulas Data Revie	w View Deve	loper			🔞 _ = 🛪 X
	sual Macros	ord Macro Relative References cro Security e	Insert Design Run Dialog XIL Map Properties Import				nent el	t
	B3	- (● f _x	1000					*
	Α	В	С		D	E		XML Source 🔹 💌
1	FAMILY	PROFILE ID	URL		SOURCE	ID	Π	XML maps in this workbook:
2	Ascaphidae	1000	http://en.wikipedia.org/wiki	/Ascaphidae	MED	1234811		-
3	Ascaphidae		http://en.wikipedia.org/wiki		PMC	PMC3428790		
4	Bombinatoridae	1000	http://en.wikipedia.org/wiki	/Bombinatoridae	MED	22800568		
5								
6								
7								
8								
9								
10								
11								
12								This workbook does not contain any XML maps. Click XML Maps to add an XML map to
13								this workbook.
14								Options 👻 XML Maps
15								Verify Map for Export
16							-	(i) Tips for mapping XML
H.	🕩 🕨 Sheet1 🖉 S	heet2 🖌 Sheet3 📈		4		•		<u> </u>
Re	ady 🎦			Average: 1000	Count: 2	? Sum: 2000 👔]]] 100% 😑 —

- 5. Click the XML Maps... button in the new panel.
- Click the Add... button in dialog that is displayed. A standard Windows file selection dialog is displayed. Use that to find and select the file named *labslink_v1.1.xsd* which can be downloaded from <u>http://europepmc.org/docs/labslink_v1.1.xsd</u>
- 7. A dialog offering a choice of "root nodes" will be displayed. Select **links** and click **OK**.

X	(ML Maps			? ×
ł	XML <u>m</u> aps in this	; workbook:		
	Name	Root	Namespace	
	links_Map	links	<no namespace=""></no>	
	•			
	<u>R</u> ename	<u>A</u> dd	Delete OK Cancel	

- 8. Click **OK** to return to the Excel spreadsheet. The structure of the XML defined in the schema will be displayed in the **XML Source** panel.
- 9. Drag the elements of the XML structure, one-by-one, from the panel to the heading of the column containing the relevant information. As you do so, the background colour of the cells in the columns that contain the data will change, to signify their inclusion in the mapping to XML. The elements in the XML structure will be emboldened as they are mapped.

C) •) • (* •		🗐 🗧 🛛 Book1 [Com	patibility Mode] - Microsoft	Ex	Table Tools	_ = >
C	Home Ins	ert Page Layout	Formulas Data	Review View De	veloper	Design	🔞 _ 📼 🗙
Tak •	ole4	ummarize with Pivot Remove Duplicates Convert to Range Tools	Table Export Refresh External Table Data			nns	Table Styles
	B1		PROFILE ID				3
	А	В		с	D	E	XML Source 🔹 🗙
1	FAMILY	PROFILE ID 🛛 🔽	URL		SOURCE	E ID	XML maps in this workbook:
2	Ascaphidae	1000	http://en.wikipedia.org	y/wiki/Ascaphidae	MED	1234811	links_Map 💌
З	Ascaphidae	1000	http://en.wikipedia.org	y/wiki/Ascaphidae	PMC	PMC3428790	🗆 🖾 links
4	Bombinatoridae	1000	http://en.wikipedia.org	r∕wiki/Bombinatoridae	MED	22800568	
5							ink
6							e 🔁 🔁 resource
7							url
8 9							title □2≸ record
9 10							source
10							- id
12							To map repeating elements, drag the
13							elements from the tree onto the worksheet where you want the data headings to
14							appear.
15							To import XML data, right click an XML
16							mapped cell, point to XML, and then click
17							Import.
	→ M Sheet1 / S idy 🞦	heet2 🖌 Sheet3 🦯		Average: 100	0 Count:	4 Sum: 3000	

After mapping all the required columns, your spreadsheet should look something like this:

0) - (-	- 🗆 📜 🖪 🛃 🗄	Book1	[Compatibility Mode] -	Microsoft Ex	Table Tools		_ = ×
C	Home	Insert Page Layout	t Formulas Dat	a Review Vie	w Developer	Design		🕜 _ = 🛪 🗙
Tat •	ole4	Summarize with Pivot Remove Duplicates Convert to Range Tools	Table Export Refresh External Table	🙊 🗹 Banded Ro	v 📄 First Colum 📄 Last Columi ws 📄 Banded Col e Style Options	n =====	Table Styles	
	E1	\bullet (• f_x	ID					8
	Α	В		С	C) E	XML Source	
1	FAMILY	PROFILE ID	URL		SOUR	CE 🔽 ID 🛛 🔽	XML maps in th	is workbook:
2	Ascaphidae	1000) http://en.wikipedi	a.org/wiki/Ascaphi	dae MED	1234811	links_Map	-
3	Ascaphidae) http://en.wikipedi	0 1 1		PMC3428790	🖃 🞽 links	
4	Bombinatorida	ie 1000) http://en.wikipedi	a.org/wiki/Bombin	atoridae MED	22800568		themaVersion
5							🗎 🗎 🗁 👹 lir	ik Ť providerId
6								j resource
7								📑 url
8 9								ittle title
9 10								
11							-	🚰 id
12							To map repeat	ing elements, drag the
13								the tree onto the ere you want the data
14							headings to ap	
15							To import XML	data, right click an XML
16							mapped cell, po	oint to XML, and then click
17	A N Charte	Charles Charles	87 /				Import.	
Rea		Sheet2 / Sheet3 /	2		1111			

10. The *title* element is optional. If left unmapped, we will simply display the URL. However, you may wish to add another column for link titles. The cells in that column could all be populated with a standard title, or constructed using a concatenate formula, incorporating both static text and variables, such as the family name in the example used here.

Exporting links to an XML file

- In some versions of Excel, the ribbon will have switched to the **Design** tab during preceding steps. Re-select the **Developer** tab and click the **Export** icon.
- 2. A standard Windows dialog for saving files is displayed. Choose/enter a filename and click **Export**. An XML file based on the contents of the spreadsheet will be created. If opened in a text editor, it should look something like this:

```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<links>
  <link providerId="1000">
    <resource>
      <url>http://en.wikipedia.org/wiki/Ascaphidae</url>
      <title>Wikipedia page for the frog family Ascaphidae</title>
    </resource>
    <record>
      <source>MED</source>
      <id>1234811</id>
    </record>
  </link>
  <link providerId="1000">
    <resource>
     <url>http://en.wikipedia.org/wiki/Ascaphidae</url>
     <title>Wikipedia page for the frog family Ascaphidae</title>
    </resource>
    <record>
      <source>PMC</source>
      <id>PMC3428790</id>
    </record>
  </link>
  <link providerId="1000">
    <resource>
      <url>http://en.wikipedia.org/wiki/Bombinatoridae</url>
      <title>Wikipedia page for the frog family Bombinatoridae</title>
    </resource>
    <record>
      <source>MED</source>
      <id>22800568</id>
    </record>
   </link>
</links>
```

Preparing a provider details XML file

Another XML file is required, containing details to associate with the links. It is not necessary to upload this every time, but no links can be displayed on Europe PMC without it. It can be created manually in a text editor, by copying the example at the end of this document, or using Excel as follows:

- 1. Open a new spreadsheet, or a blank worksheet in the same spreadsheet as the links.
- 2. In the first row, enter the following four column names:
 - ID
 - RESOURCE NAME

- DESCRIPTION
- EMAIL
- 3. In the second row, under each heading, enter values. The ID is an integer given to you when you applied. The resource name is the heading under which you wish your links to be displayed. The description should be a couple of sentences, describing the resource being linked to, or the method used to generate the links. The email is the email address of a person who can respond to, or redirect, both technical issues (e.g. invalid XML files, broken links) and enquires about the quality or legality of the linked resources.
- 4. If you are maintaining multiple profiles, each with a different provider ID, the details of those can be added in subsequent rows.
- 5. Follow the instructions above for loading the schema file. But, when offered a choice of root elements, select "providers" instead of "links".
- 6. Follow the instructions above for dragging XML elements to columns in the spreadsheet. This should result in a spreadsheet that looks something like this:

C. 2 "	· (~ - 💷 📜 🔳 🚮	📰 🖬 🔻 🛛 Bool	(1 [Comp	atibility Mode] - Mic	rosoft Ex	Table	Tools	_ = X
Home	Insert Page Layo	ut Formulas (Data	Review View	Develop	er De	sign	🙆 – 🗖 🗙
Table Name: Table5 · Properties				lumn d Columns	Quick Styles + Table Styles			
D1	- (° ;	🖅 EMAIL						*
	Research Consortium	C DESCRIPTION Frog families identific in the literature	▼ EMJ ed jblo	D AIL ggs@frogresearch.	org	To map rejelements f worksheet headings t	in this workboo Map roviders schemaVers provider id resour email peating element rom the tree of where you way o appear.	ion ceName otion its, drag the nto the and the data
Ready	1 Sheet2 Sheet3	2					XML data, riqhi	Click an XML

7. Click the "Export" icon in the "Developer" tab in the ribbon as with links. The resulting XML file should look something like the following:

xml version="1.0" encoding="UTF-8" standalone="yes"? <providers></providers>
<pre><pre>></pre></pre>
<id>1000</id>
<pre><resourcename>Frog Research Consortium</resourcename> <description>Frog families identified in the literature</description></pre>
<pre><description>riog ramifies rdentified in the fiterature</description> jbloggs@frogresearch.org</pre>