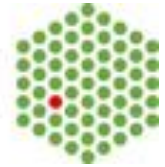




Europe
PubMed
Central

EMBL-EBI



**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.

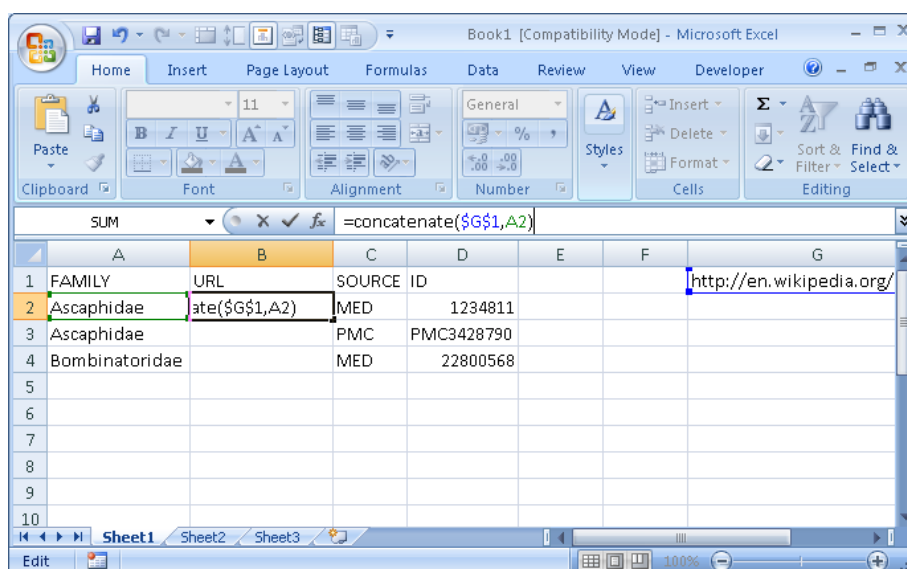
Populating the spreadsheet with required data

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:

	A	B	C	
1	FAMILY	SOURCE	ID	
2	Ascaphidae	MED	1234811	
3	Ascaphidae	PMC	PMC3428790	
4	Bombinatoridae	MED	22800568	
5				

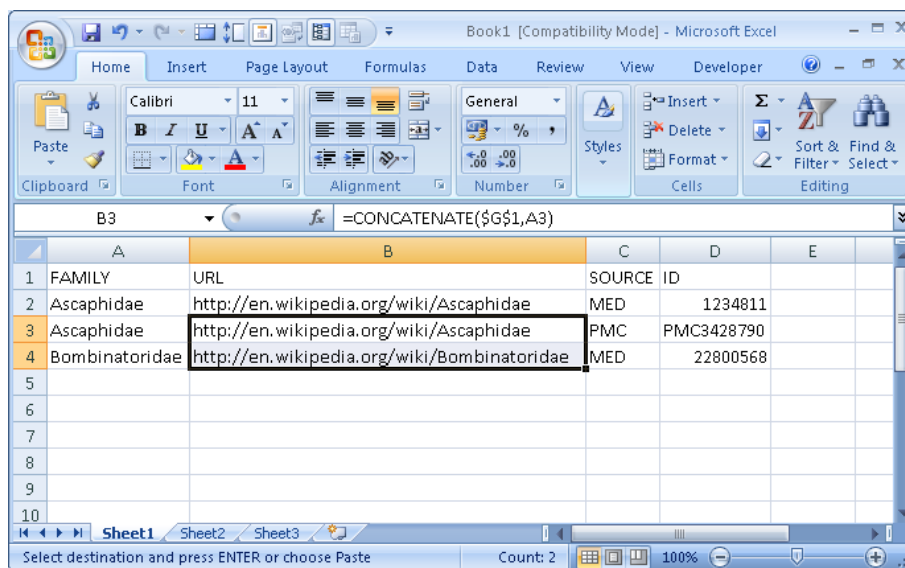
The first task is to generate URLs to the resources. To do this:

1. Enter the base URL in cell F1. In this example we will use: <http://en.wikipedia.org/wiki/>
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(G1,A2)`



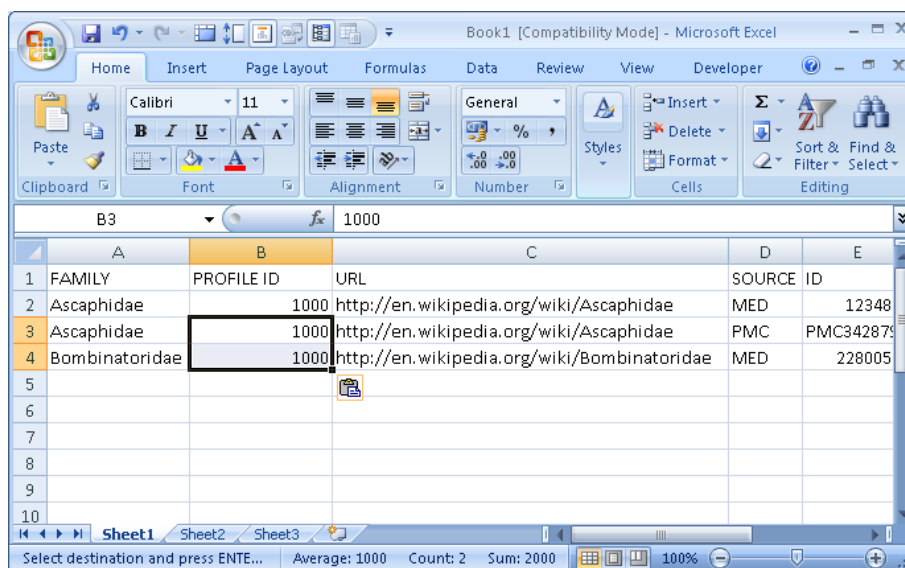
5. Right-click on cell B2 and select **Copy**.

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



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:

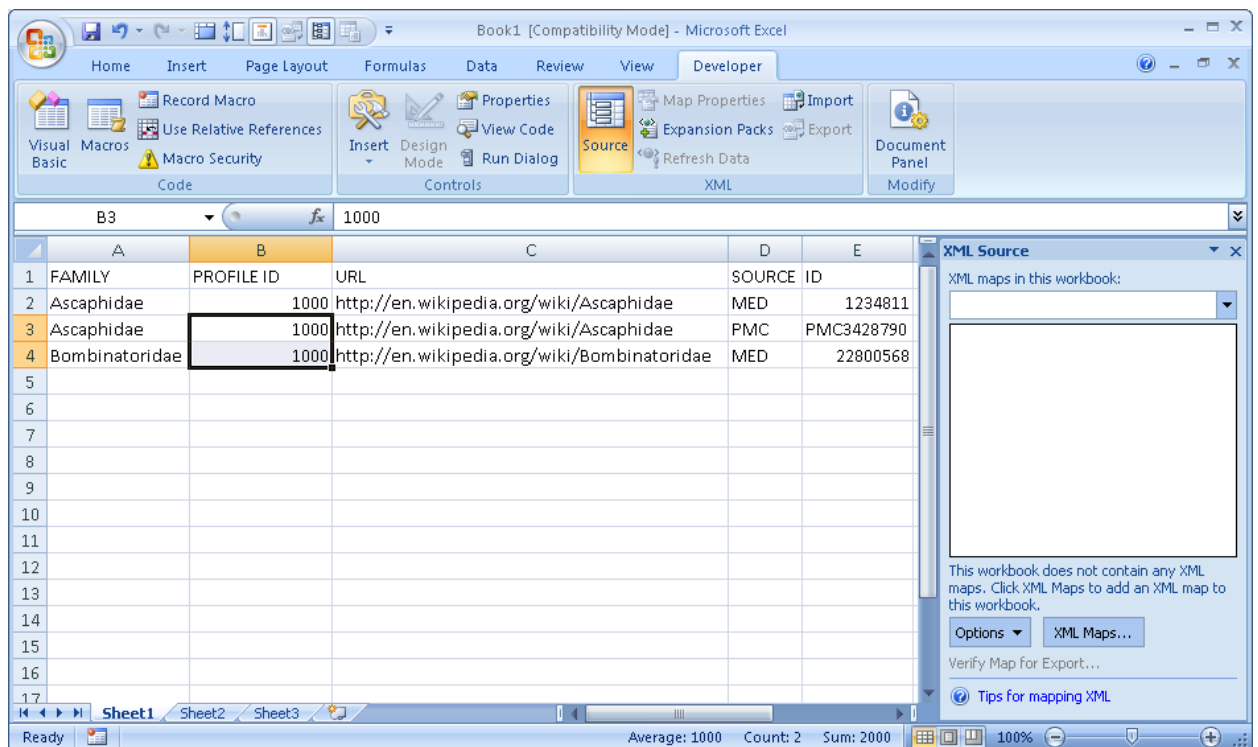
1. Right-click on the B-column and select **Insert**.
2. Give the new B-column a heading, say *PROVIDER ID*.
3. In the cell B2, enter the provider ID. This should be an integer given to you when you applied.
4. Right-click on cell B2 and select **Copy**.
5. Select all the cells in the B-column for which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



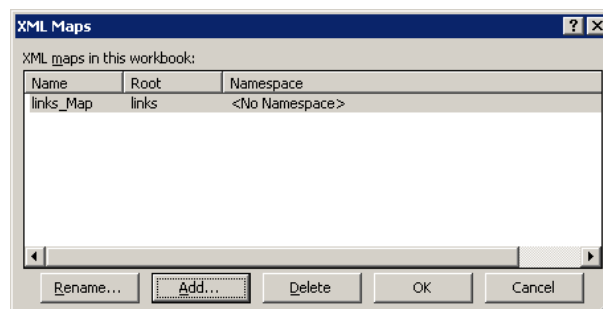
Mapping the data to the XML Schema

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:

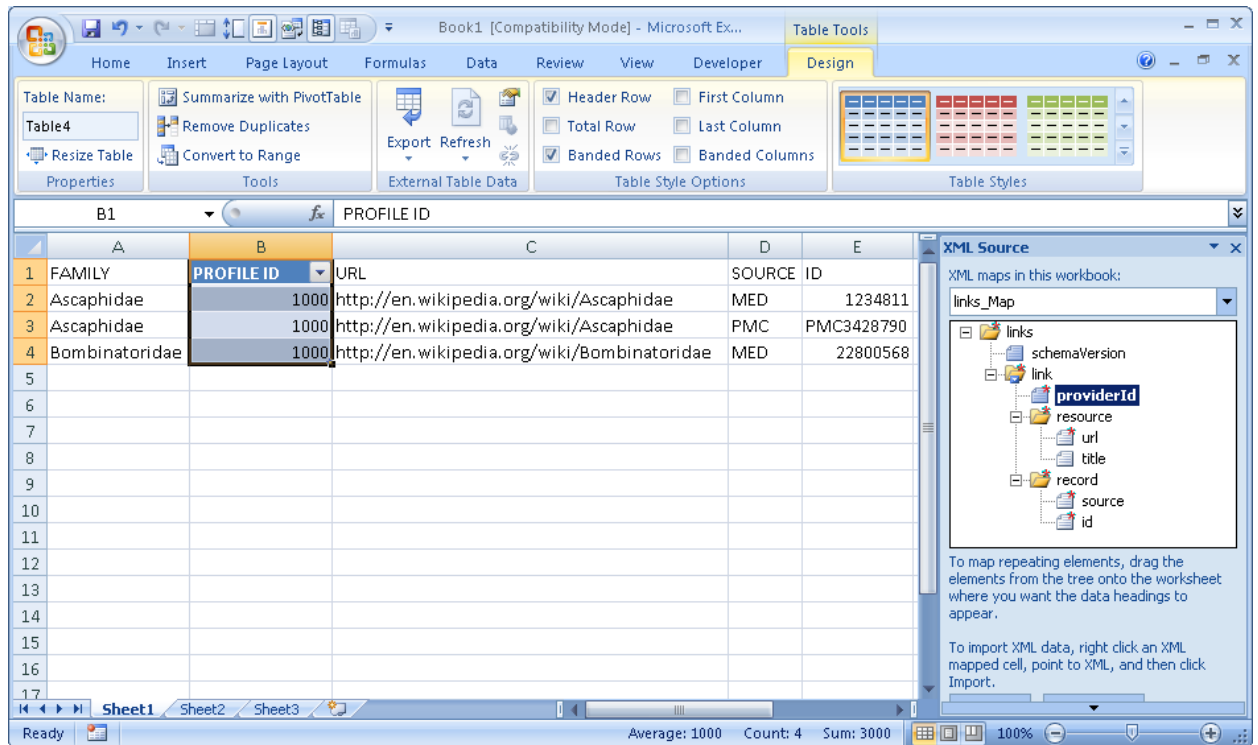
1. Click on the **Office Button** (the large round in the top-left of the window in Excel 2007) or the **File** tab (in the ribbon in Excel 2010/2013.)
2. Select **Excel Options**. In the **Popular** section of the options dialog, select the **Show Developer tab in the Ribbon** checkbox, if it is not 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.



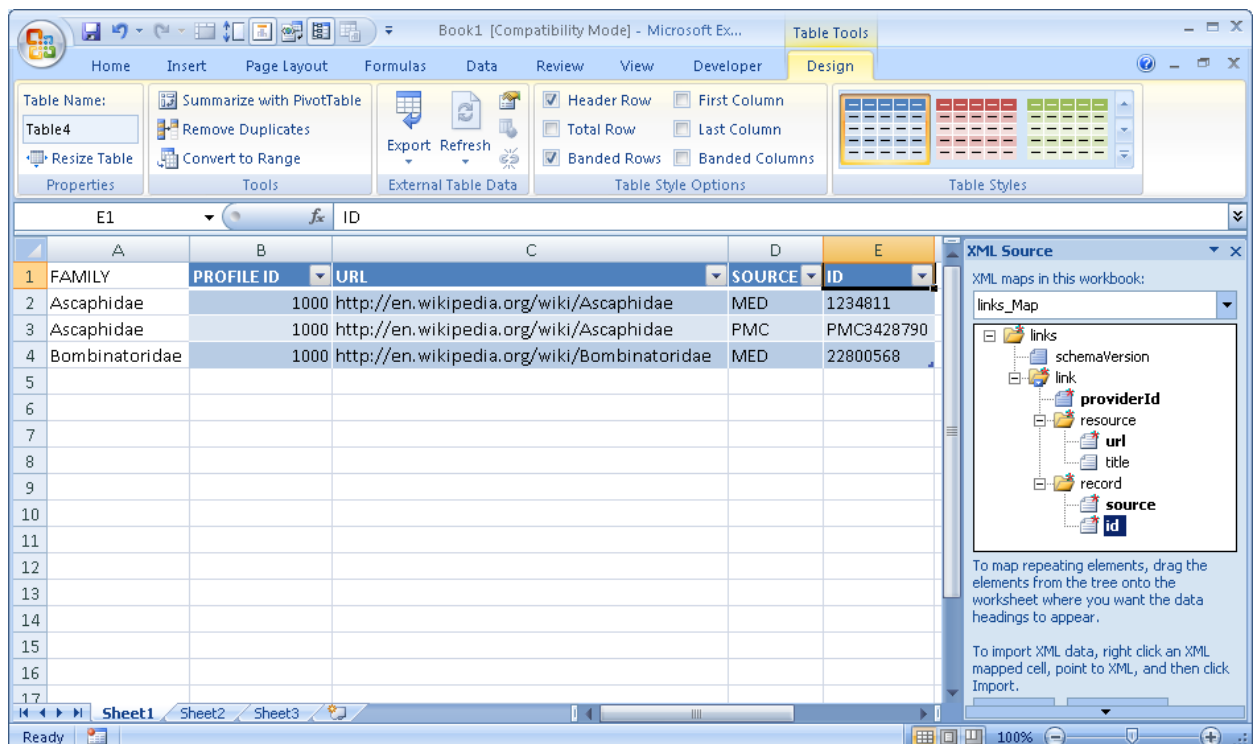
5. Click the **XML Maps...** button in the new panel.
6. Click the **Add...** button in dialog that is displayed. A standard Windows file selection dialog is displayed. Use that to find and select the file named *labslink_v1.1.xsd* which can be downloaded from http://europepmc.org/docs/labslink_v1.1.xsd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



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.



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



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

1. 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:

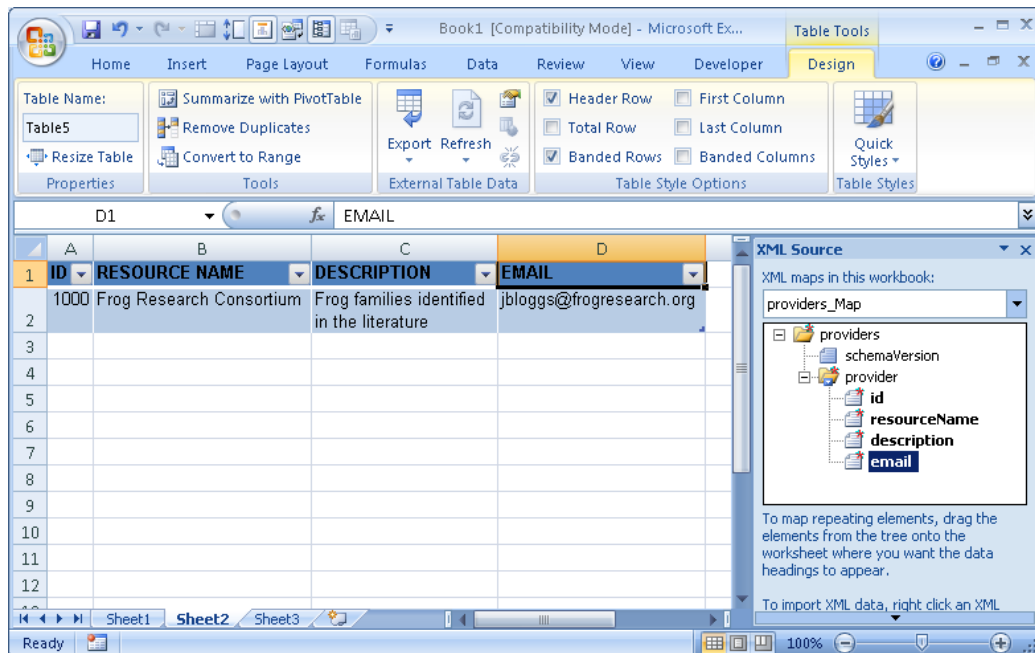
```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<links>
  <link providerId="1000">
    <resource>
      <url>http://en.wikipedia.org/wiki/Ascaphidae</url>
      <title>Wikipedia page for the frog family Ascaphidae</title>
    </resource>
    <record>
      <source>MED</source>
      <id>1234811</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
- 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.
 - If you are maintaining multiple profiles, each with a different provider ID, the details of those can be added in subsequent rows.
 - Follow the instructions above for loading the schema file. But, when offered a choice of root elements, select “providers” instead of “links”.
 - Follow the instructions above for dragging XML elements to columns in the spreadsheet. This should result in a spreadsheet that looks something like this:



- 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>
  <provider>
    <id>1000</id>
    <resourceName>Frog Research Consortium</resourceName>
    <description>Frog families identified in the literature</description>
    <email>jbloggs@frogresearch.org</email>
  </provider>
</providers>
```