GW Toolset Mass Upload training

Ally Crockford
Wikimedian in Residence, National Library of Scotland

Permissions: getting access

0 Create user account on Commons and [Commons Beta](#)

0 Request the GWToolset user right on Beta:
   0 Join [GlamTools mailing list](#) and send request

0 Before testing: whitelist your upload domain [via Phabricator](#)
   0 Work with organisation to determine domain
   0 Larger organisations may have local virtual directories where different derivatives are stored – make sure you whitelist the right domain!
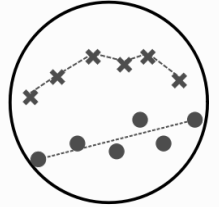
Notes on licensing

0 When releasing content on a Wiki-compatible license via the Toolset, remember that you're also releasing the metadata

0 Metadata mapping is under your control: potentially sensitive (or not useful) information can be excluded, **but!**

0 Confirm with an organisation that the metadata is theirs to license and they are happy to license it openly:

  0 Be especially clear when dealing with historical metadata or metadata from collaborative/funded digitisation projects
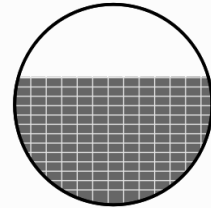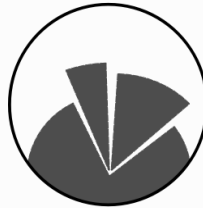
Metadata: formatting

0 How your collection data might be held:
   (Gill Hamilton, Digital Access Manager @ National Library of Scotland)

   0 In a **database**:
      i.e. MS Access, MS SQL Server,
      MySQL
   0 In a **spreadsheet**:
      i.e. MS Excel, Google Sheet
   0 In a **document**: i.e.
      MS Word, Open Office, Google
      Docs
   Remember: **ALWAYS WORK ON A BACKUP COPY OF YOUR
   METADATA!**

Metadata: formatting

0 Certain metadata fields will need to be added to existing metadata if they don't already exist:

   0 **Upload URL field**: <gwtoolset_url_to_the_media_file>

   0 A **location or institution** field: <wiki_location>

0 Certain metadata fields *can be* added or altered to make use of Wiki templates:

   0 **<wiki_permission>**: e.g. {{PD-UKGov}}{{PD-1923}}

   0 **<wiki_source_info>**: institution source template i.e. {{NLS Collections|Shelfmark|ObjectID}}

   0 *****<dimensions>**: {{Size |unit=cm|height=xx|width=yy}}
       (if XML wiz on hand)

Institutional source templates

0 Points back to institution, to their shelf mark or digital gallery: encourages traffic and can improve discovery for users

0 E.g. {{NLS Collections|sequence number|digital object ID}}

This image is available from the **National Library of Scotland** under the sequence number or Shelfmark ID **Weir.8(6)**. You can see this image in its original context, along with the rest of the Library's digital collections, in **the NLS Digital Gallery**

```
<center>
{| style="width:100%; border:1px solid black; background:#f6f6f4; padding:0; text-align:center;"
|-
| align=left | [[File:National Library of Scotland.JPG|150px]]
| This image is available from the <big>'''[[w:National Library of Scotland|National Library of
Scotland]]'''</big> {{ #if:  {{{1|}}} | under the sequence number or Shelfmark ID '''{{{1}}}'''. }} {{ #if:  {{{2|}}} |
You can see this image in its original context, along with the rest of the Library's digital collections, in
<big>'''[http://digital.nls.uk/{{urlencode:{{{2|{{{id}}}}}}}} the NLS Digital Gallery]'''</big>}}<small><br/>
| align="right" | [[File:Flag of Scotland.svg|150px]]
|}</center><includeonly>{{{category|[[Category:Images from National Library of Scotland]]}}}
</includeonly>
```
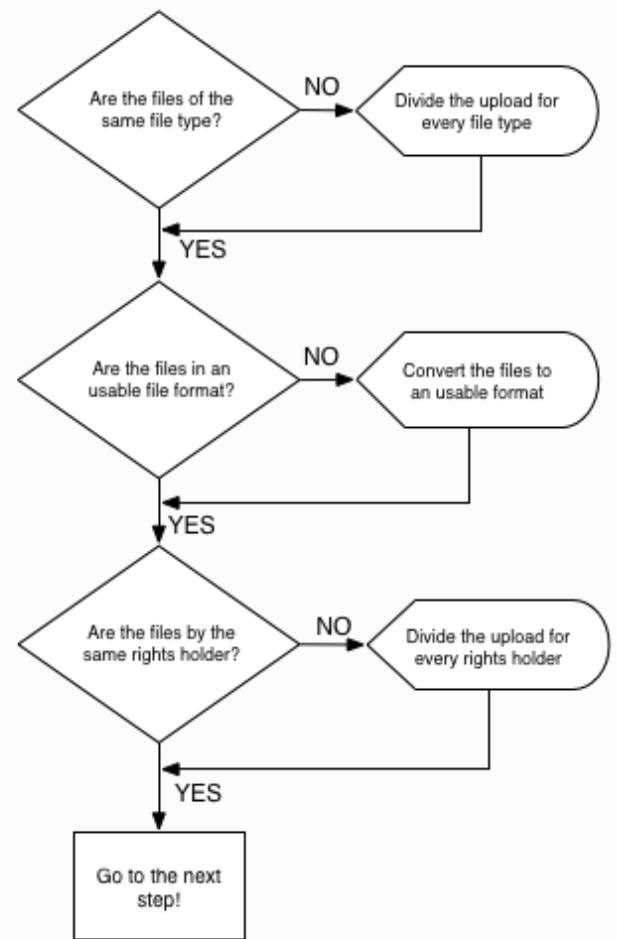
Dividing batches

0 Toolset can handle thousands of images – is this best for discovery and integration?

0 Divide by **file type**, **rights holder** for upload purposes

0 Divide by collection, creator, or theme for **discovery and integration** purposes

   0 How will users find your uploads? How will you publicise them?

Are the files of the same file type? — NO → Divide the upload for every file type

YES

Are the files in an usable file format? — NO → Convert the files to an usable format

YES

Are the files by the same rights holder? — NO → Divide the upload for every rights holder

YES

Go to the next step!

"GLAMWiki Toolset manual content selection" by Ter-burg - CC BY 4.0 via Wikimedia Commons

Mapping metadata

Special page

Search 🔍

GWToolset

- Metadata mapping

Step 2: Metadata mapping

Metadata file

Total number of records found in this metadata file: 10.

Mapping the metadata fields

Below is/are :

- A list of the fields in the MediaWiki Template:Artwork.
- Drop-down fields that represent the metadata fields found in the metadata file.
- A sample record from the metadata file.

In this step you need to map the metadata fields with the MediaWiki template fields.

1. Select a metadata field under the "Maps to" column that corresponds with a MediaWiki template field under the "Template field" column.
2. You do not need to provide a match for every MediaWiki template field.

Map your metadata

Template Artwork

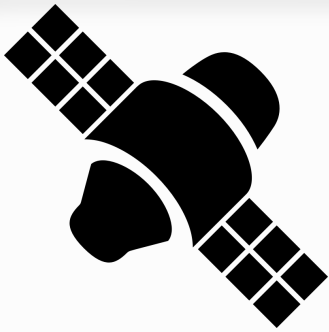| Template field		Maps to		Metadata's example record's contents.
accession number :	⊕	dc:identifier ⇕	dc:contributor	Aankoop uit het F.G. Waller-Fonds
artist :	⊕	dc:creator ⇕	dc:creator	prentmaker: Call, Jan van (I)
author :	⊕	⇕	dc:creator	vermeld op object uitgever: Schenk, Pieter (I)
credit line :	⊕	⇕	dc:creator	vermeld op object verlener van privilege: onbekend
date :	⊕	dc:date ⇕	dc:date	derde kwart 17e eeuw
demo :	⊕	dc:publisher ⇕	dc:date	vierde kwart 17e eeuw
department :	⊕	⇕	dc:date	eerste kwart 18e eeuw
description :	⊕	dc:description ⇕	dc:description	In een landschap knielt een vrouw voor een buste van Priapus, aan wie zij een offer brengt. Rech op de achtergrond lopen twee vrouwen op een weg. Illustratie voor de Ecloga 7 van Virgilius.
dimensions :	⊕	dcterms:extent ⇕	dc:description	In a landscape kneels a woman for a bust of Priapus, to whom they make an offering. Right in th background two women on a road. Illustration for the seventh Eclogue of Virgil.

Categorising

0 Create **tracking category** for all uploaded images i.e. [[Category:Images uploaded as part of NLS-WMUK collaboration]]

0 Institution tag and/or institution template should place images into **institutional category** if pre-exists
  0 If no such category exists, may want to create one

0 Adding categories based on metadata: using 'Person', 'Place' fields can add categories automatically, but **confirm formatting** to avoid linking to non-existent categories

0 Hidden category **[[Category:GWToolset batch upload]]** automatically added; should add **[[Category:Supported by Wikimedia UK]]** to individual uploads or to tracking category

Batch preview

- Failed previews often due to **improperly formatted URL** to upload or **lack of whitelisted domain**
    - In case of unidentifiable error, send information via **Toolset mailing list** (error description, file type, relevant XML details, URL domain, etc)

- Image itself will not preview
    - Use batch preview to **identify broken links** (i.e. to categories) and **formatting problems** (templates, bad XML characters, etc.)
    - Return to metadata mapping to correct any issues without uploading content, or cancel upload if metadata needs work itself

Adjustments

0 **File names**: will they be unique, discoverable, appropriate Commons file names?
  0 Before uploading, check for **duplicate file names** throughout metadata: remember, batch preview only shows first 5 files
  0 **Add identifier** to batch file names (i.e. Actor in Clown Costume - Wier Collection.jpg; Scotia Depicta - Cromarty - Plate.jpg)
  0 Institution may require or request their name in file name (i.e. NLS Haig - Big gun going up to the front.jpg)

0 **Mediafile throttle**: how many images do you want to upload per background job? How long do you want the upload to take?

After uploading

0 Watch upload on **Special:New Files** (linked from upload screen)

0 **Promote**!

   0 Blog posts, mailing lists, WikiProjects, individual users?, volunteer communities, press release, events (edit-a-thons, tag-a-thons)

0 **Track metrics**:

   0 Google Analytics for partner website
   0 Treeviews
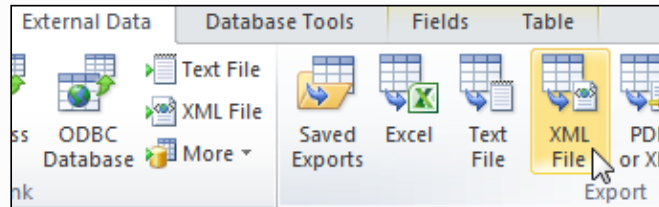   0 BaGLAMa 2 - *request in advance!*
   0 GLAMorous

Creating XML files

All advice courtesy of Gill Hamilton, Digital Access Manager at
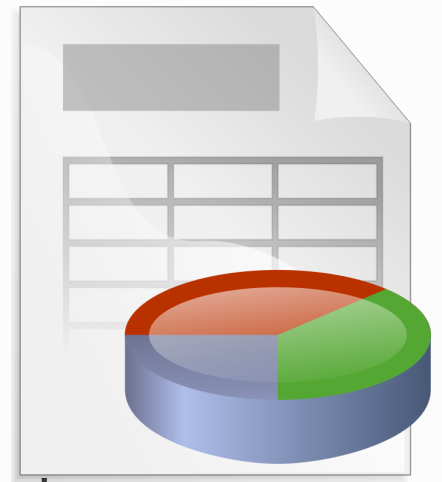National Library of Scotland

Warning: here there be monsters...

MS Access to XML

0 What you will need:

    0 Microsoft Access (this assumes 2010 version)
    0 A **COPY** of your data
    0 A little bit of patience!

0 Assuming your data is in a flat file in MS Access:

    0 Open your data in MS Access
    0 From the menu select **External Data > Export > XML File**
    0 Choose the location of your XML file
    0 Select the **Data (XML)** option. This will create an XML file that you can use in the toolset
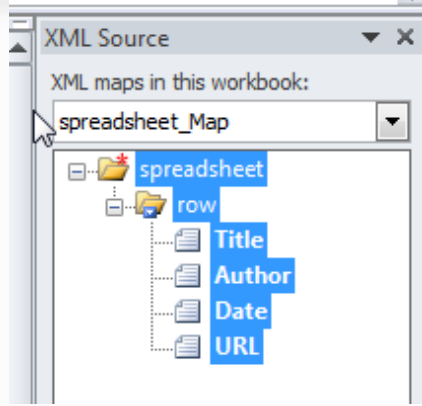   0 C'est tout!

Spreadsheets to XML

0   Getting started:

    0   Will need a **Microsoft Excel file** (this assumes 2010 version): **convert** Google sheet docs or Open Office spreadsheet files into Excel

    0   Remember: work on a **COPY** of your data!

0   Tools you will need:

    0   Microsoft Excel

    0   A **text editor** (Notepad, TextEdit, Notepad++) (If you have and know how to use an XML editor, that is ideal, but standard text editor will do fine)

    0   An eye for detail

    0   An eye for data structure
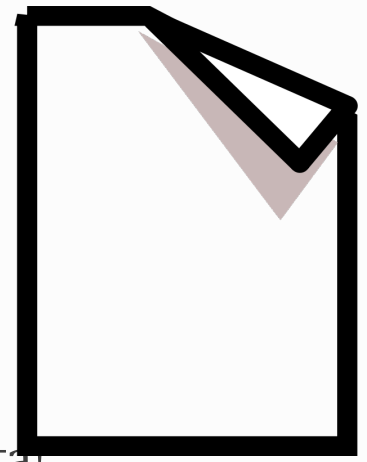
    0   Patience!

Spreadsheets to XML

1. Create an **XML schema** based on the column name and the first two rows of data
2. Go to Excel and open a **copy** of your spreadsheet containing your collection data
3. Switch on the **Excel Developer** tab (how to?)
4. From the Developer tab:
   - Select **Source > XML maps > Add**
   - Select the **XML schema** you made in **1.** above
   - Click OK (perhaps twice)
5. You should see your **XML map** in a window to the right of your spreadsheet data
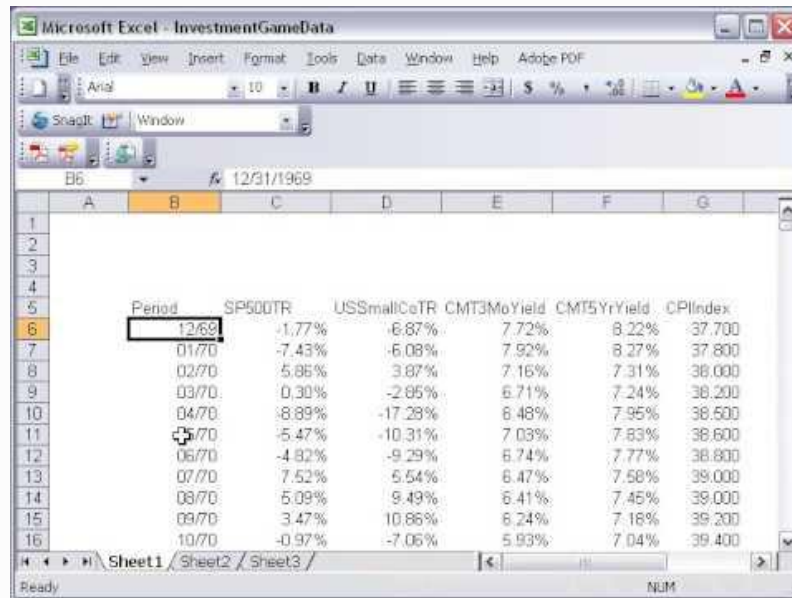
6. Click and drag the **XML map** to replace column names in your spreadsheet (grab **"spreadsheet"**)

7. You've now defined an **XML structure** for your data!

8. Click **File > Save As > [*new filename*]** and save as an **Excel Worksheet**

9. Finally, click **File > Save As > [*filename*]** and save as **XML data**

10. Et voila! The XML data file can be used with the Toolkit!

Documents to XML

0 Getting started:

    0 Will need a **Microsoft Word file** (this assumes 2010 version): **convert** Google Docs or Open Office Documents into Word

    0 Remember: work on a **COPY** of your data!

0 If data is in a **Word table**:

    0 Copy whole table and paste into empty **spreadsheet**

    0 Save the file

    0 Follow **Spreadsheet to XML** instructions (above)

0 If data is not in a Word table/is **unstructured**:

    0 Edit the document to **create a structure** (ideally a table) and continue as above

Links and tips

0   Creating an XML schema for Excel to XML:
  0   <u>YouTube video</u> using different data but a similar technique
0   Switching on the Excel Developer tab instructions:
  0   Simple instructions for <u>'how to'</u>

Collection item URLs for images

Every organisation will have different ways of constructing the URLs that point to digital files.

Here's an example:

http://digital.nls.uk/oor-wullie/images/home/oor-wullie-audio-map**.png**

http://	**protocol**, it's nearly always http://
digital.nls.uk	**domain**
/	indicates that what follows the is a **/ sub folder**
oor-wullie	a sub folder or location
images	a sub folder or location
home	a sub folder or location
Oor-wullie-audio-map	**filename**
.png	**file extension**, it's often jpg, png or gif for images

Collection item URLs for images

Example is based on a text based file name:
http://digital.nls.uk/oor-wullie/images/home/oor-wullie-audio-map**.png**

Some might be based on numbers, often a unique identifier for the file.
E.g.: **http://deriv.nls.uk/dcn3/7441/74411791.3.jpg**

If you are creating filenames for digitised images best practise is to
- make the filenames unique
- consider using numbers or mix of numbers and letters. This allows you to scale up in the future
- Use Persistent Identifiers
  - http://www.metadaten-twr.org/2010/10/13/persistent-identifiers-an-overview/