GLAMwiki Toolset (GWToolset)

Fæ's pocket guide to speedily releasing thousands or millions of files to Wikimedia so volunteers can add value to collections

Why does the tool exist?

- Demand from Museums, Archives and Libraries for a way of managing large uploads to Wikimedia Commons
- An independent end user Steering Group drove the requirements and Europeana delivered the project
- This has taken 2 years from original discussions and cost ~200k euros

What does GWT do?

- Provides a direct upload service for mass collections of files available on the internet and of any size
- Gives a visual map of metadata to Commons templates
- No reliance on programming skills
- Any Commons allowed file type (image, audio, video) up to 1GB per file

How does it do it?

- GWToolset takes your list of images, uploads them from a source website and populates Commons templates for you
- This happens on WMF computers, rather than relying on yours
- Instead of just talking, let's look at some screenshots taken today

http://commons.wikimedia.org/wiki/Special:GWToolset
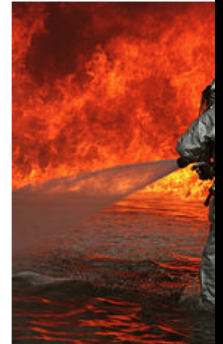
GWToolset

- Metadata mapping

GWToolset is a MediaWiki extension that allows GLAMs (Galleries, Libraries, Archives and Museums) the ability to mass upload content based on an XML file containing respective metadata about the content. The intent is to allow for a variety of XML schemas. Further information about the project can be found on its project page. Feel free to contact us on that page as well. Select one of the menu items above to begin the upload process. Please add yourself to the list of users at GWToolset users and follow the good practice of creating pages explaining your new projects, as these can help create a focus for our wider community of volunteers.
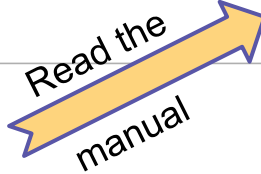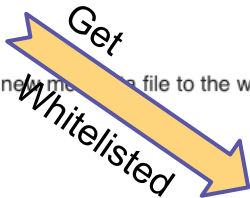
Step 1: Metadata detection

The metadata upload process consists of 4 steps:

1. Metadata detection
2. Metadata mapping
3. Batch preview
4. Batch upload

In this step, you upload a new metadata file to the wiki. The tool will attempt to extract the metadata fields available in the metadata file, which you will then map to a MediaWiki template in "Step 2: Metadata mapping".

Domain whitelist

If your media file domain is not listed below, please request ⧉ that your media file domain be added to the Wikimedia Commons domain whitelist. The domain whitelist is a list of domains Wikimedia Commons checks against before fetching media files. If your media file domain is not on that list, Wikimedia Commons will not download media files from that domain. The best example, to submit in your request, is an actual link to a media file.

- am.adlibhosting.com
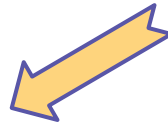- *.collectie.legermuseum.nl
- *.europeana1914-1918.eu

Upload your metadata file

1. What is the XML element that contains each metadata record: `record` *

2. Which MediaWiki template: *
   - Select a template:
     `Photograph ▲▼`
   - or enter a custom template:
     `TorontoHollarCollection`

3. Which metadata mapping: `GWToolset:Metadata_Mappings/User-name`
   GWToolset:Metadata_Mappings

4. Mediafile throttle: `20`
   After the batch preview, in step 3, GWToolset uploads the remaining records in your batch upload via background jobs. The mediafile throttle controls the number of mediafile requests Wikimedia Commons will make against your mediafile server each time a background job is run. You can set the mediafile throttle between 1-20. For example, if the total number of records in your batch upload is 100 and you set the throttle to 20, Wikimedia Commons will run 5 background jobs in order to process your entire batch upload. The time between each background upload job varies depending on server load and configuration. We anticipate that on Wikimedia Commons a GWToolset background job will run at least every 5 minutes.

5. Make sure the XML file is well-formed with this XML Validator⤢. *
   Select the metadata file source.
   - Metadata file upload: `/Volumes/1GB/RM.xml` `Browse…`
     Accepted file type: xml
     Maximum file size: 1,024,000 kilobytes

GWToolset

- Metadata mapping

Step 2: Metadata mapping

Metadata file

Total number of records found in this metadata file: 605.

Mapping the metadata fields

Below is/are :

- A list of the fields in the MediaWiki Template:Photograph.
- 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 Photograph

Template field	Maps to	Metadata's example record's contents.
accession number :	accession_number	accession_number SK-A-320
credit line :		artist anonymous
date :		category_1 Images from the Rijksmuseum
demo :	accession_number	category_2 Paintings in the Rijksmuseum Amsterdam
department :	artist	category_3 Paintings
depicted people :	category_1	date {{betweenl1550l1574}}
depicted place :	category_2	description {{nll1=Portret van een vrouw. Vroeger geïdentificeerd als Elisabeth I (1533-1603), koningin van Engeland. Buste, iets naar links, met bloemen in het haar en een parelsnoer om de hals.}}
description :	category_3	
	date	
dimensions :	description	dimensions drager "Height:" {{sizelcml44}}drager "Width:" {{sizelcml32.5}}
exhibition history :	dimensions	
	exhibition_history	exhibition_history
	filename	
inscriptions :	gwt_source	filename Portret van een vrouw Rijksmuseum SK-A-320
	institution	
institution :	makerLine	gwt_source http://www.rijksmuseum.nl/media/assets/SK-A-320
	matches	
medium :	medium	institution Rijksmuseum
	notes	
notes :	object_history	makerLine
	object_type	
object history :	permission	matches http://commons.wikimedia.org/w/index.php?search=Rijksmuseum+SK-A-320

Template Photograph

Template field	Maps to		Metadata's example record's contents.
accession number :	⊕	accession_number	accession_number SK-A-320
credit line :	⊕		artist anonymous
date :	⊕	date	category_1 Images from the Rijksmuseum
demo :	⊕		category_2 Paintings in the Rijksmuseum Amsterdam
department :	⊕		category_3 Paintings
depicted people :	⊕		date {{between|1550|1574}}
depicted place :	⊕		description {{nl|1=Portret van een vrouw. Vroeger geïdentificeerd als Elisabeth I (1533-1603), koningin van Engeland.
description :	⊕	description	Buste, iets naar links, met bloemen in het haar en een parelsnoer om de hals.}}
	⊖	makerLine	dimensions drager "Height:" {{size|cm|44}}drager "Width:" {{size|cm|32.5}}
dimensions :	⊕	dimensions	exhibition_history
exhibition history :	⊕	exhibition_history	filename Portret van een vrouw Rijksmuseum SK-A-320
inscriptions :	⊕		gwt_source http://www.rijksmuseum.nl/media/assets/SK-A-320
institution :	⊕	institution	institution Rijksmuseum
medium :	⊕	medium	makerLine
notes :	⊕	notes	matches http://commons.wikimedia.org/w/index.php?search=Rijksmuseum+SK-A-320
object history :	⊕	object_history	medium paneel; olieverf
other fields :	⊕		notes
other versions :	⊕	matches	object_history Acquisition: 1800
permission :	⊕	permission	object_type
photographer :	⊕		permission {{PD-US}}{{PD-old}} Data and photography rights: "All data and all images made available through the API
references :	⊕	references	are either in the public domain or are subject to a CC0 license." https://www.rijksmuseum.nl/en/api/terms-
source :	⊕	source	and-conditions-of-use {{cc-zero}}
strict :	⊕		place_of_creation Fontainebleau
title :	⊕	title	priref 5420
GWToolset title* :	⊕	filename	references
URL to the media file* :	⊕	gwt_source	source http://www.rijksmuseum.nl/collectie/SK-A-320
		Save mapping	title Portret van een vrouw
			wikidata Q17320079

GWToolset

- Metadata mapping

Step 3: Batch preview

Below are the results of uploading the first 3 records from the metadata file you selected and mapping them to the MediaWiki template you selected in "Step 2: Metadata mapping".

Review the templates, and if the results meet your expectations, continue with "Step 4: Batch upload", by clicking on the "Process batch" button below.

If you are not happy with the results, go back to "Step 2: Metadata mapping" and adjust the mapping as necessary.

If you need to make adjustments to the metadata file itself, go ahead and do so and re-upload it by beginning the process again with "Step 1: Metadata detection".

- Step 1: Metadata detection
- Step 2: Metadata mapping

[Process batch]

File:Portret van een vrouw Rijksmuseum SK-A-320.jpeg

Mediafile placeholder
This is where a representation of your mediafile will appear.

Summary

Title	Portret van een vrouw
Description	**Nederlands:** Portret van een vrouw. Vroeger geïdentificeerd als Elisabeth I (1533-1603), koningin van Engeland. Buste, iets naar links, met bloemen in het haar en een parelsnoer om de hals ;
Date	between 1550 and 1574
Medium	paneel; olieverf
Dimensions	drager "Height:" 44 cm (17.3 in)drager "Width:" 32.5 cm (12.8 in)
Institution	Rijksmuseum Amsterdam ✎ ▐▐▐ ▶
Accession number	SK-A-320
Object history	Acquisition: 1800
Source	http://www.rijksmuseum.nl/collectie/SK-A-320

BEFORE

- Can Commons categories be "deduced" from existing metadata/catalogues?
- Are there plans for community engagement? (e.g. village pump notice, blog post, WikiProject posts)
- Request user rights for GWToolset
- Request source website "whitelisting"
- Run tests (find the "beta cluster")

LOChh 3601-3800.xml (~/Downloads/HABS/backlog/)

```
423858    <credit>{{HABS-source}}</credit>.
423859    <record_number>3796-19530</record_number>.
423860    <source>http://lcweb2.loc.gov/master/pnp/habshaer/wv/wv0300/wv0371/photos/172032pu.tif</source>.
423861    <date>Documentation compiled after 1933</date>.
423862    <url_catalog>http://www.loc.gov/pictures/item/wv0371.photos.172032p</url_catalog>.
423863    <title>PERSPECTIVE VIEW (NORTH) OF EAST END - Storer College, Brackett Hall, Harpers Ferry, Jefferson County, WV</title>.
423864    <author></author>.
423865    <filename>PERSPECTIVE VIEW (NORTH) OF EAST END - Storer College, Brackett Hall, Harpers Ferry, Jefferson County, WV HABS
          WVA,19-HARF,32-D-3</filename>.
423866    <size>17.1mb</size>.
423867    </record>.
423868    <record>.
423869    <medium>5 x 7 in.</medium>.
423870    <place>West Virginia; Jefferson County; Harpers Ferry</place>.
423871    <repository>Library of Congress Prints and Photographs Division Washington, D.C. 20540 USA
          http://hdl.loc.gov/loc.pnp/pp.print</repository>.
423872    <latitudelongitude></latitudelongitude>.
423873    <permission>{{PD-USGov-NPS}}</permission>.
423874    <reproductionnumber>HABS WVA,19-HARF,32-E-1</reproductionnumber>.
423875    <notes>.
423876    * Significance: Constructed in 1939-40 and named after its principal benefactor, Permelia Eastman Cook Hall holds significance as the last
          new building constructed on the campus of Storer College, a historically black college in Harpers Ferry, West Virginia, which ceased
          operations in 1955.  This closure was, in part, an ironic and unfortunate outcome of the landmark Brown v. Board of Education decision.
          Storer College had been founded by northern Baptists in 1867 primarily f
          outset Storer situated its curriculum on liberal arts coursework, in t
          and women.  Such training for female students at Storer included cook
          for the few areas of paid work at the time considered appropriate fo
          funds and project oversight coming largely from women in Baptist mis
          Midwestern states, the building of Cook Hall is part of a tangled na
          education, and religion in 19th- and 20th-century America. Cook Hall
          stone used for its walls and the Colonial Revival architectural elem
          aspects of the design had precedent in other Storer College building
          by the National Park Service after a congressional appropriation in
          for an expansion of the Harpers Ferry National Monument (renamed Ha
          acquisition also included the establishment of an eastern training
          and Anthony Memorial Hall-the centerpiece of the Storer campus-were
          administration building for what was dedicated in 1964 as the 'Steph
          as a dormitory, it is still used as offices for the National Park Se
423877    * Unprocessed Field note material exists for this structure: N1213.
423878    * Survey number: HABS WV-277-E.
423879    * Building/structure dates: 1940 Initial Construction</notes>.
423880    <categoryarea>Harpers Ferry, West Virginia</categoryarea>.
423881    <callnumber>HABS WVA,19-HARF,32-E-1</callnumber>.
423882    <related>Cook, Permelia Eastman; Price, Virginia B, transmitter;
          project manager; Schara, Mark, field team; Davidson, Paul, field t
423883    <collections>Historic American Buildings Survey/Historic American
423884    <credit>{{HABS-source}}</credit>.
423885    <record_number>3796-19532</record_number>.
423886    <source>http://lcweb2.loc.gov/master/pnp/habshaer/wv/wv0300/wv0372
423887    <date>Documentation compiled after 1933</date>.
423888    <url_catalog>http://www.loc.gov/pictures/item/wv0372.photos.17255
423889    <title>PERSPECTIVE VIEW OF NORTH (FRONT) - Storer College, Cook Ha
423890    <author></author>.
423891    <filename>PERSPECTIVE VIEW OF NORTH (FRONT) - Storer College, Cook Ha
          WVA,19-HARF,32-E-1</filename>.
423892    <size>17.0mb</size>.
423893    </record>.
```

An xml file is a nicely laid out text file, with records that are highlighted in "tags" to separate the data.

These are often created automatically, but can be created or edited in most text editors, such as jEdit (a free open source tool)

See http://en.wikipedia.org/wiki/xml

GO

- Take your time, double check your mapping and additional fields
- Check your uploads were successful
- If something unexpected happened, the user email list is a good place to ask for help
  - http://lists.wikimedia.org/mailman/listinfo/glamtools

AFTER

- There should be a point of **contact** to discuss issues that arise, from metadata errors to copyright verification
- **Promote** the collection to volunteers and potential re-users
- **Plan** for periodically maintaining the collection

290,000 photographs and drawings of historic American buildings and places uploaded [edit]

144.9% — 289,782 files uploaded out of 200,000 (as of 10 August)
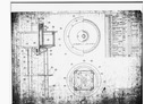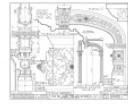
🐦 Retweet this announcement! ⧉

Examples

Collapsed "Million Dollar Bridge", Alaska

Stained glass detail, Winchester House, California

Drawing of mercury float pedestal and clock case, Block Island Southeast Lighthouse, Rhode Island

Drawings of architectural details, Dr. James Tillary House, NY, demolished in 1935

Top of flume, Electron Hydroelectric Project

Cotton mill interior, Alabama

Golden Gate Bridge, taken from the top of the south tower, 1984

I am happy to announce that my upload project of historic American buildings has completed today, perfectly timed the day before the Wikimania hackerthon ⧉! The upload has taken around 2 months of calendar time, with a major part of my time being spend "housekeeping", such as adding better geolocation tags, identifying National Register of Historic Places numbers and fixing odd category matches (such as "Mobile" being a place in the USA, rather than a phone!). This means we now have **fifty times** more images from the Historic American Buildings Survey on Commons and available to Wikimedia projects.

There is plenty to be done to improve the use and categorization of these images. For TIFFs that are too large to be rendered on Commons, I am slowly but systematically generating equivalent PNG files that are usable (please appreciate this means pulling the files one at a time down my home internet and processing the file on my ancient underpowered Macmini; it takes a while).

If you are coming to Wikimania, you may wish to join a session which explains how I uploaded this many images:

You can find me at Wikimania wearing a volunteer red shirt—no, I'm not in Star Trek security.

- **Thursday 7 August**, 14:00-15:45, ad-hoc demonstration and Q&A/workshop on mass uploading ⧉ (Frobisher 6)
- **Sunday 10 August**, 12:30-13:00 (may start 10 mins early), Using the Mass upload tool to copy GLAM collections to Commons ⧉
- Liam and I will be available at various times to support a wiki-lounge style advice space for the GLAMWiki toolset. Seek us out. 🙂

--Fæ (talk) 12:57, 5 August 2014 (UTC)

> That is awesome, Fæ! And what you call "housekeeping" (adding better geolocation tags, identifying National Register of Historic Places numbers and fixing odd category matches) is almost the most important thing... :-) --Atlasowa (talk) 07:18, 6 August 2014 (UTC)
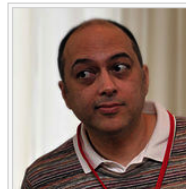>
> > It is a great and wonderful job that Fæ has done! I request to him to explain the uploading process as I am not able to join the Wikimania 2014. -- Biswarup Ganguly (talk) 08:49, 6 August 2014 (UTC)
> >
> > > If you go to mw:Help:Extension:GWToolset there have been some recent updates to the manual and there are some videos which give you an idea of how the tool works. I was in discussions today about how training materials are improving for the tool, and I'm sure that one of the benefits of talking through the tool, user experience and where the tool seems complex to use, will result both in improvement to the documentation and some good usability improvements. I probably underestimate how tricky some parts of this are, so I'm sure that the hackerthon workshop tomorrow will emphasise these barriers. Those at the hackerthon should sign up at https://wikimania2014.wikimedia.org/wiki/Mass_upload_training ⧉ (Thursday, 2pm in Frobisher 6) --Fæ (talk) 21:33, 6 August 2014 (UTC)
>
> Congrats Fæ, that is a great outcome. --99of9 (talk) 00:42, 7 August 2014 (UTC)

Improvement plan

- Training material
- Uploading collections not on the internet
- User friendly tool error reporting
- More metadata file types, so there may be no need to convert to xml
- A more formal User Group & Open GLAM programme
- Extend the toolset to reporting tools

File:Entrance - Bastrop State Park, Refectory, Bastrop, Bastrop County, TX HABS TX-3522-A-7.tif [edit]

From Wikimedia Commons, the free media repository

File	File history	File usage on Commons	Metadata

Size of this preview: 746 × 599 pixels. Other resolutions: 299 × 240 pixels | 598 × 480 pixels | 747 × 600 pixels | 956 × 768 pixels | 1,275 × 1,024 pixels.

Categories (+⁺): Bastrop, Texas (−) (±) (↓) (↑) | National Register of Historic Places in Bastrop, Texas (−) (±) (↓) (↑) | (+)

Hidden categories: Files from the Historic American Buildings Survey | National Register of Historic Places with known IDs | PD US NPS | GWToolset Batch Upload (−) (±) (↓) (↑) | Images uploaded by Fæ (−) (±) (↓) (↑)

Summary [edit]

Title	Entrance - Bastrop State Park, Refectory, Bastrop, Bastrop County, TX
Description	Civilian Conservation Corps (CCC); Works Progress Administration (WPA); National Park Service; Surte, Aparna, delineator; Benson, delineator; Campbell, Adrienne, delineator; Butman, A Elizabeth, delineator; Ruiz, Teri Asendorf, delineator; Emanuelli, Ely Merheb,
Depicted place	Texas; Bastrop County; Bastrop
Date	Documentation compiled after 1933
Dimensions	4 x 5 in.
Photographer	Related names: Civilian Conservation Corps (CCC) Works Progress Administration (WPA)
Institution	Library of Congress Prints and Photographs Division Washington, D.C. 20540 USA http://hdl.loc.gov/loc.pnp/pp.print 🔗
Accession number	HABS TX-3522-A-7
Credit line	This file comes from the Historic American Buildings Survey (HABS), Historic American Engineering Record (HAER) or Historic American Land Survey (HALS). These are programs of the National Park Service established for the purpose of documenting historic places. Records consist of measured drawings, archival photographs, and written reports. This tag does not indicate the copyright status of the attached work. A normal copyright tag is still required. See Commons:Licensing for more information. English | +/–
Notes	• 2003 Charles E. Peterson Prize, Fourth Place • Significance: The Refectory was one of the few structures in the park that was not designed by Fehr; its plan and design were approved before the National Park Service hired him. It is a load-bearing stone masonry building with an exterior treatment of random ashlar with rusticated finish and a wood shingle roof. The Refectory was constructed from 1934-36; later additions included the bathroom and kitchen wings, both wood-framed with "waney-edged" weatherboard siding. In addition, the rear terrace of flagstones was replaced by poured concrete and 3 of the doors on the northwest elevation were remodeled into recessed windows. The interior of the Refectory exhibits some of the crafts of Companies 1805 and 1811, including stone and wood carvings above the fireplaces. • Unprocessed Field note material exists for this structure: N890 • Survey number: HABS TX-3522-A • Building/structure dates: 1934-1936 Initial Construction
References	This is an image of a place or building that is listed on the National Register of Historic Places in the United States of America. Its reference number is 97001242. More images 🔗
Source	http://www.loc.gov/pictures/item/tx1036.photos.214531p 🔗
Permission (Reusing this file)	This image or media file contains material based on a work of a National Park Service employee, created as part of that person's official duties. As a work of the U.S. federal government, such work is in the **public domain**. See the NPS website 🔗 and NPS copyright policy 🔗 for more information. English | español | македонски | русский | Tiếng Việt | 中文 | +/–

Question:
Which is better, using Commons for information on an image, or going back to the source?

7. Entrance - Bastrop State Park, Refectory, Bastrop, Bastrop County, TX

About This Item	Obtaining Copies	Access to Original

View Larger

JPEG (54kb) | JPEG (139kb) | TIFF (22.9mb)

Item is from **this survey**

None
http://hdl.loc.gov/loc.pnp/hhh.tx1036/photos.214531p

Related

- Browse neighboring items by call number.
Collection: Historic American Buildings Survey/Historic American Engineering Record/Historic American Landscapes Survey

Title: 7. Entrance - Bastrop State Park, Refectory, Bastrop, Bastrop County, TX
Medium: 4 x 5 in.
Reproduction Number: HABS TX-3522-A-7
Rights Advisory: No known restrictions on images made by the U.S. Government; images copied from other sources may be restricted. (http://www.loc.gov/rr/print/res/114_habs.html)
Call Number: HABS TX-3522-A-7
Repository: Library of Congress Prints and Photographs Division Washington, D.C. 20540 USA http://hdl.loc.gov/loc.pnp/pp.print
Place:

- Texas -- Bastrop County -- Bastrop

Collections:

- Historic American Buildings Survey/Historic American Engineering Record/Historic American Landscapes Survey

Bookmark This Record:
http://www.loc.gov/pictures/item/tx1036.photos.214531p/

Rights assessment is your responsibility.

Watch out for my next big project on the history of medicine thanks to Wellcome Images and Wiki Project Medicine

Questions

- Fæ
  - http://commons.wikimedia.org/wiki/User_talk:Fae
  - faewik@gmail.com
- User email list
  - http://lists.wikimedia.org/mailman/listinfo/glamtools
- Project page
  - http://commons.wikimedia.org/wiki/Commons:GLAMwiki_Toolset