



Experimenting with IA

John Dewees, MLIS, CPL
OhioDIG – January 9, 2019
CC0 – Public Domain Dedication

Overview of Methods

- Anyone can contribute to the Internet Archive!
- Web Uploader
 - Documentation: <https://help.archive.org/hc/en-us/articles/360017788831-How-to-upload-files-to-create-a-new-item-page->
- FTP Transfer
 - Documentation: <https://archive.org/help/contrib-advanced.php>
- Uploading PDF copies of ebooks being discussed

Web Uploader

- Simple and intuitive interface, drag and drop files to be uploaded into one complex digital object
- Required metadata:
 - Page Title
 - Page URL (determines the URL slug)
 - Description (free text field)
 - Subject Tags
 - Collection (defaults to “Community Texts”)

Web Uploader, continued

- Supports Creative Commons licenses (including CC0) and Public Domain
- Creates one complex digital object from the upload
 - If uploading a collection of photographs, the entire collection will be the object with associated metadata
 - If individual images need to be described, items will have to be uploaded one at a time
- There is the ability to pre-set metadata fields in the web uploader to expedite process (though I haven't explored this)

FTP Transfer

- The simpler means of contributing to the Internet Archive because templates allowing metadata to be prepopulated
- IA documentation is robust for upload of ebooks via FTP, simple to follow with minimal technical experience
- FileZilla was FTP client used to initiate transfer, free and open source software

FTP Transfer, continued

- Transfer process is composed of:
- FTP Transfer of Book
 - PDF ebook
 - Metadata XML file (this template will be edited every time)
 - File Index XML file (this file will be the same every time, no need to edit)
 - Only for ebooks though! Audio/video/still images/etc. is a different story
- Requesting IA to ingest the content you transfer through a GET request
 - Just a URL you type into a web browser that includes the identifier for the ebook

Finding Metadata Examples

Treasures of East Toledo : a history of the schools, churches, houses, and other landmarks of the East side
by Michaels, Larry

Topics East Toledo (Toledo, Ohio) - History, Toledo (Ohio) - History
Collection opensource
Language English

A publication that gives the history of schools, churches, houses and other landmarks of the East Side area of Toledo, Ohio. The Toledo Public Schools that serve the area are named and described in detail, as well as the many churches in the area. Private homes of prominent and historical figures are described. Photos of most of the landmarks mentioned are included. Book scanned is a donation by the author.

Contact_information Please contact the Local History and Genealogy Department of the Toledo Lucas County Public Library for more information at digitization@toledolibrary.org
Identifier aalh_treasuresofeasttoledo_r97711mic
Identifier-ark ark:/13960/t9p349290
Ocr ABBYY FineReader 11.0 (Extended OCR)
Pages 174
Ppi 600
Publisher_digital Toledo Lucas County Public Library
Rights_description In copyright
Rights_uri <http://rightsstatements.org/vocab/InC/1.0/>

Reviews [+ Add Review](#)

There are no reviews yet. Be the first one to write a review.

7 Views

DOWNLOAD OPTIONS

ABBYY GZ	1 file
DAISY For print-disabled users	1 file
EPUB	1 file
FULL TEXT	1 file
ITEM TILE	1 file
KINDLE	1 file
PDF	1 file
SINGLE PAGE PROCESSED JP2 ZIP	1 file
TORRENT	1 file
SHOW ALL	10 Files 5 Original

IN COLLECTIONS

Click “SHOW ALL” to see all files associated with the record, including the metadata XML file

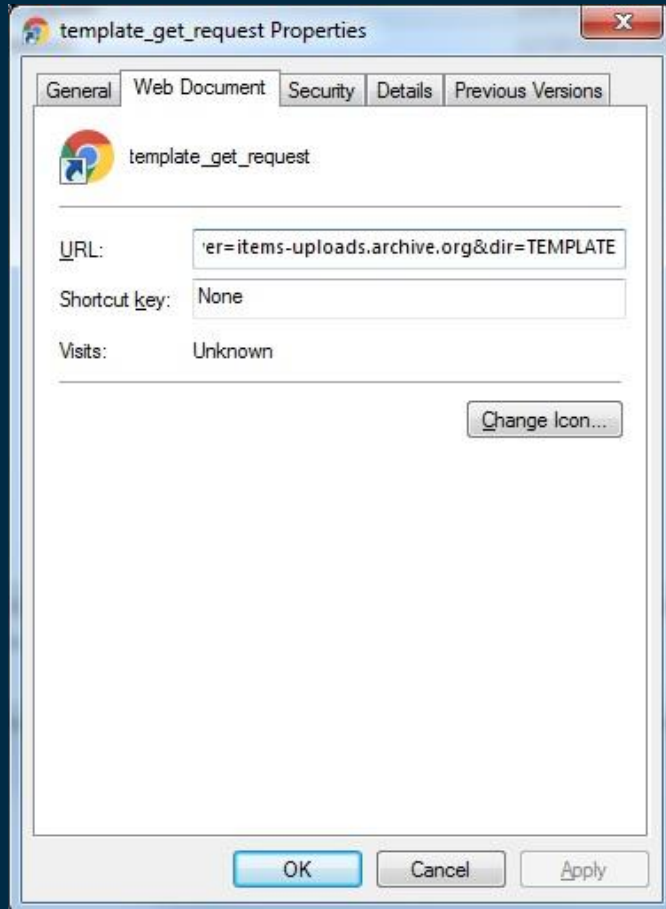
Finding Metadata Examples, continued

Files for aalh_treasuresofeasttoledo_r97711mic

Name	Last modified	Size
↑ Go to parent directory		
__ia_thumb.jpg	06-Dec-2018 15:48	14.3K
aalh_treasuresofeasttoledo_r97711mic.pdf	06-Dec-2018 14:58	288.1M
aalh_treasuresofeasttoledo_r97711mic_abbyy.gz	06-Dec-2018 15:47	3.9M
aalh_treasuresofeasttoledo_r97711mic_archive.torrent	06-Dec-2018 15:48	22.9K
aalh_treasuresofeasttoledo_r97711mic_djvu.txt	06-Dec-2018 15:48	261.4K
aalh_treasuresofeasttoledo_r97711mic_djvu.xml	06-Dec-2018 15:48	2.6M
aalh_treasuresofeasttoledo_r97711mic_files.xml	06-Dec-2018 15:48	3.4K
aalh_treasuresofeasttoledo_r97711mic_ip2.zip (View Contents)	06-Dec-2018 15:18	220.1M
aalh_treasuresofeasttoledo_r97711mic_meta.xml	06-Dec-2018 15:47	1.8K
aalh_treasuresofeasttoledo_r97711mic_scandata.xml	06-Dec-2018 15:48	51.1K

Upload one object with web uploader and edit the metadata to a point that you like, then download the ..._meta.xml file to create your template.

GET Request via URL Template



To create a template for the GET request, create a bookmark using the URL prefix supplied in IA's documentation, and then substitute the "TEMPLATE" at the end for the identifier of the item being uploaded. Click on the bookmark and it will execute the GET request.

https://archive.org/services/contrib-submit.php?user_email=digitization@toledolibrary.org&server=items-uploads.archive.org&dir=TEMPLATE

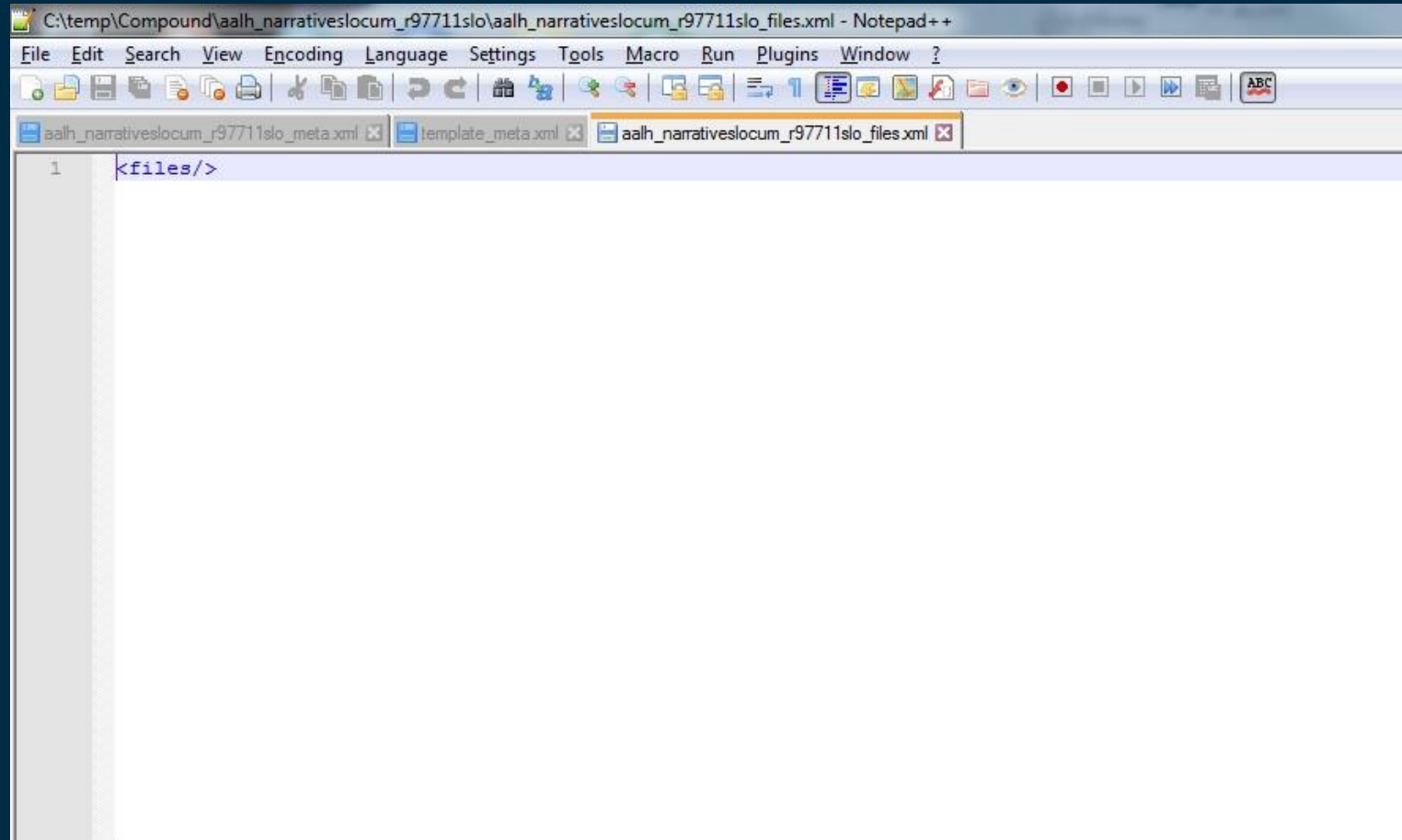
GET Request via URL, continued

This XML file does not appear to have any style information associated with it. The document tree is shown below.

```
<?xml version="1.0" encoding="UTF-8" ?>
<result type="success">
  <message>
    Queued for format conversion and move to download server.
  </message>
  <url>
    //archive.org/details/aalh_narrativeslocum_r97711slo
  </url>
</result>
```

TIP: Don't use ampersands (“&”) in XML Metadata. This will result in an error and the record will not be ingested.

template_files.xml Example



The screenshot shows a Notepad++ window with the following details:

- File name: C:\temp\Compound\aalh_narrativeslocum_r97711slo\aalh_narrativeslocum_r97711slo_files.xml
- Menu bar: File, Edit, Search, View, Encoding, Language, Settings, Tools, Macro, Run, Plugins, Window, ?
- Toolbar: Standard Notepad++ icons including Save, Undo, Redo, Copy, Paste, Find, etc.
- Tab bar: aalh_narrativeslocum_r97711slo_meta.xml, template_meta.xml, aalh_narrativeslocum_r97711slo_files.xml
- Text area: Line 1 contains the XML tag `<files/>`

This XML file will never change (so long as you are uploading ebooks)

template_meta.xml Example

```
C:\temp\template_meta.xml - Notepad++
File Edit Search View Encoding Language Settings Tools Macro Run Plugins Window ?
aalh_toleddfuturegreatcity_r977113sco_meta.xml x template_meta.xml x
1 <?xml version="1.0" encoding="UTF-8"?>
2 <metadata>
3   <mediatype>text</mediatype>
4   <collection>opensource</collection>
5   <creator></creator>
6   <creator></creator>
7   <description></description>
8   <language>eng</language>
9   <title></title>
10  <subject></subject>
11  <subject></subject>
12  <subject></subject>
13  <publisher_digital>Toledo Lucas County Public Library</publisher_digital>
14  <contact_information>Please contact the Local History and Genealogy Department of the Toledo L
15  <rights_uri>http://rightsstatements.org/vocab/NoC-US/1.0/</rights_uri>
16  <rights_description>No copyright - United States</rights_description>
17  <rights_uri>http://rightsstatements.org/vocab/InC/1.0/</rights_uri>
18  <rights_description>In copyright</rights_description>
19  <identifier></identifier>
20  <uploader>digitization@toledolibrary.org</uploader>
21 </metadata>
22
```

This XML file will change every time, but you can include presets such as rights statements URIs and descriptions, and multiple tags for subject or creator fields, and then delete what is unnecessary.