Spring 2019 POWER Library Training

PA Photos and Documents Intermediate Islandora / Hands-On Workshop

This project is made possible in part by Library Services and Technology Act (LSTA) funds from the U.S. Institute of Museum and Library Services and through Library Access Funds administered by the Office of Commonwealth Libraries, Department of Education, Commonwealth of Pennsylvania, Tom Wolf, Governor.
Learning Goals

At the end of this session you will be able to:

- **Intermediate Session:**
  - Prepare and name your digital files for inclusion in PA Photos and Documents digital collections
  - Upload multiple simple or compound digital objects to a PA Photos and Documents collection in Islandora using a spreadsheet and uploads to an SFTP server
  - Use a two-person workflow in Islandora to approve new uploads

- **Hands-on Practice Session:**
  - Prepare and name your digital files for inclusion in PA Photos and Documents digital collections
  - Upload simple or compound digital objects to a PA Photos and Documents collection using a basic description in Islandora, either individually or in batch via spreadsheet
  - Make changes to the description of the items in a PA Photos and Documents collection
  - Consider various approaches to planning digitization projects and cataloging using Islandora for PA Photos and Documents
Training Demo Files

Spreadsheet Uploads

• Organize your files for your project
  • Create folders on your computer for each project.
  
  • Filenames should be uniquely identified so that you can easily refer back to them
    • Anytown Public Library_Anniversary Party_001.tif
    • APLParty2017-10-01_0015.tif
    • NOT image001, file002, etc.

• Compound (Multipage) Objects should each have their own folder
  • Folder name is the title of the object
    • Anytown Public Library Board Meeting Minutes 1960-03-01
  • Pages should have the name followed by _Page N

• Have a backup strategy for your digital files
  • Ideally flash/hard drives or cloud-based storage, not recordable media
  • LOCKSS – “Lots of Copies Keep Stuff Safe”
Spreadsheet Uploads

• Required to upload your own materials:
  • SFTP software – free or pay (whatever your institution supports)
    • PC – Filezilla, PuTTY, WinSCP
    • Mac – CuteFTP, Cyberduck, Filezilla
  • Connection set up for the DEMO server with Training Credentials: (Not required for today’s training – test files are available.)
    • Protocol: sftp
    • Server: demo.hslc.dgicloud.com
    • Port: 22
    • Username: trainingmanager
    • PW: Provided with your handouts
    • Production server will have a different URL and dedicated credentials. Please contact HSLC Support if you wish to do spreadsheet uploads in production.
  • Institutions with strong firewall controls may need to allow connections to the SFTP server. Contact support@hslc.org if you cannot connect.
Spreadsheet Uploads

Example SFTP client software setup for Filezilla on Windows – each software is slightly different
Spreadsheet Uploads

- SFTP Server – connect with credentials provided
- Drag files from your local computer/drive onto the root of the /uploads folder
  - Generally left to right drag, depending upon software
  - Production – you can create subfolders if you wish
  - Demo [Training] – **please use the /uploads folder only**
- Remove your files when you are finished your spreadsheet loads
  - Generally right-click and delete, depending upon software
Spreadsheet Uploads

Local site: C:\Users\Diane\Downloads\Remote site: uploads

File Name: StoneBridgeRiverPostcardA_Front.jpg
Size: 8,138,629
Type: JPG File
Last Modified: 3/24/2019 12:33:33 AM
Spreadsheet Uploads

• Spreadsheet
  • Supported template sources: Excel, Numbers, Open Office, or Google Sheet
  • Must be able to save to CSV file
    • Excel—standard CSV and not UTF-8 or Macintosh CSV
  • Spreadsheet stays on local computer while digital files must be uploaded to remote server for processing (SFTP step)
  • Spreadsheet must have particular set of fields and structures
    • Must be exact! [Often takes troubleshooting which is why it is an intermediate/advanced level project]
Spreadsheet Uploads

• Key components:
  • label = human readable name to reference digital object
  • binary_file = exact name of digital file placed on SFTP server
    • Plus folder path if you are using a subfolder (not in Training)
  • Parent_object = what is the object which contains the file
    • = name of collection for a single object
    • = compound object parent for a child object
    • = always references the namespace and then the collection/object pid
      • Separates with a colon
      • E.g. islandora:training1 for Training 1 collection
      • E.g. islandora:training1_20190327_1 for a parent object
      • Production namespaces will be different [papd for main PaPD server]
Spreadsheet Uploads

• Key components:
  • parent_predicate = relationship of object to its container
    • isMemberOfCollection for single objects
      • Photos, posters, single audio, video etc.
    • isConstituentOf for child objects with a parent
      • Compound objects, audio & video with transcripts, postcards, etc.
  • For Books or Newspapers – please contact Support to discuss options
Spreadsheet Template – Images

• Step 1 – Prepare your files
• Step 2 – Connect to SFTP server
• Step 3 – Drag and drop your files to copy them from your local computer to the remote server

NB: The remainder of this exercise will use files already present on the server, so you don’t have to upload anything if you don’t wish to.
Spreadsheet Template – Images

• Step 4 – Use the provided example spreadsheet
  • Islandora_spreadsheet_photo_example.xlsx
• Edit the template spreadsheet with new values & add your name
  • Column A – label
    • Add – your name to the label [- Doreva]
  • Column B – binary_file
    • If you are using your own images, make sure the filename is EXACT
  • Column C – parent_object is the collection itself
    • Islandora:training# (training# collection) for training
  • Column D - cmodel
  • Column E – title
    • Add – your name to the title [- Doreva]
Spreadsheet Template – Images

• Step 4 – Edit the template spreadsheet with new values
  • Columns F through AR
    • Edit these with any changes your wish to make to the metadata, as relevant to your type of object(s)
  • Column AT- pid – change to reflect date and your training collection
    • For the parent of a compound object, you must enter a pid to connect to the child objects
    • Recommended: namespace:username_date_#incrementing
      • Eg.: islandora:trainingmanager1_20190328_5
    • Production namespaces will be different [papd and psa]
  • Column AU – predicate
    • isMemberofCollection for single objects
    • isConstituentOf for child objects with a parent
Spreadsheet Template – Images

• Step 5 – Save your spreadsheet as a CSV file
  • Excel – plain Comma Separated Value (CSV) – NOT CSV-UTF8 or Macintosh CSV

• Step 6 – Load your new CSV into Islandora
  • Demo: https://demo.hslc.dgicloud.com/islandora_spreadsheet_ingest
    • (Production URL will be different – contact HSLS for information if you wish to run spreadsheet loads)
  • Choose and Upload your CSV file
  • Choose Template = HSLC Spreadsheet MODS Template
  • Output DSID = MODS
  • Transform output= leave blank
Spreadsheet Template – Images

• Base Binaries Folder – should remain: /sftp/hslc_sftp/uploads
  • This is the default setting for the SFTP server location. *N.B. = The Production path will change.*
  • If you are using a subfolder of the SFTP drive you can add /yourfolder to the path
• CSV Delimiters = leave blank
• Contained paged content objects = leave blank for training
• Namespace: training = Islandora
  • Production: will be papd for main PA Photos and Documents (psa for State Archives)
• Ingest immediately = **uncheck this box**
Islandora Spreadsheet Ingest

CVL files are expected to follow the core rules defined in RFC 4180, with configurable defaults overriding some of the specifications.

If "Ingest immediately" is unchecked, the batch set ID and a link to the batch queue will be provided to you after preprocessing.

### Spreadsheet Ingest Parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>CSV File *</td>
<td>islandora_spreadsheet_photo_example.csv</td>
</tr>
<tr>
<td>Template *</td>
<td>HSCLC Spreadsheet MODS Template</td>
</tr>
<tr>
<td>Output DSID *</td>
<td>MODS</td>
</tr>
</tbody>
</table>

Select a CSV file to upload, delimited using the below-configured delimiters.

Select the template to apply the given CSV to.

ID of the datastream that generated metadata should be placed in.
Transform output
Transform the output datastream to another (ex: MODS to Dublin Core).

Base Binaries Folder
/sftp/hslc_sftp/uploads
Path to a folder on the local filesystem where binaries are being kept. Files under the object_location (formerly binary_file) column will be source contain an object_location column, or the object_location values are URLs, this path may be left blank. Files under object_location must be

CSV Delimiters

Contains paged content objects

Namespace
islandora

Ingested items will be given the next available PID in the given namespace. Leave blank to use the default configured namespace in Fedora.

Ingest immediately
Whether to process the batch immediately after preprocessing.

Batch Ingest
Spreadsheet Template – Images

• Click **Batch Ingest** to start the process.
  • You should see a green success message. If you don’t, something has gone wrong and you may need to check your CSV file. Please write down the error message and contact HSLC Support. ([support@hslc.org](mailto:support@hslc.org))
Spreadsheet Template – Images

• Step 7 – Check the batch first!
  • Click on the link to **Preprocessed Batch [Number]** in the green message.
  • You should see all your objects listed as **Ready to Ingest**.

<table>
<thead>
<tr>
<th>ID</th>
<th>STATE</th>
</tr>
</thead>
<tbody>
<tr>
<td>islandora.training1_20190330_1</td>
<td>Ready to ingest</td>
</tr>
<tr>
<td>islandora.training1_20190330_2</td>
<td>Ready to ingest</td>
</tr>
<tr>
<td>islandora.training1_20190330_3</td>
<td>Ready to ingest</td>
</tr>
<tr>
<td>islandora.training1_20190330_4</td>
<td>Ready to ingest</td>
</tr>
</tbody>
</table>
Spreadsheet Template – Images

- **Step 8 – Process the batch**
  - Click the link to **Process Set** and then **Start Batch Processing**.
  - You should receive a green success message. Please give it a few minutes for the objects to index.
  - If you receive an error message, please take note and contact Support.
Spreadsheet Template – Images

- Step 9 – Review your new digital objects!

<table>
<thead>
<tr>
<th>Title</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gettysburg Confederate Dead</td>
<td>Cadavers of confederate soldiers in the field in Gettysburg</td>
</tr>
<tr>
<td>Hanover Junction Pa.</td>
<td>Passengers waiting for a train that is pulled into the station.</td>
</tr>
<tr>
<td>Three Confederate Soldiers</td>
<td>Three confederate soldiers talking</td>
</tr>
</tbody>
</table>
Spreadsheet Template – Compound

• Same process with **some key differences** in spreadsheet:
  • Step 4 – Use the provided example spreadsheet
    • Islandora_spreadsheet_compound_example.xlsx
  • **Edit the template spreadsheet with new values & add your name**
    • Column A – label
      • Add – your name to the label [- Doreva]
    • Column B – binary_file
      • If you are using your own images, make sure the filename is EXACT
      • Empty for the compound object parent because it is being created
      • **Must match parent pid** for the child objects, or else you will create orphans
Spreadsheet Template – Compound

• Step 4 - Same process with some key differences in spreadsheet:
  • Column C: parent object
    • For the parent of a compound object, its parent is the namespace and pid of the collection to where it is being loaded:
      • Eg. islandora:training1
    • For the child objects, it is the pid of the parent object that is listed for all children
      • Eg. islandora:training1_20190331_15
    • If you assign an incorrect pid, the compound object “family” will not be created and the compound object will not look right
  • Column D – cmodel
    • Fore the compound object parent = islandora:compoundCModel
    • For the individual (basic) images (JPG) = islandora:sp_basic_image
Spreadsheet Template – Compound

• Same process with *some key differences* in spreadsheet:
  • Column E – title
    • Add – your name to the title [- Doreva]
  • Column AG – pid
    • For the parent of a compound object, you must enter a pid to connect to the child objects
    • Recommended: namespace:username_date_#incrementing
      • Eg.: islandora:trainingmanager1_20190328_5
    • For the child objects, it is best to leave this blank and let Islandora create them
    • Production namespaces will be different [papd for main PaPD server]
      • *Do not duplicate pids used – this will cause “unexpected errors” and failures*
  • Column AH – predicate
    • isMemberofCollection for the parent object
    • isConstituentOf for the child objects
Spreadsheet Template – Compound

• Step 7 – Check the batch!
  • Click on the link to **Preprocessed Batch [Number]** in the green message.
  • You should see all your objects listed as **Ready to Ingest**.

![Set 147 Batch Queue](image)

<table>
<thead>
<tr>
<th>ID</th>
<th>State</th>
<th>Message</th>
<th>Set ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>islandora:5736</td>
<td>Ready to ingest</td>
<td></td>
<td>147</td>
</tr>
<tr>
<td>islandora:5737</td>
<td>Ready to ingest</td>
<td></td>
<td>147</td>
</tr>
<tr>
<td>islandora:training1_20190331_01</td>
<td>Ready to ingest</td>
<td></td>
<td>147</td>
</tr>
</tbody>
</table>
Spreadsheet Template – Images

• Step 8 – Process the batch
  • Click the link to Process Set and then Start Batch Processing.
  • You should receive a green success message. Please give it a few minutes for the objects to index. Note any errors.
Spreadsheet Template – Compound

• Step 9 – Review your new digital objects!
Spreadsheet Load Troubleshooting

• Did you save your file as Comma Separated Values (CSV) only?
• Are there any extra data points to the right or below your data?
  - Clear contents
• Did you create a unique pid for any compound object parents?
  - Namespace:username_date_incrementing#abcdefg...
• Did you match the parent object to the parent’s pid to ”unify the family”?
• Did you assign the correct predicate in the last column?
• Take a deep breath! It will work out. Recheck your spreadsheet. Wait a few minutes and try again or contact Support for assistance.
Break Time!
Two-Step Upload Workflow

• CONTENTdm – separated the object creation from the administrative approval

• Islandora – can do the same thing if you need a 2-level workflow – potentially useful if you:
  • Supervise student workers
  • Have volunteers or interns
  • Have divided quality control responsibilities or procedures at your institutions
Two-Step Upload Workflow

• Basic User
  • Username: traininguser(#)
  • Credentials: provided to you in training
  • Has lowest level permissions to create objects
  • Objects will be inactive [non-viewable] upon creation

• Collection Manager
  • Username: trainingmanager(#)
  • Credentials: provided to you in training
  • Will have permissions to make objects active
Two-Step Upload Workflow

• Example:
  • Basic User uploads two simple objects for review
    • Follows procedures for creating 2 new basic images (photographs) in their training collection
    • Receives success message with note that the object is inactive
    • One will be approved and one will be rejected

• Manager will review inactive objects
  • Manager will publish one and leave the other unpublished (or delete it)
Two-Step Upload Workflow

**Hockey, 14**

*Hockey, 14* (ID: islandora:5730) has been ingested.

This object is not active. Metadata may not display correctly.
Two-Step Upload Workflow
Two-Step Upload Workflow

• Review:
  • Manager logs in with appropriate credentials
  • Inactive Objects List:
    • https://demo.hslc.dgicloud.com/admin/islandora/tools/simple_workflow
  • Manager reviews objects
    • Manage – Datastreams
    • Edit MODS datastream to view/edit metadata [Update]
    • Please do not edit other datastreams.
Two-Step Upload Workflow

- Review:
  - Manager checks and publishes approved objects
Hands-on time!

**PA Photos and Documents - Beginner**
- Login to Islandora using the Training credentials provided
- Upload 2 or 3 test simple objects (photos, single page text, single image, single audio or video) – use test files provided or your own
- Edit at least one simple object and make one change (e.g. subject, location covered, or description)

**PA Photos and Documents – Beginner Plus**
- Upload one compound object (postcard, multi-page text, book) - use test files provided or your own
Hands-on time!

PA Photos and Documents - Intermediate
• Login to Islandora using the Training credentials provided
• Use the sample spreadsheet to bulk load some provided test data to your test collection - edit it to reflect your name and training collection #
  • Simple images

PA Photos and Documents – Intermediate Plus
• Login to Islandora using the Training credentials provided
• Use the sample spreadsheet to bulk load some provided test data to your test collection – edit it to reflect your name, object pids, and training collection #
  • Compound Objects
Hands-on time!

**PA Photos and Documents - Intermediate**

- Login to Islandora using the Training credentials provided
- Use the traininguser(#) login to create some inactive simple objects in your test collection for review
- **In another web browser** – login to Islandora as the trainingmanager(#)
  - Go to the Inactive Object review queue:
    - [https://demo.hsic.dgicloud.com/admin/islandora/tools/simple_workflow](https://demo.hsic.dgicloud.com/admin/islandora/tools/simple_workflow)
  - Approve an object
  - Edit an object’s description [edit MODS datastream]
  - Review but do not approve an object
Hands-on time!

**PA Photos and Documents - Advanced**

- Login to Islandora DEMO using the Training credentials provided
- Login to Islandora DEMO SFTP server using the Training credentials provided
- Select multiple files of your own to upload
  - Prepare and upload your files to the Training SFTP server using Training credentials provided
- Prepare your spreadsheet according to template of type of files you are loading
  - General, Compound, Audio-Visual
  - [For Books or Newspapers talk to HSLC Support outside of training]
Consultation/Collaboration!

Digital Project Planning

• What awesome materials do you have to preserve in your own collections?
• What are your biggest challenges? Successes?
• Do you have a lot of objects to work with? Consider separating workflows.
  • Having staff or volunteers describe the scanned digital files using spreadsheets. Crowdsourcing!
  • Bulk uploading files using spreadsheets
• What collaborations and relationships are you forming with local partners?
  • Is there potential for shared volunteers, or shared funding for staffing or equipment?
  • How are you promoting your events?
• And many more questions...
Resources and Tools

• PA Photos and Documents
  • Documentation - https://www.powerlibrary.org/librarians/pa-photos-and-documents/support-resources-for-collection-managers/

• Support
  • POWER Library Technical Support Form - https://www.powerlibrary.org/librarians/librarian-support/
Resources and Tools

• Copyright and Rights Statements
  • Copyright and Fair Use Bibliography by HSLC
  • Copyright 101 by PA Digital
  • What is a Rights Statement? By PA Digital
  • Implementing RightsStatements.org by PA Digital

• Metadata Guidelines
  • PA Digital Metadata Guidelines

• Digitization Resources
  • Scan PA - https://www.powerlibrary.org/librarians/special-projects-office-of-commonwealth-libraries/digitization/scanpa/
  • Scribe Station - https://www.powerlibrary.org/librarians/special-projects-office-of-commonwealth-libraries/digitization/
Resources and Tools

• Image Editors
  • Irfanview (free) (PC) - https://www.irfanview.com/
  • Gimp (free) (PC/Mac) - https://www.gimp.org/

• File Renaming
  • Bulk File Renamer (free) - PC
    • http://www.bulkrenameutility.co.uk/Main_Intro.php
  • NameChanger (free) – MAC
    • https://mrrsoftware.com/namechanger/

• Optical Character Recognition (OCR)
  • Built in to Islandora for books – for full text searching
  • Often included with scanning software (ABBYfine)
Questions?
Thank you!

PA Photos and Documents
Intermediate Islandora / Hands-On Workshop

This project is made possible by a grant from the Institute of Museum and Library Services as administered by Pennsylvania Department of Education through the Office of Commonwealth Libraries, and the Commonwealth of Pennsylvania, Tom Wolf, Governor.