Ingesting Digitized Materials

For materials that have been previously accessioned, like scans received from a vendor.

Simple Ingest

1. Create a folder named for the collection ID in \Lincoln\Library\SPE_Processing\ingest
   a. Use Find-It to find the correct collection ID
   b. Examples:
      i. \Lincoln\Library\SPE_Processing\ingest\apap101
      ii. \Lincoln\Library\SPE_Processing\ingest\ua809
      iii. \Lincoln\Library\SPE_Processing\ingest\apap138
2. Log on to the railsdev Processing server
   a. Open a Command Line shell
   b. run ssh railsdev
3. Run: ingest <collection ID>
4. You can now type "exit" to log off the server and the command will run in the background
5. Check if an ingest is running: check ingest
6. Results will log to \Lincoln\Library\SPE_Processing\ingest\log\<collection ID> as: <timestamp>-ingest-<collection ID>.txt

Results of Ingest

1. Files will be packaged unto a SIP bag here: \Lincoln\Masters\Archives\SIP\<collection ID>\<package ID>
   a. SIP and AIP packages are here: https://github.com/UAlbanyArchives/packages
2. Processing folder for package is created in \Lincoln\Library\SPE_Processing\backlog\<collection ID>\<package ID>
3. Master files are placed in 'masters' subfolder
4. Example Processing package:
   • ua809_JxkK2VWVFu7F8VWaTe72BG
   • derivatives
   • masters
   • metadata

Examining Running Ingest

• When you run "ingest," you will see the PID or process ID:
  [1] 16994
• To list running ingest processes run: check ingest

$ check ingest
  gw234478 16994 12.9 0.2 206436 21184 pts/0   D   10:13   0:30 python3 /opt/lib/ingest-processing-workflow
  /ingest.py apap301

• if you need to stop the process (not recommended) use this command:
  sudo kill -9 <PID>

What is happening with "check"

• "check" is a function defined in /etc/profile.d/processingFunctions.sh that runs: ps aux | grep [i]ngest, etc.

If process is completed, but ingest folder being is not deleted

• Run python script \LINCOLN\Library\SPE_Processing\checkIngest.py
  • It does a compare between Ingest and Backlog to see all of the files were moved successfully.
• Check logs: \LINCOLN\Library\SPE_Processing\ingest\log
• An example of a log report where the ingest folder couldn't be deleted:
Advanced Use

**Ingesting from directory other than \Lincoln\Library\SPE_Processing\ingest**

1. Must use path accessible to the railsdev server
2. Must convert to Linux path:
   - `\Romeo\SPE\folder1\folder2` is `/media/SPE/folder1/folder2`
   - `\Lincoln\Masters\Special Collections\Electronic_Records_Library` is `/media/Masters/Special Collections/Electronic_Records_Library`
3. Run `ingest` using `-p` flag:

   ```
ingest apap101 -p */media/Masters/Special Collections/Electronic_Records_Library/apap101*  
```

...Will still log to \Lincoln\Library\SPE_Processing\ingest

**What is happening:**

- `ingest` is a function defined in: `/etc/profile.d/processingFunctions.sh`
- It runs a python script: `/opt/lib/ingest-processing-workflow/ingest.py`
  - It relies on packages:
    - `/opt/lib/ingest-processing-workflow/ingest.py`
    - `/opt/lib/ingest-processing-workflow/packages/SIP/`

  ```
  usage: ingest.py [-h] [-p PATH] [-a ACCESSION] ID
  
  positional arguments:
  ID Collection ID for the files you are packaging.
  
  optional arguments:
  -h, --help show this help message and exit
  -p PATH, --path PATH Path of files to ingest. Folder will be removed afterwards.
  -a ACCESSION, --accession ACCESSION Optional ArchivesSpace Accession ID for new acquisitions.
  ```

- Github repository for ingest-processing-workflow is here: https://github.com/UAlbanyArchives/ingest-processing-workflow
- The script is run under `nohup`, so it runs in the background

  ```
  nohup python3 /opt/lib/ingest-processing-workflow/ingest.py apap301 >> /media/SPE/ingest/apap301-ingest.log 2>&1
  ```

- No longer uses pyenv
  - Uses a pyenv called "ingest" by running: `pyenv activate ingest`
  - List all pyenv with: `pyenv versions`
Related articles

- Ingesting Digitized Materials
- Updating the website
- Subjects for bulk ingest
- Processing Ingested Digital Files
- asInventory Location Codes