DIP munger scripts

Last updated: May 7, 2025

The **dip-munger scripts** are custom command-line scripts that support upload of stored Archivematica DIP packages to AtoM. They were developed in 2018 and retired in 2025. This page documents their use during that period; the rationale for retirement is described in the last section below. For the new workflow, see the Upload DIPs to AtoM page.

For SFU Archives, there were two main use cases for the scripts (and its successor workflow):

1. **Digitization** of previously described analog materials that have existing AtoM descriptions and digital copies are ready for upload to AtoM.

2. **Access / copyright review** of materials previously ingested to Archivematica and described in AtoM and now cleared for online dissemination (no access or copyright restrictions).

Archivematica can upload digital objects to AtoM by sending them to their **parent** description record. But in cases like those described above, we want to send digital objects to an existing AtoM record itself. The dip-munger scripts did this through the following steps:

- Download Archivematica's **stored DIP** using the dip-retrieve script.
- Edit the csv file created by dip-retrieve to map DIP objects to AtoM url stubs.
- Upload the DIP with edited csv file to AtoM records (dip-upload for the digital object, dip-metadata for its metadata only).

The scripts were originally created by SFU's Alex Garnett in 2018. They were revised by Tessa Walsh (Artefactual Systems) in 2021 to work with the new security requirements of SFU Cloud, including the need to run the scripts with Multi-Factor Authentication (MFA). They were reviewed with minor revisions in 2024 following the Archives' migration of its SFU Cloud servers from RHEL7 to RHEL9.

Installation

The dip-munger scripts are compiled Python scripts that must be installed on your computer so that they can be easily run as line commands in Terminal. The scripts have a number of dependencies that can make installation difficult, especially in SFU's "managed Mac / PC" environment. In order to install the scripts, you must have **administrator** access to your machine.

Follow install instructions from the Artefactual Lab dip-mungers GitHub repository.

Notes

- You will need to install mysql and mysqlclient if they are not already installed on your computer.
- Your SFU user name must be added to the Archives' VM access control list in order to access the VMs on which AIPs and DIPs are stored.
- You must have API keys associated with your AtoM and Archivematica user accounts; you can view / generate keys from your user profile in both those applications.

Configuration

After installation, the configuration file is located at ~/.dip-mungers; open to edit defaults.

```
[GENERAL]
USERNAME = <<sfu_account_name>>
JUMP_SERVER_HOSTNAME = <<bastion_host_name>>
JUMP_SERVER_PORT = <<port>>

[STORAGE_SERVICE]
DEV_API_KEY = <<API_key_associated_with_dev_Storage_Service_account>>
DEV_URL = <<dev_Storage_Service_url>>
PROD_API_KEY = <<API_key_associated_with_production_Storage_Service_account>>
PROD_URL = <<production_Storage_Service_url>>

[ATOM]
DEV_API_KEY = <<API_key_associated_with_dev_AtoM_account>>
DEV_HOSTNAME = <<dev_AtoM_server_name>>
DEV_URL = <<dev_AtoM_url>>
PROD_API_KEY = <<API_key_associated_with_production_AtoM_account>>
PROD_HOSTNAME = <<production_AtoM_server_name>>
PROD_URL = <<production_AtoM_url>>
```

Note

- HOSTNAME must use the server name, not its url alias.

dip-retrieve

Run the dip-retrieve script to download a stored DIP along with an accompanying csv file.

```
$ dip-retrieve <<aip_uuid>>
```

The DIP will download to your Desktop folder.

- You need only specify the AIP uuid.
- Stored DIPs can also be downloaded from the Storage Service interface, but Storage Service will not create / download the required csv file.

Edit the csv file

Before running the upload scripts, you must edit the csv file included in the DIP package.

- The csv file will be included in the objects folder and it will have the same name as the parent DIP folder, e.g. ACN2018-025_Fellman_EmailAttachments.csv.

The csv file has two columns: **filename** and **slug**.

- The **filename** column (A) lists the names of the digital objects in the DIP; the names include the UUID that Archivematica assigned to each object.
- The **slug** column (B) will be blank; this is where you enter the AtoM slugs.

The AtoM slug is the last part of the AtoM url that follows **server_name.archives.sfu.ca/.**

- E.g. in cottonwood.archives.sfu.ca/f-10-4-0-0-0-1 the slug is f-10-4-0-0-0-1.
- In SFU AtoM, the slug is typically (but not always) identical to the unit's reference code, with the f- prefix in lowercase.

Find the AtoM slugs of the existing descriptions and enter them into column B of the csv file, matching the slug to the appropriate filename.

Be aware that the order of DIP objects in the filename column is not necessarily alphabetical.

Depending on how the original DIP objects were named, you may be able to use OpenRefine to extract slugs via calculation.

Save the csv file with changes and leave it in the same location.

dip-upload

Run the dip-upload script when you want to upload a copy of the DIP objects (not just metadata).

```
dip-upload <<dip_file_path>>
```

You will be prompted to enter your SFU computing password, an OTP code (get from your MFA app or device), and your password again.

- The two passwords reflects that you need to jump through SFU's bastion host and then to the AtoM server.

Any error (mismatch) in the **slug** column will cause the entire operation to quit.

- Correct the data on the csv file and re-run the upload script.

dip-metadata-upload

This script is similar to the dip-upload script, except that it sends only the minimal file metadata about the DIP objects rather than the full object itself.

- This is the equivalent of doing a bulk **metadata-only DIP upload**.

```
dip-metadata <<dip_file_path>>
```

Again, you will be prompted for password (twice) and OTP code.

The Archives rarely uses this script. The main use case is where you digitize previously described analog materials, have not yet cleared the records for access or copyright restrictions but you would like to send the digital object metadata to the existing Atom records to indicate to users that digital copies exist.

Links

The scripts were created by Alex Garnett, but re-written by Tessa Walsh from Artefactual Systems. Source code and documentation are available on the Artefactual Labs GitHub site at: https://github.com/artefactual-labs/dip-mungers.

2025 update: retirement of dip-munger scripts

In May 2025, the Archives decided to retire the dip-munger scripts.

- Scripts were developed by Alex Garnett based on similar work by the City of Vancouver Archives.
- Around 2021, the scripts needed to be rewritten to work with SFU's new security environment; Alex was no longer at SFU and Artefactual (Tessa Walsh) did the rewrite.
- In 2024 the Archives migrated its servers from RHEL7 to RHEL9 and the scripts needed to be updated / reinstalled; Tessa Walsh no longer worked at Artefactual.
- Working with Artefactual in 2025, the Archives was not able to get the scripts working in the new environment.

In the meantime, Artefactual had developed a standardized command-line tool (Digital object load task) to handle the same problem.

- It works in a broadly similar way to the SFU dip-mungers scripts, using a csv file to map digital objects to existing AtoM descriptions.

The dip-mungers scripts:

- Had to been installed on the archivist's computer, a task made complicated by SFU's managed Mac requirements.
- Used a config file to pass user-specific information allowing the scripts to connect to SFU servers to download / upload data.
- Had to be run from on-campus (on the SFU network).

The CLI tool:

- Does not need to be installed on a local computer.
- Can be run from off-campus (with the SFU VPN).
- Requires the archivist to manually create the csv file and separately upload the DIP objects + csv file to the AtoM server before running the upload command.

Given the difficulties in maintaining and running our own custom scripts, the Archives decided to adopt the standard AtoM CLI tool.

- The main downside is that the SFU security environment requires the archivist to first copy DIP objects + csv to the Archives' bastion host, then again to the AtoM server (two separate commands).
- The Archives is working (as of May 2025) with Artefactual to streamline this process so that only one command is needed.

- In the interim, the current two-step process is documented on the Upload DIPs to AtoM page.