LungMAP Ingest Solution Instructions

Below find instructions for how LungMAP Consortium members can share datasets within the consortium (**Interim Ingest Solution**) and formally submit datasets and metadata to the Data Coordination Center (**Long-Term Ingest Solution**). See instructions for each of these solutions below.

[Interim Ingest Solution Instructions](#)
[Long-Term Ingest Solution Instructions](#)

# Long-Term Ingest Solution Instructions

The long-term ingest solution has similar underlying infrastructure as the interim solution, but should accompany inclusion of detailed metadata. Such metadata need not be immediately added to a project, but can be supplied over a time until the user officially submits the project to the DCC (internal or external sharing).

Signing Up

To request an account (if your institutional login is not already recognized) send an email to:

- Michal.Kouril@cchmc.org, Stephanie.Loos@cchmc.org

First Steps
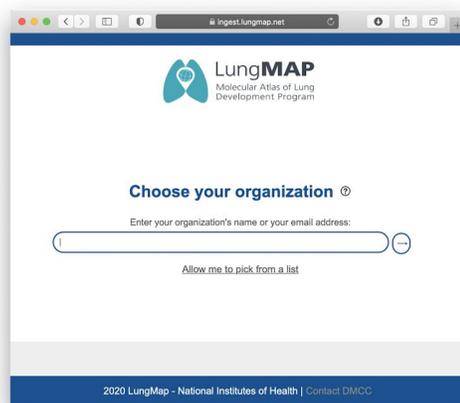
1. Go to: https://ingest.lungmap.net/
2. Select your institution or "Google"
3. Login

4. After login, your name should appear in the upper right hand corner of the interface.
5. To create a new project, select the "New Submission" button and then click the "details" button in the newly created entry.

Submissions

Submission	Owner	Created	Valid	Options
Bronchiolitis Obliterans Syndrome - 9 year old - 10x Genomics	Nathan Salomonis	2020-10-14 12:24 PM	false	details
Identification of the lymphangioleiomyomatosis cell and its uterine origin	Nathan Salomonis	2020-10-28 12:30 PM	true	details
Single cell transcriptomic profiling of pluripotent stem cell-derived SCGB3A2+ airway epithelium reveals fate plasticity *- 10x Chromium: bronchospheres vs. alveolospheres day 41*P	Nathan Salomonis	2020-11-06 08:57 AM	false	details
mTORC1 activation in lung mesenchyme drives sex- and age-dependent pulmonary structure and function decline	Michael Morley	2020-11-20 11:08 AM	false	details

New Submission

6. In the new window, select "Edit" to begin supplying project-level metadata.

Submission: Edit   Submit

Submission Details

Name:	New Submission Created Mon Nov 23 22:41:34 2020
Valid Forms:	**Contains Missing or Invalid Data** - The form data must be valid before submitting.
Owner:	Nathan Salomonis, Lungmap DCC
UUID:	3c35d1e7-7f36-4f78-bdf6-a5b15fec6d1a
Date Created:	2020-11-23 05:41 PM
Date Last Updated:	2020-11-23 05:41 PM

File Summaries

Type	Count	Total Size

Web Upload   CLI Upload

Authorized Users   Edit

Institution	Name	eMail	Role

7. Note, there are different tabs in the forms on the website, with required fields having a red asterisk next to them. When completed for each window select "Save" and proceed to the next tab to supply additional information. The interface may require you to remove additional fields not needed.

Project

Project | Contributors | Publications | Funders | Accessions | Supplementary data

Project title *

Single cell transcriptomic profiling of pluripotent stem cell-derived SCGB3A2+ airway epithelium reveals fate plasticity *- 10x Chromium: bronchospheres vs. alveolospheres day 41*P ✔

Project label *

PSC Airway Mouse ✔

Project description

Many lung diseases involve alterations in the cellular identity of the lung epithelium. Improved insight into airway development and the changes to cellular identity that result from abnormal signaling may therefore improve understanding of the etiology of these complex diseases. We have previously described a protocol to generate epithelial airway spheres from human pluripotent

Submission date *

2020-11-06T13:57:24.570795Z ✔

Submitter ID

Nathan.Salomonis@cchmc.org ✔

Document ID *

E9FBE758-DEEF-4D20-891B-D7B0010(✔

Accession

Update date

2020-11-06T14:28:25.320597Z ✔

Updater ID

Nathan.Salomonis@cchmc.org ✔

[Save]

8. Once completed, all principle metadata will be displayed on the project page as shown below. If you want to share the project with other people (e.g., clinical collaborator to add metadata), select the Edit button under "Authorized Users" and find their profile.

9. After saving, you can upload files through the GUI or CLI. After selecting which option of choice, see the below steps.

Uploading in the GUI

1. Select the option "Web Upload" (see above) to bring up S3 GUI
2. Upload option
   a. Drag and drop folders. **You can also create folders to organize the project data by sample, data-type, or experiment**.
   b. Select individual files (**NOTE: IT IS RECOMMENDED THAT YOU UPLOAD INDIVIDUAL FILES RATHER THAN ZIP FOLDERS (e.g., fastq.gz, h5, mtx, TIFF, ND5). THESE FILE EXTENSIONS WILL BE AUTOMATICALLY RECOGNIZED TO ALLOW ANNOTATION BY FILE TYPE**).
2. Once complete, return to the LungMAP ingest broker web page where you began. Here, you should see a list of the file types and total sizes in the same File Summaries menu.

3. Further down on the web page, you will find a section named "Files" that contains a list of all uploaded files. From here, you will be able to select the "View Files" option to begin adding file-level metadata as needed (IN DEVELOPMENT).

Accessing, uploading and sharing files through CLI

See additional info at AWS CLI:
https://docs.aws.amazon.com/cli/latest/userguide/cli-chap-install.html

1. First, ensure the AWS CLI is installed as outlined in the above link for your desired operating system (requires administrative access / support to install).
2. Open a terminal window from the location in which you want to upload files locally or remotely.
3. Copy and paste access variables obtained from the LungMAP provided S3 link in the webs interface (**AWS_ACCESS_KEY=**, …)
   (or setup in the config file)
4. list content: **aws s3 ls s3://lungmap-share/**
5. upload/sync content:
   aws s3 sync localfolder/ s3://lungmap-share/site-private/folder/
6. remove/move/copy: **aws s3 rm … , aws s3 mv …, aws s3 cp ... , …**

Interim Ingest Solution Instructions

The interim solution should be used when sharing files without associated metadata initially provided (e.g., project information, sample IDs). These include zip files, presentations or files to specifically share with groups within but not outside the LungMAP consortium. For projects with sample-level metadata, we recommend using the Long-Term Ingest Solution described below.

Signing Up

To request an account (if your institutional login is not already recognized) send an email to:

● Michal.Kouril@cchmc.org, Elaine.Wirrig@cchmc.org

First Steps

1. Go to the AWS URL: https://lungmap.awsapps.com/start/
2. Select your institution or "Google" (or use one of the highlighted selections)

3. After selecting your institution you'll be redirected to your institutional login page.
4. Following the successful login you'll see AWS Landing page and after choosing "AWS Account" and "LungMap Admin" you are presented with two links
   a. Link to the Management console (your site's storage area -- S3 bucket)
   b. Link to command line credentials
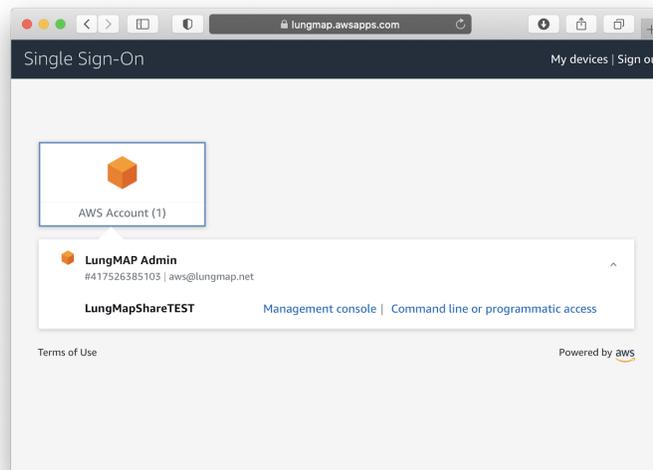
5. Within either option (UI-based sharing or CLI-based sharing) you'll have an option to select where you would like to share your data (within your institutional bucket only or with all other LungMAP Research Centers - not publicly). Next section describes UI-based data sharing and CLI-based data sharing section follows.

Uploading in the GUI

4. After logging in (as described in the previous section) clicking on Management Console link will bring up AWS S3 user interface
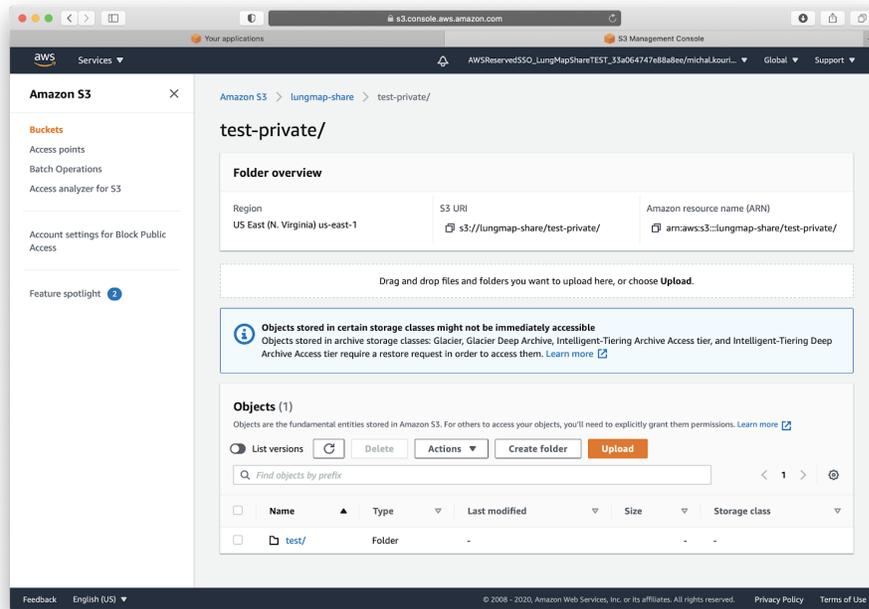
By default the console will open in your private space (not the one shared with other sites). At the top of the screen you see your current location (e.g. "Amazon S3 > lungmap-share > test-private/"). Clicking on these breadcrumbs allows the user to go to the root of the S3 bucket " "Amazon S3 > lungmap-share" and review files in other sites' folders.

5. Upload options
    a. The screen allows the user to drag and drop files or entire folders.
    b. When using the "Upload" button, users can select individual files to upload.

Users can manage their own private or shared folders. Functions such as create folder, move, copy and delete files and folders are available. Please note -- download is limited to individual files only. In order to download entire folders please use CLI as outlined in the next section.

Accessing, uploading and sharing files through CLI

In order to use AWS CLI please follow the login process outlined in the first section and copy the command line credential from the AWS LungMap Landing page by clicking on "Command line or programmatic access". You will be presented with three options for each of the two main platforms (Mac/Linux and Windows). Option 1 is typically the most convenient as it allows simply cut-and-paste the credentials into a script or a terminal window directly. Please note that these credentials do expire between 1 and 12 hours.
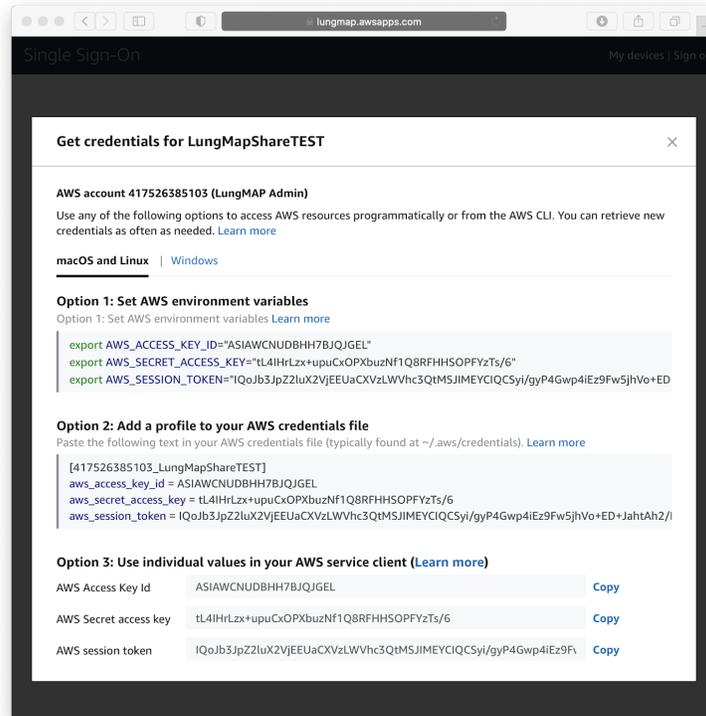
In order to use command line access please make sure AWS CLI is installed (for additional information please visit AWS CLI userguide: https://docs.aws.amazon.com/cli/latest/userguide/cli-chap-install.html)

1. First, ensure the AWS CLI is installed as outlined in the above link for your desired operating system (requires administrative access / support to install).
2. Open a terminal window from the location in which you want to upload files locally or remotely.
3. Copy and paste environmental variables obtained from the LungMAP provided S3 link in the webs interface (**AWS_ACCESS_KEY=**, …)
   (alternatively you can use option 2. and setup the config file -- please note the credentials will expire within hours)
4. With your credentials available in the environmental variables you may interact with the AWS S3 bucket. Examples
   a. List S3 bucket content:

      aws s3 ls s3://lungmap-share/

   b. List S3 bucket subfolder content (in this case CCHMC shared area):

      aws s3 ls s3://lungmap-share/cchmc-shared/

c. Upload/sync content to S3:

aws s3 sync localfolder/ s3://lungmap-share/cchmc-private/myproject/

d. Remove file

aws s3 rm s3://lungmap-share/cchmc-private/myproject/myfile.txt

e. Move

aws s3 mv s3://lungmap-share/cchmc-private/myproject/myfile.txt s3://lungmap-share/cchmc-private/myproject1/myfile.txt

For full command reference please see https://docs.aws.amazon.com/cli/latest/reference/s3/.

Reference:

1. Available top folders:
    a. List of private folders for each site (only the owner site has access -- read-write):

       cchmc-private, pnnl-private, rochester-private, ucsd-private and upenn-private

    b. List of shared folders for each site (all sites have read-only access, owner site has read-write access):

       cchmc-shared, pnnl-shared, rochester-shared, ucsd-shared and upenn-shared

    c. List of folders for each site that hold submitted data -- this is a temporary setup that will be deprecated (only the owner site has access -- read-only):

       cchmc-submitted, pnnl-submitted, rochester-submitted, ucsd-submitted and upenn-submitted