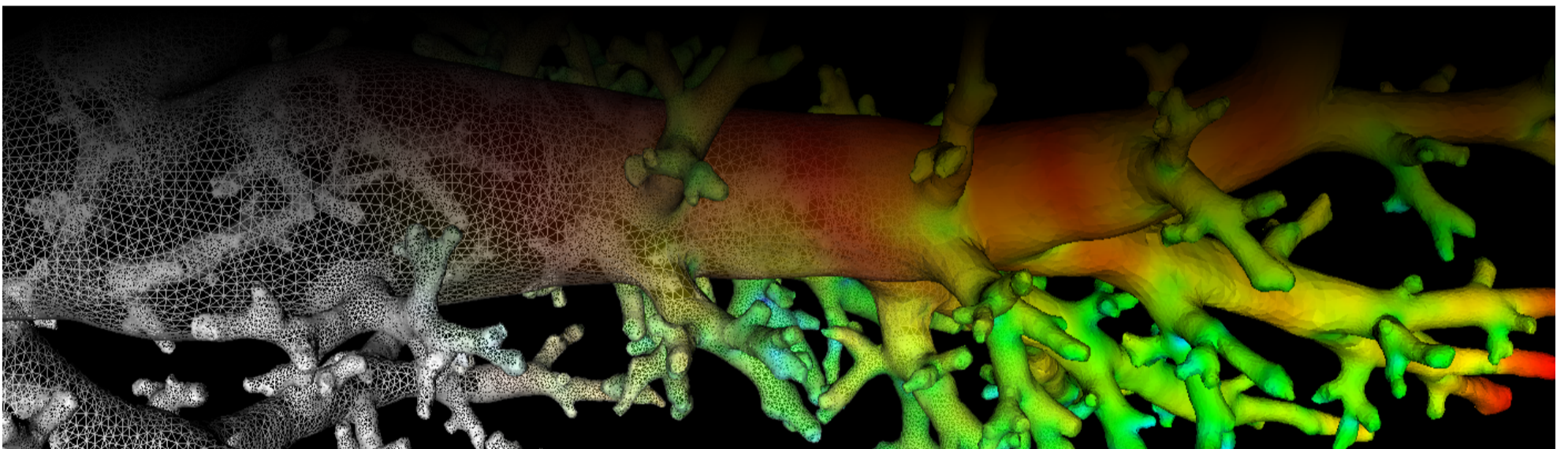


Lung Anatomy + Particle Deposition (lapd) Mouse Archive for Modeling and Computational Toxicology



lapdMouseDBUtil

lapdMouseDBUtil is a command line tool to systematically list, search, and download files from the lapdMouse archive. Using lapdMouseDBUtil you can for example download all `AirwayWallDeposition.vtk` files for all datasets in the archive using one command:

```
python lapdMouseDBUtil.py --pattern=/*AirwayWallDeposition.vtk --action=download
```

Getting started

lapdMouseDBUtil is a [Python](#) script. It does not require installation but it requires Python to be installed on the user's system.

Prerequisite - Python: If Python has not yet been installed on your system, go to <https://www.python.org/downloads>, download and install a version suitable for your operating system. lapdMouseDBUtil has been tested with Python 3.10.13 and 3.6.8 on Windows 10, Ubuntu 16.04, RHEL 8, and MacOS Sierra.

Download lapdMouseDBUtil: Download [lapdMousePythonExamples.zip](#) and unzip.

Test: Go to the folder containing `lapdMouseDBUtil.py` and call `python lapdMouseDBUtil.py`

Step-by-step usage guide

Using lapdMouseDBUtil to obtain files from the lapdMouse data archive involves the following steps:

- [List datasets](#)

- [List program options](#)
- [Select files by specifying file search pattern](#)
- [Specify local download folder](#)
- [Download data \(or update data\)](#)
- [Verify download status](#)

All steps will be explained in detail below.

List datasets

Calling `lapdMouseDBUtil.py` without any parameters lists all datasets in the lapdMouse data archive.

```
$ python lapdMouseDBUtil.py
DB access status: available
DB query: root=., depth=0, pattern=*
m01 -> ./m01 (folder)
m01 -> ./m01 (folder)
m03 -> ./m03 (folder)
...
Matching files/folders: total=16(0.0 B), downloaded=0, require download=16(0.0 B), require update=0
```

It provides in addition the following information:

- **Database access status**, whether the tool was able to connect to the lapdMouse database.
- The **file search pattern** utilized in querying the data archive.
- The **list of matching files/folders** in the data archive. Each line follows the pattern `nameInArchive -> localFileName (download status, size)`, where **download** status is one of:
 - `require download`: the file is available on the server but has not yet been downloaded
 - `downloaded`: the file has been downloaded and matches the version on the server
 - `require update`: the file has been downloaded but a newer version is available on the server
 - `folder`: folder, not a file
- A **summary** of files/folders in the data archive matching the query pattern organized by their download status and their total size.

List program options

To obtain a list of options supported by `lapdMouseDBUtil.py`, call

```
$ python lapdMouseDBUtil.py --help
usage: lapdMouseDBUtil.py [-h] [-p [PATTERN]] [-l LOCALDIR]
                        [-a {list,download,update,none}]
```

lapdMouse data archive access tool to list, search, and download files.

optional arguments:

```
-h, --help            show this help message and exit
-p [PATTERN], --pattern [PATTERN]
                        file pattern
-l LOCALDIR, --localDir LOCALDIR
                        local base directory for download
-a {list,download,update,none}, --action {list,download,update,none}
                        action to perform
```

lapdMouse project webpage including documentation and support:

This work was supported by NIH project R01ES023863.

Select files by specifying file search pattern

First, specify a file search pattern (`--pattern`) describing the files you are interested in. E.g. list all aerosol deposition images

`Aerosol*.mha` for dataset `m01`:

```
$ python lapdMouseDBUtil.py --pattern=m01/*Aerosol*.mha
DB access status: available
DB query: root=m01, depth=0, pattern=*Aerosol*.mha
m01/m01_Aerosol.mha -> ./m01/m01_Aerosol.mha (require download; 28.9 GB)
m01/m01_AerosolSub2.mha -> ./m01/m01_AerosolSub2.mha (require download; 3.6 GB)
m01/m01_AerosolSub4.mha -> ./m01/m01_AerosolSub4.mha (require download; 463.1 MB)
Matching files/folders: total=3(33.0 GB), downloaded=0, require download=3(33.0 GB), require update=0
```

Common search patterns to specify interesting files include:

- all datasets: `python lapdMouseDBUtil.py --pattern=*`
- all files for a specific dataset (m01): `python lapdMouseDBUtil.py --pattern=m01/*`
- all aerosol deposition images for a specific dataset (m01): `python lapdMouseDBUtil.py --pattern=m01/*Aerosol*.mha`
- all .vtk meshes for a specific dataset (m01): `python lapdMouseDBUtil.py --pattern=m01/*.vtk`
- all raw cryomicrotome data for a specific dataset (m01): `python lapdMouseDBUtil.py --pattern=m01/m01_RawCryomicrotomeData/*`
- raw cryomicrotome aerosol deposition image data for a specific dataset (m01): `python lapdMouseDBUtil.py --pattern=m01/m01_RawCryomicrotomeData/m01_RawCryoImages_rd.tar`
- Info.md files for all datasets: `python lapdMouseDBUtil.py --pattern=/*/Info.md`

Specify local download folder

After specifying the files you are interested in, select a local output folder where to download the files to (`--localDir`). The default output directory is the current working directory.

```
$ python lapdMouseDBUtil.py --pattern=m01/*Aerosol*.mha --localDir=/home/christian/data/lapdMouseDB
DB access status: available
DB query: root=m01, depth=0, pattern=*Aerosol*.mha
m01/m01_Aerosol.mha -> /home/christian/data/lapdMouseDB/m01/m01_Aerosol.mha (require download; 25.4 GB)
m01/m01_AerosolSub2.mha -> /home/christian/data/lapdMouseDB/m01/m01_AerosolSub2.mha (require download; 3.2 GB)
m01/m01_AerosolSub4.mha -> /home/christian/data/lapdMouseDB/m01/m01_AerosolSub4.mha (require download; 405.2 MB)
Matching files/folders: total=3(28.9 GB), downloaded=0, require download=3(28.9 GB), require update=0
```

This file selection contains 3 files with ~28.9GB of data listed as `require download`.

Download data (or update data)

Once the file selection has been completed and a local output folder specified, trigger the actual download (`--action=download`):

```
$ python lapdMouseDBUtil.py --pattern=m01/*Aerosol*.mha --localDir=/home/christian/data/lapdMouseDB --operation=download
DB access status: available
DB query: root=m01, depth=0, pattern=*Aerosol*.mha
m01/m01_Aerosol.mha -> /home/christian/data/lapdMouseDB/m01/m01_Aerosol.mha (require download; 25.4 GB)
  Downloading ...[DONE] time: 02:31(mm:ss)
m01/m01_AerosolSub2.mha -> /home/christian/data/lapdMouseDB/m01/m01_AerosolSub2.mha (require download; 3.2 GB)
  Downloading ...[DONE] time: 00:21(mm:ss)
m01/m01_AerosolSub4.mha -> /home/christian/data/lapdMouseDB/m01/m01_AerosolSub4.mha (require download; 405.2 MB)
  Downloading ...[DONE] time: 00:03(mm:ss)
Download statistics: files=3(28.9 GB), time=00:02:56(hh:mm:ss), speed=167.8 MB/second
```

The download of the requested 3 files with 28.9 GB took ~3 minutes. To update files that have been previously downloaded but have a newer version on the server, use `--action=update`.

Verify download status

After download, querying the data archive for the same files again, will list the files with their updated download status (`downloaded` instead of `require download`):

```
$ python lapdMouseDBUtil.py --pattern=m01/*Aerosol*.mha --localDir=/home/christian/data/lapdMouseDB
DB access status: available
DB query: root=m01, depth=0, pattern=*Aerosol*.mha
m01/m01_Aerosol.mha -> /home/christian/data/lapdMouseDB/m01/m01_Aerosol.mha (downloaded; 25.4 GB)
m01/m01_AerosolSub2.mha -> /home/christian/data/lapdMouseDB/m01/m01_AerosolSub2.mha (downloaded; 3.2 GB)
m01/m01_AerosolSub4.mha -> /home/christian/data/lapdMouseDB/m01/m01_AerosolSub4.mha (downloaded; 405.2 MB)
Matching files/folders: total=3(28.9 GB), downloaded=3(28.9 GB), require download=0, require update=0
```