ARECIBO OBSERVATORY
PUERTO RICO

Work Step Instruction

How to Download Data on TACC Example

Process Overview

This Work Step Instruction (WSI) will take you through the steps necessary to download data. It will go through the **Browsing**, **Requesting Data** and **Download** of the data. Browsing the data is where you explore what data is on TACC and decide what data you want to be made available for download. Requesting data is the stage in which you can request that data be made available for your download (this is done via email). Downloading is the stage in which you can download the data you requested once it has been made available for download. This example can help with understanding how to get data from TACC.

Requirements

A. Access to a Computer/Server.

B. Optional

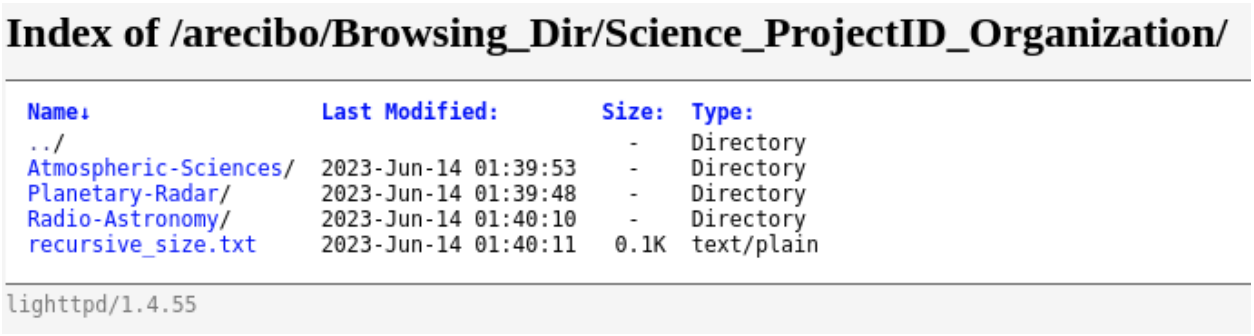
   a. Wget tool for linux terminal (pre-installed on most linux distributions)

   b. Wget tool for windows teminal

   c. Curl tool for MAC terminal (pre-installed on many MAC systems)

Example Browsing and Making Data List

A. Go to the following link:
   a. https://web.corral.tacc.utexas.edu/arecibo/Browsing_Dir/Science_ProjectID_Organization/
   b. Something similar to the photo below will come up.



Index of /arecibo/Browsing_Dir/Science_ProjectID_Organization/

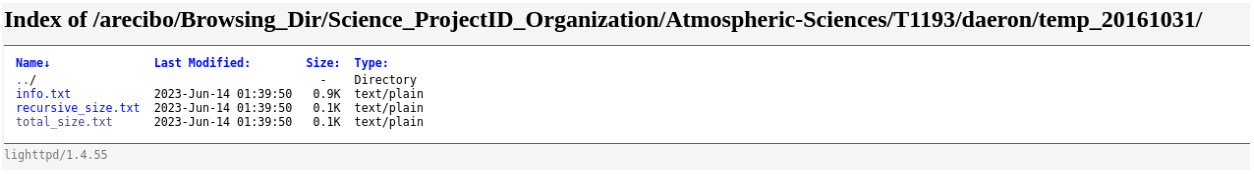
Name↓	Last Modified:	Size:	Type:
../		-	Directory
Atmospheric-Sciences/	2023-Jun-14 01:39:53	-	Directory
Planetary-Radar/	2023-Jun-14 01:39:48	-	Directory
Radio-Astronomy/	2023-Jun-14 01:40:10	-	Directory
recursive_size.txt	2023-Jun-14 01:40:11	0.1K	text/plain

lighttpd/1.4.55

   c. In this directory you can browse the available data in something similar to a directory structure.
   d. **Recursive_size.txt** is included in each directory and contains the estimate of space needed if that directory (and all its contents) were to be made available.
   e. **info.txt** files have the filenames and filesizes of the immediate files in the directory.
   f. **total_size.txt** is the s size of the files in the immediate directory (not recursive)
   g. **NOTE**: That directory contains an organization of the data, but could not include all the data as not all the data has been properly classified in that manner. Work is being done to correctly classify the data and add it to the browsing directory. There are other organizations that include all the data, but are not as easy to browse.

B. Choose the directories you want to stage
   a. For this example, we'll go to Atmospheric-Sciences -> T1193 -> daeron -> temp_20161031

b. It should show something like the picture below:



c. From the recursive_size.txt, I can see that 41.45 GB is the estimate of the space it occupies.

d. To pick this folder to be staged, we save the link of the directory we want staged. In this example, we'll pick two directories:

1. https://web.corral.tacc.utexas.edu/arecibo/Browsing_Dir/Science_ProjectID_Organization/Atmospheric-Sciences/T1193/daeron/temp_20161031/
2. https://web.corral.tacc.utexas.edu/arecibo/Browsing_Dir/Science_ProjectID_Organization/Radio-Astronomy/VLBI/rd08a/

    i. First link is atmospheric data, second is astronomy data.

    ii. From the recursive sizes, it can be seen that this data will take around 150GB of space.

    iii. **NOTE:** For the moment, only **directories** and their **full contents** can be staged, however feedback and suggestions can be sent to the contact below.

C. With these steps finished, we have a list of the directories we want data from, next stage would be to request the data.

Example Data Request

A. Requesting the data is as simple asopening a ticket with the TACC team.

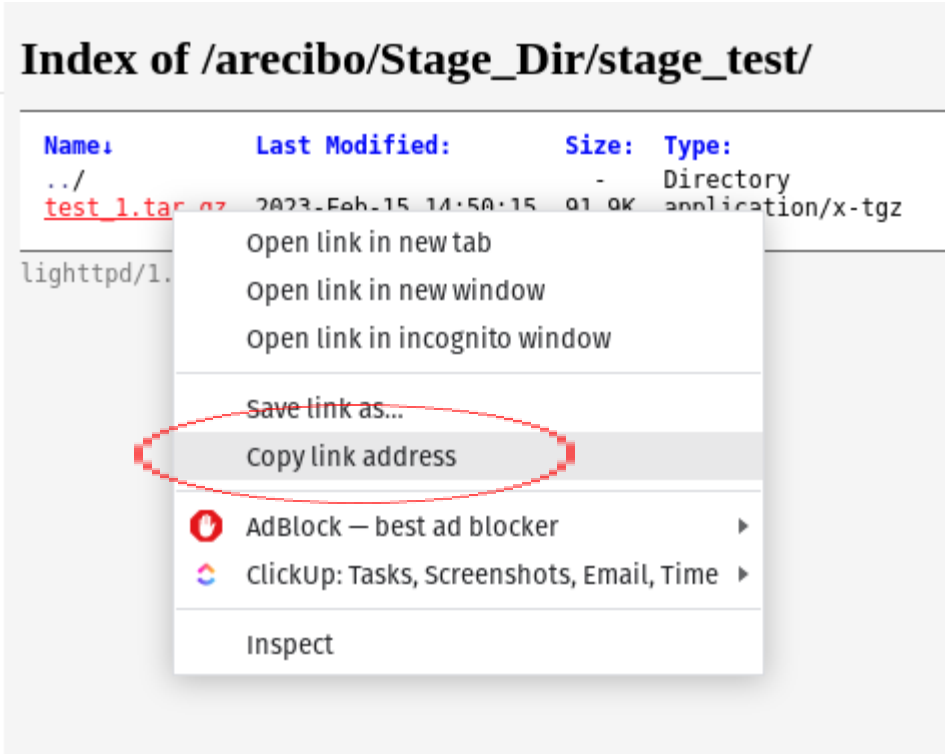
B. Navigate to:

a. https://www.tacc.utexas.edu/research/tacc-research/arecibo-observatory/

C. Click on Submit Ticket to Request Data
D. Fill in the information required
    a. For category, select Arecibo Data
E. In the message section provide the following:
    a. Provide your name
    b. Provide the affiliated institution
    c. The purpose in using the data
    d. The specific folder's paths or links as gotten in the previous section.
F. The TACC team will then walk you through their process to make the data accessible through the help ticket interactions.

Data Download

A. There are dirrefent ways of downloading the data.
    a. Through **browser** (The browser default download folder would have to be changed)
    b. Through terminal on MAC, Linux or Windows
    c. For browser instructions, see linked document
B. To download through terminal, wget or curl can be used.
C. Given the example of staged data found here, you can copy the link address by right clicking the tar archive and clicking on "Copy link address".



D. Wget terminal download

a. Open a terminal and type wget -c {copied_link_here}

```
~$ wget -c https://web.corral.tacc.utexas.edu/arecibo/Stage_Dir/stage_test/test_1.tar.gz
```

i. The -c option is to continue the download if it was interrupted

ii. **Optional**: You can limit the download rate using the **--limit-rate** option. Example:

```
:~$ wget -c https://web.corral.tacc.utexas.edu/arecibo/Stage_Dir/stage_test/test_1.tar.gz --limit-rate=10K
```

b. If the download was interrupted, you can re-run the command to resume the download.

E. Curl terminal download

a. Open a terminal and type curl -OC - {copied_link_here}

```
curl -OC - https://web.corral.tacc.utexas.edu/arecibo/Stage_Dir/stage_test/test_1.tar.gz
```

i. The -O option is to use the same filename as the source's link

ii. The -C option is to continue the download (if it was interrupted)

iii. The "-" character is so that the computer determines where you left off to start downloading. As opposed to being specified as a part of the command

iv. **Optional:** "**-- limit-rate**" can be added to the end to limit the download rate. G,M,K or B can be used to specify the unit. More info here.

1. Example:

```
curl -OC - https://web.corral.tacc.utexas.edu/arecibo/Stage_Dir/stage_test/test_1.tar.gz --limit-rate 100K
```

F. With these steps you know how to download data.

G. There's further documentation in this directory that goes into further detail of the data sharing system.