Long Term Archive Howto

This is a short manual on how to search for and retrieve data from the LOFAR Long Term Archive.

To access the LTA, go to: lta.lofar.eu

For background information and in case of problems, please refer to the Frequently Asked Questions page.

Release Notes

Release	Description
July 2018	1) All data is now searchable, not only released data or data of projects you're a member of. Note that for downloading data (staging) the proprietary restrictions still apply. 2) All projects can now be selected by all users, not only by project members. This provides the means to filter data based on project. 3) The user interface has been changed to improve user friendliness. Key points are a more intuitive menu structure, improved project selection/deselection and restyling. 4) Implemented cone search algorithm based on the Haversine formula for angular distance calculation. The calculated angular distance to the reference coordinates is now displayed in the search results. 5) Search keys added for “time resolution”, “Frequency resolution” and “nr. of subbands per subband group” 6) Search keys removed: “sky model database”, strategy name“

Public vs Proprietary data

User Access

I forgot my password

Please visit https://webportal.astron.nl/pwm/public/ForgottenPassword to request a new one.

Searching / retrieving data

The LTA catalogue can be searched directly without needing any account. Access to all projects and search queries will return results of the entire catalogue because metadata are public for all LTA content.

Staging and subsequent downloading of **public** data always requires an account with LTA user access privileges. This automatically happens if you were a member of the original project proposal in Northstar/MoM. If you do not have an account yet, you can register with ”Create account“. LTA access privileges will need to be granted by Science Operations and Support. Please, once you have created

your account, submit a support request to the Radio Observatory helpdesk asking for LTA privileges.

To stage and retrieve project-related data in the LTA which are **proprietary** you need to have an account in MoM that is enabled for the archive and coupled with the projects of interest. To this aim you can request Science Operations & Support to be added to the list of co-authors of the project. When you send such a request, you must add the project's PI in cc. After Science Operations and Support adds you to the project, you might get an email asking you to set a new password in ASTRON Web Applications Password Self Service. Please note that this will set a new password not just for the LTA *but for MoM (LOFAR/WSRT) and Northstar as well*.

Step-by-step guide to search and retrieve data

Basic search:

- log in to https://lta.lofar.eu/
- click SEARCH DATA in the top menu
- specify the data product types of interest and a target name or coordinated
- click on “Search” button at the bottom
- from the screen that follows, you should be able to stage the data products

Advanced search:

- log in to https://lta.lofar.eu/
- click SEARCH DATA in the top menu
- click on the side Advanced Search drop-down list
- specify the data product types of interest from the drop down list
- select products features and specify a target name or coordinated
- click on “Search” button at the bottom
- from the screen that follows, you should be able to stage the data products

Project search (to restrict all data searches to that project only):

- log in to https://lta.lofar.eu/
- click BROWSE PROJECTS in the top menu
- at this level membership can be checked, with the first column showing if you are a member of the project or for finding public projects. Available options are:
    - click on the project name to view the project details
    - use the 'search' button to select the project and go to the search page
    - use the 'show data' button to select the project and to show all data in it
- from the screen that follows, you should be able to either search / select / stage the data products

How to find data in the archive

Once your account is set up, or as anonymous user you can navigate the catalogue. In the former case you can login by clicking on the top right LOGIN button shown below.

Page navigation

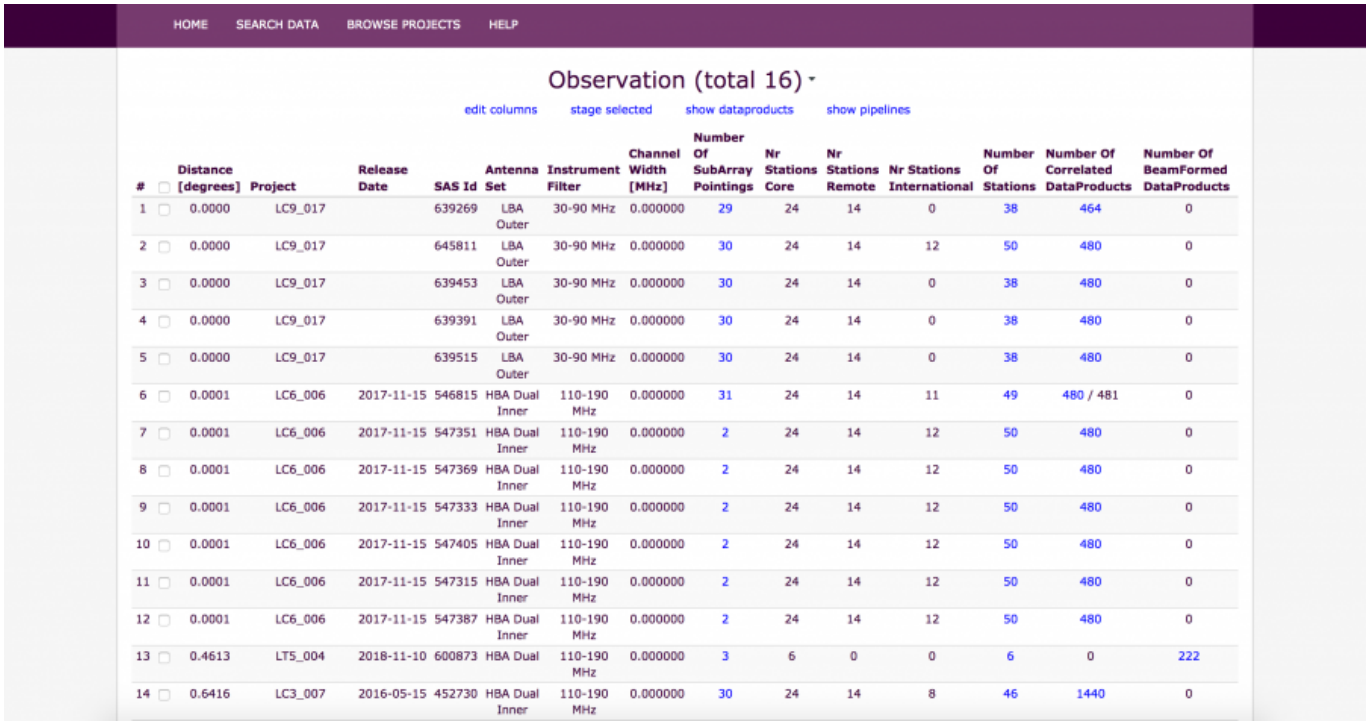
The LTA menu, as shown below, gives access to the main functionality.

A search in the LTA catalogue can be initiated by clicking on the SEARCH DATA button on the menu. At this point a default basic search is setup, where users can select the data product type of interest and perform a cone search. An advanced search mode, with more advanced parameters per data type, can also be selected by clicking on the drop menu on the left side.

A “project” can be selected by clicking on the BROWSE PROJECTS button on the menu. This is particularly useful for project related searches, as well as for checking either public projects or user's co-author membership. At this stage several actions are allowed

- Click on the project name to view the project details and eventually select it.
- Use the 'search' button to select the project and go to the search page. Subsequent searches will then be done within that project.
- Use the 'show data' button to select the project and to show all data in it.

Finding Data



#	Distance [degrees]	Project	Release Date	SAS Id	Antenna Set	Instrument Filter	Channel Width [MHz]	Number Of SubArray Pointings	Nr Stations Core	Nr Stations Remote	Nr Stations International	Number Of Stations	Number Of Correlated DataProducts	Number Of BeamFormed DataProducts
1	0.0000	LC9_017		639269	LBA Outer	30-90 MHz	0.000000	29	24	14	0	38	464	0
2	0.0000	LC9_017		645811	LBA Outer	30-90 MHz	0.000000	30	24	14	12	50	480	0
3	0.0000	LC9_017		639453	LBA Outer	30-90 MHz	0.000000	30	24	14	0	38	480	0
4	0.0000	LC9_017		639391	LBA Outer	30-90 MHz	0.000000	30	24	14	0	38	480	0
5	0.0000	LC9_017		639515	LBA Outer	30-90 MHz	0.000000	30	24	14	0	38	480	0
6	0.0001	LC6_006	2017-11-15	546815	HBA Dual Inner	110-190 MHz	0.000000	31	24	14	11	49	480 / 481	0
7	0.0001	LC6_006	2017-11-15	547351	HBA Dual Inner	110-190 MHz	0.000000	2	24	14	12	50	480	0
8	0.0001	LC6_006	2017-11-15	547369	HBA Dual Inner	110-190 MHz	0.000000	2	24	14	12	50	480	0
9	0.0001	LC6_006	2017-11-15	547333	HBA Dual Inner	110-190 MHz	0.000000	2	24	14	12	50	480	0
10	0.0001	LC6_006	2017-11-15	547405	HBA Dual Inner	110-190 MHz	0.000000	2	24	14	12	50	480	0
11	0.0001	LC6_006	2017-11-15	547315	HBA Dual Inner	110-190 MHz	0.000000	2	24	14	12	50	480	0
12	0.0001	LC6_006	2017-11-15	547387	HBA Dual Inner	110-190 MHz	0.000000	2	24	14	12	50	480	0
13	0.4613	LT5_004	2018-11-10	600873	HBA Dual	110-190 MHz	0.000000	3	6	0	0	6	0	222
14	0.6416	LC3_007	2016-05-15	452730	HBA Dual Inner	110-190 MHz	0.000000	30	24	14	8	46	1440	0

Depending on the search parameters, e.g., which data products were requested (observation, pipeline), lists of observations and/or pipelines will be returned (see observations example above). From this point there are several options:

1. select observations/pipelines and stage (prepare for download) all data related to the selection

2. select observations and “show pipelines” related to the observations, then select pipelines and stage
3. select observations/pipelines, “show dataproducts” related to the selection, possibly filter the data products (to have smaller selections) and then select and stage the data products

Note that observations often have no raw data in the archive, but the metadata is visible because subsequent pipelines have processed the raw data further. To get to the pipelines related to observations, use “Show Pipelines”.

To see whether observations or pipelines have data products in the LTA, look for the “Number of Correlated/BeamFormed DataProducts” column. These colums, as also a few others, can also be used to navigate to the relevant dataproducts.

Once you have a list of dataproducts on your screen, the “Release Date” will tell you when the data are available for public download. If data is public, or you are a member of the project you are looking in, the “checkbox” column will be selectable and staging can proceed. You can also hover with your mouse over the checkbox and get more information, like the size, location and checksums.

There is a separate page with **more detailed information** and advanced tricks to help you find and download your data. You will find examples of how to **browse the archive with your own python script**.

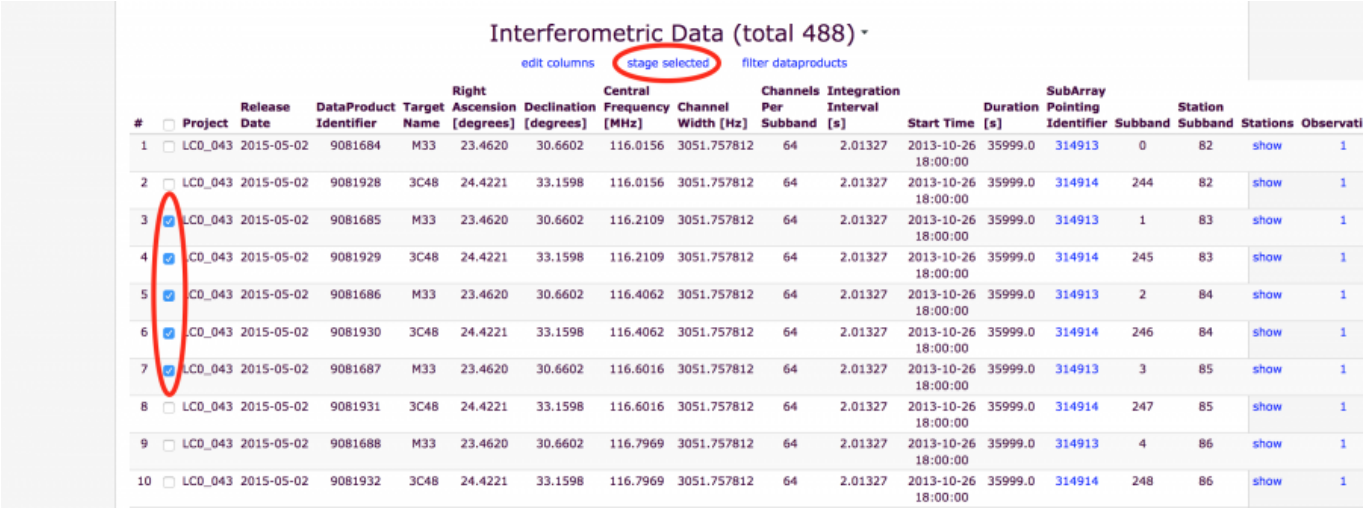
Unspecified Data/Process

Some data has had problems somewhere in the automation and control part of the LOFAR software during observation or processing. Sometimes a few subbands might be affected, sometimes an entire observation. Science Operations and Support will check the data, (re)run things manually or fix things if needed and then archive the data. This does mean that the automation and control sometimes loses track of the files and the archiving process has no information beyond the Observation ID and filename itself. In such cases a few subbands or an entire observation might end up under “Unspecified Process”. We do attempt to fix things at a later date, but that is not always feasible. If the files were archived, the data itself is usable. What is missing is the information that the LTA needs to properly label and query the data.

If an Observation is missing, or is missing subbands, please check if it ended up under Unspecified.

Staging data (Prepare for download)

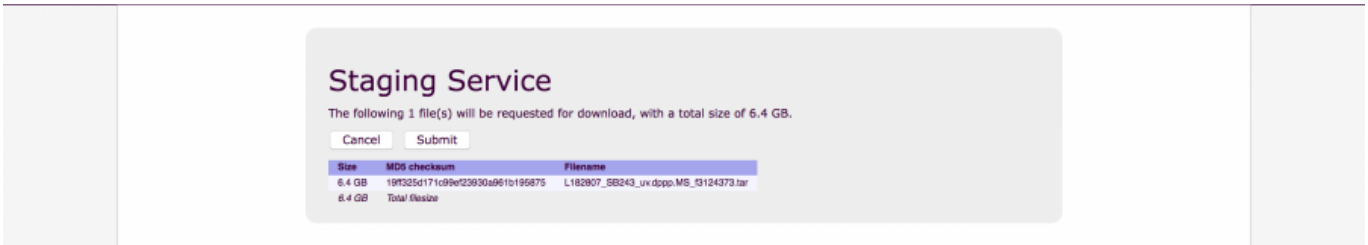
Once you have a list of dataproducts, observations or pipelines, you can use the check boxes to select which files you want to download. The first check box can be used to select or deselect all files or observations on a page.



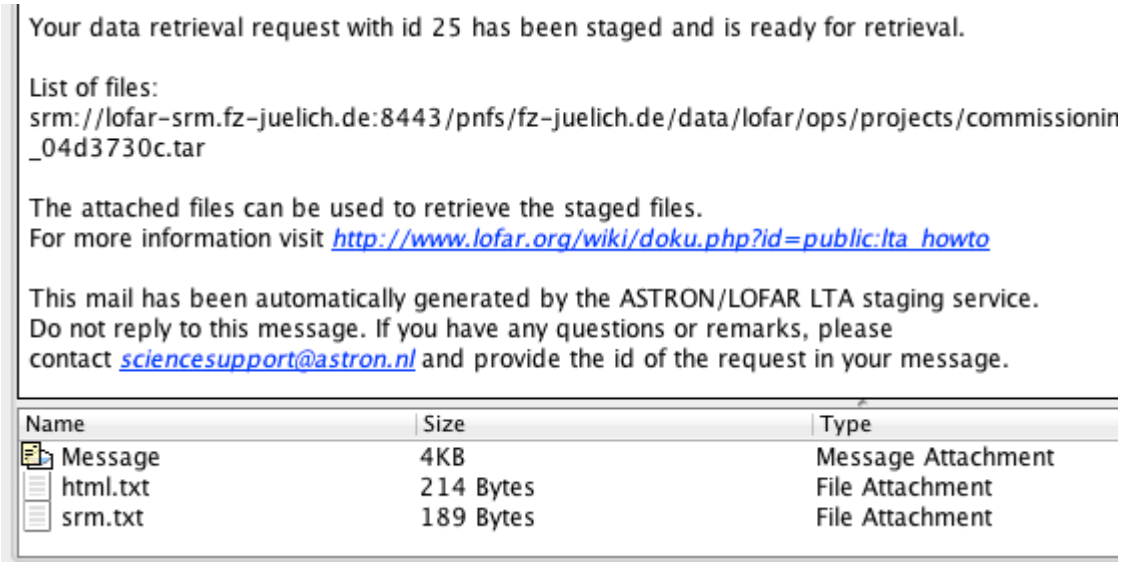
#	Project	Release Date	DataProduct Identifier	Target Name	Right Ascension [degrees]	Declination [degrees]	Central Frequency [MHz]	Channel Width [Hz]	Channels Per Subband	Integration Interval [s]	Start Time	Duration [s]	SubArray Pointing Identifier	Subband	Station Subband	Stations	Observati
1	LC0_043	2015-05-02	9081684	M33	23.4620	30.6602	116.0156	3051.757812	64	2.01327	2013-10-26 18:00:00	35999.0	314913	0	82	show	1
2	LC0_043	2015-05-02	9081928	3C48	24.4221	33.1598	116.0156	3051.757812	64	2.01327	2013-10-26 18:00:00	35999.0	314914	244	82	show	1
3	LC0_043	2015-05-02	9081685	M33	23.4620	30.6602	116.2109	3051.757812	64	2.01327	2013-10-26 18:00:00	35999.0	314913	1	83	show	1
4	LC0_043	2015-05-02	9081929	3C48	24.4221	33.1598	116.2109	3051.757812	64	2.01327	2013-10-26 18:00:00	35999.0	314914	245	83	show	1
5	LC0_043	2015-05-02	9081686	M33	23.4620	30.6602	116.4062	3051.757812	64	2.01327	2013-10-26 18:00:00	35999.0	314913	2	84	show	1
6	LC0_043	2015-05-02	9081930	3C48	24.4221	33.1598	116.4062	3051.757812	64	2.01327	2013-10-26 18:00:00	35999.0	314914	246	84	show	1
7	LC0_043	2015-05-02	9081687	M33	23.4620	30.6602	116.6016	3051.757812	64	2.01327	2013-10-26 18:00:00	35999.0	314913	3	85	show	1
8	LC0_043	2015-05-02	9081931	3C48	24.4221	33.1598	116.6016	3051.757812	64	2.01327	2013-10-26 18:00:00	35999.0	314914	247	85	show	1
9	LC0_043	2015-05-02	9081688	M33	23.4620	30.6602	116.7969	3051.757812	64	2.01327	2013-10-26 18:00:00	35999.0	314913	4	86	show	1
10	LC0_043	2015-05-02	9081932	3C48	24.4221	33.1598	116.7969	3051.757812	64	2.01327	2013-10-26 18:00:00	35999.0	314914	248	86	show	1

The LOFAR Archive stores data on magnetic tape. This means that it cannot be downloaded right away, but has to be copied from tape to disk first. This process is called 'staging'.

When you have made your selection of files, click on *stage*. This shows you the following message. It means that a request has been sent to the LTA staging service to start retrieving the requested files from the tape and make them available on disk. You will get a confirmation e-mail, to acknowledge that your staging request was received and the process was queued. When the files are staged, you will get a notification email informing you that your data are ready for retrieval.



The e-mail that you get when the staging on disk is complete gives you a list of files and has several attachments. Amongst them are two files `html.txt` and `srm.txt`:



There are two different ways to download your files with these attachments: http and srm

We also attach plain lists of the files/SURLs that were scheduled for staging (in the confirmation mail), those that were successfully staged, and (if any) those that could not be staged (in the success /

partial success notifications).

Please take note of the following

1. Unless you have an extremely fast connection (10 Gbit/s or more), **it is in general advisable to stage no more than 5 TB at a time** (see also point 4). At maximum efficiency a 1 Gbit/s connection will already take 12 hours to retrieve 5 TB of data, in practice it will often take quite a bit more.
2. On a 1 Gbit/s connection as a general rule of thumb, you should be able to retrieve data at about 100-500 GB/hour, especially if you try to retrieve 4-8 files concurrently. If you see speeds much lower than this, you might have some kind of network problem and should in general contact your IT staff.
3. Staging the data from tape to disk might take quite a bit of time. In the large data centres that the LTA uses, the tape drives are shared with all users and requests are queued. This is not just users of LOFAR but large data other projects like the LHC. This might mean that it takes anywhere from a few hours to a day or more to stage a copy of your data from tape to disk.
4. The amount of space available for staging data is limited although quite large. This space is however shared between all LOFAR LTA users. This includes LTA operations for buffering data from CEP to the LTA before it gets moved to tape. If many users are staging data at the same time, and/or LOFAR operations is transferring large amounts of data, the system might temporarily run low on disk space. You might then get a message that your request was only partially successful. In general the request will still finish 1-2 days later and we do monitor if requests don't get stuck and restart if needed.
5. We strive to keep a copy of data that was staged on disk for 1-2 weeks so you have some time to download it. After that it might get removed to make space for more recent requests. The copy of the data on tape is only read and will still be available if you need to access the data again at a later stage but you might need to stage a copy to disk again.
6. We are continuously trying to improve the reliability and speed of the available services. Please contact Science Operations and Support if you have any problems or suggestions for improvement.
7. The data centres the LTA uses also have maintenance or small outages sometimes. Science Operations and Support can advice you if this is the case and when it is planned to end, if you are having trouble accessing data. In general this will not be at the same dates as the LOFAR stop days.

Download data

You can download your requested data with the files from your e-mail notification. There are different possibilities and tools to do this. If you're unsure, which one to use, please refer to the according FAQ Answer.

HTTP download

If you open `html.txt` this file contains a list of http links that you can feed to a unix commandline tool like `wget` or `curl` or even use in a browser.

For wget you can use the following command line:

```
wget -i html.txt
```

This will download the files in `html.txt` to the current directory (option '-i' reads the urls from the specified file).

Preferrably, especially when downloading large files, you should also use option '-c'. This will continue unfinished earlier downloads instead of starting a fresh download of the whole file. (Make sure to first delete existing files that contain error messages instead of data, if you use this option):

```
wget -ci html.txt
```

Do not set the username and password on the wget command line because this allows other users on the system to view them in the process list. Instead you should create a file ~/.wgetrc with two lines according to the following example:

```
user=lofaruser
password=secret
```

Note: This is only an example, you have to edit the file and enter your own personal user name and password!

Set access authorizations of the .wgetrc file to user only so that the credentials are not exposed to anybody else, e.g.:

```
chmod 600 .wgetrc
```

There is no easy way to have wget rename the files as part of the command directly. It does not accept the -O flag inside a file it gets with -i. You can either rename files afterward, e.g. using the following command:

```
find . -name "SRMFifoGet*" | awk -F %2F '{system("mv "$0" "$NF)}'
```

or add the -O option to each line in html.txt but then feed each line to wget separately like this: cat `html.txt | xargs wget`. By default the html.txt file does not contain such options.

The following Python script will take care of renaming and untarring the downloaded files:

```
#M.C. Toribio
#toribio@astron.nl
#
#Script to untar data retrieved from the LTA by using wget
#It will DELETE the .tar file after extracting it.
#
#Notes:
#When using wget, the files are named, as an example:
#SRMFifoGet.py?surl=srm:%2F%2Fsrm.grid.sara.nl:8443%2Fpnfs%2Fgrid.sara.nl%2F
data%2Flofar%2Fops%2Fprojects%2Flofarschool%2F246403%2FL246403_SAP000_SB000_
uv.MS_7d4aa18f.tar
# This scripts will rename those files as the string after the last '%'
# If you want to change that behaviour, modify line
# outname=filename.split("%")[-1]
```

```
#
# Version:
# 2014/11/12: M.C. Toribio

import os
import glob

for filename in glob.glob("*SB*.tar*"):
  outname=filename.split("%")[-1]
  os.rename(filename, outname)
  os.system('tar -xvf '+outname)
  os.system('rm -r '+outname )

  print outname+' untarred.'
```

Another Python script for renaming the downloaded (and previously untarred) files. It removes the random part of the filename before the .tar extension:

```
import os
import sys
import glob


# AUTHOR: J.B.R. OONK  (ASTRON/LEIDEN UNIV. 2015)
# - changes LTA retrieval filename to standard filename
# - run in the directory where LTA files are located


# FILE DIRECTORY
path = "./"  #DIRECTORY

filelist = glob.glob(path+'*.tar')
print 'LIST:', filelist

#FILE STRING SEPARATORS
sp1d='%'
sp2d='2F'
extn='.MS'
extt='.tar'

#LOOP
print '#####  STARTING THE LOOP  #####'
for infile_orig in filelist:

  #GET FILE
  infiletar  = os.path.basename(infile_orig)
  infile     = infiletar
  print 'doing file: ', infile

  spl1=infile.split(sp1d)[11]
  spl2=spl1.split(sp2d)[1]
```

```
  spl3=spl2.split(extn)[0]
  newname = spl3+extn+extt

  # SPECIFY FILE MV COMMAND
  command='mv ' + infile + ' ' +newname
  print command

  # CARRY OUT FILENAME CHANGE !!!
  # - COMMENT FOR TESTING OUTPUT
  # - UNCOMMENT TO PERFORM FILE MV COMMAND
  #os.system(command)

  print 'finished rename of: ', newname
```

Note that wget does not overwrite existing files. If you use the continue option ('-c') it will append any missing parts to the existing file. If you don't use the continue option and there is a file present (e.g. from a stopped earlier download), wget creates a new file by appending a number (e.g., '.1') to the filename.

There are some small example links if you browse to https://lofar-download.grid.sara.nl/ where you can test with for example the file1M (which is 1 MB) if your setup is correct.

SRM download

If you open the file `srm.txt` this file contains a list of srm locations which you would feed to `srmcp`. SRM is a GRID specific protocol that is currently supported for data at the SARA and Jülich locations. It is faster, especially if you have significantly more than 1 Gbit/s bandwidth. It requires a valid GRID certificate and installation of the GRID srm software. NB There is an alternative installation that does not require root privileges. Contact Science Operations and Support if you think you might need a GRID account but it can not be provided by your own institute. An example command line would be:

```
srmcp -server_mode=passive -copyjobfile=srm.txt
```

to retrieve all requested files contained in srm.txt or e.g.

```
srmcp -server_mode=passive srm://lofar-srm.juelich.de:8443/pnfs/fz-
jeulich.de/data/lofar/ops/projects/commissioning2012/file.tar
file:////data/files/file.tar
```

to retrieve a single file. You need `–server_mode=passive` if you are behind a firewall or on an internal network. Omitting this option may result in improved transfer speed as it will attempt to use multiple streams when retrieving a file. An alternative strategy to improve the overall transfer speed is to run multiple srmcp requests in parallel, e.g. by splitting the provided srm.txt file and feeding the partial lists to separate srmcp commands.

If you do experience insufficient transfer speeds with srmcp, you may want to look into using srmcp with a globus-url-copy copy script.

Troubleshooting

- There is a LTA FAQ page that should help with the common difficulties.

From:
https://www.astron.nl/lofarwiki/ - **LOFAR *Wiki***

Permanent link:
https://www.astron.nl/lofarwiki/doku.php?id=public:lta_new

Last update: **2018-07-10 15:11**