

LOFAR DATA RETRIEVAL TUTORIAL II

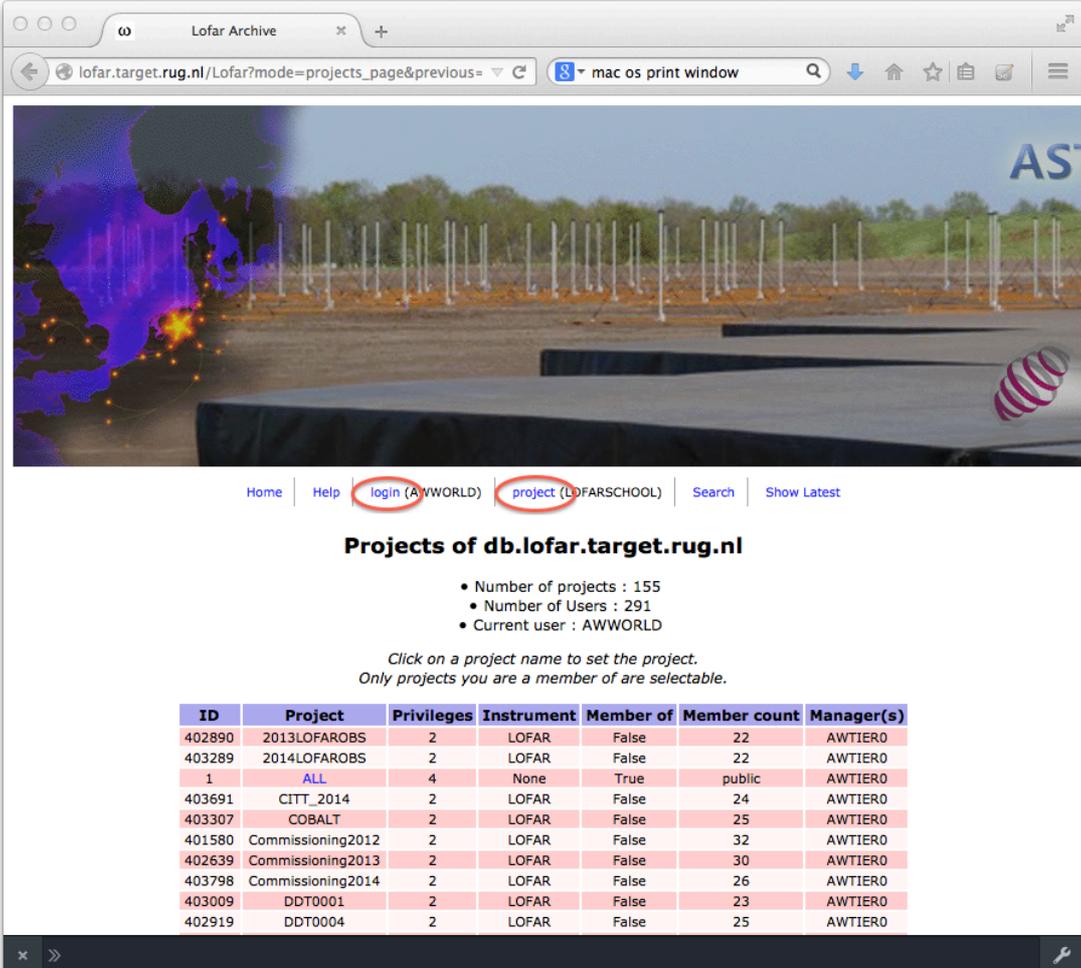
With your LOFAR webservices account and, optionally, your personal grid certificate set up as described in part I of this tutorial, we will now look at retrieving LOFAR data from the LTA.

For this tutorial you will be downloading a small ‘public’ dataset from the LTA to the CEP3 processing cluster. Two approaches will be taken: using the HTTP based download service and using the Grid client toolset.

NB The LTA user interface is under development and several changes to the query interfaces are to be expected as the LOFAR LTA is being prepared for public availability early 2015.

1) Find data and prepare for retrieval

1. Go to <http://lofar.target.rug.nl> and log in with your LOFAR web services credentials.



Projects of db.lofar.target.rug.nl

- Number of projects : 155
- Number of Users : 291
- Current user : AWWORLD

Click on a project name to set the project.
Only projects you are a member of are selectable.

ID	Project	Privileges	Instrument	Member of	Member count	Manager(s)
402890	2013LOFAROBS	2	LOFAR	False	22	AWTIERO
403289	2014LOFAROBS	2	LOFAR	False	22	AWTIERO
1	ALL	4	None	True	public	AWTIERO
403691	CITT_2014	2	LOFAR	False	24	AWTIERO
403307	COBALT	2	LOFAR	False	25	AWTIERO
401580	Commissioning2012	2	LOFAR	False	32	AWTIERO
402639	Commissioning2013	2	LOFAR	False	30	AWTIERO
403798	Commissioning2014	2	LOFAR	False	26	AWTIERO
403009	DDT0001	2	LOFAR	False	23	AWTIERO
402919	DDT0004	2	LOFAR	False	25	AWTIERO

2. Select the project “LOFARSCHOOL”: Click on the menu item “project” and next click on the project “LOFARSCHOOL” (scroll down if needed).

ID	Project	Privileges	Instrument	Member of	Member count	Manager(s)
403614	LC2_027	2	LOFAR	False	25	AWTIERO
403616	LC2_028	2	LOFAR	False	25	AWTIERO
403618	LC2_029	2	LOFAR	False	27	AWTIERO
403620	LC2_030	2	LOFAR	False	26	AWTIERO
403622	LC2_031	2	LOFAR	False	26	AWTIERO
403624	LC2_032	2	LOFAR	False	25	AWTIERO
403626	LC2_033	2	LOFAR	False	26	AWTIERO
403630	LC2_034	2	LOFAR	False	26	AWTIERO
403632	LC2_035	2	LOFAR	False	26	AWTIERO
403634	LC2_036	2	LOFAR	False	26	AWTIERO
403636	LC2_037	2	LOFAR	False	24	AWTIERO
403638	LC2_038	2	LOFAR	False	32	AWTIERO
403640	LC2_039	2	LOFAR	False	26	AWTIERO
403643	LC2_040	2	LOFAR	False	26	AWTIERO
403645	LC2_041	2	LOFAR	False	25	AWTIERO
403651	LC2_042	2	LOFAR	False	26	AWTIERO
403843	LC3_001	2	LOFAR	False	35	AWTIERO
403833	LC3_002	2	LOFAR	False	26	AWTIERO
403837	LC3_014	2	LOFAR	False	25	AWTIERO
403845	LC3_030	2	LOFAR	False	24	AWTIERO
401573	LEA032	2	LOFAR	False	25	AWTIERO
401591	LEA066	2	LOFAR	False	29	AWTIERO
403826	LOFARSCHOOL	4	LOFAR	True	public	AWTIERO
403655	LT2_001	2	LOFAR	False	28	AWTIERO
403657	LT2_002	2	LOFAR	False	25	AWTIERO
403659	LT2_003	2	LOFAR	False	26	AWTIERO
401551	MSSS	2	LOFAR	False	28	AWTIERO
402641	MSSS_HBA_2013	2	LOFAR	False	49	AWTIERO
403485	MSSS_LBA_2014	2	LOFAR	False	23	AWTIERO
403312	Pulsars	2	LOFAR	False	32	AWTIERO
402506	Pulsars2	2	LOFAR	False	30	AWTIERO
402514	Solar	2	LOFAR	False	23	AWTIERO
401563	test-lofar	2	LOFAR	False	27	AWTIERO

powered by

There are several approaches you can take for finding data you are interested in. We will look at some of them.

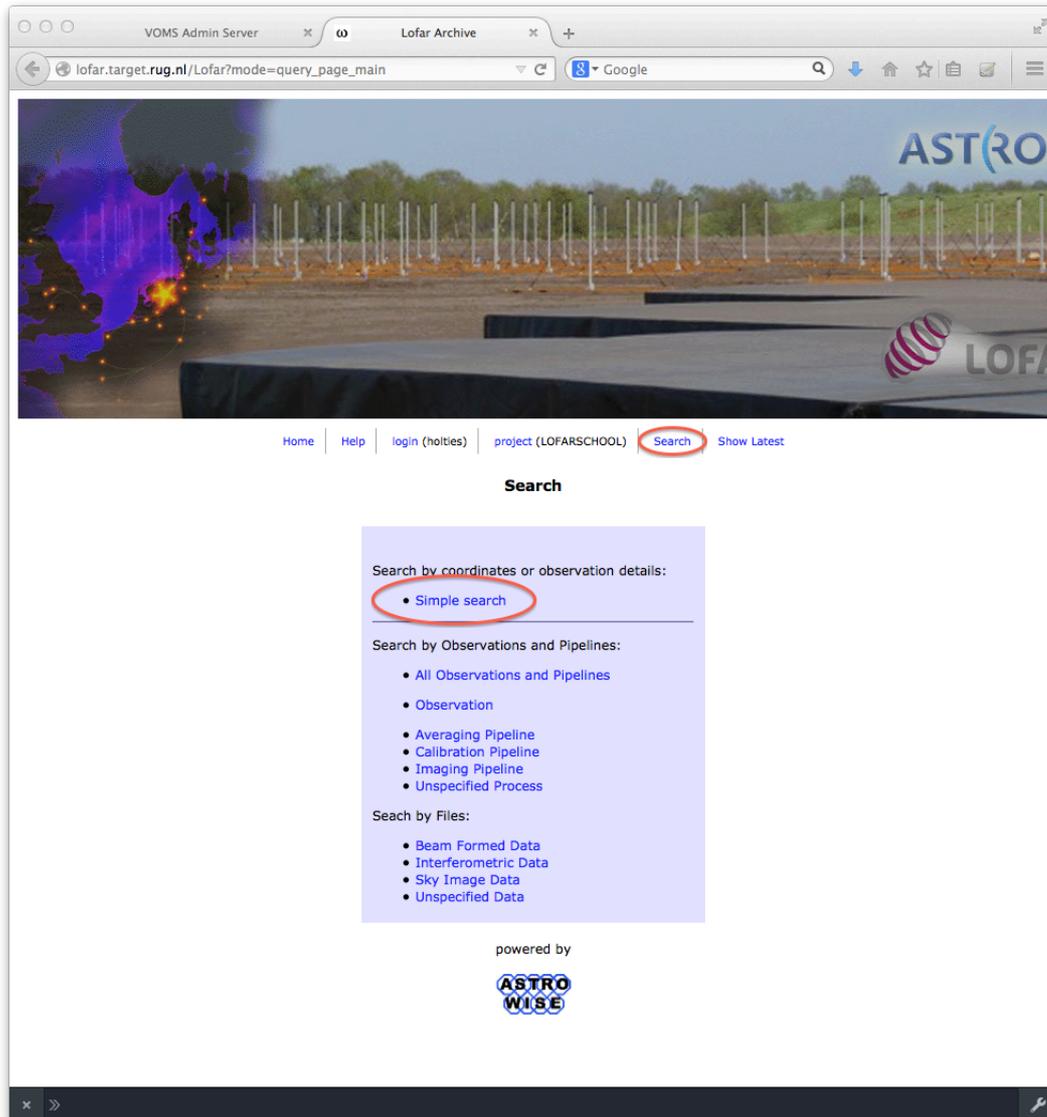
1a) Searching

There are a number of search options that are presented to you when you click on the “Search” menu item.

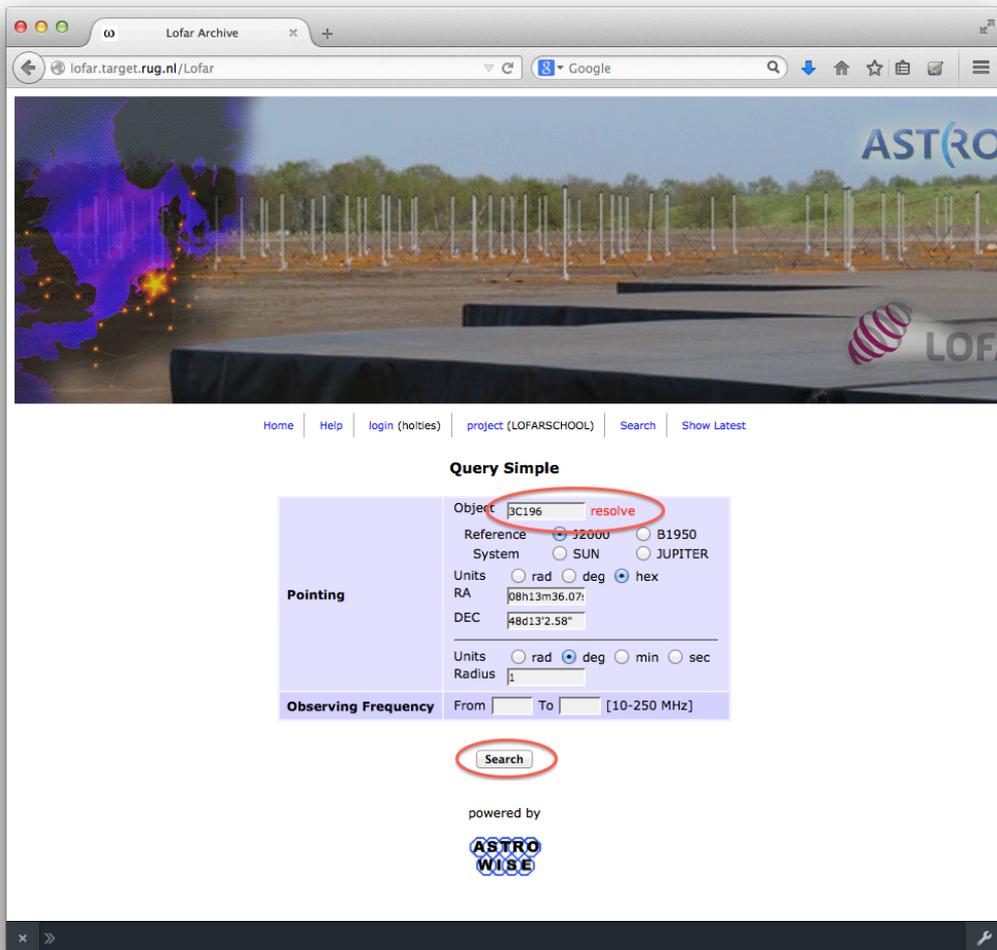
3. Click on the “Search” menu item and look at the various search options

The first option is a so-called cone search targeted at scientists that are interested in all data available for a particular object or a particular area in the sky. It provides name resolving and options to look for results in an area around the central object/coordinates. We will be looking for an observation of 3C196.

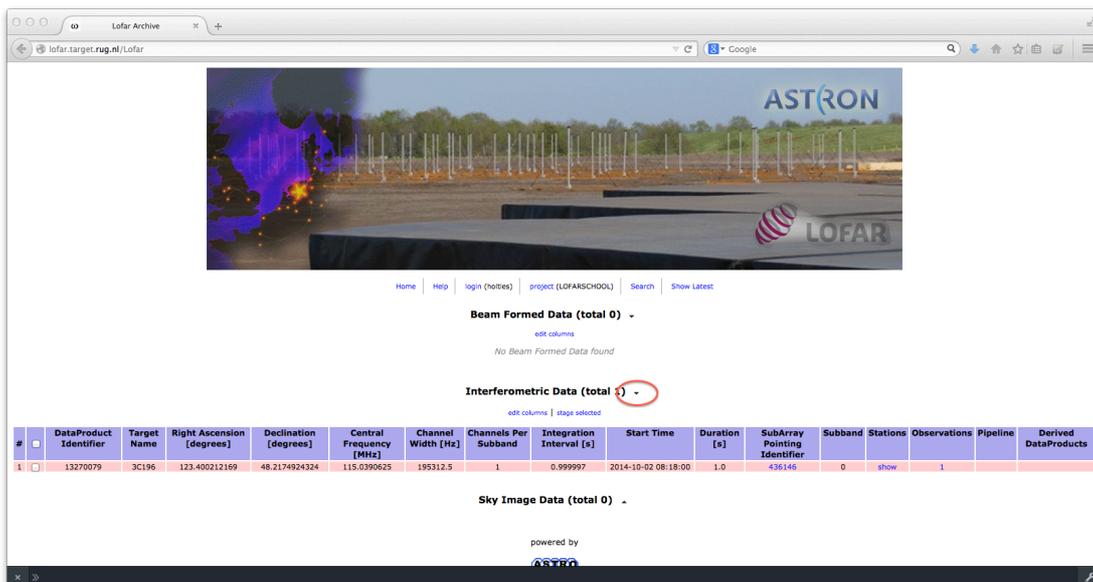
4. Click on “Simple Search”



5. Fill in 3C196 in the Object field and click “resolve” (not required)
6. Click “Search”



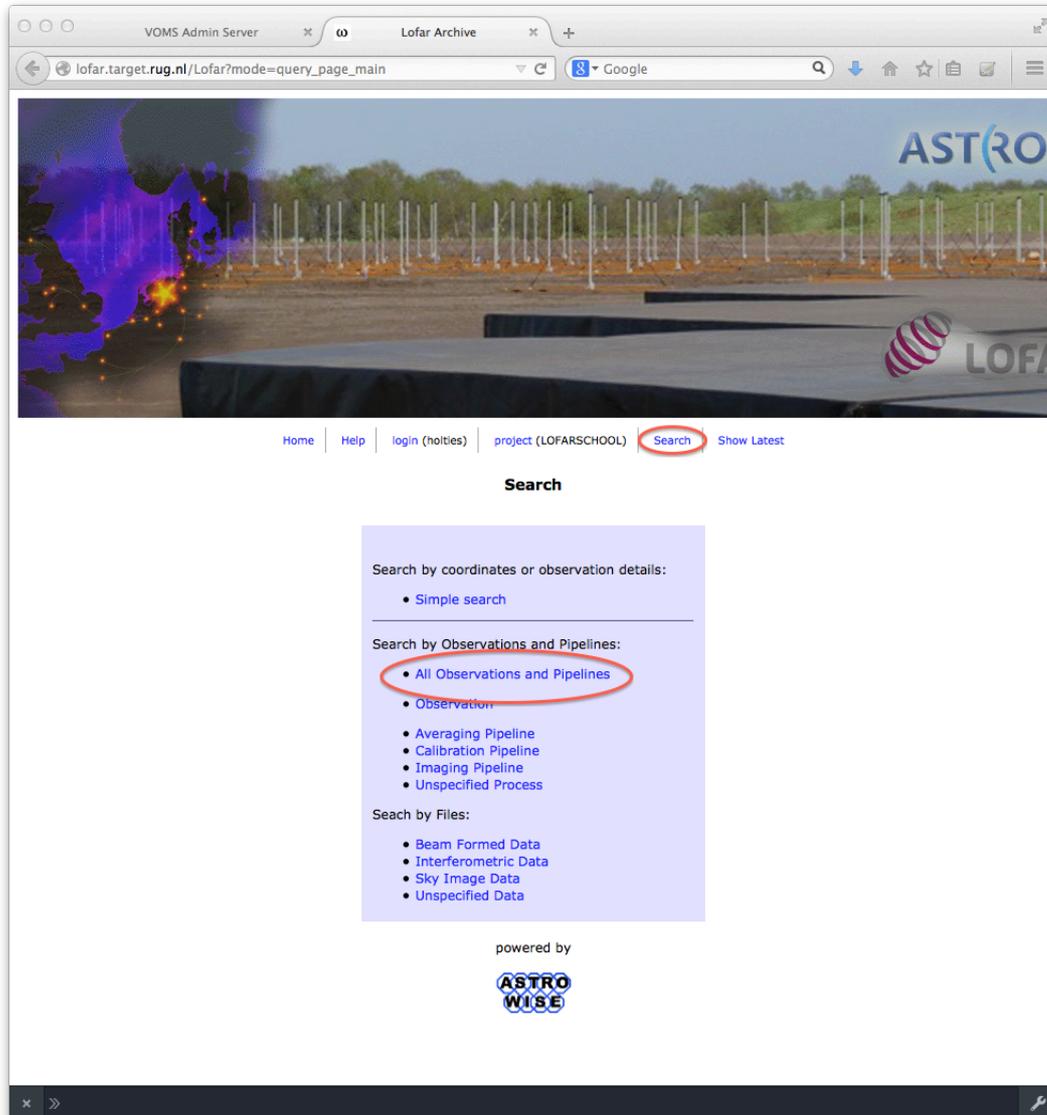
7. If needed, scroll down to the header “Interferometric data” and click on the small arrow beside it to view the interferometric dataproducts that have been found



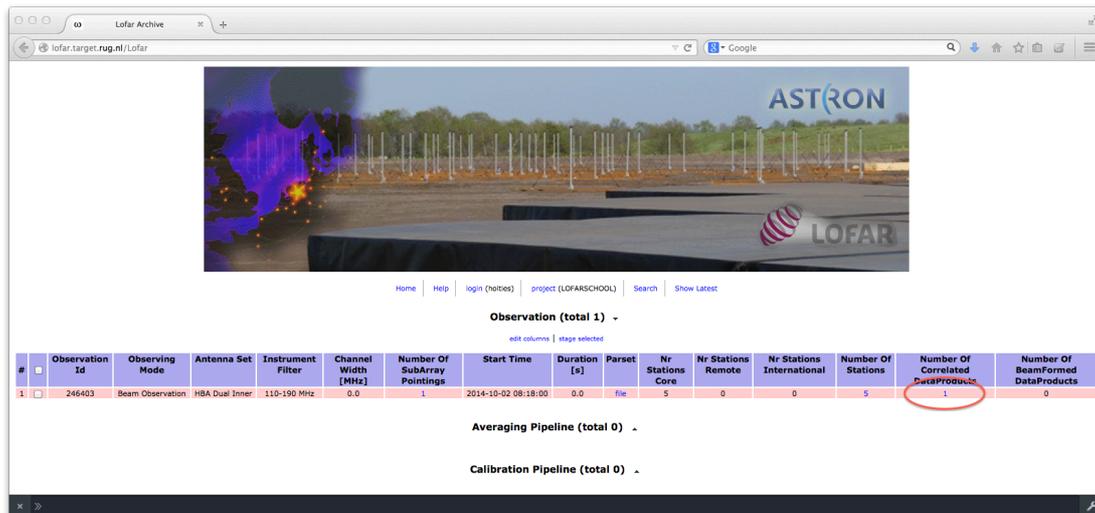
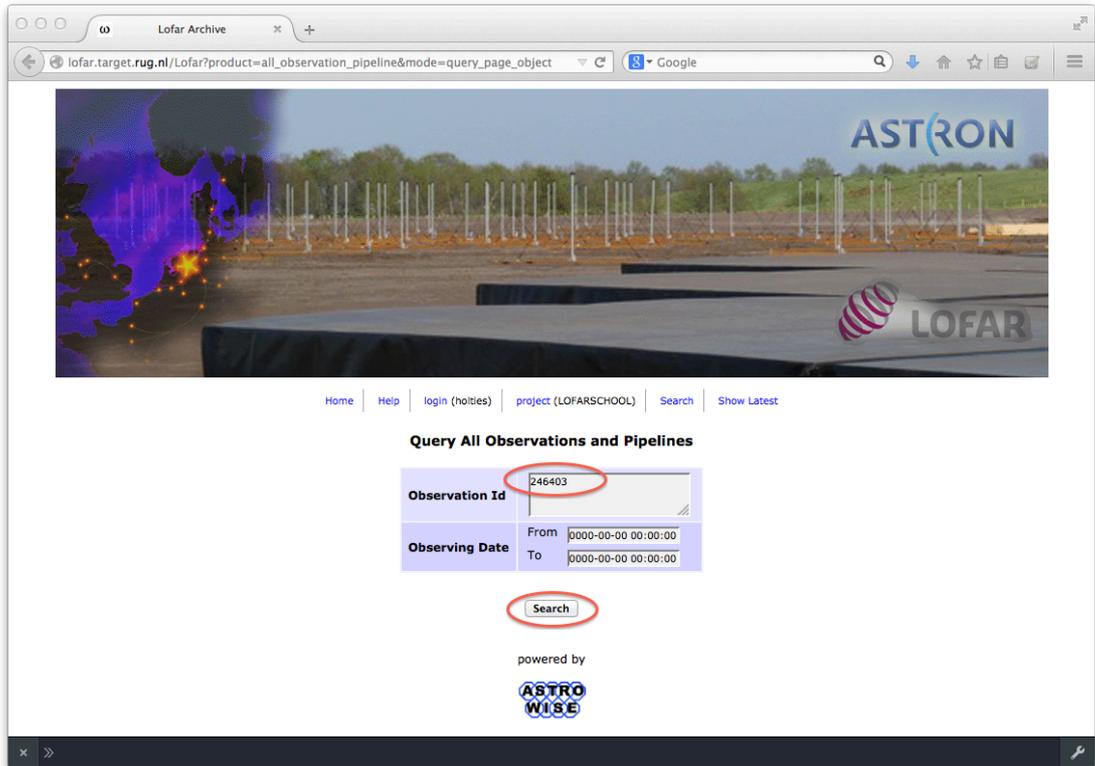
Another option is to search based on observation or processing pipeline details you may know (for example because you have been notified of it being available)

for your LOFAR project). We are looking for visibility data generated through a pipeline with LOFAR ID “246403”

8. Click on the “Search” menu item and than on “All Observations and Pipelines”



9. Fill in “246403” in the “Observation ID” field (if you know a date/date range, you could use that as well) and click “Search”



10. In the result list, click on the entry in the column “Number of Correlated Dataproducts”

Home | Help | login (hitites) | project (LOFARSCHOOL) | Search | Show Latest

Interferometric Data (total 1) -

edit columns | stage selected

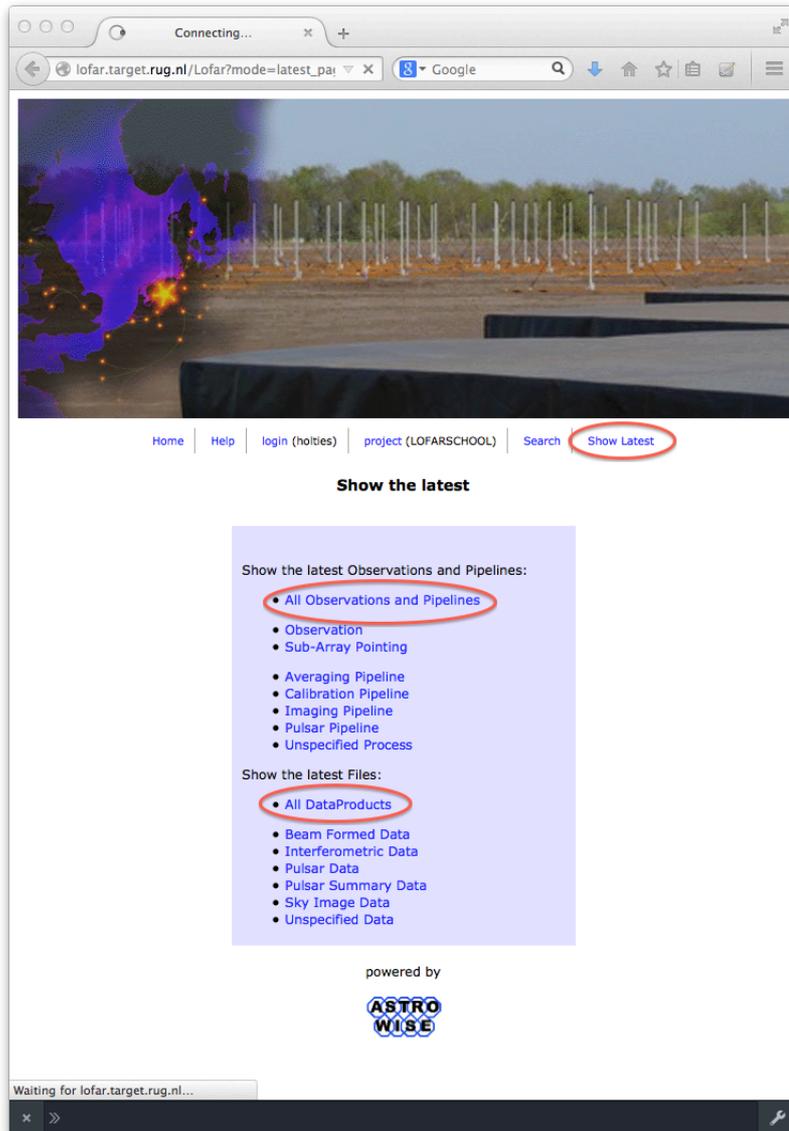
#	DataProduct Identifier	Target Name	Right Ascension [degrees]	Declination [degrees]	Central Frequency [MHz]	Channel Width [Hz]	Channels Per Subband	Integration Interval [s]	Start Time	Duration [s]	SubArray Pointing Identifier	Subband	Stations	Observations	Pipeline	Derived DataProducts
1	13270079	3C196	123.400212169	48.2174924324	115.0390625	195312.5	1	0.999997	2014-10-02 08:18:00	1.0	436146	0	show	1		

powered by
ASTRO WISE

If you are interested in dataproducts generated specifically by an observation or a specific pipeline, you can use the applicable search form.

One more option, useful for recently ingested data for a specific project you may be interested in, is to browse for it.

11. Click on the “Show latest” menu item. Similar options are provided as for searching. We will be taking the most straightforward approach for recent ingests (but not very practical for projects where there is a lot of activity):



12. Click “All Observations and Pipelines” and browse to the dataproduct. Alternatively, Click on “All Dataproducts”

Once you have the appropriate result for a query, you can select data and ask for it to be prepared for retrieval. Since the LOFAR LTA is huge (exceeding 10 petabyte and growing with 6 petabyte each year) most of the archived data ends up on tape relatively soon after observing although there is a good chance the data will still be on disk if you request retrieval immediately after being notified that your data is in the LTA.

Preparing data for retrieval goes through a process called “staging”. The staging service handles, if needed, the migration of data from tape to disk, and provides details needed for transferring data to the user system.

13. From one of the result lists generated in the previous steps, select the entry with dataproduct id 13270079 and click “stage selected”

Interferometric Data (total 1) -

#	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	Stations	Observations	Pipeline	Derived DataProducts
1	33270079	3C196	123.400212169	48.2174924324	115.0390625	195312.5	1	0.999997	2014-10-02 08:18:00	1.0	436146	0	show	1		

powered by
ASTRO WISE

14. Click "Submit" in the confirmation screen.

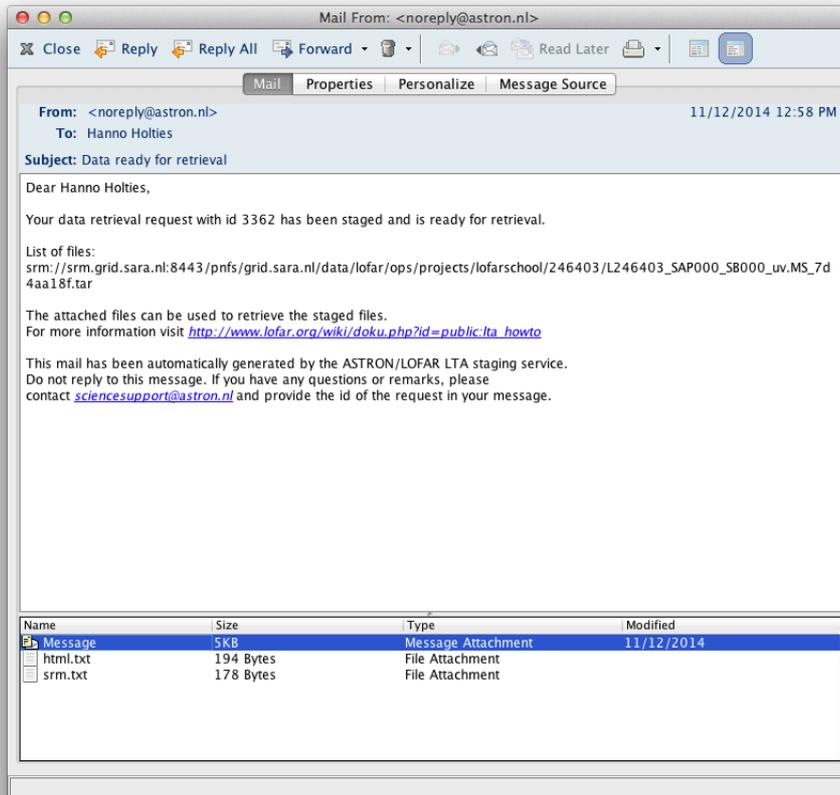
Staging Service

The following file(s) will be requested for download (count 1, total size 610.0 KB).

Size	MD5 checksum	Filename
610.0 KB	4d397ee367e949c4040f471dd429bee7	L246403_SAP000_SB000_uv.MS_7d4aa18f.tar
610.0 KB	Total filesize	

powered by
ASTRO WISE

15. Wait for a mail notifying that the dataproduct has been staged successfully. NB For the data request in this tutorial, the mail should be expected within a minute or so. If data is requested that is currently only available on the tape-backend, in particular if there are many files or a large volume that needs to be retrieved, it may take a day or more for the data to become available.



Ila) Data retrieval using wget

The mail received in the last step of the previous section contains some instructions and two file attachments that can be used to transfer the staged data to your system. We will start with the low-threshold “HTTP download” option.

1. Copy the (content of) html.txt to the system where you are going to transfer the data to.

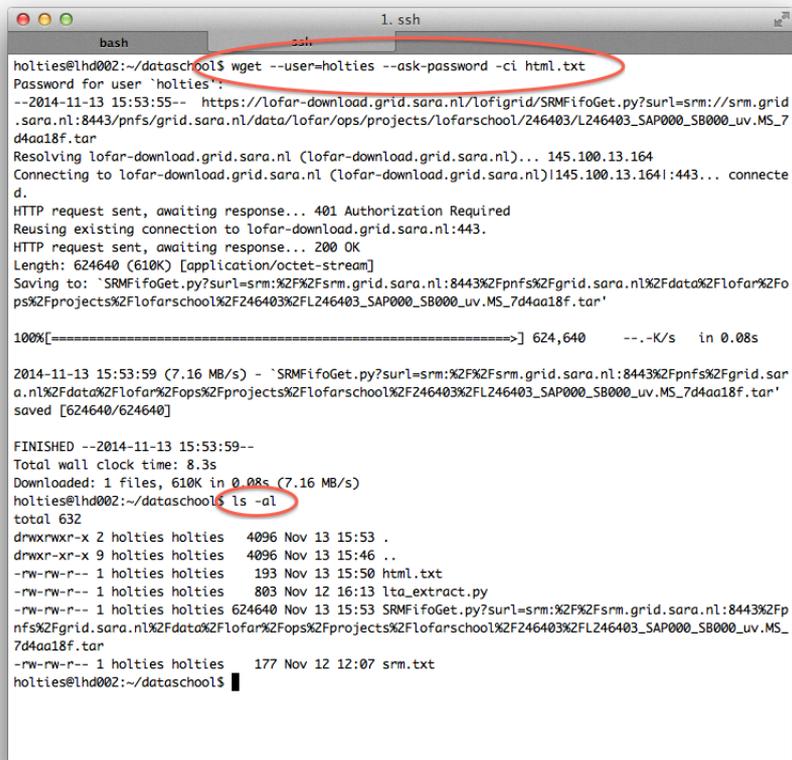
```

holties@lhd002:~/dataschool$ ls -al
total 16
drwxrwxr-x 2 holties holties 4096 Nov 12 12:07 .
drwxr-xr-x 8 holties holties 4096 Nov 12 12:07 ..
-rw-rw-r-- 1 holties holties 193 Nov 12 12:07 html.txt
-rw-rw-r-- 1 holties holties 177 Nov 12 12:07 srm.txt
holties@lhd002:~/dataschool$ cat html.txt
https://lofar-download.grid.sara.nl/lofigrid/SRMFileGet.py?surl=srm://srm.grid.sara.nl:8443/pnfs/grid.sara.nl/data/lofar/ops/projects/lofarschool/246403/L246403_SAP000_SB000_uv.MS_7d4aa18f.tar
holties@lhd002:~/dataschool$

```

2. Now retrieve the files using wget:

```
> wget --user=my_username --ask-password -ci html.txt
```



```
bash
holties@lhd002:~/dataschool$ wget --user=holties --ask-password -ci html.txt
Password for user 'holties':
--2014-11-13 15:53:55-- https://lofar-download.grid.sara.nl/lofigrid/SRMFifoGet.py?surl=srm://srm.grid.sara.nl:8443/pnfs/grid.sara.nl/data/lofar/ops/projects/lofarschool/246403/L246403_SAP000_SB000_uv.MS_7d4aa18f.tar
Resolving lofar-download.grid.sara.nl (lofar-download.grid.sara.nl)... 145.100.13.164
Connecting to lofar-download.grid.sara.nl (lofar-download.grid.sara.nl)|145.100.13.164|:443... connected.
HTTP request sent, awaiting response... 401 Authorization Required
Reusing existing connection to lofar-download.grid.sara.nl:443.
HTTP request sent, awaiting response... 200 OK
Length: 624640 (610K) [application/octet-stream]
Saving to: 'SRMFifoGet.py?surl=srm:%2F%2Fgrid.sara.nl:8443%2Fpnfs%2Fgrid.sara.nl%2Fdata%2Flofar%2Fops%2Fprojects%2Flofarschool%2F246403%2FL246403_SAP000_SB000_uv.MS_7d4aa18f.tar'

100%[=====] 624,640  --.-K/s  in 0.08s

2014-11-13 15:53:59 (7.16 MB/s) - `SRMFifoGet.py?surl=srm:%2F%2Fgrid.sara.nl:8443%2Fpnfs%2Fgrid.sara.nl%2Fdata%2Flofar%2Fops%2Fprojects%2Flofarschool%2F246403%2FL246403_SAP000_SB000_uv.MS_7d4aa18f.tar' saved [624640/624640]

FINISHED --2014-11-13 15:53:59--
Total wall clock time: 8.3s
Downloaded: 1 files, 610K in 0.08s (7.16 MB/s)
holties@lhd002:~/dataschool$ ls -al
total 632
drwxrwxr-x 2 holties holties 4096 Nov 13 15:53 .
drwxr-xr-x 9 holties holties 4096 Nov 13 15:46 ..
-rw-rw-r-- 1 holties holties 193 Nov 13 15:50 html.txt
-rw-rw-r-- 1 holties holties 803 Nov 12 16:13 lta_extract.py
-rw-rw-r-- 1 holties holties 624640 Nov 13 15:53 SRMFifoGet.py?surl=srm:%2F%2Fgrid.sara.nl:8443%2Fpnfs%2Fprojects%2Flofarschool%2F246403%2FL246403_SAP000_SB000_uv.MS_7d4aa18f.tar
-rw-rw-r-- 1 holties holties 177 Nov 12 12:07 srm.txt
holties@lhd002:~/dataschool$
```

wget allows for providing credentials on the command line as well using the option “--password” but these will be visible for other users on the system as long as wget is active retrieving files. Using the “--ask-password” option as described prevents this security issue to occur.

Alternatively, on systems where you have a personal home directory that you trust is not accessible by others, you can create a “.wgetrc” file in your home that contains the following two lines:

```
user=<LOFAR webservice account>
password=<Your password>
```

This file contains credential information so limit access to owner only:

```
> Chmod 600 $HOME/.wgetrc
```

Using the .wgetrc file, it is not required to provide credentials each time you run the wget command to retrieve files.

The wget ‘-c’ option is mostly useful for large/long transfers as it will allow continuing a partial transfer after it was broken off, for example resulting from network issues or unforeseen system reboots.

One drawback of wget is that by default it uses the complete URL as destination file name and it does not allow for destination file renaming if an input file is used. Many LOFAR dataproducts consist of a set of files that are made available through the archive as a tar-file. Therefore, in general you need to unpack the downloaded dataproduct anyway:

```
3. tar -xf ./<location of downloaded dataproduct>
```

If you would like to just rename the files, use e.g. the following command:

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

A script that takes care of renaming files and untarring them is available from http://www.lofar.org/wiki/doku.php?id=public:lta_howto&#http_download

It should be noted that at the LOFAR LTA sites, special purpose HTTP download servers handle this type of data retrieval. These servers only have limited capacity and the sites allow for far better performing data retrieval when you are using the Grid client toolset. Therefore, the following procedure is the recommended way of transferring any other than a limited number of files with a limited total volume:

Iib**) Data retrieval using srmcp (optional)**

This section assumes you have successfully applied for a personal grid certificate and have it installed on the system where you want to transfer the files to. It also assumes that an appropriate Grid client toolset is installed and your environment is set up to use it.

The installation of the client toolset itself is outside of the cope of this tutorial but if you want to install this on a system of your choosing, there is a page providing details for a portable installation (not requiring root access) available at:

<http://www.lofar.org/wiki/doku.php?id=public:srmclientinstallation>

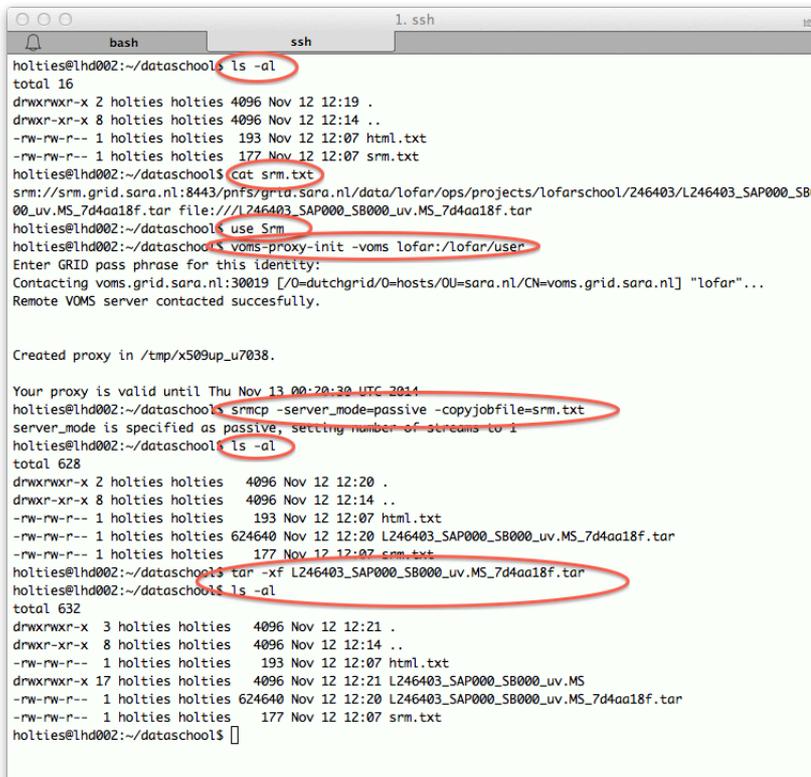
This page includes a walkthrough. It should be noted that for best performance it is highly recommended that globus-url-copy is installed separately from one of the linux repositories provided on that page and this does require an administration account.

NB Step 2 is specific for the CEP3 cluster and basically replaces calling the init script from the portable installation described in the previous paragraph.

1. Copy the (content of the) srm.txt from the staging mail to the system where you are going to transfer the data to.
2. Set up the user environment by running:
> use Srm
3. Generate a so-called proxy certificate. This is what the grid client tools use to authenticate to the grid servers:
> voms-proxy-init -voms lofar:/lofar/user

You will be requested to provide the password you used to encrypt the certificate key.

4. Retrieve the file(s):
srmcp -server_mode=passive -copyjobfile=srm.txt
5. Untar the retrieved file:
tar -xf <location of downloaded dataproduct>



```
holties@lhd002:~/dataschools$ ls -al
total 16
drwxrwxr-x 2 holties holties 4096 Nov 12 12:19 .
drwxr-xr-x 8 holties holties 4096 Nov 12 12:14 ..
-rw-rw-r-- 1 holties holties 193 Nov 12 12:07 html.txt
-rw-rw-r-- 1 holties holties 177 Nov 12 12:07 srm.txt
holties@lhd002:~/dataschools$ cat srm.txt
srm://srm.grid.sara.nl:8443/pnfs/grid.sara.nl/data/lofar/ops/projects/lofarschool/L246403/L246403_SAP000_SB0
00_uv_MS_7d4aa18f.tar file:///L246403_SAP000_SB000_uv_MS_7d4aa18f.tar
holties@lhd002:~/dataschools$ use Srm
holties@lhd002:~/dataschools$ voms-proxy-init -voms lofar:/lofar/user
Enter GRID pass phrase for this identity:
Contacting voms.grid.sara.nl:30019 [/O=dutchgrid/O=hosts/OU=sara.nl/CN=voms.grid.sara.nl] "lofar"...
Remote VOMS server contacted successfully.

Created proxy in /tmp/x509up_u7038.

Your proxy is valid until Thu Nov 13 00:20:30 UTC 2014
holties@lhd002:~/dataschools$ srmcp -server_mode=passive -copyjobfile=srm.txt
server_mode is specified as passive, setting number of streams to 1
holties@lhd002:~/dataschools$ ls -al
total 628
drwxrwxr-x 2 holties holties 4096 Nov 12 12:20 .
drwxr-xr-x 8 holties holties 4096 Nov 12 12:14 ..
-rw-rw-r-- 1 holties holties 193 Nov 12 12:07 html.txt
-rw-rw-r-- 1 holties holties 624640 Nov 12 12:20 L246403_SAP000_SB000_uv_MS_7d4aa18f.tar
-rw-rw-r-- 1 holties holties 177 Nov 12 12:07 srm.txt
holties@lhd002:~/dataschools$ tar -xf L246403_SAP000_SB000_uv_MS_7d4aa18f.tar
holties@lhd002:~/dataschools$ ls -al
total 632
drwxrwxr-x 3 holties holties 4096 Nov 12 12:21 .
drwxr-xr-x 8 holties holties 4096 Nov 12 12:14 ..
-rw-rw-r-- 1 holties holties 193 Nov 12 12:07 html.txt
drwxrwxr-x 17 holties holties 4096 Nov 12 12:21 L246403_SAP000_SB000_uv_MS
-rw-rw-r-- 1 holties holties 624640 Nov 12 12:20 L246403_SAP000_SB000_uv_MS_7d4aa18f.tar
-rw-rw-r-- 1 holties holties 177 Nov 12 12:07 srm.txt
holties@lhd002:~/dataschools$ []
```