Long Term Archive Frequently Asked Questions

This page is targeted at users who want to learn more on the LOFAR LTA (Long Term Archive). It should answer the most common questions and provide help in case of difficulties in data retrieval. If the information on this page did not succeed to solve any issues you may encounter, please send a mail to Science Support.

Questions

- **Q**: I never worked with the LTA before. Where is a good starting point?
- **Q**: Why is data retrieval so difficult?
- **Q**: What is an appropriate amount of data to retrieve?
- **Q**: What is all this SRM/staging stuff about?
- **Q**: I did not receive a mail notification that my request was scheduled!
- **Q**: I did not receive a mail notification that my data is ready for retrieval! Has my request gone lost?
- **Q**: There are different ways to download. Which one is the best?
- **Q**: My download speeds are too slow. What can I do?
- **Q**: My files only contain some error message instead of data
- **Q**: My data files are corrupted

Answers

I never worked with the LTA before. Where is a good starting point?

New users are advised to read through our tutorial page, the LTA How-To, which guides you through the whole process from getting a user account, over finding your data, to the actual download.

Why is data retrieval so difficult?

It is important to understand the the data volumes of LOFAR are pretty huge and handling them requires different technologies than what we all know and use in our everyday life. For instance, LTA data is stored on magnetic tape and has to be copied to a hard drive (getting 'staged') before it can be retrieved. To transfer these amounts of data within reasonable time requires careful consideration and special tools. We try to make the LTA as convenient to use as possible, e.g. by providing http downloads for users without Grid certificate and portable Grid tools for those who want or need the extra performance. We are aware of the fact that data retrieval is quite close to the backend technology and we hope to be able to provide solutions with higher abstraction in the future.

What is an appropriate amount of data to retrieve?

This depends. There are two things to consider: The capabilities of your own system and the capabilities of the LTA services. The most important thing to know about LTA capabilities, is that the

disk pool that temporarily holds your data and from where it can be downloaded, is of limited capacity. This means that the data you requested is only available for download for a limited time (since the space is needed for new requests at some point). Your data is only guaranteed to stay available for 7 days. It can be re-requested after that, but you should never request more data than you can download within a few days. In most cases, this is limited by the capabilities of your own system, especially your network connection. (And available local storage space, of course.)

As a rule of thumb, we ask you to keep your requests below 5 TB in volume and smaller than 1'000 files. The larger your request, the longer it takes until you can retrieve the first file. Also, please limit the number of requests running in parallel to a few, especially when they contain many files. In principle, we avoid introducing hard limits, but rely on reasonable user behavior. This also means that you can block the system for a long time or, in the worst case, even bring it down. So please act responsibly or we might have to enforce some limits in the future to keep the system available for other users. Be aware, that we may cancel your request(s) in excessive cases to maintain LTA operation.

If you, by accident, staged some 100'000 files or 100 TB of data, please contact Science Support, so that we can stop these requests, thanks!

What is all this SRM/staging stuff about?

stay tuned

I did not receive a mail notification that my request was scheduled!

If the LTA catalog did not show any error when you submitted your request, then it is safe to assume that your request was registered in our staging system. Usually, you should get a notification mail that this has happened within a few minutes. If you did not receive the notification within an hour, then our staging service may be down. Note that your request is not lost in this case and will be picked up after the service is back online. In urgent cases or if you are not sure that something went wrong while submitting your request, please contact Science Support.

I did not receive a mail notification that my data is ready for retrieval! Has my request gone lost?

After you got a notification that your requests was scheduled, it is in our database and there's hardly a possibility that it got lost. Staging requests can take up to a day or two, but will finish a lot sooner in most cases. This depends on your request's size but also on how busy the storage systems are by other user's requests at the moment. Sometimes, the LTA storage systems are down for maintenance and this can delay the whole procedure. You can check for downtimes here.

It is not alarming when your request did not finish in 24 hours, even when your last request finished within 10 minutes. In urgent cases or if you did not receive a notification after 48 hours, please contact Science Support.

There are different ways to download. Which one is the best?

That depends. In short terms: Http downloads are the easiest (e.g. via wget), downloads via SRM tools can be faster and are encouraged for large amounts.

The SRM systems which the LTA sites operate are integrated in the Grid. To work with them directly, you need a Grid certificate. To allow users without a Grid certificate to download LTA data, we operate webservers as a frontend to the SRM backend. These webservers provide the requested data via http downloads. The webservers are not excessively capable machines and meant for occasional users. If you retrieve huge amounts of data on a regularly timescale, please work with SRM directly, especially if you own a Grid certificate. We provide a portable toolkit to make it as easy to set up as possible.

For instructions, refer to the LTA How-To page.

My downloads are too slow. What can I do?

First of all, you have to check how slow your download really is. If wget shows an estimated time of arrival of several hours, this does not necessarily mean that the download is 'slow': some files in the LTA are also just really huge. In most cases, your local network connection will be the bottleneck. For instance, a standard 'Fast Ethernet' network connection allows download speeds of around 12 MB/s at a maximum. Our systems are able to handle that. In case you can rule out your network connection as the bottleneck: there are different ways to download your data and not all provide the same performance. By our experience, this is the order of performance:

- Http downloads are the slowest option. The speed is limited by the server's network connection (~120 MB/s), which is shared by all users, and an upper limit per download (around ~30 MB/s) for technical reasons. If your download maxes out at the per-download limit, you may try to start up to four downloads in parallel. **Note:** There is no performance benefit to expect from more than four parallel transfers! However, there is a connection limit, which you may trigger if you start too many parallel downloads.
- SRMCP is the faster option in most cases, since you work with the SRM backend directly.
- SRMCP + copy-script seems to be the fastest solution available. It uses globus-url-copy, which is reported to have superior performance over the default srmcp transfer.

For instructions, refer to the LTA How-To page.

You may also want to read this FAQ Answer for further explanation: There are different ways to download. Which one is the best?

My files only contain some error message instead of data

Please read the error message carefully. In many cases, it should give you some indication of what went wrong. If this does not help you, please contact Science Support or retry after a few hours.

Important: If you use wget with option '-c', please note the following: wget does not check the contents of an existing file, so when restarting wget with option '-c' (continue) to retrieve the failed files, it will append the later data chunk to the existing file that contains the error message (and not the first section of you data). Make sure to delete the existing error files (should be obvious by the small file size) before calling 'wget -ci' again, to avoid corrupted data. If you already ended up with a

corrupted file, you have to delete that and re-retrieve the whole file.

My data files are corrupted

Please check the beginning of your files. If there is an error message, please refer to this answer. Otherwise, please try to re-retrieve an affected file. If this does not help, please contact Science Support.

From:

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

Permanent link:

https://www.astron.nl/lofarwiki/doku.php?id=public:lta_faq&rev=1424707714

Last update: **2015-02-23 16:08**