# LOFAR DATA PRODUCTS AND MANAGEMENT: TOWARDS THE SKA

R. F. Pizzo

Head of LOFAR and WSRT/Apertif Science Support

Trieste, December 3rd 2015

ASTRON

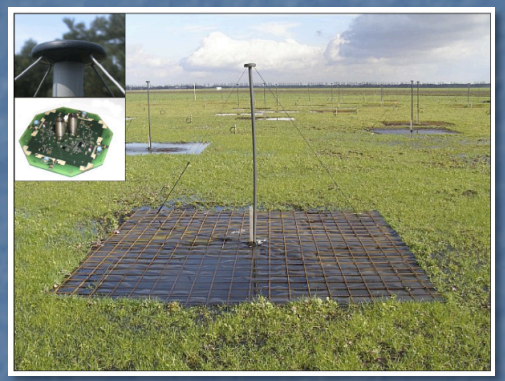
LOFAR

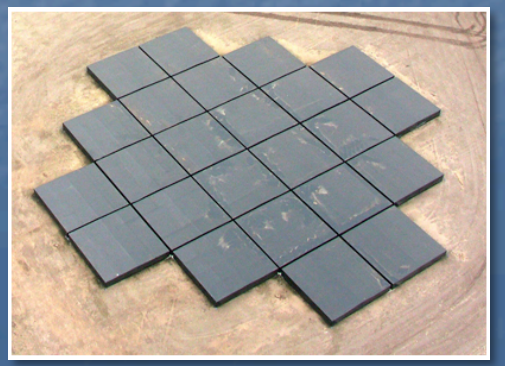
# THE LOW FREQUENCY ARRAY – KEY FACTS

LOFAR

ASTRON

- The International LOFAR telescope (ILT) consists of an interferometric array of dipole antenna stations distributed throughout the Netherlands, Germany, France, UK, Sweden (+ Poland, …)
- Operations started in December 2012
- Operating frequency is 10-250 MHz
- 1 beam with up to 96 MHz total bandwidth, split into 488 sub bands with 64 frequency channels (8-bit mode)
- < 488 beams on the sky with ~ 0,2 MHz bandwidth
- Low band antenna (LBA; Area ~ 75200 $m^2$; 10-90 MHz)
- High Band Antenna (HBA; Area ~ 57000 $m^2$; 110-240 MHz)



# THE LOFAR SYSTEM: DATA FLOW

LOFAR

ASTRON

Station signals collected in the station cabinets

Signal sent to COBALT for correlation

CEP2

CEP3

Data sent to CEP2 for initial RO processing – products might get copied to CEP3


Products sent to the long-term archive

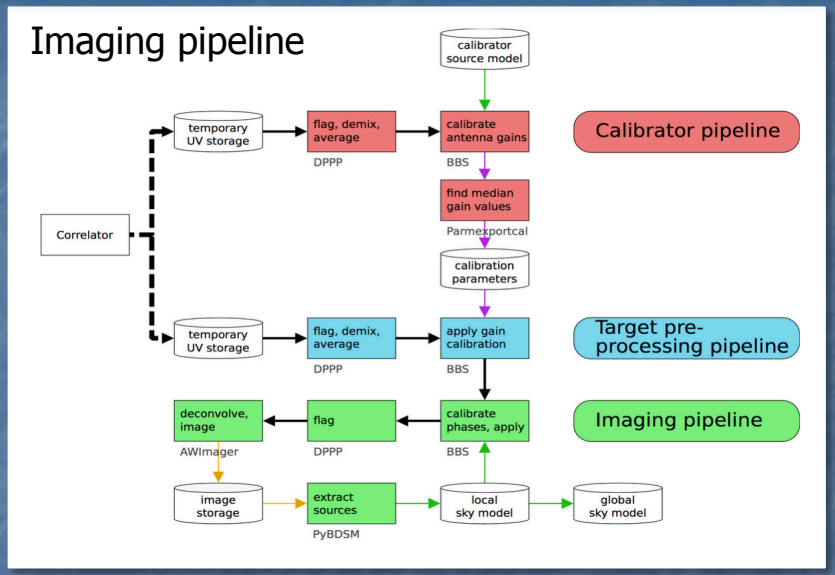
- Large data transport rates ➔ data storage challenges (35 TB /h)
- LOFAR is the first of a number of new astronomical facilities dealing with the transport, processing and storage of these large amounts of data and therefore represents an important technological pathfinder for the SKA

# LOFAR DATA PROCESSING

LOFAR

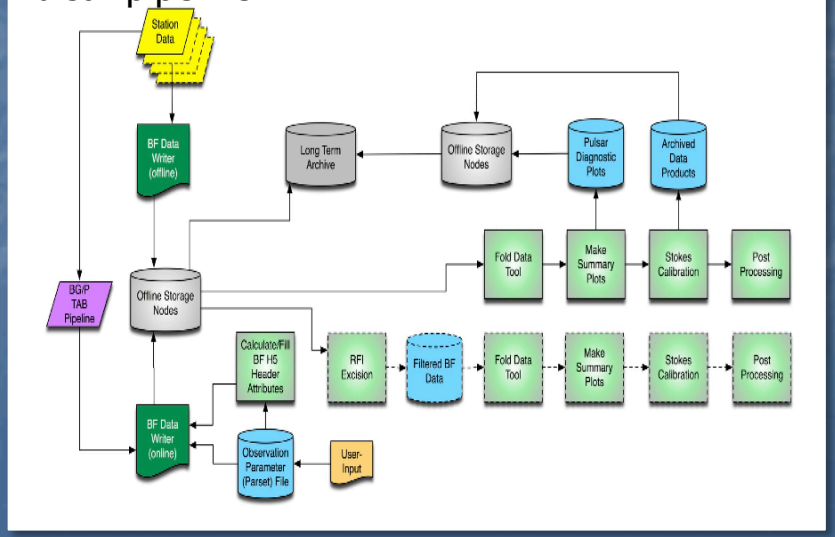
ASTRON

## Imaging pipeline




- Visibility data
- RFI removal
- Removal of brightest sources in the sky contaminating science in the field center
- Averaging
- Calibration
- Imaging + selfcalibration + source extraction
- Final images + cubes

## Pulsar pipeline




- Beam-formed data serve a variety of science cases -> several pipeline exist
- RFI masking
- dedispersion
- Searching of the data for single pulses and periodic signals

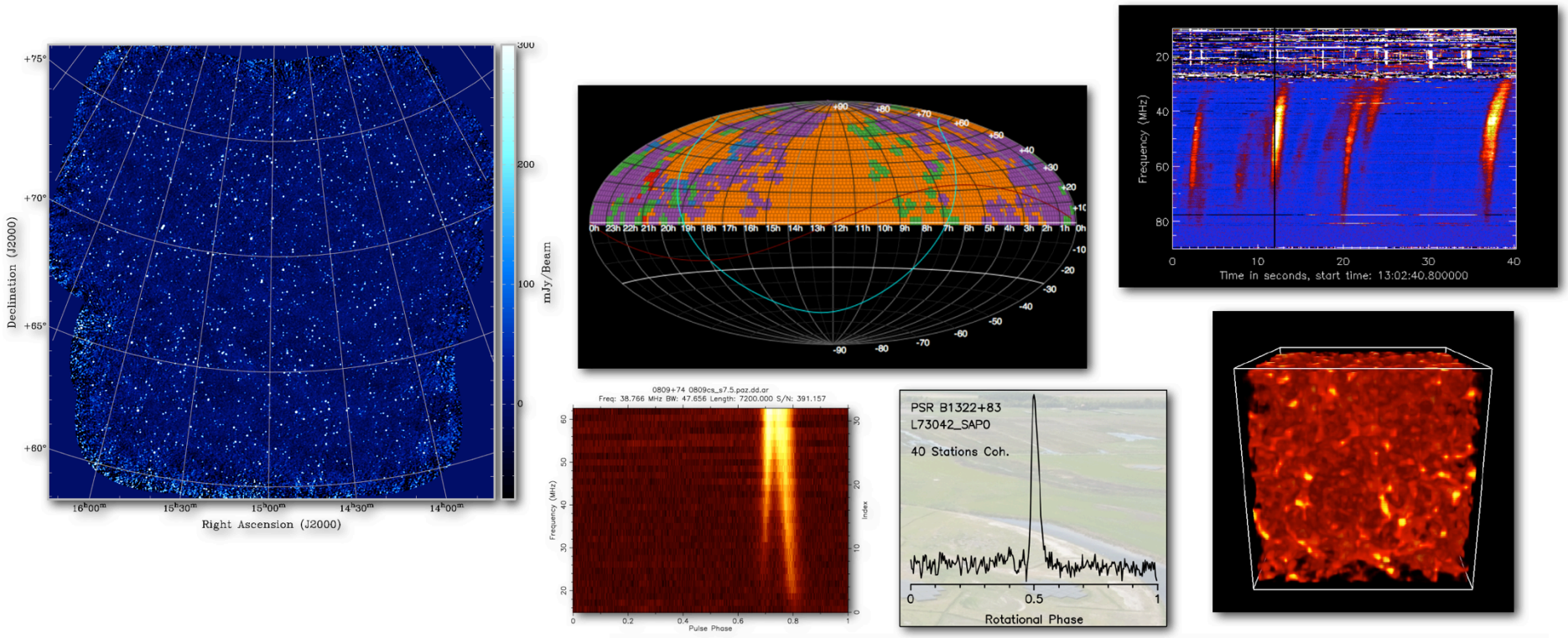
More pipelines in an advanced state of development (solar, transient, long-baselines, selfcalibration, extreme peeling…)

# LOFAR DATA PRODUCTS

LOFAR

ASTRON

- Velocity (raw data rates of 13 Tbits/s, correlated ~ 15 TB/hr)
- Volume ( 100 TB visibilities, 1 TB cubes, 1 PB catalogues)
- Variety (raw telemetry, uv data, beam-formed data, 2D-3D-4D-5D cubes, RM cubes, light-curves, catalogues, etc.)

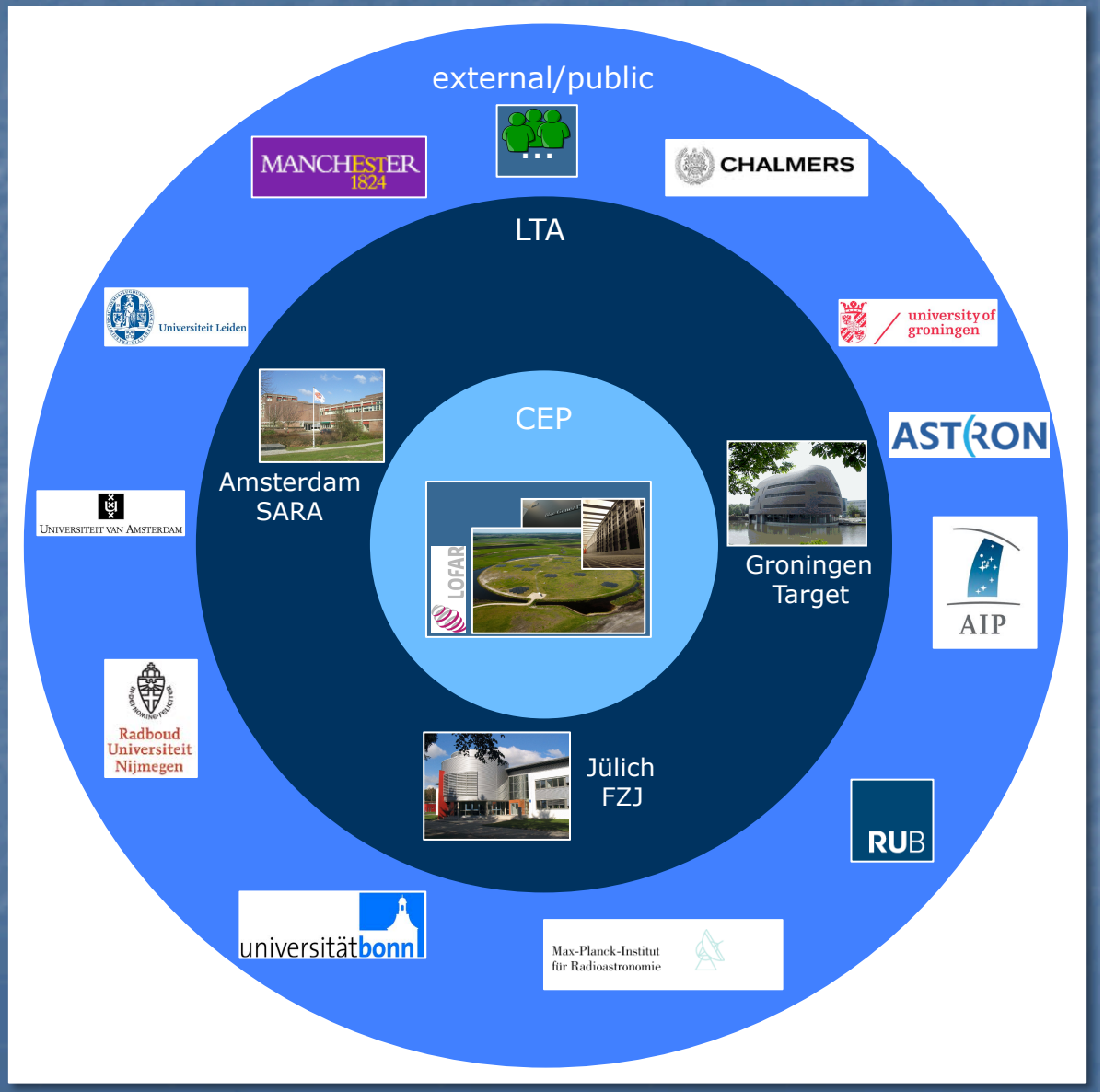


# LTA: LONG-TERM ARCHIVE

LOFAR

ASTRON

- Distributed information system created to store and process the large data volumes generated by the LOFAR radio telescope
- Currently involves sites in the Netherlands and Germany (1 more to come in Poland in 2016)
- Each site involved in the LTA provides storage capacity and optionally processing capabilities.
- Network consisting of light-path connections (utilizing 10 GbE technology) that are shared with LOFAR station connections and with the European eVLBI network

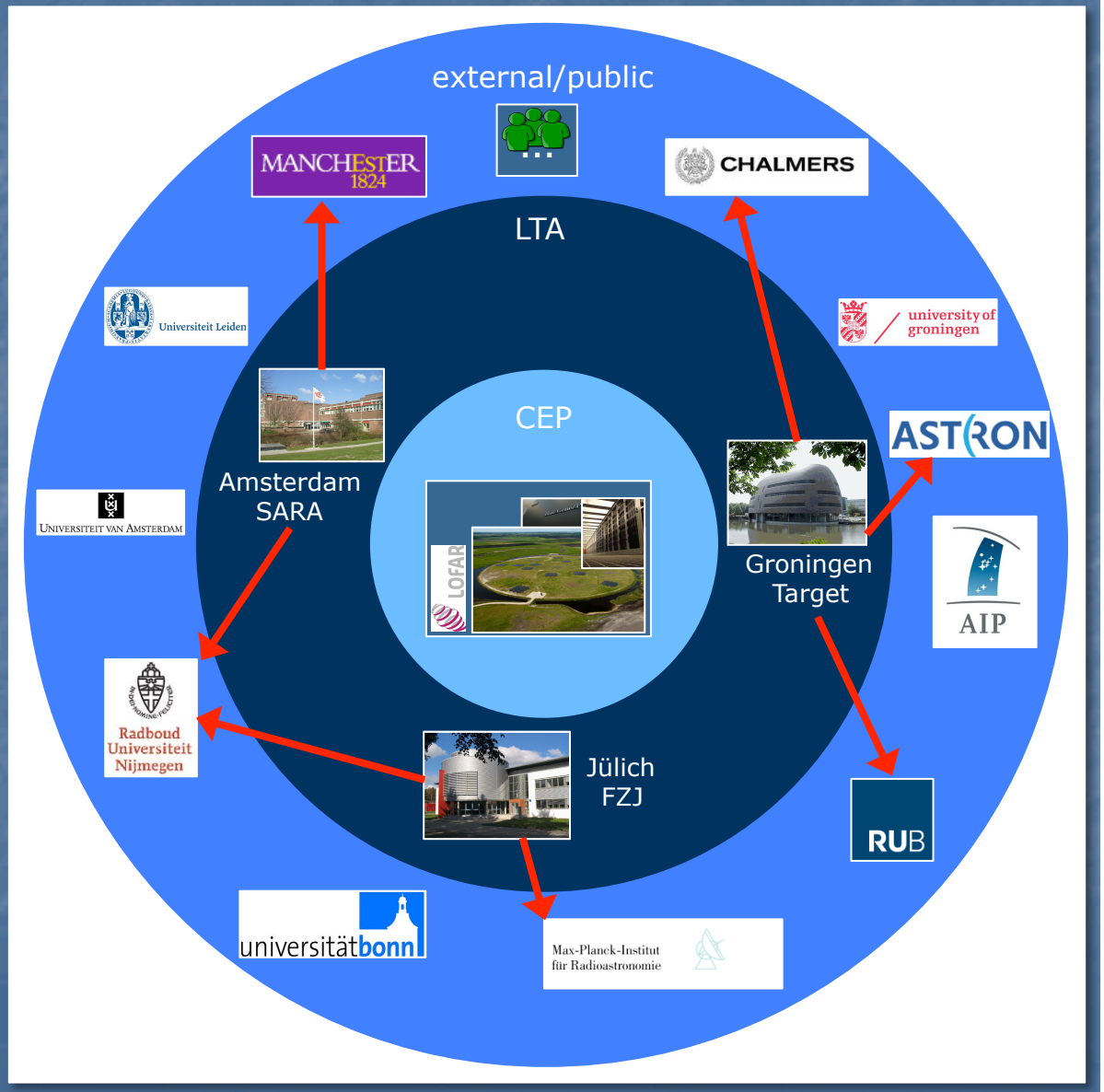


# DATA DOWNLOAD

LOFAR

ASTRON

- **Web based download server**
  - 'LTA enabled' ASTRON/ LOFAR account
  - Low threshold
  - Primarily for few files & smaller volumes
- **GridFTP**
  - Requires grid user certificate
  - More robust; superior performance
  - Requires grid client installation



# LTA: ASTROWISE

- Interface to query the LTA database and retrieve data to own compute facilities
- Public data – data that has passed the proprietary period become public and can be retrieved by anyone

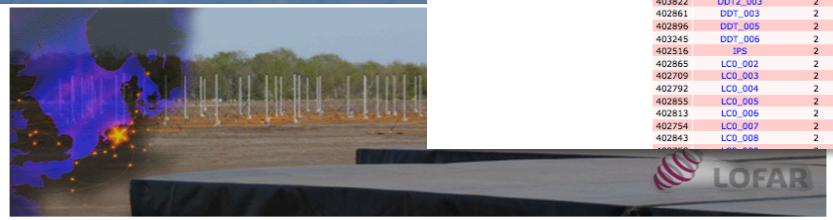

Home | Help | login (pizzo) | project (LC2_014) | Search | Show Latest

**Projects of db.lofar.target.rug.nl**

- Number of projects : 150
- Number of Users : 284
- Current user : pizzo

*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 | True | 21 | AWTIER0 |
| 403289 | 2014LOFAROBS | 2 | LOFAR | True | 21 | AWTIER0 |
| 1 | ALL | 4 | None | True | public | AWTIER0 |
| 403691 | CITT_2014 | 2 | LOFAR | True | 23 | AWTIER0 |
| 403307 | COBALT | 2 | LOFAR | True | 24 | AWTIER0 |
| 401580 | Commissioning2012 | 2 | LOFAR | True | 31 | AWTIER0 |
| 402639 | Commissioning2013 | 2 | LOFAR | True | 29 | AWTIER0 |
| 403798 | Commissioning2014 | 2 | LOFAR | True | 25 | AWTIER0 |
| 403009 | DDT0001 | 2 | LOFAR | True | 22 | AWTIER0 |
| 402919 | DDT0004 | 2 | LOFAR | True | 24 | AWTIER0 |
| 402921 | DDT0007 | 2 | LOFAR | True | 23 | AWTIER0 |
| 403146 | DDT0012 | 2 | LOFAR | True | 25 | AWTIER0 |
| 402892 | DDT002 | 2 | LOFAR | True | 26 | AWTIER0 |
| 403167 | DDT1_001 | 2 | LOFAR | True | 26 | AWTIER0 |
| 403211 | DDT1_002 | 2 | LOFAR | True | 23 | AWTIER0 |
| 403806 | DDT2_001 | 2 | LOFAR | True | 28 | AWTIER0 |
| 403822 | DDT2_003 | 2 | LOFAR | True | 24 | AWTIER0 |
| 402861 | DDT_003 | 2 | LOFAR | True | 24 | AWTIER0 |
| 402896 | DDT_005 | 2 | LOFAR | True | 25 | AWTIER0 |
| 403245 | DDT_006 | 2 | LOFAR | True | 22 | AWTIER0 |
| 402516 | IPS | 2 | LOFAR | True | 21 | AWTIER0 |
| 402865 | LC0_002 | 2 | LOFAR | True | 30 | AWTIER0 |
| 402709 | LC0_003 | 2 | LOFAR | True | 29 | AWTIER0 |
| 402792 | LC0_004 | 2 | LOFAR | True | 24 | AWTIER0 |
| 402855 | LC0_005 | 2 | LOFAR | True | 30 | AWTIER0 |
| 402813 | LC0_006 | 2 | LOFAR | True | 29 | AWTIER0 |
| 402754 | LC0_007 | 2 | LOFAR | True | 34 | AWTIER0 |
| 402843 | LC0_008 | 2 | LOFAR | True | 27 | AWTIER0 |




Home | Help | login (pizzo) | project (LC2_014) | Search | Show Latest

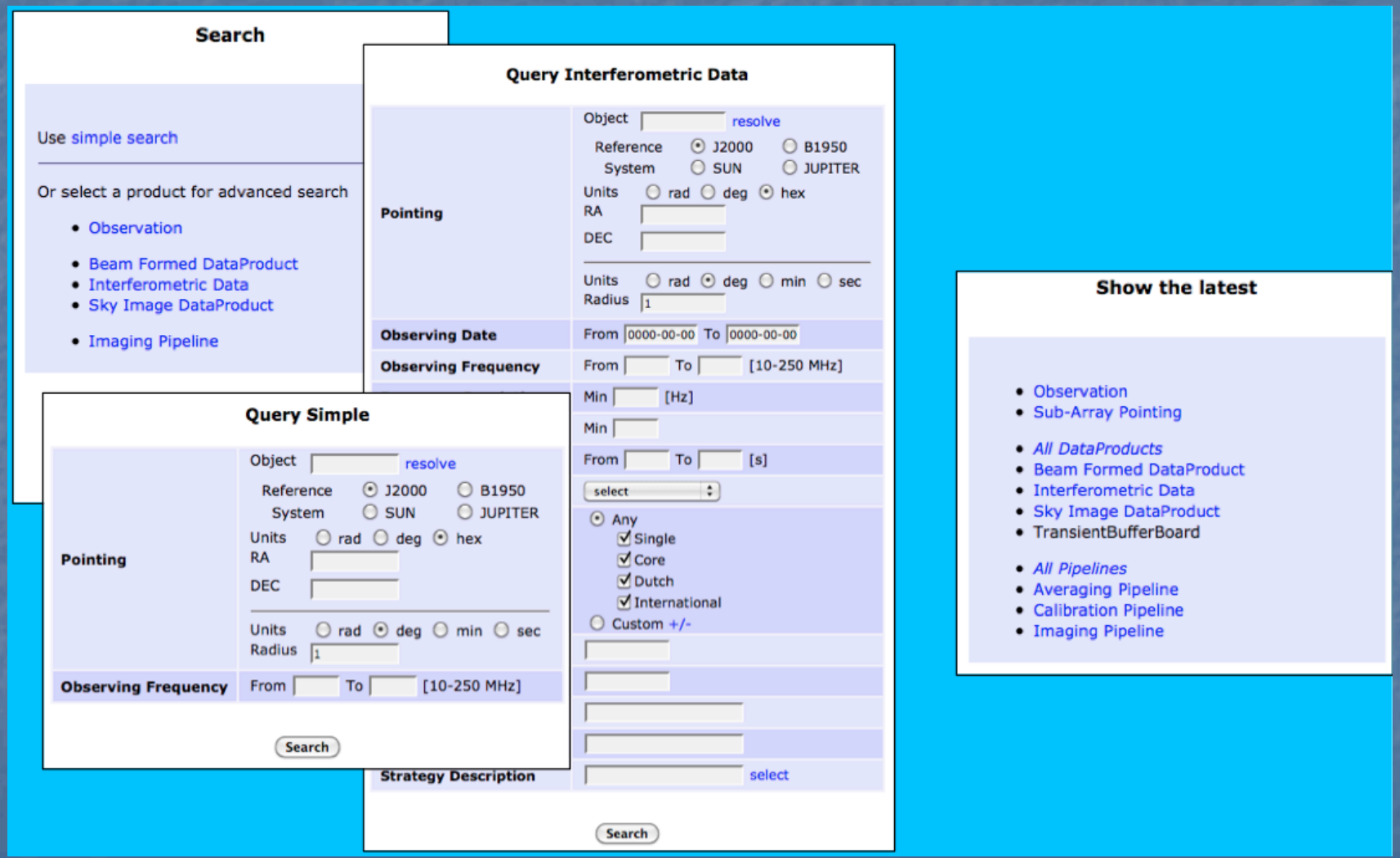
**Observation 1 to 100 (showing 100 of total 128)**

edit columns | stage selected

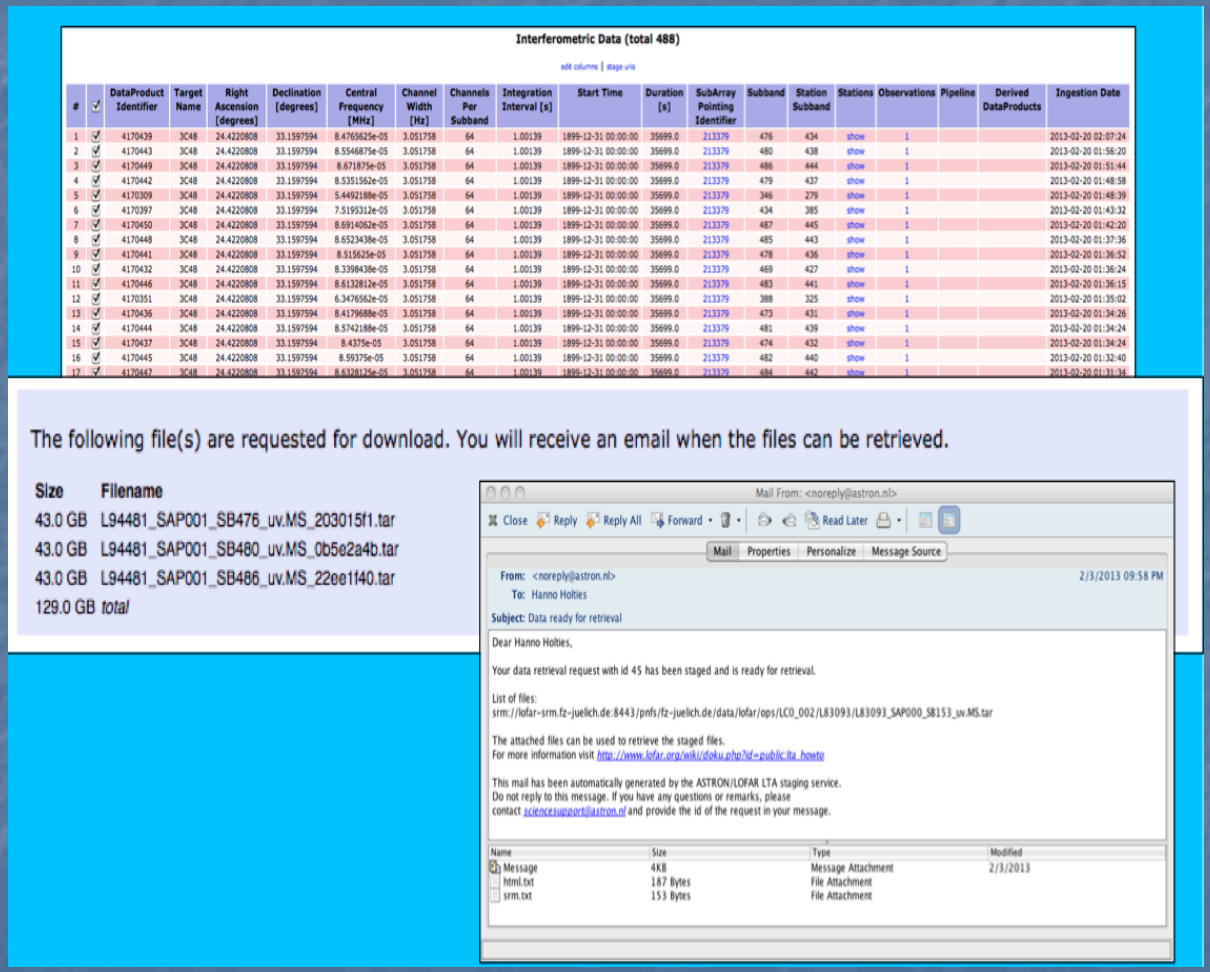
*first | previous | 1 | 2 | next | last*

| # | Observation Id | Observing Mode | Antenna Set | Instrument Filter | Channel Width [MHz] | Number Of SubArray Pointings | Start Time | Duration [s] | Parset | Nr Stations Core | Nr Stations Remote | Nr Stations International | Number Of Stations | Number Of Correlated DataProducts | Number Of BeamFormed DataProducts |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 100 | 240850 | Beam Observation | HBA Dual Inner | 110-190 MHz | 0.0030517578125 | 1 | 2014-08-09 14:42:21 | 1840.0 | file | 24 | 14 | 0 | 38 | 0 / 488 | 0 |
| 99 | 240852 | Beam Observation | HBA Dual Inner | 110-190 MHz | 0.0030517578125 | 1 | 2014-08-09 15:13:21 | 151.0 | file | 24 | 14 | 0 | 38 | 0 / 488 | 0 |
| 98 | 240854 | Beam Observation | HBA Dual Inner | 110-190 MHz | 0.0030517578125 | 1 | 2014-08-09 15:16:21 | 1842.0 | file | 24 | 14 | 0 | 38 | 0 / 488 | 0 |
| 97 | 240856 | Beam Observation | HBA Dual Inner | 110-190 MHz | 0.0030517578125 | 1 | 2014-08-09 15:47:21 | 160.0 | file | 24 | 14 | 0 | 38 | 0 / 488 | 0 |
| 96 | 240858 | Beam Observation | HBA Dual Inner | 110-190 MHz | 0.0030517578125 | 1 | 2014-08-09 15:50:21 | 1840.0 | file | 24 | 14 | 0 | 38 | 0 / 488 | 0 |
| 95 | 240862 | Beam Observation | HBA Dual Inner | 110-190 MHz | 0.0030517578125 | 1 | 2014-08-09 16:24:22 | 1832.0 | file | 24 | 14 | 0 | 38 | 0 / 488 | 0 |
| 94 | 240864 | Beam Observation | HBA Dual Inner | 110-190 MHz | 0.0030517578125 | 1 | 2014-08-09 16:55:21 | 151.0 | file | 24 | 14 | 0 | 38 | 0 / 488 | 0 |
| 93 | 240866 | Beam Observation | HBA Dual Inner | 110-190 MHz | 0.0030517578125 | 1 | 2014-08-09 16:58:21 | 1841.0 | file | 24 | 14 | 0 | 38 | 0 / 488 | 0 |
| 92 | 241336 | Beam Observation | HBA Dual Inner | 110-190 MHz | 0.0030517578125 | 1 | 2014-08-14 14:00:00 | 151.0 | file | 24 | 14 | 0 | 38 | 0 / 488 | 0 |
| 91 | 241338 | Beam Observation | HBA Dual Inner | 110-190 MHz | 0.0030517578125 | 1 | 2014-08-14 14:03:00 | 1830.0 | file | 24 | 14 | 0 | 38 | 0 / 488 | 0 |
| 82 | 241340 | Beam Observation | HBA Dual Inner | 110-190 MHz | 0.0030517578125 | 1 | 2014-08-14 14:34:01 | 150.0 | file | 24 | 14 | 0 | 38 | 0 / 488 | 0 |
| 81 | 241342 | Beam Observation | HBA Dual Inner | 110-190 MHz | 0.0030517578125 | 1 | 2014-08-14 14:37:01 | 1830.0 | file | 24 | 14 | 0 | 38 | 0 / 488 | 0 |
| 80 | 241344 | Beam Observation | HBA Dual Inner | 110-190 MHz | 0.0030517578125 | 1 | 2014-08-14 15:08:01 | 150.0 | file | 24 | 14 | 0 | 38 | 0 / 488 | 0 |
| 79 | 241346 | Beam Observation | HBA Dual Inner | 110-190 MHz | 0.0030517578125 | 1 | 2014-08-14 15:11:01 | 1840.0 | file | 24 | 14 | 0 | 38 | 0 / 488 | 0 |
| 78 | 241348 | Beam Observation | HBA Dual Inner | 110-190 MHz | 0.0030517578125 | 1 | 2014-08-14 15:42:00 | 151.0 | file | 24 | 14 | 0 | 38 | 0 / 488 | 0 |
| 77 | 241350 | Beam Observation | HBA Dual Inner | 110-190 MHz | 0.0030517578125 | 1 | 2014-08-14 15:45:00 | 1841.0 | file | 24 | 14 | 0 | 38 | 0 / 488 | 0 |
| 76 | 241352 | Beam Observation | HBA Dual Inner | 110-190 MHz | 0.0030517578125 | 1 | 2014-08-14 16:16:00 | 150.0 | file | 24 | 14 | 0 | 38 | 0 / 488 | 0 |

# LTA CATALOG QUERIES



# LTA CATALOG DATA RETRIEVAL



Interferometric Data (total 488)

edit columns | stage urls

| # | 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 | Observations | Pipeline | Derived DataProducts | Ingestion Date |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 4170439 | 3C48 | 24.4220808 | 33.1597594 | 8.4765625e-05 | 3.051758 | 64 | 1.00139 | 1899-12-31 00:00:00 | 35699.0 | 213379 | 476 | 434 | show | 1 | | | 2013-02-20 02:07:24 |
| 2 | 4170443 | 3C48 | 24.4220808 | 33.1597594 | 8.5546875e-05 | 3.051758 | 64 | 1.00139 | 1899-12-31 00:00:00 | 35699.0 | 213379 | 480 | 438 | show | 1 | | | 2013-02-20 01:56:20 |
| 3 | 4170449 | 3C48 | 24.4220808 | 33.1597594 | 8.671875e-05 | 3.051758 | 64 | 1.00139 | 1899-12-31 00:00:00 | 35699.0 | 213379 | 486 | 444 | show | 1 | | | 2013-02-20 01:51:44 |
| 4 | 4170442 | 3C48 | 24.4220808 | 33.1597594 | 8.5351562e-05 | 3.051758 | 64 | 1.00139 | 1899-12-31 00:00:00 | 35699.0 | 213379 | 479 | 437 | show | 1 | | | 2013-02-20 01:48:58 |
| 5 | 4170309 | 3C48 | 24.4220808 | 33.1597594 | 5.4492188e-05 | 3.051758 | 64 | 1.00139 | 1899-12-31 00:00:00 | 35699.0 | 213379 | 346 | 279 | show | 1 | | | 2013-02-20 01:48:39 |
| 6 | 4170397 | 3C48 | 24.4220808 | 33.1597594 | 7.5195312e-05 | 3.051758 | 64 | 1.00139 | 1899-12-31 00:00:00 | 35699.0 | 213379 | 434 | 385 | show | 1 | | | 2013-02-20 01:43:32 |
| 7 | 4170450 | 3C48 | 24.4220808 | 33.1597594 | 8.6914062e-05 | 3.051758 | 64 | 1.00139 | 1899-12-31 00:00:00 | 35699.0 | 213379 | 487 | 445 | show | 1 | | | 2013-02-20 01:42:20 |
| 8 | 4170448 | 3C48 | 24.4220808 | 33.1597594 | 8.6523438e-05 | 3.051758 | 64 | 1.00139 | 1899-12-31 00:00:00 | 35699.0 | 213379 | 485 | 443 | show | 1 | | | 2013-02-20 01:37:36 |
| 9 | 4170441 | 3C48 | 24.4220808 | 33.1597594 | 8.515625e-05 | 3.051758 | 64 | 1.00139 | 1899-12-31 00:00:00 | 35699.0 | 213379 | 478 | 436 | show | 1 | | | 2013-02-20 01:36:52 |
| 10 | 4170432 | 3C48 | 24.4220808 | 33.1597594 | 8.3398438e-05 | 3.051758 | 64 | 1.00139 | 1899-12-31 00:00:00 | 35699.0 | 213379 | 469 | 427 | show | 1 | | | 2013-02-20 01:36:24 |
| 11 | 4170446 | 3C48 | 24.4220808 | 33.1597594 | 8.6132812e-05 | 3.051758 | 64 | 1.00139 | 1899-12-31 00:00:00 | 35699.0 | 213379 | 483 | 441 | show | 1 | | | 2013-02-20 01:36:15 |
| 12 | 4170351 | 3C48 | 24.4220808 | 33.1597594 | 6.3476562e-05 | 3.051758 | 64 | 1.00139 | 1899-12-31 00:00:00 | 35699.0 | 213379 | 388 | 325 | show | 1 | | | 2013-02-20 01:35:02 |
| 13 | 4170436 | 3C48 | 24.4220808 | 33.1597594 | 8.4179688e-05 | 3.051758 | 64 | 1.00139 | 1899-12-31 00:00:00 | 35699.0 | 213379 | 473 | 431 | show | 1 | | | 2013-02-20 01:34:26 |
| 14 | 4170444 | 3C48 | 24.4220808 | 33.1597594 | 8.5742188e-05 | 3.051758 | 64 | 1.00139 | 1899-12-31 00:00:00 | 35699.0 | 213379 | 481 | 439 | show | 1 | | | 2013-02-20 01:34:24 |
| 15 | 4170437 | 3C48 | 24.4220808 | 33.1597594 | 8.4375e-05 | 3.051758 | 64 | 1.00139 | 1899-12-31 00:00:00 | 35699.0 | 213379 | 474 | 432 | show | 1 | | | 2013-02-20 01:34:24 |
| 16 | 4170445 | 3C48 | 24.4220808 | 33.1597594 | 8.59375e-05 | 3.051758 | 64 | 1.00139 | 1899-12-31 00:00:00 | 35699.0 | 213379 | 482 | 440 | show | 1 | | | 2013-02-20 01:32:40 |
| 17 | 4170447 | 3C48 | 24.4220808 | 33.1597594 | 8.6328125e-05 | 3.051758 | 64 | 1.00139 | 1899-12-31 00:00:00 | 35699.0 | 213379 | 484 | 442 | show | 1 | | | 2013-02-20 01:31:34 |

The following file(s) are requested for download. You will receive an email when the files can be retrieved.

| Size | Filename |
|---|---|
| 43.0 GB | L94481_SAP001_SB476_uv.MS_203015f1.tar |
| 43.0 GB | L94481_SAP001_SB480_uv.MS_0b5e2a4b.tar |
| 43.0 GB | L94481_SAP001_SB486_uv.MS_22ee1140.tar |
| 129.0 GB | *total* |

Mail From: <noreply@astron.nl>

Close | Reply | Reply All | Forward | Read Later

Mail | Properties | Personalize | Message Source

From: <noreply@astron.nl> 2/3/2013 09:58 PM

To: Hanno Holties

Subject: Data ready for retrieval

Dear Hanno Holties,

Your data retrieval request with id 45 has been staged and is ready for retrieval.

List of files:
srm://lofar-srm.fz-juelich.de:8443/pnfs/fz-juelich.de/data/lofar/ops/LC0_002/L83093/L83093_SAP000_SB153_uv.MS.tar

The attached files can be used to retrieve the staged files.
For more information visit http://www.lofar.org/wiki/doku.php?id=public:lta_howto

This mail has been automatically generated by the ASTRON/LOFAR LTA staging service.
Do not reply to this message. If you have any questions or remarks, please
contact sciencesupport@astron.nl and provide the id of the request in your message.

| Name | Size | Type | Modified |
|---|---|---|---|
| Message | 4KB | Message Attachment | 2/3/2013 |
| html.txt | 187 Bytes | File Attachment | |
| srm.txt | 153 Bytes | File Attachment | |

- The LOFAR Archive stores data on magnetic tape. Data cannot be downloaded right away, but has to be copied from tape to disk first. This process is called 'staging'
- Limitations:
  - stage no more than 5 TB at a time and no more than 20000 files
  - Staging data from tape to disk might take some time since drives are shared with all users (also non-LOFAR) and requests are queued
  - Staging space is limited and shared between all LOFAR users – system might temporarily run low on disk space
  - Data copy remains on disk for 2 weeks
  - Maintenance and small outages experienced regularly

# PROCESSING IN THE LTA

LOFAR

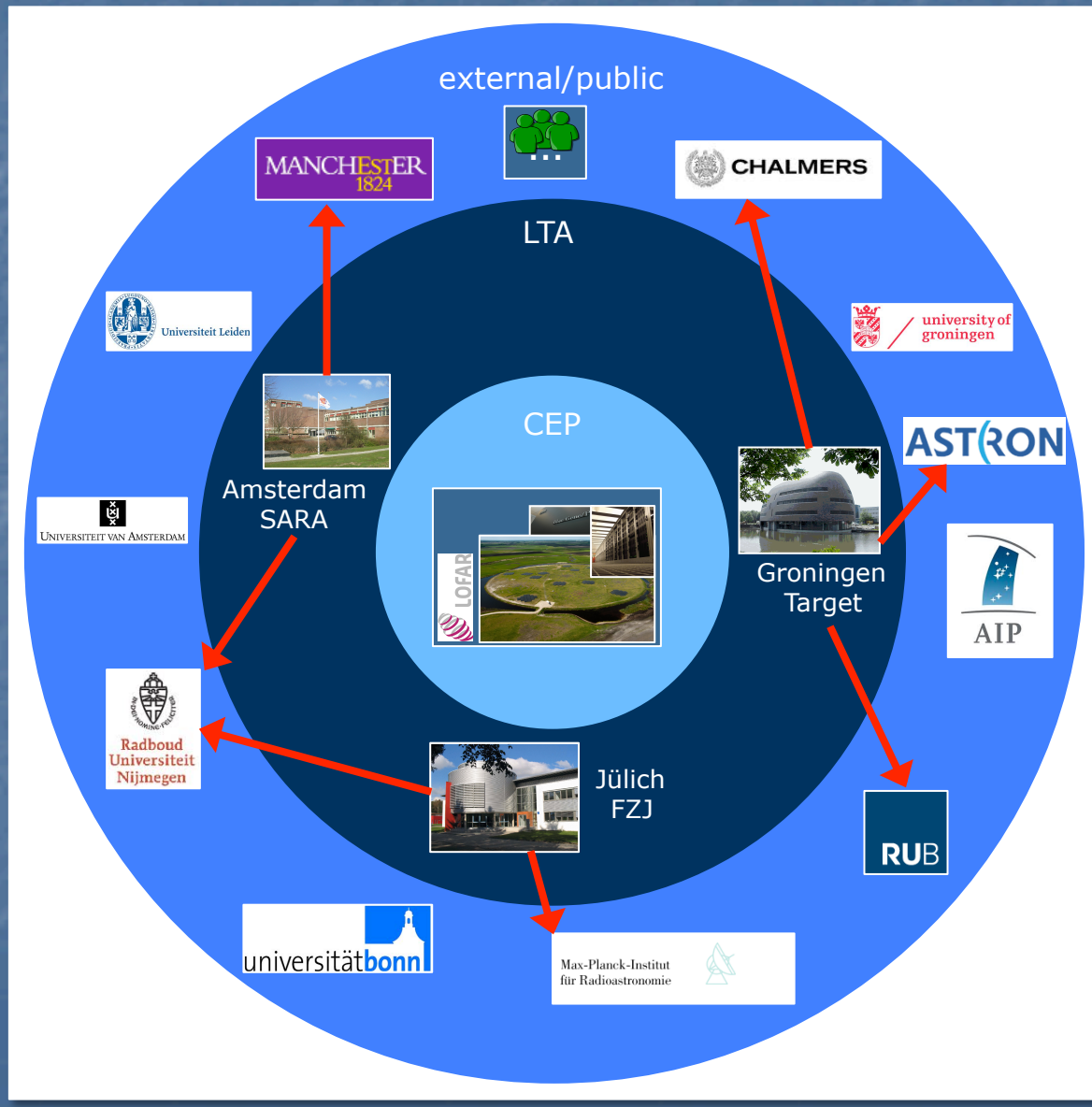
ASTRON

- **Use Processing resources at the LTA**
- **Service to LOFAR users**

  Standardized pipelines
  Integration with catalog & user interfaces
  Processing where the data is
  Hide complexity & inhomogeneity

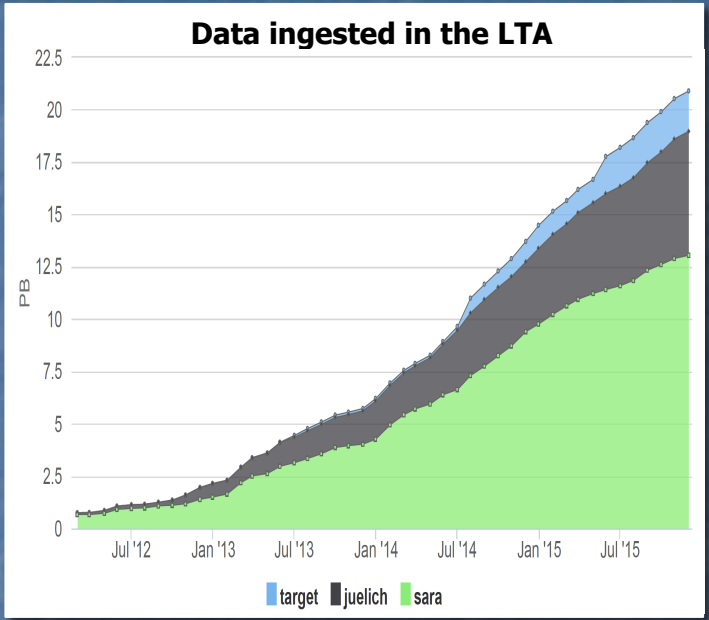
- **Expert users can**

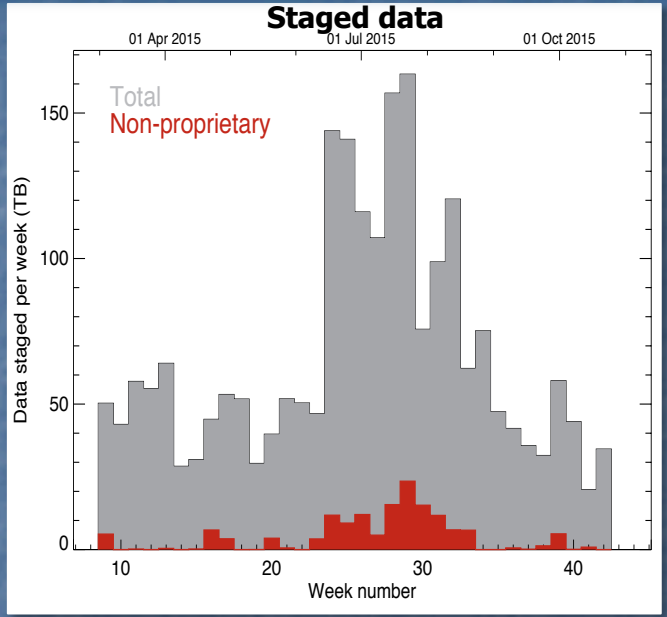
  Run custom software
  Use native protocols
  Optimize workload
  Build on integration with catalog
  - Queries
  - Ingest output including data lineage

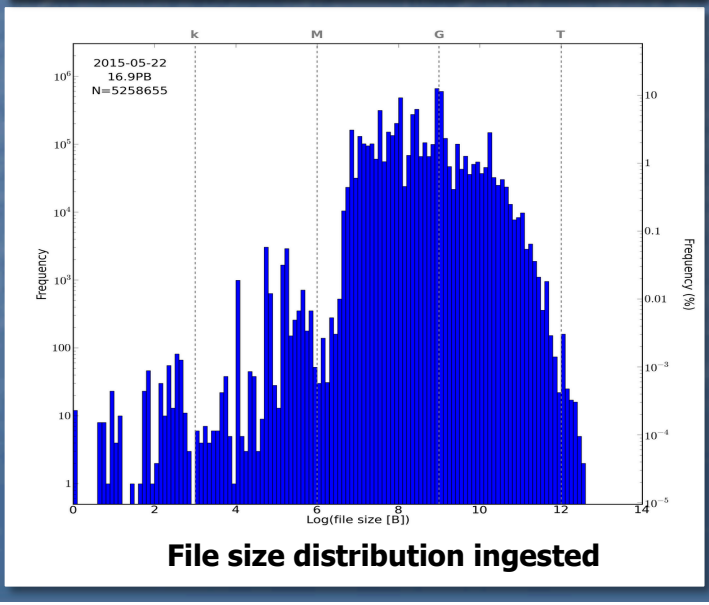


# DATA AT THE LTA

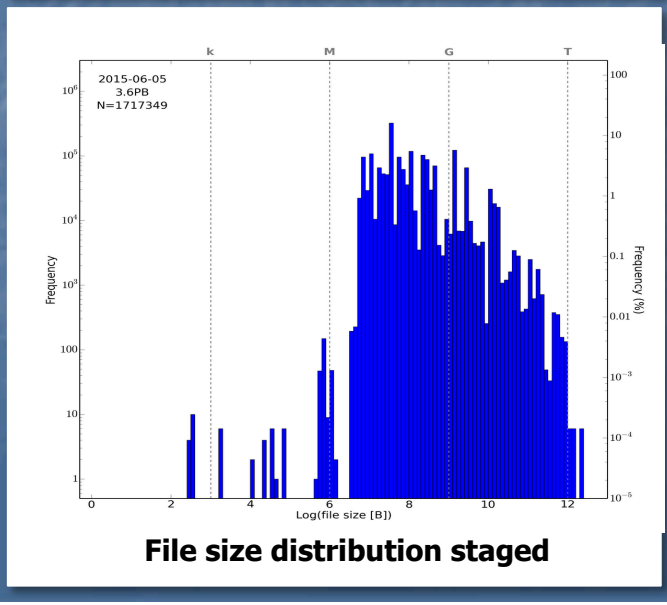
ASTRON




File size distribution ingested




File size distribution staged

- Exceeded 20 PB of data in the LTA!
- Current growth per year: 6 PB (and increasing!!)
- 5.5 million data products
- > 1 billion files

Courtesy of LOFAR LTA team: L. Cerrigone, J. Schaap, H. Holties, W. J. Vriend, Y. Grange

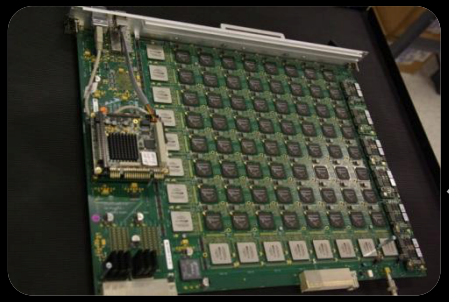
# SKA: A LEADING BIG DATA CHALLENGE FOR 2020

LOFAR

ASTRON

**Antennas**

**Digital Signal Processing (DSP)**



Transfer antennas to DSP
2020: 20,000 PBytes/day
2028: 200,000 PBytes/day

Over 10's to 1000's kms

**HPC Processing**
**2020: 300 PFlop**
**2028: 30 EFlop**

To Process is HPC
2020: 100 PBytes/day
2028: 10,000 PBytes/day

Over 10's to 1000's kms

**High Performance Computing Facility (HPC)**

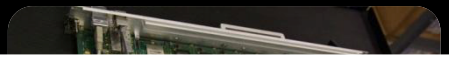
# SKA: A LEADING BIG DATA CHALLENGE FOR 2020

LOFAR

ASTRON

**Antennas**

**Digital Signal Processing (DSP)**



To Process is HPC
2020: 100 PBytes/day
2028: 10,000 PBytes/day

000's kms

| | LOFAR | SKA |
|---|---|---|
| Raw Telescope | 112 PB/yr | 60 EB/yr |
| Archive Rate | 6 PB/yr | 100 PB/yr |

**2020: 300 PFlop**
**2028: 30 EFlop**

**erformance Computing Facility (HPC)**

# SCIENCE REGIONAL CENTERS

LOFAR

ASTRON

## *Data oriented operations:*

- Data archiving and curation
- Data management, discovery, and access (VO compliant)
- Automated processing and reprocessing (calibration, imaging etc)
- Generation and storage of derived science products
- Continued pipeline and algorithmic development

## *Science oriented operations:*

- Portal-based data product access (raw and science-ready)
- Interface to data processing and reprocessing pipelines
- Interface to VO multi-wavelength discovery and analysis tools
- Support for development of custom user analysis (HPC-enabled)
- Development of new algorithms, science enabling tools, data analytics (visualisation, etc.)
- End-to-end astronomer support (proposal => science exploitation)
- Community education & outreach
- Face-to-face user support
- 24/7 help desk

# THANKS