LOFAR2.0 Large Programme Proposals
Data Management
Capabilities

	Organisatie / Organization	Datum / Date
Auteur(s) / Author(s):		
Roberto Pizzo	ASTRON	2023-06-07
John D. Swinbank	ASTRON	
Irene Bonati	ILT	
Controle / Checked:		
...	...	...
Goedkeuring / Approval:		
...	...	...
Autorisatie / Authorization:		
...	...	...
Handtekening / Signature:		

Doc. Nr.:   SDC-021
Rev.:   1.4
Date: 2023-06-07
Doc. Type: Note

Document History

Revision	Date	Description
1.4	2023-06-07	Incorporate more ILT-requested changes
1.3	2023-06-07	Incorporate further ILT-requested changes
1.2	2023-06-06	Incorporate feedback from the ILT Board
1.1	2023-05-31	Incorporate feedback from Irene Bonati
1.0	2023-05-22	Version to accompany call for proposals
0.1	2023-03-24	Initial draft

List of Abbreviations

CEP CEntral Processing.

ERIC European Research Infrastructure Consortium.

FAIR Findable, Accessible, Interoperable, Reusable.

HBA High-Band Antenna.

ICD Interface Control Document.

ILT International LOFAR Telescope.

LBA Low-Band Antenna.

LINC LOFAR INitial Calibration.

LOFAR LOw Frequency ARray.

LTA Long Term Archive.

OSI Open Source Initiative.

PULP PULsar Pipeline.

SDC Science Data Centre.

1 Introduction

This document, which accompanies the call for LOFAR2.0 Large Programme proposals, summarizes the services that will be allocated by the ILT Foundation — and, in future, LOFAR ERIC — to process, archive, and distribute LOFAR2.0 Data Products. These services, deriving from development effort, operational activities, and infrastructure capacity contributed by various partners, will be provided to end users under the management of the ASTRON Science Data Centre (SDC). It is a practical application, for the LOFAR2.0 Large Programme Call, of the LOFAR ERIC Data Policy, and is intended to act as a reference for those teams submitting proposals; it is expected that those proposals will refer directly to the services described here, and proposal teams are asked to indicate which of these capabilities they expect to build upon or otherwise use.

This document is structured in two parts. First, we establish some basic obligations and expectations regarding data and software management in the LOFAR2.0 scientific community. These principles are intended to ensure fair access to resources for all LOFAR users (or potential users), and to maximise the overall scientific impact of the infrastructure.

The second part of this document then describes the capabilities that will be available to Large Programmes and other users of the LOFAR2.0 system. We address these in four parts: data archiving and curation; scientific pipelines; data processing; and data discovery and access. Note that the development of these systems is still underway: this document describes a system that we believe is both plausible and useful in unlocking the scientific capabilities of the instrument, but that requires ongoing effort to realise. We actively solicit assistance from the wider community in achieving these goals.

2 Obligations & Expectations

All users of LOFAR2.0 data are expected to comply with the following principles:

- Software that is is developed to process or otherwise analyse LOFAR data which results in a publication should be made available under an OSI-approved licence.[1] Where possible, we encourage (but do not require) compliance with the FAIR Principles for Research Software.[2]

- Advanced or science-ready data products produced by Large Programme teams, or resulting from open skies observations, should be made available to the widest possible scientific community by publishing them through the LOFAR Long Term Archive (LTA), insofar as this is compatible with the applicable proprietary period or other data rights issues. If those products are generated outside LTA-managed infrastructure, the ILT aims to enable and assist the science team with LTA ingest, including ensuring that they comply with applicable ICDs or other standards.

---

[1]https://opensource.org/licenses/
[2]https://zenodo.org/record/6623556#.YqCJTJNBwlw

SDC

Doc. Nr.:	SDC-021
Rev.:	1.4
Date:	2023-06-07
Doc. Type:	Note

3 System Capabilities

3.1 Data Archiving

The LTA will be made available in support of LOFAR2.0 operations. It will be extended to provide for the curation for a range of advanced data products (e.g. calibrated visibilities, image cubes).

In order to maximise the utility of the archive, the LTA, as a scarce resource, will be allocated on a merit basis. Advanced data products (e.g. images, calibration solutions) are envisaged to be stored indefinitely. For most observing modes, raw data will not be stored. Instrumental and intermediate products (e.g. flagged and compressed visibilities, together with raw data for expert observing modes), will be retained for a period of around 18 months, depending on resource availability; the ILT will work with its partners and the Large Programme teams to ensure that they collectively provide the resources needed to optimise the scientific output of LOFAR. Exceptions to this policy will be considered in special cases (e.g. data with unique scientific value that cannot be reproduced through re-observation).

3.2 Availability of Scientific Pipeline

A core set of imaging pipeline components — such as DP3, WSClean, etc — will (continue to) be provided and supported. The list of components available may evolve with time, as new components are delivered and obsolete components retired; such changes will be carried out in consultation with, and with appropriate notice to, the widest possible user community.

Pipelines for performing preprocessing (flagging, averaging, compression, subtraction of bright sources) and direction-independent calibration and imaging (the LINC pipeline) will be provided. These pipelines will be based on the core pipeline components. In addition to being available for download and use on other systems, these pipelines will be integrated with the LOFAR data processing and archiving system in such a way that operators can use them to process data coming from the telescope (see next section) and/or stored in the archive.

A pipeline for performing direction-dependent calibration and imaging (the Rapthor pipeline) will be provided. This pipeline will aim to provide imaging of both Dutch and international baselines in HBA and LBA. However, we regard this as an ongoing research problem, and are aware that different science goals may require a range of imaging fidelity vs. performance or other trade-offs. As such, the Rapthor pipeline will prioritise addressing the widest possible range of science cases and providing a default "science grade" imaging capability to the general user community; it may be necessary for some communities with specialist needs to provide their own tooling beyond Rapthor. We welcome joint development efforts, and hope that those communities will work with Team Rapthor to incorporate their changes into the mainstream Rapthor pipeline wherever possible.

The PULP pipeline, used for processing pulsar timing data, will remain a community developed and maintained project. Best-effort support will be provided for it by SDC and ILT staff. No explicit support will be available for the processing of pulsar survey data.

Increased engagement with developing and testing pipelines from the wider user community will dramatically improve the prospects of successfully delivering pipelines that meet the goals described

above. While ad-hoc contributions may have value, we actively invite and encourage programme teams to dedicate developers and scientists with appropriate skills (e.g. radio astronomy algorithms, C++ development) to work as part of our established agile development process.

3.3   Data Processing

The preprocessing pipeline will be executed on the LOFAR central processing system (CEP), and the resulting intermediate made available for archiving (subject to retention policies described above).

The LOFAR data processing system at the LTA will have the capability to execute the LINC and Rapthor pipelines described above. While no dedicated compute time can be guaranteed for executing these pipelines at the LTA, the ILT will use various resources and budget acquisition opportunities accessible to partners, and, in future, to LOFAR ERIC to obtain structural processing allocations at LTA sites and other sites. We will proactively engage with infrastructure providers and science teams and attempt to coordinate processing in a way that maximises the total benefit to both Large Programme teams and open skies observations.

Pulsar timing data will be processed under SDC management on CEP, as is currently the case for LOFAR1. No infrastructure will be provided for pulsar survey data processing.

3.4   Data Discovery and Access

Advanced data products — including advanced data products ingested from external processing by science teams, as described above — will be re-ingested and made available through the LTA.

All data products stored in the LTA will be made available through Virtual Observatory-compliant interfaces.

The LTA system will be upgraded to support the FAIR data management principles.[3]

A new data discovery front-end, provisionally known as ADEX, will provide an improved and more accessible user experience for accessing science-ready and advanced products than is available through the current LTA portal. This will provide data discovery and visualization facilities that are broadly comparable to those provided by other major archives.

---

[3]https://www.nature.com/articles/sdata201618