

SCIENCE DATA POLICY OF LOFAR ERIC

This version was approved by the Interim LOFAR ERIC Council on 25 April 2023.

Version history:

Version	Date	Comment	Prepared by
0.1	26-04-2022	<i>Bullets and principles - Presented to and endorsed by the ILEC on 11-05-2022</i>	<i>I. Bonati R. Vermeulen</i>
0.2	15-03-2023	<i>Full document</i>	<i>I. Bonati</i>
0.3	23-03-2023	<i>Addressed comments by ILT ERIC WG</i>	<i>I. Bonati</i>
0.4	07-04-2023	<i>Addressed comments by the ILT Board</i>	<i>I. Bonati R. Vermeulen</i>
1.0	25-04-2023	<i>Approved by ILEC</i>	

SCIENCE DATA POLICY OF LOFAR ERIC

This version was approved by the Interim LOFAR ERIC Council on 25 April 2023.

PREAMBLE

This document details the principles of the science data policy of LOFAR ERIC, in line with Article 32 of the LOFAR ERIC Statutes. The policy outlines the ownership and access to scientific data acquired, managed, or created through research by or involving LOFAR ERIC. It applies to users contributing to LOFAR ERIC or using data supplied through LOFAR ERIC.

The aim of this policy is to ensure open and easy access to the stored LOFAR ERIC science data and to conserve these data so as to maximise their overall long-term science yield within reasonable technical and budgetary means.

DEFINITIONS

Open Science: Practice of providing unhindered access to scientific articles and data from public research through information and communication technology (ICT) tools and incentives.

Open Access: Practice of providing online access to data that is free to and reusable by the end-user.

FAIR data: Data meeting the principles of findability, accessibility, interoperability, and reusability (FAIR).

LOFAR Long Term Archive (LOFAR LTA): The overarching archive infrastructure under which the science data owned by or under the stewardship of LOFAR ERIC are conserved and made available to the community with the aim of maximal science use. The LOFAR LTA has a single logical database and a uniform interface, and comprises data stored at several physical locations, typically data centres where also processing capacity is available next to storage capacity (see the LTA Uniform Conditions). As of March 2023, these are Samenwerkende Universitaire RekenFaciliteiten (SURF) in Amsterdam (the Netherlands), Forschungszentrum Jülich (FZJ) in Jülich (Germany), and Póznán Supercomputing and Networking Center (PSNC) in Póznán (Poland).

Metadata: A set of descriptive information providing the context, content, and structure of the science data and their management through time. Observational metadata include, for example, information on sky coverage, frequency coverage, antenna sets, observation date(s) and start/end times, and information about the user group and the project that originated the data.

Standard LOFAR pipeline: A set of software modules documented and maintained by LOFAR ERIC, for use in LOFAR ERIC operations, and/or offered to the science user community for processing LOFAR ERIC (observing) data to a more advanced state of science readiness.

Standard observed and processed data: Data products resulting from allocated LOFAR ERIC science observations, as well as processed data products obtained by standard LOFAR pipelines running on any LOFAR ERIC (post-)processing machine.

Derived data: Data obtained by using tools beyond the standard LOFAR pipelines. These data allows for significant added or modified science use, and are obtained through further post-processing of LOFAR data using non-standard software developed and maintained by a user or user group, and/or making use of non-LOFAR-ERIC facilities.

Stand-alone observing data: Data originating from non-LOFAR ERIC observations, during stand-alone operation of a LOFAR station by its owner (as regulated in the Uniform Conditions for LOFAR stations and in the Collaboration Agreements between LOFAR ERIC and its Collaborating Organisations).

Optional Activity and add-on-facility data: Data originating in the course of LOFAR ERIC Optional Activities (see the Financial Plan of LOFAR ERIC) or resulting from observations made possible through a separate agreement.

Sub-array mode: Observing mode where a group of one or more LOFAR stations is assigned to a specific data observing project, for which the station data are transmitted to LOFAR ERIC processing resources.

Single-station mode: Observing mode where observations are carried out under central LOFAR ERIC control, in which the data are initially collected and processed on a per-station basis, as opposed to either being cross-correlated between stations in real time (for interferometry), or being used online for (sub)array beam-forming. This observing mode refers to all modes using standard LOFAR functionality as well as any add-on facilities available for a particular station, as contractually agreed with LOFAR ERIC.

Stand-alone mode: Private use of LOFAR stations by their owners, guaranteed for a specified minimum fraction of time (as regulated in the Uniform Conditions for LOFAR stations and in the Collaboration Agreements between LOFAR ERIC and its Collaborating Organisations). In this mode, operational control of a specific station is explicitly passed to the local owner, and handed back to LOFAR ERIC Operations at the end of a pre-arranged time block. Use of a station in its stand-alone periods is fully outside of LOFAR ERIC control.

1. GENERAL PRINCIPLES

- (1.1) LOFAR ERIC supports the integrity, transparency, and openness of science data. With this policy, LOFAR ERIC therefore undertakes to adhere to the principles of Open Science, Open Access, and FAIR data, as offered by the EU open science

policy (https://research-and-innovation.ec.europa.eu/strategy/strategy-2020-2024/our-digital-future/open-science_en).

- (1.2) LOFAR ERIC recognises that science must be grounded in a common culture of data stewardship, with appropriate curation of data throughout and after the period during which the research is conducted.
- (1.3) LOFAR ERIC promotes data interoperability and standardisation, which facilitates working with large volumes of diverse science data.
- (1.4) LOFAR ERIC shall provide guidance to users to ensure that research carried out using material made accessible through LOFAR ERIC is undertaken within a framework that recognizes the rights of science data owners and privacy of individuals.
- (1.5) For the purposes of ensuring and measuring the impact of LOFAR ERIC on the full (science) community in accordance with the mission of LOFAR ERIC, all persons accessing the LOFAR LTA will need to register, and agree to this Data Policy the LOFAR LTA terms and conditions governing access.
- (1.6) Following the principle in Article 1.4 of the LOFAR ERIC Statutes, stating LOFAR ERIC pursues its activities on a non-economic basis, public access to data regulated in this policy is granted for non-commercial use only. If in line with its principal activities, commercial use of LOFAR ERIC data will be regulated through the enforcement of special licenses.

2. THE LOFAR LONG TERM ARCHIVE (LTA)

- (2.1) The distributed LOFAR LTA, operated by LOFAR ERIC, stores science data obtained through observations carried out using LOFAR ERIC facilities or obtained by standard LOFAR pipeline processing, as well as other operational and development data.
- (2.2) Science data shall be stored and made accessible alongside the relevant metadata, subject to General Data Protection Regulation (GDPR) limits, for optimal traceability and usage by users.
- (2.3) The overall available LOFAR LTA resources include capacity for data storage, data retrieval, and associated connectivity. Access to LOFAR LTA services, including use of LTA capacity, is governed by the LOFAR ERIC User Access policy. Typically, specific parts of the available LTA resources will be allocated to enable processing of data from specific allocated observing (user) projects, while other parts of the LTA capacity are assigned to long-term goals, that transcend specific (user) projects and relate to maximal science impact and conserving data for open science, open access usage. Detailed data storage policies and practices, such as data retention in various steps of curation, are set and reviewed as needed based on evolving technical capabilities, and on the assembled resources, taking into account a total operations

budget balanced for maximal science productivity of LOFAR ERIC. These practices are documented online.

- (2.4) Science datasets and relevant metadata stored in the LOFAR LTA are catalogued and become publicly accessible on a pre-defined timescale, in accordance with specific proprietary rights conferred during allocation of the observing time, or as part of the agreement on the storage of these data.
- (2.5) The LOFAR LTA will have suitable security arrangements in place regarding internal storage and handling. In case of allegations of security breaches and confidentiality disclosures regarding research data, the LOFAR ERIC Director will set up a suitable investigation and conclude on appropriate measures in accordance with international scientific practices for such cases, with ultimate recourse possible to the LOFAR ERIC Council.

3. DATA OWNERSHIP AND ACKNOWLEDGEMENT

3.1 Acknowledgement and notification

- (3.1.1) Any resulting publication using LOFAR ERIC science data of any kind (standard observed and processed data, derived data, stand-alone data, or Optional Activity and add-on facility data) must acknowledge the dependence on LOFAR ERIC standard data or facilities, as provided on the website of LOFAR ERIC.
- (3.1.2) When publishing research papers or producing any output, whose results were obtained by making use of LOFAR ERIC science data of any kind or LOFAR ERIC facilities, users or user groups will notify LOFAR ERIC to ensure the efficient tracking of publications.

3.2 Standard observed and processed data

- (3.2.1) All standard data products resulting from allocated LOFAR ERIC science observations (whether interferometric, beam-formed, single-station, or otherwise) remain the property of LOFAR ERIC.
- (3.2.2) User groups may not disseminate, beyond their specific collaboration, standard observed or processed data stored in the LOFAR LTA, neither within the proprietary time nor later, apart from exceptional circumstances to be agreed in advance. In particular, user groups should avoid having any form of separate public repository for such data, but should always refer to the LOFAR LTA.
- (3.2.3) If LOFAR ERIC decides to terminate the storage of some standard observed or processed datasets in the form necessary to enable further science analysis, LOFAR ERIC will contact the relevant user group and offer an opportunity for them to arrange alternative storage under their own responsibility. In such cases, they may

disseminate these data as they see fit; however, a way for the data to be referenced in and accessible through the LOFAR LTA interface must be arranged with LOFAR ERIC.

3.3 Derived data

- (3.3.1) Derived data is property of the user(s) who produced the data. A user or user group may choose to make this data public, either under their own arrangements or by making arrangements with LOFAR ERIC on access through the LOFAR LTA, or to keep it private, unless delivery to the LOFAR LTA has been stipulated as a condition for granting observing time, proprietary access to standard data, or LOFAR ERIC processing time.
- (3.3.2) When the LOFAR LTA accommodates the storage of derived data, this will by default become public immediately; a proprietary period may however have been stipulated during allocation of the observing that produced the input data, or when the arrangement on the storage of the derived data was agreed.
- (3.3.3) If the derived data is not stored in the LOFAR LTA, an appropriately framed referral from the LOFAR LTA logical database to data stored and maintained elsewhere by the user or user group which derived the data may be arranged.

3.4 Stand-alone observing data

- (3.4.1) For any data originating from non-LOFAR-ERIC observing, during stand-alone operation of a LOFAR station by its local owner, the station owner decides on the ownership of stand-alone observing data.

3.5 Optional Activity and add-on facilities data

- (3.5.1) Data ownership and storage of Optional Activities and add-on facilities data will by default follow the regulations tied to a given Optional Activity or a separate agreement.

4. DATA USE AND PROPRIETARY RIGHTS

Exclusive "proprietary use" rights to specific data in the LOFAR LTA may be granted to specific users for specific lengths of time for specific purposes, in accordance with the regulations below.

4.1 Standard proprietary period

- (4.1.1) By default, the team to which observing resources are granted as a result of a successful proposal shall have the exclusive right to use the data within the confines of the science purposes for which it was collected during a standard proprietary period of 1 year after observation.

- (4.1.2) The standard proprietary period starts by default after the last data instalment of a full project in the period over which observing was allocated has been made available to a project's Principal Investigator and are ready to be retrieved in the LOFAR LTA.
- (4.1.3) The standard proprietary period applies to data observed in the period over which observing was allocated, even for projects with long-term allocations.
- (4.1.4) Further data post-processing may happen at some later time on LOFAR ERIC facilities, subject to allocation of processing resources. By default this does not extend the proprietary period on the original data already stored in the LOFAR LTA, and the derived data fall under the stipulations of 3.3.2.

4.2 Non-standard proprietary period

- (4.2.1) Any proposal may contain a motivated request for non-standard data use rights. Extensions to the standard proprietary period may be granted:
- On its own initiative or prompted by a proposal, in order to promote the widest possible science use of LOFAR data and to optimise the overall science output from LOFAR ERIC, taking due account of the interests of the proposing user group;
 - In exceptional scientific circumstances or personal reasons, such as maternity or parental leave.
- (4.2.2) Time allocators may consider making LOFAR data for larger projects public without proprietary period or with a reduced proprietary period. The time allocators may also stipulate that data must be shared, even while still in its proprietary period, by different science projects proposed simultaneously or successively. More generally, the time allocators may make stipulations that delimit the scope or nature of observations and/or use of data to a specific science topic, when these are clearly beneficial to optimise the overall science output from LOFAR ERIC, or to create viable options or Open Skies and Reserved Access Shares.
- (4.2.3) Once the proprietary period has elapsed, and unless otherwise decided, LOFAR data shall be publicly accessible.

5. DATA HANDLING DURING OBSERVATIONS

5.1 Regular observing (sub-array, single-station)

- (5.1.1) For observations carried out using regular configurations (including sub-array and single-station mode), LOFAR ERIC allocates and schedules the observing time, and arranges for data transport, processing, and storage.

5.2 Stand-alone observing

- (5.2.1) For stand-alone observations, local station owners must make private arrangements for stand-alone data transport, processing, and storage. Data from stand-alone

operation periods are not by default entered into the LOFAR LTA; their distribution and use rights remain at the discretion of the local station owner.

6. REVIEW

This Data Policy shall be reviewed as needed and at least once every 5 years.

7. ENTRY INTO FORCE

This Data Policy is established and maintained as needed by the LOFAR ERIC Council.