

DATA MANAGEMENT PLAN of the INFNO-LNGS research infrastructure

DMP_INFNO-LNGS_RI_v1.0

Effective from dd.mm.yyyy

Distribution list: INFNO-LNGS

Author	Role - Affiliation
Francesca Marchegiani	Technologist - INFNO-LNGS
Reviewer/s	
Carlo Bucci	Head of Research Division - INFNO-LNGS
Sandra Parlato	Head of Computing and Networking Service - INFNO-LNGS
Marcello Messina	CSN2 Coordinator - INFNO-LNGS
Federico Ferraro	CSN3 Coordinator - INFNO-LNGS
Giulia Pagliaroli	CSN4 Coordinator - INFNO-LNGS
Matthias Laubenstein	CSN5 Coordinator - INFNO-LNGS
Approver	
Ezio Previtali	Director - INFNO-LNGS

Table of contents

1. Document Control Sheet	3
2. Revision history	3
3. List of acronyms and abbreviations	4
4. Definitions	5
5. Introduction.....	12
6. FAIR data.....	14
7. Data and metadata	15
7.1 Data and metadata sources.....	15
7.1.1 Data and metadata coming from the INFN-LNGS research division.	15
7.1.2 Data and metadata coming from the INFN-LNGS technical and general services division.	15
7.1.3 Data and metadata coming from physics experiments.....	15
7.1.4 Data and metadata coming from activities performed with third parties.....	15
7.2. Data types.....	16
7.3. Data processing levels.....	16
8. Agreements on rights of use	17
9. Ownership of data	17
10. Data curation	17
11. Data handling, storage, backup, and preservation strategies	18
12. Data sharing and access.....	21
13. Data integrity.....	22
14. Data archiving.....	23
15. Personal data.....	24
16. Responsibilities and resources	24
17. Relevant policies and guidelines	25
Figure 1. Data lifecycle	13
Figure 2. FAIR data principles and data compliance check	14
Figure 3. Structure of data storage.....	20

DocID	Rev.	Status
DMP_INFNO-LNGS_RI	1.0	Draft

1. Document Control Sheet

Document	
Title	Data management plan of INFNO-LNGS research infrastructure
ID	DMP_INFNO-LNGS_RI
Version	V1.0
Effective from	dd.mm.yyyy
Available at	https://www.infn.it/.....
File name	DMP_INFNO-LNGS_RI_v1.0.pdf

2. Revision history

Version	Date	Author	Main changes
v1.0	06.09.2024	Francesca Marchegiani	Initial version

3. List of acronyms and abbreviations

Acronyms/abbreviations	Full text
CC-BY	Creative Commons attribution
CCR	National INFN Computing & Networking Committee
DMP	Data Management Plan
DOI	Digital Object identifier
DS	Data steward
EU	European Union
FAIR	Findable, Accessible, Interoperable, Reusable
INFN	National Institute for Nuclear Physics
IT	Information Technology
LNGS	Gran Sasso National Laboratory
MoU	Memorandum of Understanding
OAR	Open Access Repository
PI	Principal Investigator
PID	Persistent identifier
RI	Research Infrastructure
VM	Virtual Machine

DocID	Rev.	Status
DMP_INFNO-LNGS-RI	1.0	Draft

4. Definitions

Access

“Access” refers to the legitimate and authorised physical, remote and virtual admission to, interactions with and use of Research Infrastructures and to services offered by Research Infrastructures to Users. Such Access can be granted, amongst others, to machine time, computing resources, software, data, data-communication services, trust and authentication services, sample preparation, archives, collections, the set-up, execution and decommissioning of experiments, education and training, expert support and analytical services. Access authorization to IT resources is granted by the Structure Director or by his delegate for a limited period of time, not exceeding the duration of their professional duties within INFN.

Copyright owner

Co-ownership agreements could be signed in case of research teams or activities with third party companies. Otherwise specified the copyright owner is INFN.

Data

Research data (hereinafter “data”) are any information that has been collected, observed, generated or created to validate original research findings¹. Data can come in various forms and types according to the discipline of research. For example, data can be quantitative or qualitative statements. Data can include (but are not limited to): documents, spreadsheets, e-logbook, laboratory notebooks, field notebooks, diaries, questionnaires, transcripts, codebooks, audiotapes, videotapes, photographs, films, test responses, slides, artefacts, specimens, samples, collections of digital outputs, data files, database contents (video, audio, text, images), software package, models, algorithms, scripts, contents of an application (input, output, logfiles for analysis software, simulation software, schemas), methodologies and workflows, standard operating procedures and protocols, scientific papers, README files. The definition includes also blueprints, topographical, architectural and structural plans, technical reports able to describe plants, buildings and infrastructure.

In this context, “data” pertains to data collected using the INFN-LNGS-RI within the activities performed by both research division and technical & general services division or within the physics experiments hosted in underground or aboveground laboratories, as well as activities with third parties.

Database

¹

https://library.leeds.ac.uk/info/14062/research_data_management/61/research_data_management_explained

DocID	Rev.	Status
DMP_INFNO-LNGS_RI	1.0	Draft

Collection of independent works, data or other materials arranged in a systematic or methodical way and individually accessible by electronic or other means².

Data creator

The staff that creates or generates data.

Data steward (DS)

A person providing disciplinary support (develops appropriate solutions, answers questions, provides advice and help) for research data management and sharing before, during and after a research project by applying the FAIR data principles.

Digital object identifier (DOI)

Unique, long-term identifier allowing the identification of a data record. This identifier will be created by INFN Open Science Working Group as routinely done for the sharing of products on the Open Access Repository (OAR).

Embargo

Period after which the data sharing and access is allowed to the public according to a licence of use. Some kinds of data may not be immediately sharable due to the nature of records themselves or to comply with (third parties) agreements. This refers to materials necessary to be held confidential by a researcher until they are published or information which is protected under law or the disclosure of which would constitute an infringement of agreement.

Facility

Scientific research facility (hereinafter facility) means buildings, services, and structures, including equipment and laboratories, used or to be used to conduct scientific research and to collaborate with the business community, industries, and other knowledge institutions. The INFN-LNGS-RI has at its disposal the following research facilities: Bellotti Ion Beam Facility, Cryo-Platform facility, STELLA laboratory and NOA clean room Radon free facility.

FAIR data principles

Set of principles, guidelines, and best practices to ensure that research data is Findable, Accessible, Interoperable and Re-usable, in compliance with ethical, commercial and confidentiality restrictions and the principle "as open as possible and closed only when necessary".

INFN Open Access Repository (INFN OAR)

² Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases.

DocID	Rev.	Status
DMP_INFNO-LNGS_RI	1.0	Draft

It is the INFN's institutional repository able to collect, preserve and disseminate digital copies of the research outputs of INFN research institution.

Information technology resources

Information technology resources include computing facilities and network services considered as a whole. The terms computing facilities and network services refer to:

- computers and similar electronic devices, printers and other devices (e.g. scanner and storage systems) belonging to the Institute or anyhow connected to its network.
- equipment and networking infrastructure belonging to the Institute or anyhow connected to the Institute's network.
- local and geographical networking service apart from the geographical access guaranteed by agreements between Institutions and Federations (e.g. Eduroam).
- virtual machine instances or networking equipment.
- software and data purchased, produced, or published by INFN.

Individuals using information technology resources belonging to INFN can be identified as:

- User
- User group contact person
- System administrator
- Computing and Networking service
- Structure director

Licence

Licence³ is an agreement by means of which the copyright owner defines the rights granted to the different stakeholders. In a digital environment, standardised licences based on a set of predefined reuse conditions, such as Creative Commons, are used. Licences for digital objects are machine readable.

The practice of open science is closely linked to the dissemination of the research outputs to other scientists or stakeholders so open licenses like CC-BY, CC-BY-SA, CC0 or equivalent are required, while metadata deposited must be open under the CC0 or equivalent licence⁴.

With creative commons licences all rights are retained by the owner of data, but permission to others to distribute and share the scientific outputs is given.

Memorandum of understanding (MoU)

³ <https://www.openaire.eu/guides/>

⁴ Creative Commons licenses. <https://creativecommons.org/share-your-work/> (URL accessed on November 17, 2023)

DocID	Rev.	Status
DMP_INFNO-LNGS_RI	1.0	Draft

A memorandum of understanding (MoU) is defined as a document that outlines the terms and understanding between two (bilateral) or more parties (multilateral) who intend to enter into a formal agreement or partnership. Objectives, parties involved, expectations, roles and responsibilities, timeline, confidentiality, intellectual property, are some of the key elements to include in this document.

Metadata

“Metadata” describes information pertaining to data collected from experimental instruments, including (but not limited to) the context of the experiment, the research team, experimental conditions, and other logistical information to support data discovery, identification and facilitate interoperability. They provide details about who, what, where, when, why and how the data were collected, processed and interpreted.

Metadata may include, for instance, title, abstract, keywords, language, author, subject, description, publisher, contributor, publication date, research context, analytical method, details about the experiment, file formats, sizes and types, version, persistent identifier, copyright information, licensing information, access privileges. To ensure correct and proper use and interpretation of the data by the owner and users, the use of metadata standards is required.

Different disciplines develop and adopt various metadata standards and/or practices for the management of their research data and materials.

A curated registry of commonly used domain-specific metadata standards can be found at FAIRsharing.org⁵ or at the Digital curation centre (DCC) website⁶.

Some examples are reported below:

OpenAIRE (OAI-PMH v2.0): the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) provides an application-independent interoperability framework based on metadata harvesting⁷.

Dublin Core: a basic, domain-agnostic standard which can be easily understood and implemented, and as such is one of the best known and most widely used metadata standards. Sponsored by the Dublin Core Metadata Initiative, Dublin Core was published as ISO Standard 15836-1 in 2017 and ISO Standard 15836-2 in 2019.

DataCite Metadata Schema: a set of mandatory metadata that must be registered with the DataCite Metadata Store when minting a DOI persistent identifier for a dataset. The domain-agnostic properties were chosen for their ability to aid in accurate and consistent identification of data for citation and retrieval purposes. Sponsored by the DataCite consortium, version 4.5 was recently released in 2024.

⁵ <https://fairsharing.org/search?fairsharingRegistry=Standard>

⁶ <https://www.dcc.ac.uk/guidance/standards/metadata/list>

⁷ <https://www.openarchives.org/OAI/2.0/openarchivesprotocol.htm>

Monte Carlo simulation

It is a mathematical technique used to estimate the possible outcomes of an uncertain event.

Moral rights

Moral rights are rights exclusively connected to the author of the work, in this context to the data creator.

Persistent identifier (PID)

A persistent identifier⁸ is a long-lasting reference to a resource. That resource might be a publication, dataset, or person. Equally it could be a scientific sample, funding body, set of geographical coordinates, unpublished report, or piece of software. Whatever it is, the primary purpose of the PID is to provide the information required to reliably identify, verify, and locate it. A PID may be connected to a set of metadata describing an item rather than to the item itself.

There are different PID types for different kinds of resources. In the current research environment, we most commonly see two varieties: those for objects (research organization, publications, data, software, such as RORs, URNs, DOIs, ARKs, Handle) and those for people (researchers, authors, contributors, such as ORCIDs, ISNIs).

Principal investigator (PI)

The PI is the researcher identified in the research proposal with scientific responsibility for the project (for EU projects – for sponsored projects) or the spokesperson of an experiment.

Proprietary research

Proprietary research is defined as that for which users request confidentiality of proposal, data and results for a certain period. This research is executed by signing an agreement in which clear guidelines on how data and metadata will be managed are reported. In particular, the agreement between parties (third party or research team and INFNO-LNGS-RI) should provide a description about copyright, embargo, licence, long-term storage, and data curation.

Raw data

Data acquired automatically or manually, without any processing level.

Research data management policy

A research data management policy defines data management roles and responsibilities of the different stakeholders within the INFNO-LNGS-RI.

⁸ <https://www.openaire.eu/guides/>

DocID	Rev.	Status
DMP_INFNO-LNGS_RI	1.0	Draft

Research Infrastructures

“Research Infrastructures”⁹ are facilities, resources and services that are used by the research communities to conduct research and foster innovation in their fields. They may include major scientific equipment (or sets of instruments), knowledge-based resources such as collections, archives and data, e-infrastructures, such as data and computing systems and communication networks and any other tools that are essential to achieve excellence in research and innovation. They may be 'single-sited', 'virtual' and 'distributed'.

Research team

The term “Research Team” includes the PI/PIs, all the staff involved in research as well as international collaborations.

Service

Operating units constituting a division (e.g. LNGS Computing and Networking service belongs to the Research Division as reported in the INFNO-LNGS organization chart).

System administrator

A system administrator is a professional who manages data processing system, including database systems, local area networks and security equipment. System administrators can be:

- LNGS Computing and Networking service staff
- A user group contact person who has administrator privileges and is appointed by the Structure Director
- The Structure Director, who appoints a user group contact person for each experiment or working group and also serves as the administrator of the experiment’s computing and network resources at LNGS. Moreover, the Structure Director grants access authorization to the IT resources.

Storage

Data storage is a computing technology that enables saving data in a digital format on computer components and recording media, including cloud services.

In the context of Research Data Management, it is necessary to ensure that data are stored securely until the end of the project and throughout the minimum retention period.

User

⁹ European charter for access to research infrastructures – Principles and guidelines for access and related services
<https://op.europa.eu/en/publication-detail/-/publication/78e87306-48bc-11e6-9c64-01aa75ed71a1/language-en>

“User” is anyone who has access to INFN computing facilities and network services, according to the functions and professional duties they perform in the Institute.

The term includes all of those who operate within the INFN, therefore employees (researchers, technologists, technical staff) and associates (e.g., university or other research bodies employees), those with research grants for cooperation in research activities and scholarships, those who operate within the Institute as part of several contracts or agreements, as well as temporary guests and contract workers.

User group contact person

The user group contact person is a person who coordinates the users and the usage of local computing resources belonging to one or more groups, experiments, or services, according to the guidelines provided by the Computing and Networking Service.

Virtual machine (VM)

A virtual machine (VM) is the virtual representation of a physical computer and composed of an operating system and one or more application/s.

DocID	Rev.	Status
DMP_INFNN-LNGS_RI	1.0	Draft

5. Introduction

The Gran Sasso National Laboratories (LNGS) of the National Institute of Nuclear Physics (INFN), is one of the worldwide largest underground experimental facility, housing large and medium scale experiments focused on astroparticle, nuclear and neutrino physics. They are unique thanks to their size, ease of access, geographical location and their intrinsic low radioactive concentration of the surrounding dolomitic rocks.

This data management plan (DMP) offers guidelines to staff and users of the INFN-LNGS research infrastructure (INFN-LNGS-RI), in compliance with INFN policies, on how the INFN-LNGS-RI research data and related metadata will be handled during their entire lifecycle, as shown in Figure 1.

This document aims to:

- outline the data-management plan for INFN-LNGS-RI staff and
- provide guidelines to the research teams to formulate their own DMP in collaboration with the INFN-LNGS-RI and confirming INFN-LNGS-RI regulations.

It offers a perspective by outlining how data will be generated, collected, preserved, accessed and shared to guarantee reproducibility, interoperability and re-use according to the FAIR principles¹⁰. Data produced have to be FAIR, that means findable, accessible, interoperable and reusable, according to European guidelines on Open Science focused on developing transparent processes, enhancing research activity, its verifiability, the integrity of research results and proper scientific communication^{11, 12, 13}.

INFN endorses the principle of *“as open as possible, as closed as necessary”*.

Within the INFN-LNGS-RI, are foreseen the following four categories of data and related metadata that need to be managed:

1. Data and metadata coming from the INFN-LNGS research division.
2. Data and metadata coming from the INFN-LNGS technical and general services division.
3. Data and metadata coming from physics experiments hosted in the aboveground and/or underground INFN-LNGS laboratories.
4. Data and metadata coming from activities performed with third parties.

This DMP will be regularly updated, when required and at least every two years. All INFN-LNGS-RI users, included research teams of the physics experiments, should agree on this document and implement it during the entire data lifecycle.

¹⁰ Science Europe (2021). Practical Guide to the International Alignment of Research Data Management, <https://doi.org/10.5281/ZENODO.4915862>

¹¹ Horizon Europe: open science, <https://op.europa.eu/en/web/eu-law-and-publications/publication-detail/-/publication/9570017e-cd82-11eb-ac72-01aa75ed71a1>

¹² The Italian portal dedicated to Open Science, <https://open-science.it/es/english>

¹³ Italian National Plan for Open Science (PNSA 2021-2027), <https://www.mur.gov.it/it/news/lunedì-20062022/pubblicato-il-piano-nazionale-della-scienza-aperta>

DocID	Rev.	Status
DMP_INFNO-LNGS_RI	1.0	Draft

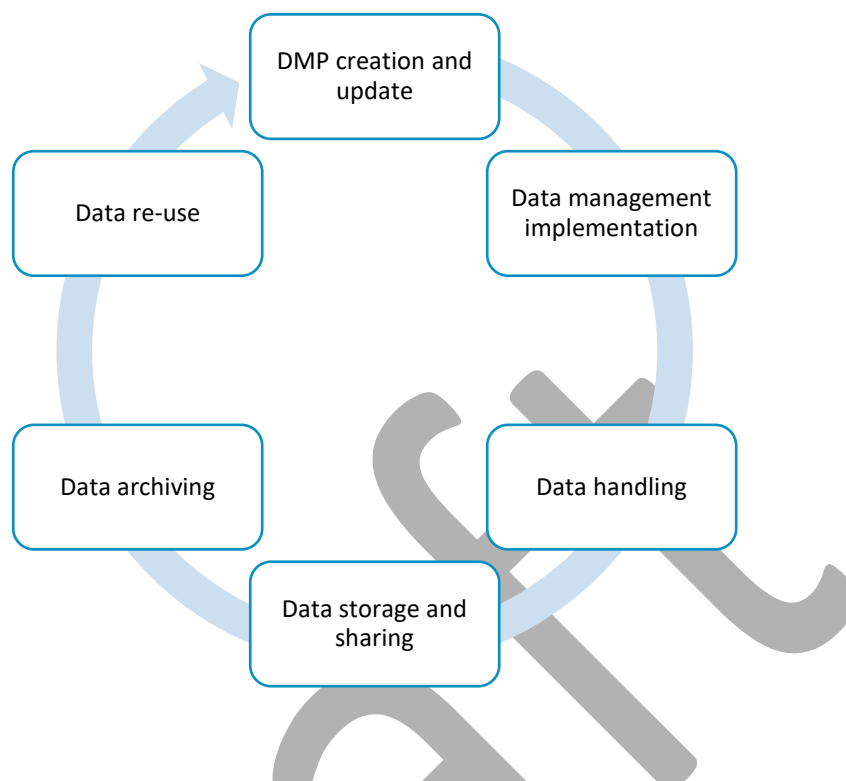


Figure 1. Data lifecycle

DocID	Rev.	Status
DMP_INFNO-LNGS_RI	1.0	Draft

6. FAIR data

The FAIR data principles¹⁴ aim to improve the findability, accessibility, interoperability and reusability of digital resources. They have become part of national and international research funding programs and at the same time, they pose a prerequisite for data for their reliability, trustworthiness, and quality. In Figure 2 a schematic description of the principles and relative action points able to assess the FAIRness of data have been reported.



Figure 2. FAIR data principles and data compliance check ¹⁵

¹⁴ How to make your data FAIR. Basic information with links to resources. <https://www.openaire.eu/how-to-make-your-data-fair>

¹⁵ Source: <https://ccafs.cgiar.org/open-access-and-fair-principles>

DocID	Rev.	Status
DMP_INFEN-LNGS_RI	1.0	Draft

7. Data and metadata

7.1 Data and metadata sources

Within INFN-LNGS RI, four categories of data and related metadata to be managed are foreseen and described below.

7.1.1 Data and metadata coming from the INFN-LNGS research division.

Data and associated metadata coming from services belonging to the INFN-LNGS research division include both those acquired by analytical instruments and technical data generated to fit a purpose (e.g. technical reports, analytical measurements, software code, simulations). Metadata should include details on the methodology used, analytical and procedural information, definitions of variables, vocabularies, units of measurement, any assumptions made as well as the format and file type of the data.

7.1.2 Data and metadata coming from the INFN-LNGS technical and general services division.

Data and associated metadata coming from services belonging to the technical and general service division include both data and metadata coming from environmental monitoring activities and blueprints, topographical, architectural, and structural plans, 3D representations, drawings, technical reports able to describe plants, buildings and infrastructure.

These latter are managed by official document management systems already used by INFN (e.g. Alfresco, Pandora, Indico, ...).

7.1.3 Data and metadata coming from physics experiments.

Data and associated metadata coming from research activities within the experiments hosted in the aboveground and/or underground laboratories of the INFN-LNGS-RI follow the lifecycle described in their own DMP, which in any case must be compliant with the present one.

7.1.4 Data and metadata coming from activities performed with third parties.

Experimental projects using the INFN-LNGS-RI span many different fields: from nuclear astrophysics, neutrino physics and studies of dark matter to biology experiments, studies and diagnostics applied to materials in the cultural heritage framework, geographical origin and provenance studies as well as application of physics in biomedical science.

Data and associated metadata coming from research activities with third parties are managed according to the related agreement and however in compliance with this DMP and the INFN

DocID	Rev.	Status
DMP_INFNO-LNGS-RI	1.0	Draft

policies. Only a short-term storage for this kind of data is assured, so data will be preserved for no longer than 3 years unless otherwise required and agreed.

7.2. Data types

A definition of different types of data belonging to the INFN-LNGS-RI is provided below:

- **Data once generated and/or published, are no longer updated**
e.g. scientific paper, raw data
- **Data which may be modified or extended**
e.g. such as data subject to revision and modifiable for the purpose of improving their quality/reliability. For example:
 - **Experimental data**
data originating from institutional activities, acquired or produced recurrently and systematically. For instance, chemical and radiopurity screening to evaluate the intrinsic contamination of the detector components and their related raw materials, and more generally data generated in the framework of experimental activities.
 - **Monitoring data**
data captured in real-time for example, data produced by sensors that generate a time series.
 - **Simulation data**
Data generated from simulation (e.g. Monte Carlo simulation).
 - **Technical data**
Drawings, blueprints, topographical, architectural, and structural plans, maps, sketches, tables, 3D representations.

7.3. Data processing levels

Data will potentially be processed through several levels, all of which may be important. The plan for processing between levels may therefore be specified. Concerning physics experiments, four levels of data are commonly recognized¹⁶:

Level 0: raw data. Data acquired automatically or manually, without any processing.

Level 1: reconstructed data. It includes quality-controlled data, simulations, and software able to perform a complete analysis. Typically, these data are derived from raw data by applying calibrations, pattern-finding algorithms, or any other pre-processing step.

Level 2: reduced data. Data at this level have a simplified format and often, as a subset of reconstructed data, are used for outreach and dissemination activities.

¹⁶ eXtreme-DataCloud, Enhance data management best practices, Deliverable D2.3 page 29; <https://nextcloud.extreme-datacloud.eu/s/TgTeQYPiXx3r7Ht#pdfviewer>

DocID	Rev.	Status
DMP_INFNO-LNGS-RI	1.0	Draft

Level 3: published data. Level 3 data have the greatest amount of processing applied and are directly linked to scientific papers.

Level 0 data are fundamental, as all other data may, in principle, be derived from them by re-running the reconstruction.

However, each physics experiment should preserve the appropriate data processing level to facilitate re-analysis, reuse, and verification of results in the future.

8. Agreements on rights of use

The MoU between the INFNO-LNGS-RI and research teams must include agreements about ownership and user rights of the experimental data.

Concerning third party activities, collaboration agreements between the parties will be set out.

9. Ownership of data

Research teams are required to collaborate with the INFNO-LNGS-RI to develop the own DMP (experiment DMP) that defines the data ownership and detail the adopted strategy. This document must be submitted to LNGS starting from the Approval phase of the experiment lifecycle, in conjunction with the submission of the Technical Design Report.

The DMP needs to assess and take into account INFNO regulations, compliance with the existing regulatory framework on intellectual property, industrial property and the protection of personal data, as well as any agreements concluded with third parties including funding bodies. It is essential to be aware of legal issues and any specific details before the research process begins, and to clearly present the selected data strategy within the research proposal's DMP.

People involved in the data generation process are acknowledged as "data creator".

Data and associated metadata resulting from in-house research will be made open access after an initial embargo period, during which access is restricted to the research team.

Data and associated metadata resulting from proprietary research will be managed according to the agreement signed, that anyway has to be compliant with INFNO policies.

Co-ownership agreements could be signed as for a research team or activities with third party.

Unless differently specified, INFNO is the owner of data and associated metadata produced at INFNO-LNGS-RI by the users of the infrastructure.

10. Data curation

The PI or a delegate, supported by INFNO-LNGS Computing and Networking Service staff, will handle data curation over time. All data and metadata will be curated in well-defined formats: reading and understanding procedures must be provided to INFNO-LNGS.

DocID	Rev.	Status
DMP_INFNN-LNGS_RI	1.0	Draft

A rough estimate of the size of data produced per day/month or year should be defined and revised every six month or whenever necessary.

Only data with relative metadata will be archived unless otherwise stated.

The research team (through the PI/ PIs or a delegate) needs to ensure that experiments' metadata are complete because the possibilities for everybody to search for, retrieve and interpret the data in the long term depends on it. The research team (through the PI/ PIs or a delegate) is responsible for the data lifecycle until the end of experiment and has to verify data and metadata completeness and integrity and to make them available. Formats and software enabling sharing and long-term access to data are preferred.

Data and metadata will have read-only access for the duration of their lifetime and they will be migrated or copied to archival facilities for long-term storage.

Each experiment and dataset will have a unique permanent identifier. Anybody publishing results based on open access data must quote the same identifier. This Digital Object Identifier (DOI) will be managed by INFN and registered with the DataCite organization¹⁷. This is a task of INFN OpenScience group.

High level metadata such as Title, Authors, Abstract, will be made public as soon as possible using DataCite. This information will be available via the persistent identifier landing page on the web. Details will be retrievable using <https://search.datacite.org/>.

11. Data handling, storage, backup, and preservation strategies

Research teams are required to collaborate with the INFN-LNGS-RI to develop their own DMP (experiment DMP) that clearly defines data management responsibilities. Specifically, each data management task (e.g. backup, data transfer, data sharing, data preservation) needs to be explicitly assigned to either the INFN-LNGS IT staff or the research team. This strategy should be documented in detail and submitted to LNGS beginning with the Approval phase, in conjunction with the submission of the Technical Design Report and maintained throughout the experiment lifecycle.

A member of the INFN-LNGS IT staff, in collaboration with the user group contact person, will conduct regular audits to monitor the implementation of the experiment DMP while also ensuring compliance with the proposed guidelines.

LNGS-based physics experiments are encouraged to use the INFN-LNGS computing infrastructure as the primary platform for data storage, analysis, computing and simulation. For instance, the Unified LNGS IT Environment (U-LITE) computing platform, directly connected with the data acquisition systems of the experiments located in the underground laboratories of LNGS, is used to store, and analyse such data and for modelling and simulation of data acquisition apparatus. In U-LITE, each experiment has its own storage server, and raw data are transferred and written to the storage

¹⁷ <https://www.datacite.org/>

DocID	Rev.	Status
DMP_INFNO-LNGS_RI	1.0	Draft

system. Moreover, a new HPC cluster recently installed at LNGS, will be an additional platform for data and simulation analysis, including the use of machine learning algorithms.

Data will be stored in standardized open formats when possible. Proprietary data formats should be converted to open formats prior to storage or a full set of software routines able to read and analyse data should be provided, along with appropriate documentation. The research teams are encouraged to share software, code, metadata, documents created during the experiment lifecycle. The structure of data storage at the INFN-LNGS-RI is shown in Figure 3.

Upon agreements with the research team, Computing and Network service can provide data storage and backup on tape. For short-term storage, data preservation is guaranteed for up to 3 years after data collection. For long-term storage and backup, data can be preserved on tape libraries at different locations within LNGS to minimize the risk of data loss. Additionally, to reduce tape usage, incremental backups are performed alongside full backup.

Experimental data archived for long-term storage are kept off-line on magnetic tapes.

Data produced and archived by the INFN-LNGS-RI will be available for at least twice the formal duration of the project (for at least 15 years after the end of the project).

The long-term data preservation is the process enabling a continued access to digital objects and information contained in them.

A sustainable implementation of long-term data preservation practices and policies has to face themes such as the roles and responsibilities of the different stakeholders, the financial aspects of long-term preservation and the necessary service infrastructure. The research teams are encouraged since the beginning of the research to develop a long-term strategic planning.

The effectiveness of long-term data preservation and therefore their findability and accessibility to reproduce, replicate and re-use scientific results depends not only on bitstream preservation, but also on format migration, emulation, software preservation, and the maintenance of algorithms, scripts and libraries.

Unless otherwise specified, data transfer to new tape generations to prevent technological obsolescence and ensure long-term access to data can be agreed with the INFN-LNGS Computing and Networking service.

However, it is important to emphasise that a long-term data storage implies high management costs due to the manpower and resources required. To mitigate these costs, a good strategy could be to properly select the data to be preserved, such as prioritizing level 1 data over all raw data.

DocID	Rev.	Status
DMP_INFNO-LNGS_RI	1.0	Draft

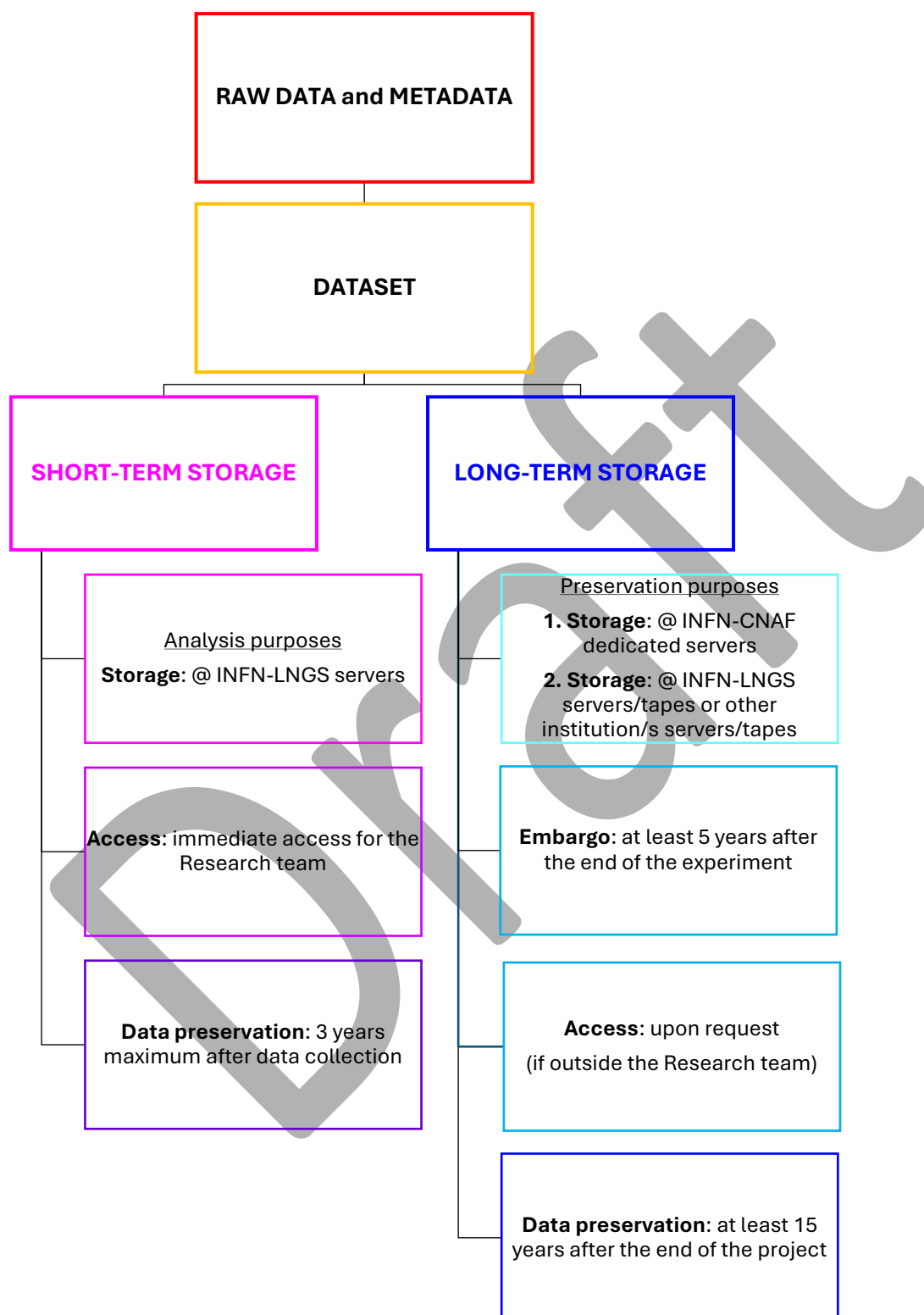


Figure 3. Structure of data storage

DocID	Rev.	Status
DMP_INFNN-LNGS_RI	1.0	Draft

12. Data sharing and access

INFN has chosen to follow the national plan for open science implementing this DMP and assuring that it complies with the FAIR data principles.

Within INFN, licenses (data access rules) are a top-down decision: unless specified, data are released under the terms of the Creative Commons license CC-BY 4.0, so that the users of such data must cite INFN, the corresponding experiments and scientific teams as the source of the data. However, Creative Commons Share-alike (CC-BY-SA) or Creative Commons No Derivative Works (CC-BY-ND) could be used if needed.

Data are accessible, upon request, by the community at large, and protected by CC open licenses according to the scientific agreements signed and the embargo if applicable.

INFN will provide open access to data at different points in time, with appropriate delays, to allow collaborators to fully exploit the scientific potential of the data before open access is triggered. As default, access to data and the associated metadata obtained from a research team is restricted to the research team for an embargo of five years after the end of the experiment. Thereafter, data will become accessible upon request. Any PI that wishes data to retain restricted access for a period longer than five years can renew the embargo period by submitting a written request, specifying the reasons for the proposed prolongation, to the Structure Director who can either accept or reject the request. Data can be made openly accessible before the end of the embargo period if the PI informs the Structure Director to do so.

Anyway, each research team is expected to produce policies with respect to open data sharing and access. These are the result of an agreement between the full set of international partners of each experiment. The PI has the possibility to transfer parts or the totality of their rights during the embargo period to INFN-LNGS or another registered person of the research team. Moreover, the PI or a delegate has the possibility to create and distribute copies of the raw data within the research team provided they use trusted applications.

INFN-LNGS cannot be held responsible for any data misuse resulting from unauthorized access, uncontrolled copies distribution, or data loss.

Authorized INFN-LNGS staff (e.g. scientists, computing and networking staff) have access to any curated data or metadata for facility-related purposes (IT management). The INFN-LNGS-RI is committed to respect the confidentiality of such data during the embargo.

It should be noted that after a few years, any software is no longer usable as it is developed by obsolete operating systems and computer drivers incompatible with later versions. A strategy supplied by the INFN-LNGS IT facility is to freeze the software environments used at the time of data production in a virtual machine (VM). This VM, along with the data and the analysis software will be provided to recreate on the scientist's machine the same software and firmware environment of the original system.

DocID	Rev.	Status
DMP_INFNO-LNGS-RI	1.0	Draft

INFN advocates for and encourages open scientific communication. Therefore, it expects significant findings from research it supports to be promptly submitted for publication, with authorship that accurately reflects the contributions of all involved.

Scientific papers related to data coming from physics experiments carried out at the INFN-LNGS-RI must cite the DOI of the data in their publication. Once known, the DOI of the paper must be sent to the spokesperson of the INFN OpenScience group.

13. Data integrity

Data integrity is the maintenance of data completeness, accuracy, quality, consistency, accessibility, and security over time. It encompasses aspects of data security, protection, and privacy.

- ◆ Data security refers specifically to measures taken to protect the integrity of the data itself against corruption, and malware ^{18,19}.
- ◆ Data protection involves safeguarding information from loss through backup and recovery procedures.
- ◆ Data privacy pertains to controlling access to the data.

Since no sensitive or personal data will be processed, security and protection requirements mainly focus on data backup and preventing data corruption throughout the entire lifecycle. Data are subject to regular backup procedures as detailed in the “Data handling, storage, backup and preservation strategies” section.

Data privacy will be guaranteed through authentication/authorization process, as well as secure connections. Security patches will be regularly applied to data servers and separate systems will be used for data backup.

Authentication and authorization play a significant role in securing systems and data: authentication verifies the users’ identities, while authorization checks their access rights. Therefore, anyone with access to data or IT resources has been authenticated and authorized beforehand.

Individual user IDs and passwords are used; these are personal, valid for a limited period according to CCR guidelines, and cannot be shared or transferred. This allows identification and traceability of the activities carried out by each user. Additionally, a specific password policy has been implemented to enhance information security and protection of research infrastructure data.

Access to data is regulated according to the user’s role within the project and the project’s data policy.

The PI or a delegate is encouraged to implement various checks to assure data integrity. For instance, during transfer from one storage medium to another, an integrity check should be performed to verify that transmitted bitstreams have not been altered along the way. Integrity

¹⁸ <https://www.fortinet.com/it/resources/cyberglossary/data-security>

¹⁹ <https://www.techtarget.com/searchdatabackup/definition/data-protection>

DocID	Rev.	Status
DMP_INFNO-LNGS_RI	1.0	Draft

information of a file can be used to verify over time that its content remains complete and unchanged (no loss, alteration or corruption). Typically, at the bitstream level, this is achieved by calculating checksums (such as SHA256) for a bitstream and comparing them to a stored value for the same file in the past.

If the checksum matches, the integrity has been maintained; if it differs, the bitstream has been modified due to undetermined reason. In such cases, the corrupted file is expected to be replaced restoring the content from a backup copy. As a best practice, every check should be documented in a report and saved in a dedicated folder.

Obsolescence occurs when digital content can no longer be used because the software and/or hardware it relies on is no longer available or accessible with current technologies. Typically, the lifespan of media and systems is around 5 to 10 years, which means that a migration is necessary to prevent them from becoming obsolete.

To address this, an obsolescence monitoring plan supported by specific technical assistance should be set up to assess and minimize risk. All data related to this verification should be detailed in a report for future reference.

Keeping backup copies in geographically separate locations is a crucial strategy to mitigate the risk of simultaneous damage from disasters. Therefore, all copies must be complete, up-to-date, and ready for recovery in case of any contingency, such as disaster recovery or replacement of corrupted content. A correctness check plan should be implemented, and the results summarised in a report. Since members of a research team may come from non-EU countries, data, metadata, and other outputs might be transferred outside the EU for research and development purposes. This scenario should be clearly described in the own MoU and the experiment's DMP.

14. Data archiving

Data have to be archived in a proper data repository that operates according to the FAIR principles. It is advised to choose a trusted institutional or domain-specific data repository with clear terms and conditions. For instance, some essential criteria have to be fulfilled: the repository chosen has to be recognized in the own research community, it has to provide persistent and unique identifiers to make data findable and citable, it has to offer standard licenses for data and code that determines terms and conditions regarding sharing and reuse and it has to use common metadata standards to help others identify and discover the data. [Re3data.org](https://re3data.org) and FAIRSharing catalog (<https://fairsharing.org/>) offer an overview of data repositories.

This choice has to be reported in each experiment DMP created and continuously updated by each research team during the experiment's lifecycle.

15. Personal data

Personal data will be processed, stored, and archived by INFN-LNGS in compliance with the principles set out in art. 6 of the EU Regulation 2016/679. The INFN Data Protection Officer (DPO) must be contacted and review this document.

16. Responsibilities and resources

Responsibilities and resources	Contact person
INFN-LNGS-RI DMP: implementation and revision	To be defined
Data compliance with INFN policies	INFN (Servizio professionale supporto giuridico) and DPO
Data owner	One or more institution belonging to the research team as explained in the MoU or any other agreement
Data co-owner	One or more institution belonging to the research team as explained in the MoU or any other agreement
Data creator	People involved in the data generation process
Data lifecycle	PI or a delegate
Data quality assurance/control	PI or a delegate
Metadata creation and management	PI or a delegate
Data storage and backup	PI or a delegate and INFN-LNGS Computing and Networking service staff
Data long-term preservation	PI or a delegate and INFN-LNGS Computing and Networking service staff
Data sharing and access	PI or a delegate and authorized INFN-LNGS staff
Data repository	PI or a delegate and INFN-LNGS Computing and Networking service staff

17. Relevant policies and guidelines

Relevant policies and guidelines	Details
Regulation on the use of INFN computing resources (24/January/2020 – Delibera del Consiglio Direttivo n. 15442)	https://www.ac.infn.it/normativa/DISCIPLINARI%20NUOVI%20E%20RETTIFICHE%202019/modifiche%20al%20Disciplinare%20per%20l%20uso%20delle%20risorse%20informatiche%20dell%20INFN.pdf
Documenti del progetto “Harmony – linee guida per la sicurezza informatica” emanati dalla CCR	https://web.infn.it/CCR/index.php/sito-utenti-del-calcolo/sicurezza-informatica/56-progetti-dei-gruppi-di-lavoro/documentazione-progetti/81-documenti-progetto-harmony
Documenti del progetto “Harmony – note per l’attuazione delle misure antiterrorismo” emanati dalla CCR	https://web.infn.it/CCR/index.php/it/sito-utenti-del-calcolo/sicurezza-informatica/56-progetti-dei-gruppi-di-lavoro/documentazione-progetti/82-documenti-harmony-antiterrorismo
Disciplinare per la tutela, lo sviluppo, la valorizzazione delle conoscenze dell’INFN (Delibera Consiglio Direttivo n. 14898 del 26 ottobre 2018)	https://sfe.inl.infn.it/wp-content/uploads/2022/04/Disciplinare-TT-INFN-emendato-2018.pdf
Disciplinare sugli spin off dell’INFN (Delibera del Consiglio Direttivo n. 14540 del 27 settembre 2017)	https://www.ac.infn.it/normativa/DISCIPLINARI%20NUOVI%20DA%20PUBBLICARE/DISCIPLINARI/nuovo%20disciplinare%20spin%20off.pdf
Disciplinare per l’accesso aperto ai prodotti della ricerca dell’Istituto Nazionale di Fisica Nucleare (Delibera Consiglio Direttivo n. 16717 del 21/luglio/2023)	https://pandora.infn.it/public/277a68
INFN Open Access Repository	https://www.openaccessrepository.it/
Code of Conduct (8/Dec/2014)	https://home.infn.it/images/cug/Codice_Etico_INFNO_EN.pdf
Code of conduct concerning anticorruption of INFN personnel	https://home.infn.it/images/cug/cod_anticorrupting_EN.pdf
Code of conduct for the protection of the dignity of people who work and operate within the National Institute for Nuclear Physics	https://home.infn.it/images/cug/Codice_Comportamento_Tutela_Dignita_EN.pdf
DMP repository	Specify the repository involved (it could be Alfresco, Open Access Repository, Fairdata repository, Zenodo, ...)
Accessibility to INFNO-LNGS-RI DMP	Open access (CC-BY 4.0)