Consiglio Nazionale delle Ricerche

Policy for the Management of Research Data acquired during the activities of the R/V Gaia Blu

Version 1.0 – October 2024

Summary

1) Aims and Objectives

The National Research Council has formulated the Open Science Roadmap[1] with which it is committed "to ensure that every step of its research activity, from planning to data collection, from analysis to the definition of procedures, from the discussion of hypotheses to the definition of the thesis, is managed in such a way as to make the results obtained easily Findable, Accessible, Interoperable and Reusable, applying the so-called FAIR-by-design". Furthermore, the CNR adopts and applies one of the principles underlying open science, that of "as open as possible as closed as necessary". Openness and sharing therefore become the "default" in scientific communication. Making the results available, in every form and at every step of the scientific process, becomes the standard to be adopted in every research activity.

CNR recognises the fundamental importance of the Research Data acquired during the activities of the R/V Gaia Blu as a valid scientific result for the progress of knowledge, constituting a heritage to be valorised and a long-term resource for scientific research and whole society. Therefore, it is committed to applying the highest standards for their collection, archiving and conservation, and publication, ensuring the maintenance of the quality and integrity values of scientific research.

Through this Policy, the Institution intends to define the guiding principles for the correct Management of Research Data acquired during the activities of the R/V Gaia Blu, in accordance with Open Science practices, FAIR principles and international and disciplinary standards of reference by specifically encouraging the sharing of scientific data and recognizing that reliable and easily accessible research data are the foundation of every research project. They are crucial for verifying the reliability and accuracy of the project's conduct and results, as well as for its reproducibility.

2) Scope of application

This policy applies to all research activities that produce data, conducted on board and using the R/V Gaia Blu. All individuals involved in the creation, collection, and/or management of Research Data acquired during the activities of the R/V Gaia Blu must adhere to this Policy.

If the research is co-financed by third parties (national or international projects), the respective agreements should also regulate the collection, processing, and management of data, their access, and storage, in accordance with this Policy and, in general, with data protection and intellectual property laws (including copyright). The specific agreement must still be drafted based on the present one, appropriately justifying any differences.

If the R/V Gaia Blu is used on behalf of third parties (*"conto terzi"*), the policy established at the contractual stage will be applied. The specific policy must still be drafted based on the present one, appropriately justifying any differences.

[1] D. Castelli, G. De Simone, F. Cancedda, L. Candela, V. Colcelli, R. Conte, F. Di Donato, S. Giannini, S. Mangiaracina, R. Puccinelli, M.A. Ranchino (2023) *Roadmap Scienza Aperta* doi:10.57665/BICE_ROADMAP2023

3) Research data management

The Research Data Management will be conducted according to the guidelines of a specific Data Management Plan (the ***Gaia Blu DMP***), which is an integral part of this policy. The Gaia Blu DMP identifies the types of data and the management methods promoted by Gaia Blu.

Additionally, it is required that in the proposal for the R/V Gaia Blu's ship time request, the Principal Investigator (PI) outlines a specific Data Management Plan (DMP) for the campaign using a dedicated template. This DMP must complement and integrate with the Gaia Blu DMP, including detailed information on the types of data to be managed and the methods. Any differences from the Gaia Blu DMP must be adequately justified and explicitly approved.

3.1) FAIR principles

Research data, as identified and defined by the Gaia Blu DMP, are managed, archived, and made freely available for scientific research or public interest purposes. Every Researcher must manage the Research Data according to the **FAIR Principles**. CNR recognizes that research data, even after the conclusion of the scientific project that produced them, constitutes the Institution's heritage and represents a long-term resource for research and societal progress. In general, research data must be collected, managed, and archived in accordance with current regulations, ensuring they are correct, complete, reliable, preserving their integrity, and documenting as exhaustively as possible the methods of their acquisition to promote reproducibility. They must also be made findable, accessible, interoperable, and, where possible, available for subsequent uses (*FAIR principles*).

3.2) Data preservation

During the research activity, Data are managed to ensure their quality and integrity, in accordance with this policy. The correct storage of Data during the active phases of research must utilize infrastructures that ensure backup procedures and protect, when necessary, the sensitive nature of the Data, in compliance with current regulations.
Data are organized into Datasets, accompanied by appropriate metadata and documentation that explain the methodologies of their creation and/or collection, as well as the protocols and tools adopted to promote their accessibility and reuse.

3.3) Embargo, storage and archiving of data

Experimental data, used for studying environmental phenomena and dynamic models and identified as research data, may be subject to an embargo period determined by the data type. This period is specified in the Gaia Blu DMP and in the DMP specific of the campaign and lasts for a maximum of **2 years** (except for projects where the research vessel is under contract by other entities for which the specific project embargo rules apply) starting from the end of the campaign, during which the research group may keep the data confidential. After the embargo period, the data will be made available to the scientific community through trusted repositories recognized as valid by the Institution.

Physical samples that will be processed onshore will follow the specifications of the Gaia Blu DMP.

At the end of the campaign, the acquired Data must be archived in a manner that is correct, complete, and reliable, ensuring their integrity. The minimum duration for data storage is 10 years.

If it becomes necessary to destroy or delete all or part of the research data (e.g., for ethical or legal reasons), such action must be traceable, and the related documentation and justification must be transparent and accessible. Therefore, part of the associated metadata will be preserved.

3.4) Licenses

Data acquired during the activities of the R/V Gaia Blu, once deposited, must be accompanied by the necessary information for their identification and, if possible, made Open Access, following the principle "*as open as possible, as closed as necessary",* and distributed with licenses that ensure free use to promote open and collaborative research. Exceptions include cases where there are opportunities for commercial exploitation, third-party rights, or other legal constraints that prevent Open Access dissemination. In such cases, Data relevant for research reproducibility should still be deposited in a Trusted Repository that allows controlled access to the Data while ensuring Open Access to the Metadata and supporting documentation.

The reference license identified is the 'Creative Commons Attribution 4.0 International' (CC-BY 4.0 International https://creativecommons.org/licenses/by/4.0/legalcode). Data reuse is permitted only if the user accepts the license associated with the data. However, the rights granted, as described in the license associated with the data, do not imply any transfer of ownership rights of the databases, data, and public information. Users may use and reuse the database and the data contained therein only in accordance with the terms of use of the license; any use not expressly authorized by the license or current law is prohibited.

4) Roles and Responsibilities

The responsibility for the proper collection, management, and preservation of Data acquired during the activities of the R/V Gaia Blu is shared between the PI and the Institution, each for the areas of their own competence.

1.1) **Role and responsibility of the PI**

The PI is responsible for:

- manage the Research Data acquired during the activities of the R/V Gaia Blu in accordance with this Policy, the Data Management Plans (Gaia Blu DMP and campaign specific DMP), the applicable regulations, and any contractual obligations with third parties, including ethics, privacy, and intellectual property rights protection;
- understand and apply the FAIR Principles for Data Management, also utilizing the information and support services provided by the Institution;
- define its own strategy for data collection, documentation, archiving, and preservation of

campaign data, including the definition of protocols and responsibilities within the research group, which should be included and implemented in the campaign specific Data Management Plan.

1.2) **Role and responsibility of CNR**

CNR is responsible for:

- providing the PI with support to access resources and infrastructure necessary to achieve proper management of the Research Data acquired during the activities of the R/V Gaia Blu and to comply with the provisions of this Policy and the Gaia Blu DMP;
- organizing and providing support services to the PI for the proper Research Data Management (including assistance for the design and completion of the DMP for the specific campaign), addressing the Data Management Working Group and other working groups, as necessary;
- organize and provide services to support the deposit and access to institutional repositories;
- ensure the management and maintenance of the Institution's infrastructure for Research Data.

5) Validity

This policy comes into effect on 1st January, 2025, and will be updated as necessary.

For any non-compliant, unforeseen, or unspecified needs not addressed in this document, please contact the email address: navegaiablu@cnr.it

Appendix - Definitions

The term "*Research Data*", or simply "*Data*", refers to information, regardless of format, collected and used following a defined protocol within a specific research activity conducted by the Researcher. These data are essential for confirming the research results and include information used in scientific publications. Examples of Research Data include, but are not limited to, results, observations, texts, images created and/or collected in digital format, along with other digital outputs such as 3D models and source code.

The term "*Dataset*" refers to a structured collection of data related to each other, created or gathered for a common purpose, and organized in a way that represents the results of a research activity.

The term "*Research Data Management (RDM)*" refers to the implementation of best practices, guidelines, and standards aimed at ensuring quality, integrity, comprehensibility, and security in data preservation during research activities. Additionally, it addresses long-term preservation and accessibility of data, in compliance with current regulations and considering open access policies.

The term "*Data Management Plan (DMP)*", also known as "*Research Data Management Plan*", is a document that provides an overview of the management of research data, whether created, collected, or reused during the research activity. The DMP details the nature and origin of the research data, data formats, their size in terms of storage space required, procedures to ensure quality and security, involvement of individuals in their creation and processing, organization and management of data during and after the research activity, as well as strategies adopted to comply with FAIR Principles while adhering to regulations and any agreements with third parties.

The "*FAIR Principles*" refer to a set of guidelines designed to make scientific data Findable, Accessible, Interoperable, and Reusable. These principles aim to improve the quality and efficiency of scientific data management, enabling greater sharing and reuse of data, often in an automated manner. Below is a brief definition of each principle:
1. Findable: Data should be easily discoverable and accessible, using rich metadata and unique identifiers.
2. Accessible: Data should be easily accessible, with detailed information on the terms and conditions for access.
3. Interoperable: Data should be structured so that they can be effectively combined and used together, sharing formats, standards, protocols, and semantics.
4. Reusable: Data should be prepared and documented to be easily understood and used by others, with clear information on their source, context, and usage restrictions (licenses).
By following these principles, the goal is to promote the maximum utility and exploitation of scientific data, facilitating collaboration, reproducibility, and innovation.

"*Open Access to Research Data*" refers to the process by which research data are deposited and shared in a manner that allows anyone to access, use, modify, and share them freely, with the fewest possible restrictions that preserve data attribution and openness.

The term "*Metadata*" refers to a structured description that provides additional information about a set of data. Metadata helps to understand the content, context, structure, and source of the data itself. They may include information such as dataset title, authors, creation date, data source, file format, collection and analysis methods, keywords, and usage rights. Metadata is essential for enabling data search, discovery, and interpretation, as well as for facilitating their management, storage, and sharing. Metadata

is typically structured according to standard schemas, often discipline-specific, internationally recognized, and implemented by long-term storage and access infrastructures.

The term "*Research Vessel (R/V)*" refers to a ship specifically designed and equipped to conduct scientific studies at sea. These vessels are used for a variety of purposes, including the collection of oceanographic, biological, geological, and environmental data. They are equipped with advanced instrumentation and onboard laboratories to perform analyses and experiments directly at sea. Research vessels can operate in diverse environmental conditions and are often utilized by research institutions, universities, and government agencies to explore and better understand marine ecosystems and oceanic processes.

The term "*Principal Investigator (PI)"* refers to the primary leader of a research project. Typically, this is a senior researcher, often with significant experience in the specific field of study of the project. The PI is responsible for the overall project design, its direction and supervision, as well as the management of associated financial and human resources. Additionally, the PI is tasked with coordinating research activities, ensuring deadlines and objectives are met, and ensuring the achievement of expected results. He is also responsible for data acquisition and management activities conducted during the project.

The term "*Repository*" refers to a platform or digital infrastructure designed for the collection, storage, and management of data and metadata. A repository provides a reliable infrastructure for securely storing data and making them accessible for the long term. These data can be of various types, such as documents, datasets, source code, images, and more. Repositories can be public or private and may be managed by academic institutions, research organizations, governments, or private entities. They play a fundamental role in promoting the sharing, preservation, reuse, including in automated form, and reproducibility of scientific and research data.

The term "*Researchers*" refers to members of the scientific community of the CNR and/or third parties who create, collect, and/or manage Research Data (such as researchers, technical-administrative staff, grant holders, research fellows, contract researchers, professional appointees, etc.).

Bibliographical references

FAIR principles definition: Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. *The FAIR Guiding Principles for scientific data management and stewardship*. Sci Data 3, 160018 (2016). https://doi.org/10.1038/sdata.2016.18

Open Definition: https://opendefinition.org/od/2.1/en/

D. Castelli, G. De Simone, F. Cancedda, L. Candela, V. Colcelli, R. Conte, F. Di Donato, S. Giannini, S. Mangiaracina, R. Puccinelli, M.A. Ranchino (2023) *Roadmap Scienza Aperta* doi:10.57665/BICE_ROADMAP2023

Istituto di Scienze Marine (ISMAR) del Consiglio Nazionale delle Ricerche (CNR), *Politica dei dati CNR ISMAR*; **DOI:** 10.26383/CNR-ISMAR.2023.6

Università di Bologna - *Policy di Ateneo per la gestione dei dati della ricerca*; https://www.unibo.it/it/allegati/policy-di-ateneo-per-la-gestione-dei-dati-della-ricerca/@@download/file/Policy-Ateneo-per-la-gestione-dei-dati-della-ricerca.pdf

Università degli Studi di Milano - *Policy sulla gestione dei dati della ricerca*; https://unimibox.unimi.it/index.php/s/8JceEcoDgQY2Z2z

European Global Ocean Observing System - *EuroGOOS Data Policy 2023*; **DOI**: https://doi.org/10.25607/OBP-1980

IOC Data Policy and Terms of Use (2023); https://iode.org/data-policy/

Istituto Superiore per la Protezione e la Ricerca Ambientale – *ISPRA Data Policy;* https://www.isprambiente.gov.it/files2024/attivita/delibera-41-cda_regolamento_data_policy.pdf

SeaDataNet I3 – *SeaDataNet Data Policy;* https://www.seadatanet.org/content/download/1695/file/SeaDataNet%20Data%20Policy.pdf