



CIIMAR Institutional Policy for Research Data Management

Version: 1.0

Date: July 2026

Authors: CIIMAR Data Management and Open Science Office; TwinDEEPS Data Steward

Approved by: CIIMAR Board of Directors

Index

1. Purpose and scope
2. Guiding principles
3. Ownership and collaboration
4. Responsibilities
5. Data Management Plans
6. Storage, documentation, onboarding and offboarding
7. Sharing, publication, embargo, and retention
8. Support, tools, and infrastructure
9. Definitions
10. References

1. Purpose and scope

This document sets CIIMAR's Institutional Policy for Research Data Management (RDM). It defines requirements for managing, sharing, and preserving research data produced in publicly funded projects developed by, hosted by, or carried out on behalf of CIIMAR.

CIIMAR is committed to Open Science and to research that is rigorous, transparent, and reproducible. This policy aligns with key European and national frameworks, including Horizon Europe, the UNESCO Recommendation on Open Science (2021), the EU Data Act (Regulation (EU) 2023/2854), Portugal's Science Law (Decree-Law No. 63/2019), and the Fundação para a Ciência e a Tecnologia Open Access Policy (FCT Open Access Policy) on immediate open access with rights retention ^[1-5].

As a research unit within the University of Porto (U.Porto) ecosystem, CIIMAR's RDM policy is aligned with U.Porto's Open Science Policy ^[6], notably its commitment to open communication, transparency, and to making research results accessible and reusable. In this context, CIIMAR supports Open Access to publications and, where appropriate, to the underlying datasets, and promotes deposit in interoperable repositories aligned with U.Porto requirements and connected to national and European infrastructures, including RCAAP (Repositório Científico de Acesso Aberto de Portugal) and EOSC (European Open Science Cloud) ^[7-8].

This policy applies to all CIIMAR researchers and collaborators (including staff, visiting researchers, and students) who create or handle research outputs such as datasets, dataproducts, metadata, documentation, code, and workflows. CIIMAR follows the principle “*as open as possible, as closed as necessary*”. Data should be shared by default, but access may be limited when required to protect privacy, confidentiality, security, ethical obligations, or Intellectual Property Rights (IPR). When funder, legal, contractual, or regulatory requirements are more specific, they take precedence.

2. Guiding principles and objectives

CIIMAR considers publicly funded research data a public good and promotes good practice from data collection to long-term preservation. CIIMAR promotes the Findable, Accessible, Interoperable and Reusable (FAIR) Guiding Principles - Findable, Accessible, Interoperable and Reusable- to make data easier to discover, access, use, and reuse ^[9].

When research involves Indigenous people, local communities, or culturally sensitive data, CIIMAR follows the Collective Benefit, Authority to Control, Responsibility, and Ethics (CARE) Principles to ensure respectful and appropriate data governance ^[10]. Personal data management should follow the General Data Protection Regulation (GDPR) ^[11]. Overall, sensitive data must be managed legally, ethically, and securely, with safeguards that match the level of risk. When data are shared, they must include enough metadata and documentation to support verification and reuse.

This policy makes open and responsible data practices part of CIIMAR's routine research work. It supports compliance and strengthens research quality, reproducibility, and impact, in line with the principles set out in The European Code of Conduct for Research Integrity ^[12]. The objectives of this policy are to:

- Promote transparency, reproducibility, and reusability of research data;
- Ensure compliance with national and international open data and open science requirements;
- Enhance data discoverability and accessibility through the use of standardised metadata and recognised repositories;
- Support a more open, inclusive, and collaborative scientific environment that fosters integrity, innovation, and public trust

3. Ownership and collaboration

All data generated by CIIMAR researchers in the course of CIIMAR activities is owned by CIIMAR, and CIIMAR will remain the custodian of that data if a researcher leaves the Centre (unless otherwise required by law or agreed in writing). For publicly funded projects and collaborations with external institutions, the ownership of the data, access and use conditions, stewardship and management responsibilities, and any applicable IPR must be defined in a formal agreement established before the project begins consistent with this Policy. Where a project uses data owned or controlled by a third party, researchers must comply with the relevant licences and/or contracts and ensure that any use, storage, sharing, or publication does not place CIIMAR in breach of legal, contractual, or funding obligations.

4. Responsibilities

The Principal Investigator (PI) (and, where applicable, Co-Principal Investigators and Team Leaders) is accountable for RDM within the project. Tasks may be delegated (e.g., to a project manager, lab manager, or data manager), but accountability remains with the PI. All project members (including students) are responsible for applying the project's data practices in day-to-day work, including documentation, secure storage, version control, preparation for sharing including archiving and publication.

At institutional level, the Data Management and Open Science Office coordinates RDM support, provides guidance, templates and training, and supports researchers in preparing and updating Data Management Plans (DMPs) and datasets for deposit in trusted repositories.

5. Data Management Plans

All publicly funded projects must have a project-specific Data Management Plan (DMP) prepared as early as possible (ideally at proposal stage) and finalised at the start of the project if not required at submission. The DMP is a living document and must be updated when significant changes occur.

At minimum, the DMP must describe:

- Data description and collection: types of data (e.g., observational, experimental, simulation), data formats (e.g., CSV, TXT), estimated data volumes, and detailed methods for data collection or generation, including instruments, protocols, or software used.
- Documentation and quality: metadata standards to be applied, comprehensive documentation to be produced (e.g., README files), quality assurance and quality control procedures to ensure data accuracy, consistency, and reproducibility.
- Storage, backup, and security: storage locations, backup routines, access control mechanisms, and technical and organisational measures for protecting sensitive or personal data, including anonymisation and pseudonymisation where appropriate.
- Legal and ethical requirements: applicable legal and ethical obligations (e.g., data protection and privacy regulations), data ownership and intellectual property rights (IPR), contractual or institutional constraints, and adherence to relevant codes of conduct and research integrity standards.
- Sharing and long-term preservation: what data will be shared, when and through which repository, under which access and licensing conditions; use of persistent identifiers (e.g., DOIs); and any embargoes or access restrictions, with clear justification.
- Roles and resources: clearly defined roles and responsibilities (including data stewardship support), and the resources required (time, budget, and infrastructure) to ensure effective data management and FAIR-aligned outcomes.

6. Storage, documentation, onboarding and offboarding

Projects must ensure data integrity and traceability throughout the project and after its lifespan. Digital data should be stored in a secured project storage space aligned with CIIMAR data policy and project DMP. Non-digital materials that cannot be digitised must be stored securely

and indexed, with a reference recorded in the project space so their existence and location are discoverable.

All research data must be accompanied by adequate documentation and descriptive metadata. Projects must apply consistent file naming and structure (e.g., project acronym, date, version). Version control should be used for datasets and code; where this is not feasible, changes must be recorded in a change log linked to the relevant outputs.

Data management onboarding and offboarding at CIIMAR

To prevent data loss, protect sensitive information, and ensure continuity, CIIMAR applies a data management onboarding and offboarding procedure at project/team level. The PI/Team Leader is accountable for ensuring it happens, with coordination and support from the Data Management and Open Science Office and, when needed, Information Technology (IT), Human Resources (HR) and Technology Transfer & Business Development Office.

When someone joins a project/team, onboarding ensures they know where the project's authoritative data, code, and documentation are kept, have the right access permissions for their role (including restrictions for sensitive data), and understand the team's working practices and responsibilities under the DMP.

When someone leaves a project/team or CIIMAR, offboarding is a planned handover. Before departure, project-relevant data, code, and documentation must be transferred to the designated project/CIIMAR custodians (not left in personal accounts or devices), key locations and handover notes must be recorded in the project space, the DMP must be updated if responsibilities or plans change, and access rights must be adjusted or revoked promptly, especially for restricted or third-party data.

7. Sharing, publication, embargo, and retention

Sharing should follow FAIR practices and support reuse. At minimum:

- Metadata must be provided at deposit, ideally using controlled vocabularies where applicable;
- Data should be shared in open formats when feasible; when proprietary formats are required, the necessary software (including versions) must be documented;

- Methods and protocols must be included or linked to enable correct interpretation and validation; and
- Where relevant for reproducibility, custom code and workflows should be shared with dependency and version notes.

Datasets must be deposited in a trusted repository. Where possible, use a domain or discipline-specific repository that provides appropriate curation and quality checks for data and metadata. Otherwise, use a repository listed in the Global Registry of Research Data Repositories: re3data ^[13] and/or aggregated by OpenAIRE ^[14]. Projects should prioritise repositories that demonstrate recognised trustworthiness (e.g., CoreTrustSeal ^[15] or equivalent).

Timing and embargo: Unless otherwise required or justified, data supporting publications must be available at the time of publication. Remaining project data should be made public within two years after the project end date.

Retention: Research data must be retained for at least 10 years. This period may be extended to meet contractual or legal obligations, support patent processes, or address public interest needs.

8. Support, tools, and infrastructure

CIIMAR is committed to supporting researchers in meeting these requirements.

CIIMAR Data Steward service and helpdesk: The Data Management and Open Science Office provide practical support across the data lifecycle, including DMP guidance, repository selection, metadata and documentation support, licensing advice, and submission checks. They also provide templates, short guidelines, and tutorials, and operate a helpdesk function for RDM questions and workflows.

Electronic Lab Notebooks (ELNs): CIIMAR encourages the use of Electronic Lab Notebooks (ELNs) or equivalent tools to record experimental data, protocols, decisions, and contextual metadata in a secure and traceable way. ELNs support consistent documentation during data collection and analysis and facilitate later curation, sharing, and repository deposit.

National services (FCT/FCCN): CIIMAR promotes the use of the national Digital Services for the Scientific Community provided through the Foundation for Science and Technology's digital services unit (FCCN services ^[16]), including resources such as POLEN Services and

infrastructure ^[17], providing the community with resources for effective management and sharing of research data (e.g. POLEN Blueprint, POLEN Sync, POLEN DataHub). CIIMAR maintains active contact with FCCN to facilitate researcher support and, where possible, to enable the pilot use of relevant services (e.g., DMP tools, repositories, and active data storage solutions).

European and domain infrastructures: Where appropriate to the discipline and data type, CIIMAR encourages the use of relevant European and international research data infrastructures and community repositories - through collaborations and agreements (e.g. domain-specific repositories such as PANGAEA, Data Publisher for Earth & Environmental Science ^[18]; ELIXIR ^[19] research infrastructure for life science data) - to strengthen data publication, interoperability, long-term preservation, and reuse.

9. Definitions

Research data: All recorded materials (digital or non-digital) collected, observed, measured, or created to support research and to understand, validate, or build upon research findings. Research data include raw data, primary data, and derived data products. Research data must be accompanied by descriptive metadata to support documentation, discoverability, and reuse.

Dataset: A coherent, versioned collection of data files and associated metadata/documentation that can be cited, accessed, and reused.

CIIMAR dataset: A dataset is considered a CIIMAR dataset when it is generated, collected, compiled, substantially processed, or curated within CIIMAR-hosted or CIIMAR-led activities; by CIIMAR-affiliated researchers acting within their CIIMAR role; using CIIMAR resources/infrastructure/services; and/or under agreements where CIIMAR is responsible for stewardship, preservation, or dissemination. Where multiple institutions contribute, stewardship, access, and ownership follow applicable agreements and funder/legal requirements.

Raw data: Data collected directly from instruments, sensors, field observations, surveys, or manual recordings. Raw data are unprocessed (e.g., not calibrated, corrected, cleaned, reformatted, encoded, or compressed beyond what is produced by the acquisition system).

Primary data: Data processed from raw data through steps such as calibration, error checking, cleaning, and quality control. Primary data provide a reliable basis for analysis and for producing additional derived outputs.

Data products: Datasets derived from raw or primary data, often integrated with other datasets or processed through additional analytical, modelling, aggregation, quality assurance, or calibration procedures (e.g., harmonised datasets, indicators, model outputs, gridded products).

Metadata: Structured information describing data and providing context (e.g., who created the data, what they represent, when and where they were collected, how they were processed, what formats and units are used, what access conditions apply). Metadata supports discovery, interpretation, reuse, and long-term archiving. Every dataset or database must include metadata.

Documentation: Human-readable materials that explain how data were generated and how they should be interpreted and reused (e.g., README files, protocols, standard operating procedures, codebooks, data dictionaries, instrument settings, calibration logs, workflow notes).

Research software: Code, scripts, notebooks, workflows, or software tools developed or adapted to collect, transform, analyse, simulate, or visualise research data. To support reproducibility, research software should be version-controlled and archived where possible in recognised repositories suitable for long-term access.

Personal data: Information relating to an identified or identifiable person; an identifiable person is one who can be identified, directly or indirectly by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that person.

Sensitive data: Data requiring enhanced protection due to legal, ethical, contractual, security, or confidentiality considerations (including, where applicable, personal data, confidential information, culturally sensitive data, precise locations of Vulnerable Marine Ecosystems (VMEs), marine mammal nurseries/breeding areas, and detailed migratory routes or aggregation hotspots).

Trusted repository: A repository with appropriate governance, sustainability, security, and preservation planning, and that supports persistent identifiers and adequate metadata for long-term access and reuse.

Persistent identifier (PID): A stable identifier used to reference digital objects over time (e.g., DOI).

License: A legal instrument defining permitted reuse of data and associated materials.

10. References

- [1] European Commission. (n.d.). Horizon Europe – Open Science. https://rea.ec.europa.eu/open-science_en ; Council of the European Union. (2022). Council conclusions on research assessment and implementation of open science. <https://www.consilium.europa.eu/media/56958/st10126-en22.pdf>
- [2] UNESCO. (2021). Recommendation on Open Science. <https://unesdoc.unesco.org/ark:/48223/pf0000379949>
- [3] European Union. (2023). Regulation (EU) 2023/2854 (Data Act). <https://eur-lex.europa.eu/eli/reg/2023/2854/oj>
- [4] Portugal. (2019). Decree-Law No. 63/2019 (Science Law). <https://diariodarepublica.pt/dr/detalhe/decreto-lei/63-2019-122317422>
- [5] Fundação para a Ciência e a Tecnologia (FCT). (2025). Nova política sobre acesso aberto a publicações científicas. <https://www.fct.pt/en/nova-politica-sobre-acesso-aberto-a-publicacoes-cientificas-da-fct-promov-e-a-visibilidade-da-producao-cientifica/>
- [6] University of Porto. (n.d.). Open Science Policy of the University of Porto. https://sigarra.up.pt/icbas/pt/LEGISLACAO_GERAL.ver_legislacao?p_nr=2966
- [7] RCAAP - Repositório Científico de Acesso Aberto de Portugal. About RCAAP. <https://www.rcaap.pt/>
- [8] European Commission. (n.d.). European Open Science Cloud (EOSC). <https://eosc.eu/>

- [9] Wilkinson, M. D., et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*. <https://www.nature.com/articles/sdata201618>
- [10] Global Indigenous Data Alliance (GIDA). CARE Principles for Indigenous Data Governance. <https://www.gida-global.org/care>
- [11] European Parliament and Council of the European Union. (2016, April 27). Regulation (EU) 2016/679 of the European Parliament and of the Council on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). EUR-Lex. <https://eur-lex.europa.eu/eli/reg/2016/679>
- [12] ALLEA. (2017). The European Code of Conduct for Research Integrity. <https://www.allea.org/wp-content/uploads/2017/03/ALLEA-European-Code-of-Conduct-for-Research-Integrity-2017-1.pdf>
- [13] FCCN/FCT. (n.d.). POLEN - Polo de Inovação Digital para o Conhecimento. <https://www.fccn.pt/areas-tecnologicas/conhecimento/polen/>
- [14] Felden, J; Möller, L; Schindler, U; Huber, R; Schumacher, S; Koppe, R; Diepenbroek, M; Glöckner, FO (2023): PANGAEA – Data Publisher for Earth & Environmental Science. *Scientific Data*, 10(1), 347, <https://doi.org/10.1038/s41597-023-02269-x>
- [15] ELIXIR. (n.d.). ELIXIR: Europe's distributed research infrastructure for life science data. <https://elixir-europe.org/>

MATOSINHOS, 13 de JULHO 2026

THE DIRECTION

Vitor Vasconcelos