



Open Data Policy



Summary

Glossary.....	2
1. Introduction and Strategic Motivation	3
2. Definition of Research Data.....	4
3. Scope and Guiding Principles	4
4. INESC TEC Research Data Repository	5
5. Research Data Publication and Institutional Commitments	6
5.1 General Principles.....	6
5.2 Guidance for Researchers.....	7
5.3 Roles and Responsibilities	8
6. Conclusion	9

Glossary

Data Management Plan (DMP)

A document that describes how research data will be managed throughout the research lifecycle, including aspects such as data collection, organisation, storage, access conditions, sharing, publication, preservation and associated responsibilities.

Data Stewardship

The set of practices, responsibilities and processes that support the effective management, documentation, preservation, sharing and reuse of research data throughout their lifecycle, in accordance with the FAIR principles and applicable legal, ethical and contractual requirements.

Dataset

A unit of research data, consisting of one or more files, organised and documented in a manner that enables discovery, validation, interpretation, reuse or integration into further research activities.

Digital Object Identifier (DOI)

A widely used type of persistent identifier assigned to datasets and other research outputs to support their unique identification, citation, discovery and long-term access.

FAIR Principles

A set of guiding principles intended to make research data Findable, Accessible, Interoperable and Reusable, thereby enhancing their visibility, accessibility and potential for reuse.

Metadata

Structured information that describes a dataset and facilitates its discovery, interpretation, management, citation and reuse.

Open Data

Research data that are made available for access, use and reuse with minimal restrictions, subject to applicable legal, ethical, contractual or intellectual property constraints.

Research Data

Factual records and information generated, collected, observed or processed during research activities. Research data may be quantitative or qualitative, raw or derived, structured or unstructured, and include any materials necessary to support the validation, interpretation and reuse of research results.

1. Introduction and Strategic Motivation

The growing emphasis on transparency, reproducibility, and public accountability has made open data a core priority in international science policy. In Europe, initiatives such as the European Open Science Cloud (EOSC), the FAIR principles (Findable, Accessible, Interoperable, Reusable), and evolving guidelines from the European Commission reflect a broad consensus: publicly funded research data should be treated as a public good.

These developments are also consistent with emerging frameworks that shape the future of research. Through the Coalition for Advancing Research Assessment (CoARA), research data are increasingly recognised as a first-class scientific output, relevant for evaluation and career progression. At the same time, open and FAIR-aligned data practices contribute to the United Nations Sustainable Development Goals (SDGs), by enhancing the societal value of science and supporting evidence-based solutions to global challenges.

For INESC TEC, aligning with these principles is not only a matter of compliance, but also a strategic opportunity to strengthen its scientific reputation, attract international collaborations, and increase the societal value of its research.

The institutional commitment is further reinforced by the recognition of INESC TEC as one of the national centres for Research Data Management. This role strengthens the consolidation of robust, sustainable practices in research data stewardship across the institution.

Proper management, preservation, and sharing of datasets - *as openly as possible and as close as necessary* - are now considered essential. These practices underpin scientific integrity, and foster collaboration and innovation.

In this context, INESC TEC recognises research data as a relevant scientific output and, as a general principle, a public good, while acknowledging that legitimate legal, ethical, contractual or intellectual property constraints may require restricted access.

This Open Data Policy sets out the institution's approach to the stewardship and dissemination of research data, providing FAIR-aligned guidance to ensure their responsible sharing, long-term accessibility, and reuse. It serves as a key instrument to support and strengthen INESC TEC's Open Science strategy, as set out in the 2023-2030 Strategic Plan.

As a Research and Technology Organisation, INESC TEC also regards research data as a strategic institutional asset. Their careful management enhances the value, visibility and long-term usability of the institution's scientific outputs, strengthening not only compliance and openness but also the broader mission of maximising the societal, scientific and organisational impact of the data produced.

By embedding the FAIR principles in data workflows, INESC TEC aims to:

- Enhance the visibility, accessibility, and impact of its research outputs;
- Strengthen data integrity, traceability, and reproducibility;
- Align with European and international requirements for responsible data sharing;
- Build a sustainable institutional culture of openness and collaboration;
- Increase institutional recognition and competitiveness in the global research landscape.

2. Definition of Research Data

For the purpose of this policy, research data are understood as factual records and information that are generated, collected, observed, or processed during research activities. In line with international definitions, such as those of the OECD and the European Commission, research data may be quantitative or qualitative, raw or derived, structured or unstructured. Examples include experimental measurements, survey responses, interviews, images, audio-visual materials, models, or simulation outputs.

Within this policy, a dataset refers to a unit of research data - whether a single file or multiple files - that has been organised and documented in a way that enables discovery, validation, reuse, or integration into further research. Datasets are considered scientific outputs in their own right, supporting transparency, reproducibility, and collaboration.

This definition also covers derived outputs that function as datasets, such as benchmark datasets, training or evaluation datasets (including corpora in areas such as natural language processing), curated or annotated materials, or structured results from simulations. What these outputs share is their potential to be reused by others to validate findings, compare results, replicate experiments, or conduct further analysis.

Datasets must be accompanied by all documentation, metadata, and, where relevant, auxiliary materials, including scripts or code, that are necessary to enable their validation, interpretation, and reuse.

3. Scope and Guiding Principles

This policy defines INESC TEC's institutional approach to the responsible management, preservation, and sharing of research data, in line with European standards and the FAIR principles. It applies to:

- Datasets generated, collected, or significantly processed by INESC TEC researchers as part of research activities, whether externally funded or internally supported;
- Datasets that underpin research publications or constitute standalone research outputs;
- Primary and secondary datasets, structured and unstructured, quantitative and qualitative;
- Datasets that cannot be made openly available, provided that access restrictions are based on legitimate ethical, legal, or contractual considerations.

This policy does not override applicable legislation or funder-specific requirements. Rather, it seeks to harmonise INESC TEC's internal data practices with these external obligations, providing researchers with clear support for compliance, quality, and impact.

The implementation of this policy is guided by the FAIR principles, which serve as a foundation for data management and sharing practices at INESC TEC.

Findable: Datasets must be assigned persistent identifiers, include standardised and rich metadata, and registered in searchable repositories;

Accessible: Datasets should be retrievable through open and standardized communication protocols, with clearly stated access conditions;

Interoperable: Datasets should be prepared and shared using structured formats and descriptive metadata that follow recognized standards;

Reusable: Datasets must be well documented, include clear licensing information, and meet domain-relevant community practices.

As a research-performing organisation, INESC TEC aims to:

- Support the long-term preservation and responsible reuse of research data;
- Facilitate open, FAIR-aligned access to datasets whenever possible;
- Promote good data stewardship practices and capacity building across its research community;
- Foster a culture of open and collaborative science through clear guidance and institutional infrastructure.

Effective implementation of this policy requires institutional mechanisms that enable researchers to manage and share data throughout the research lifecycle. The following section outlines the platforms and services available to support these practices, followed by guidance on documentation, licensing, repository use, and long-term access.

For additional information or individual support, researchers are encouraged to contact the Management Support Service.

4. INESC TEC Research Data Repository

INESC TEC provides an institutional Research Data Repository (RDM) to support the implementation of this policy and to promote the responsible sharing of research data.

The [RDM INESC TEC](#) reflects the institutional commitment to open and FAIR-aligned practices and hosts datasets produced by INESC TEC researchers and collaborators across diverse research activities. The repository is primarily oriented towards the publication, preservation and discovery of datasets intended for sharing and reuse, and does not aim to replace specialised infrastructures designed for dynamic, operational or large-scale data environments.

The repository is built on the CKAN open-source platform and supports the assignment of persistent identifiers (DOIs) via the *DataCite Fabrica* service and ensures long-term preservation through daily backups and includes citation information as part of the metadata.

The data repository also serves as a key institutional component in supporting Data Management Plans (DMPs), which are commonly prepared at project level to document the management of research data throughout its lifecycle. It is registered in the *re3data.org* and indexed in *FAIRsharing.org*, ensuring visibility and integration in the global landscape of research data repositories.

Data publication (deposit) in the RDM is a mediated process overseen by a designated data steward. This role guarantees that datasets meet institutional standards for metadata quality, documentation, licensing, and compliance with the FAIR principles before publication.

5. Research Data Publication and Institutional Commitments

5.1 General Principles

The publication and long-term stewardship of research data at INESC TEC are guided by principles of responsible research conduct and compliance with both external and institutional requirements. Accordingly:

- a) Research data resulting from publicly funded R&D projects must follow the data sharing requirements established by the relevant funding agencies.
- b) Data supporting published research outputs must comply with the data policies defined by journals or publishers, when applicable.
- c) The Management Support Service provides dedicated assistance to researchers in the preparation of DMPs, supporting early alignment with funder requirements and the adoption of suitable research data management practices. When required by funders or publishers, researchers must prepare and maintain a DMP, with institutional support provided to ensure compliance and consistency.
- d) Whenever appropriate, datasets generated within the context of INESC TEC research activities should be made available through the institutional data repository (RDM INESC TEC), as part of the institution's approach to research data stewardship, visibility and long-term accessibility.
- e) INESC TEC recognises that datasets may be deposited in external repositories or specialised infrastructures, whenever required by funders, partners, disciplinary standards, or by the technical and operational nature of the data, provided that appropriate institutional identification and contextualisation are ensured.
- f) All published datasets must be assigned persistent identifiers to ensure traceability, citation, and long-term accessibility.

- g) As a general rule, datasets should be made publicly available, except when legal, ethical, or contractual restrictions apply (e.g. protection of personal data, industrial confidentiality, intellectual property subject to patenting, or binding consortium agreements).
- h) The [INESC TEC's Intellectual Property Regulation](#) and any written agreements concluded between INESC TEC and the institutions seconding human resources to INESC TEC shall take precedence over this Policy
- i) INESC TEC commits to ensuring the digital preservation and long-term accessibility of published datasets, including secure storage, backups and metadata maintenance. INESC TEC provides institutional support throughout the data publication process.

5.2 Guidance for Researchers

Open and FAIR-aligned sharing depends on structured, well-documented data prepared throughout its lifecycle. Researchers should consider data management needs from the outset and, when applicable, prepare a DMP to guide this process. DMPs play an important role in documenting how research data will be managed throughout the data lifecycle. Depending on project requirements, DMPs may address aspects such as data organisation, storage, access conditions, repository selection, publication, preservation and responsibilities associated with data management.

To support the publication and sharing of research data, INESC TEC provides institutional mediation, technical validation, and advisory services. The following guidelines apply:

- The Management Support Service is responsible for mediating the data publication process in the RDM INESC TEC. Researchers should submit publication requests through the appropriate internal support services.

The Management Support Service also assists with publication in external repositories, in accordance with funder or publisher policies. INESC TEC maintains an official community on Zenodo and may support the creation of project-specific communities to facilitate the organisation, visibility and dissemination of datasets associated with particular research initiatives, while ensuring adequate institutional identification and attribution.

- As a general rule, datasets deposited in the institutional repository are made available under an open licence, preferably Creative Commons Attribution (CC BY) or Attribution Share-Alike (CC BY-SA). Other licences may be applied when required by funders, publishers, or disciplinary standards, provided that conditions of reuse are clearly stated in the metadata. In joint ownership situations, licensing must result from agreement among partner institutions.
- All datasets intended for publication must include:
 - the relevant files;

- comprehensive metadata;
 - sufficient documentation (e.g. a README file);
 - licensing information.
- Upon publication, the dataset is assigned a persistent identifier (DOI) through the *DataCite* service, managed by the Management Support Service.
 - The publication of jointly generated datasets should follow the agreements established among the partner institutions and be reflected in the DMP.
 - When datasets involve human participants, personal or sensitive data, researchers must consult the Data Protection Officer (DPO) during the planning phase - before data collection - and again prior to publication, if any significant changes to what has been planned occur, to ensure lawful processing, appropriate anonymisation and responsible dissemination.
 - In exceptional cases, such as those involving personal data, confidentiality agreements or security concerns, datasets may be deposited under restricted access. In such cases, metadata and any non-sensitive associated files should remain openly accessible whenever possible. Such exceptions should be based on clear justification, following principles of proportionality and risk.

5.3 Roles and Responsibilities

The table below summarises the roles and responsibilities of the main stakeholders in the publication and stewardship of research data at INESC TEC. These roles are complementary and apply across the research data lifecycle, operating within the institutional mediation process to ensure compliance with this policy and alignment with the FAIR principles.

Stakeholder	Roles and Responsibilities	Lifecycle Involvement
Researchers / Research Team	Ensure that datasets comply with this policy; provide complete data, metadata, and documentation.	Applies throughout the entire lifecycle; responsible for early planning, data organisation, documentation, and ensuring compliance with ethical, legal, or contractual restrictions.
Principal Investigator (when applicable)	Holds overall responsibility for the project's compliance with this policy; ensures the research teams fulfils its obligation.	Involved across all stages of the project; ensures overall compliance with this policy and that the research team fulfils its obligations, including attention to Data Management Plan (DMP) requirements where

		relevant, particularly during the planning and dataset preparation phases.
Data Steward (as part of the Management Support Service)	Mediates the publication process in the RDM INESC TEC; validates metadata quality and overall FAIR compliance. Supports the publication in external repositories when required.	Primarily involved in the documentation and publication phases; mediates the deposit process and validates metadata, structure, licensing and FAIR alignment before release, while also providing guidance on DMP preparation and continuous support throughout the data lifecycle.
Data Protection Officer	Reviews datasets that involve human participants or personal data; ensures compliance with GDPR and applicable legal requirements.	Engaged when datasets involve personal or sensitive data; must be consulted before data collection to assess lawful basis and compliance requirements, and reviews processing and publication conditions, if significantly different from what has been planned, to ensure conformity with GDPR and legal obligations.
Systems Administration Service	Ensures the reliable operation, security, and long-term preservation of the institutional research data infrastructure (RDM INESC TEC), including backups. Provides technical support in collaboration with the Management Support Service.	Provides continuous support across the lifecycle by ensuring secure storage, infrastructure reliability, backups, and long-term preservation of datasets.
INESC TEC	Institutional commitment to open and FAIR data practices.	Overarching and strategic role.

6. Conclusion

The INESC TEC Open Data Policy reaffirms the institution's commitment to strengthening the quality, openness, and impact of its research through responsible management and sharing of research data. By aligning with European and international standards, and embedding the FAIR principles, INESC TEC seeks to foster a consistent and sustainable culture of data stewardship. This policy provides clear guidance to support researchers in making their data discoverable,

accessible, interoperable, and reusable, thereby enhancing transparency, reproducibility, and long-term value.

Its implementation will be sustained through dedicated infrastructure, institutional expertise, and continuous support. Researchers are invited to contribute actively to this collective effort, ensuring that the data produced at INESC TEC serves not only the scientific community but also society at large. This policy will be reviewed at least every three years, or earlier if required by changes in institutional priorities, funder requirements, or legal frameworks.

Periodic follow-up may be undertaken to support continuous alignment with this policy and to ensure coherence between DMPs and published datasets.

Contacts

Campus da FEUP
Rua Dr. Roberto Frias
4200 - 465 Porto, Portugal

+351 222 094 000
www.inesctec.pt
info@inesctec.pt

Contacts

Campus da FEUP
Rua Dr. Roberto Frias
4200 - 465 Porto, Portugal

+351 222 094 000
www.inesctec.pt
info@inesctec.pt

+351 222 094 000
www.inesctec.pt

