DADOS.IPB REPOSITORY PRESERVATION POLICY

[Version 1.0, May 16th 2023]

CONTENTS

1. Goals

The main goal of this document is to describe the preservation policy related to the Polytechnic Institute of Bragança's Repository of Research Data, known as Dados.IPB, which ensures that its data are reliably maintained and accessible to members of the IPB as well as external users related to the Dados.IPB. This document describes how the Polytechnic Institute of Bragança (IPB) will support sustainable access to datasets, data collections (i.e. dataverses), metadata and related documents deposited in the Dados.IPB. Detailed strategies for preservation actions will be developed along with this policy. Detailed strategies for preservation actions will be developed along with this policy.

The full understanding of the strategies and preservation plan adopted does not dispense the need of reading the complementary documentation, namely:
- Mission and goals of the repository;
- Manuals and guides related to Open Access and data management;
- Terms of use;
- Ethics Committee Regulation of the Polytechnic Institute of Bragança;
- Regulation of Personal Data Protection of the Polytechnic Institute of Bragança;
- Privacy Policy Statement;
- Personal Data Protection Policy;
- Technology Infrastructure Usage Policy;
- Passwords Policy;
- General Dataverse Community Guidelines;
- Integration and Outsourcing;
- Legislation, standards and recommendations.

This policy is subject to change as standards, best practices and available technologies affect the Institute and data preservation. The target audience of this policy includes teachers, students, internal and external researchers, who are collaborating with the IPB, funders, users, and other interested parties with a connection to this Institute and who want to make use of the services of Dados.IPB.

The deposit of datasets, data collections, their respective metadata and additional documents are conditional on the acceptance by the authors of the Terms of Use of the Dados.IPB, which clearly delimit the rights and responsibilities of the depositor and the repository, giving it the necessary authority to ensure the best long-term preservation of the materials in the best way and make them accessible on the one hand, and guaranteeing ownership of them to their authors, on the other. The transfer of responsibility, for the preservation of data submitted to Dados.IPB of the Polytechnic Institute of Bragança, only takes place from the moment when the Services of Documentation and Libraries of the IPB (SDBIPB) in its curatorial process, evaluates and formalizes the complete incorporation of the information, i.e. the responsibility is transferred when, collectively, the following conditions are met:

1. The data depositor formally receives an e-mail from the repository staff confirming that his deposit has been accepted by the repository;

2. The depositor is able to access and download the deposited data through the graphic interfaces made available by the Polytechnic Institute of Bragança.

When the conditions above are met cumulatively, the repository takes on responsibility for the long-term preservation of the submitted dataset or collection.

2. Repository Mission

The mission of the research data repository of the Polytechnic Institute of Bragança is to provide researchers with a reliable platform (Dataverse) for storage - according to national and international guidelines - related to FAIR principles and open access policies. Encourages the publication, management, and preservation of scientific data generated by IPB researchers, promoting transparency, collaboration, and re-use of data, opening research results to companies and other audiences, thus promoting innovation and the reproducibility of science. The protection of personal data, copyrights, and industrial property rights is also guaranteed.

3. Responsibilities

Depositors: responsible for uploading data and metadata files to Dados.IPB in accordance with the repository's Terms of Use, as well as any policies and procedures governing the use of the service.

Research Units/Research Centres: responsible for administering the use of the Dados.IPB. Storage and processing spaces are allocated, as well as administrative rights for staff to manage their specific collection within the repository by creating dataverses and datasets. Units are responsible for the supervision of data uploaded into their specific collection, following the collection management policies and deposit guidelines of Dados.IPB. They can also validate data deposits for quality and integrity, through curatorial activities, including definition of the preferred file formats for the deposit or supporting depositors with advice on file format conversions.

The Documentation and Library Services of IPB (SDBIPB) and the IPB's Communication Centre (SI/CCOM): they will manage the users and their permissions, providing support in copyright decisions; they also are responsible for the validation of the datasets deposited in Dados.IPB, their quality and the associated metadata. They are also responsible for the preservation of the datasets and collections, with the responsibility to preserve and maintain them for a period of ten years after their publication, without loss in their significant properties, thus guaranteeing their intelligibility and ensuring the ability to demonstrate their authenticity. Long-term preservation is subject to the selection criteria described below in **section 6**. In addition, the SDBIPB and the SI/CCOM are responsible for the technical maintenance and administration of the repository software and service. The SI/CCOM ensures that the Dataverse application is functional, secure, and up-to-date. The SI/CCOM also maintains the storage, processing, and preservation infrastructure attached to the datasets and collections. The SDBIPB liaises with previously designated contacts in the Organic Units and Research Centers, with the curator of the specific collection and with the depositors in general, providing guides and training to teachers, students, internal and external researchers who are collaborating with IPB, users and other related interested parties who want to make use of the services of the Dados.IPB. The SDBIPB and SI/CCOM do not maintain any oversight on the formats of files uploaded by users, but will advise the use of recommended formats (described in more detail in **section 6** and will assist in identifying and mediating correctness issues in collaboration with depositors. The SDBIPB will also assist depositors in improving the organization of data, as well as improving the quality of metadata, and clarifying any questions related to deposit and publication.

The SI/CCOM: responsible for the technical management of the IT part of the Dados.IPB, guaranteeing the same level of service quality and operational security that it offers to all the other application services at the Polytechnic Institute of Bragança. The SI/CCOM is responsible for running Dados.IPB, with respect to performance and ensuring sufficient storage space for the production servers. Moreover, in case of an incident, the IT support staff is responsible for the recovery of Dados.IPB, the datasets and collections

already published and their respective metadata, following the incident policy (available for internal use only). As the repository is a critical service of the Institute, the IT support staff should assign a priority status to resolve Dados.IPB-related incidents within a maximum of 4 working hours.

4. Community of Interest

The following entities constitute the repository's community of interest:
- The direction and management bodies of the Polytechnic Institute of Bragança and their respective members;
- Documentation and Library Services of the IPB;
- Informatics Services and IPB's Communication Centre;
- Professors, internal and external researchers, grantees, employees, students, users and other related stakeholders (MORE; AQUAVALOR, for example) who want to make use of Dados.IPB services;
- General Public

5. Objectives

The repository's main objectives are:
- Contributing to the visibility and increasing the impact of research development in IPB;
- Depositing files (data sets) in organized collections (dataverses);
- Preserving files, datasets and dataverses;
- Providing a sustainable preservation environment where deposited research datasets are available to support the historical record of research, and accessible for use and reuse;
- Providing users with the means to discover and access the datasets and metadata deposited at Dados.IPB in the long term;
- Supporting compliance with funders' requirements;
- Registering persistent identifiers (DOI);
- Allowing controlled access and under prior identification;
- Associating licenses for reuse and making the terms of use explicit;
- Documenting the datasets with disciplinary metadata;
- Performing control of deposited datasets with MD5 checksum and UNF mechanism;
- Facilitating the dissemination of research results;
- Promoting the reuse of research results;
- Enabling data citation and credit to producers;
- Making data available to publishers.

To achieve these goals, the Polytechnic Institute of Bragança is responsible for ensuring the following valencies:

- Establish and ensuring the necessary procedures for the preservation of datasets and collections deposited and published in Dados.IPB, ensuring continued access to them (for a minimum period of 10 years after their publication), so that they remain accessible and understandable;
- Seek to comply with the Open Archival Information System (OAIS) reference model and other digital preservation standards and practices as they evolve;
- Seek to meet the requirements of ISO Standard 16363 for Trusted Digital Repositories;

- Establish adequate and secure backup and disaster recovery safeguards;
- Update the technological infrastructure repeatedly;
- Create the contingency plans to cope with emergencies, major failures, or catastrophic situations;
- Support and clarify issues related to the dataset deposit process;
- Assist in clarifying issues related to copyright, intellectual property, ethical issues, personal data protection, among others;
- Provide datasets usage statistics (queries and downloads) deposited in the IPB's Dados.IPB;
- Validate the metadata of datasets, ensuring their quality for information retrieval, authenticity, ease of understanding, and reuse;
- Ensure the interoperability of Dados.IPB with other information systems and services (RCAAP Portal, CienciaVitae, OpenAIRE, ORCID, RCTSaai, Polén, etc.);
- Promote dissemination and training sessions for users;
- Support relevant institutional initiatives aimed at maximizing the public benefit of scientific knowledge;
- Maintain public access to datasets and additional documents according to prevailing best practices, including adherence to environmental standards, quality control specifications, and security requirements;
- Regularly review risks related to data preservation;
- Reformat datasets and additional documents to avoid format obsolescence, according to the preference for open and recommended formats (section 6) (Whenever the need is detected);
- Keep the Dados.IPB informative page updated and guides related to Open Access and data management.

6. Scope

Long-term preservation is subject to the selection criteria. The IPB is committed to preserve and provide access to datasets, collections, associated metadata, and documents related to these data, properly deposited in the Dados.IPB that meet at least one of the following criteria:

- the datasets associated with journals, newspapers, dissertations, theses, other academic publications in perpetuity, which comply with the Terms of Use, the policies established in the Dataverse Project community and the IPB regarding Open Access, intellectual property, copyright, and property rights, and personal data protection;

- stand-alone datasets that comply with the Terms of Use, the policies set forth in the Dataverse community and the IPB regarding Open Access, intellectual property, copyright, and proprietary rights, and personal data protection;

- the datasets with high research/teaching value or difficult to reproduce (e.g. high reproduction value) that are identified by faculty, librarians, SDBIPB staff, the SI/CCOM and other IPB staff, that comply with the Terms of Use, the policies established in the Dataverse community and in the IPB regarding Open Access, intellectual property, copyright, and property rights, and personal data protection;

- other data files and additional documents resulting from the scientific and technical production of the IPB, that comply with the Terms of Use, the established policies in the Dataverse community and in the IPB regarding Open Access and intellectual property, copyright, property rights and personal data protection, and which have passed the selection made by the SDBIPB collaborators.

The datasets, collections, metadata, and additional documentation for these data should be produced by professors, students, internal and external researchers who are collaborating with the Polytechnic Institute of Bragança, and other related stakeholders who want to make use of Dados.IPB services.

Dados.IPB generally accepts data submission in any format, although it is preferable to use open formats and formats recommended for preservation (e.g. *.pdf, *.txt, *.png). The list of recommended file formats is available in the file management instructions in the user guide. Each dataset goes through a stage of content evaluation and metadata quality checking performed by SDBIPB collaborators, resulting in either acceptance for publication or the return of the data for improved description. SDBIPB staff make suggestions to depositors to convert their data files if they are not in the recommended format (whenever necessary). However, Dados.IPB accepts data in all formats, making it clear to depositors that continued access, proper preservation and reproducibility are not guaranteed for data with the non-recommended formats.

There are no size limits for datasets; however, it is advisable not to exceed 3GB per dataset file, according to the upload timeout setting for the server.

Any data stored outside of Dados.IPB (e.g., off-site storage) is not covered by the Dados.IPB Preservation Policy and is the responsibility of the researcher.

7. Metadados

The metadata of datasets is also considered relevant for preservation. The data must be described in a clear and detailed way to facilitate its understanding and reuse by others. For metadata creation, metadata schemas are used, such as Dublin Core, DDI, and DataCite that contain the following elements:

Metadata element	Description	Obligatoriness	Notes
Dataset Persistent ID	DOI	Mandatory	Automatically Assigned
Publication Date	Data de publicação de dataset	Mandatory	Automatically Assigned
Title	Dataset title	Mandatory	
Author	Author of the document	Mandatory	
Identifier Scheme	Author ID (e.g., ORCID)	Advisable	
Contact	Name of the contact person	Mandatory	
Description	Description of the dataset	Mandatory	
Subject	Domain specification	Mandatory	
Keyword	Keywords	Advisable	
Related Publication	Publication related to the deposited data	Advisable	
Notes	Additional information	Define case by case...	
Depositor	Person who deposits data	Advisable	
Deposit Date	Date of deposit	Mandatory	

Grant Information	Funder and grant information	Mandatory	
Software	Information about software used to collect, process or reuse data	Advisable	
Data Sources	Information about the source of the data collection	Advisable	
Others		Define case by case...	

8. Values

A set of values underlies the definition of the digital preservation policy and plan that the IPB adopts, namely:

- **Trust** on the part of all stakeholders, namely the authors and consumers of the datasets of the Dados.IPB. The trust is materialized in the certainty of the continuity of the service provision, in its quality and in the reliability of the provided information.

- **Transparency** in the relationship with stakeholders, through the disclosure of procedural information, as well as system-related documents, preservation metadata, and audit reports.

- **Strictness** in monitoring the legislation that constitutes the supporting legal framework, the standards, recommendations and good practices of digital preservation in order to adapt the Dados.IPB to new conjunctures.

- **Authenticity** of additional preserved data sets and documents, and ability to demonstrate this to any interested party under the terms of their preservation commitment.

- **Accessibility** of additional preserved data sets and documents, and ability to demonstrate this to any interested party under the terms of their preservation commitment.

- **Usability** of additional preserved data sets and documents allowing their reuse and elaboration of derived works.

- **Security** in preventive and corrective capacity in the event of failure or attempted intrusion, both in safeguarding the conceptual, physical and logical integrity of additional data sets and documents, and in ensuring the access rights, reserve duties and intellectual property rights to which the repository is committed.

- General **quality** of the services provided, with a commitment to constant monitoring of the different components of the system and regular internal audits

- **Innovation** applied to the deposit, administration and access functions, in a dual commitment to updating the system to technological advances, and active participation, where appropriate, in research partnerships and development of new resources

9. Description of architecture and technological infrastructure

The Dados.IPB is developed and built on open-source software, Dataverse, which enables easy and fast deposit, description, distribution and preservation of datasets, their metadata and additional documents. The Dataverse software is developed by the Institute for Quantitative Social Science at Harvard University and supported by a team dedicated to the ongoing development of the application, with the international community of Dataverse users, supported by programmers, experts in data curation and preservation, interaction and user experience, including among others the Research Data Alliance, Force11 and DataCite. The IPB is structured in Organic Units (Schools) and Research Centers, which serve as the

basis for a collection (dataverse) in the repository. In other words, the repository is organized by different Dataverses, which consist of different "sub-communities" with different collections and datasets.

For each Dataverse, user roles are assigned to specify the permissions that will define the actions they can perform, namely the administrator, the curator, and the depositor.

The administrator is the person who has all permissions in managing datasets, collections, and files.

The curator is the person with permissions to deposit, edit and publish datasets, edit License and Terms, edit Permissions, and validate the quality of the datasets and their metadata. In addition, the curator can create and edit new collections.

The depositor is the person who can deposit and edit datasets within a collection.

The Dados.IPB allows the integration of new communities and collections, at the proposal of the interested parties.

The preservation of the datasets, collections, documents and associated metadata depends on an adequate IT infrastructure, so it is monitored and reviewed periodically to ensure timely updates.

The IPB is committed to taking all necessary precautions to ensure the stability and security of system administration and operations. This includes vulnerability management, patch management, and backup and restore backup procedures. The SI/CCOM provides the necessary network and communications security equipment, providing adequate connectivity with firewall configurations.

In addition, SI/CCOM maintains on-site backups, with daily full system backups, with captures of all virtual servers. The backups are stored in a secure location and are only available to authorized employees. In this way, the state of the virtual machine (VM) can be restored from the last backup.

The VM is monitored by monitoring systems that automatically failover in case of malfunction, and continuous operation is ensured by default settings.

The technical infrastructure, including the operational servers, is in a secure Data Center, where only authorized employees have access to the equipment after identification. All systems are supplied with redundant power supplies that are in separate power groups, powered by (separate) UPS and generators, even in the event of a power failure. The space where the equipment is located has climate control, and a gas extinguishing system, and is located above sea level.

All access to the management interfaces is restricted, through network segmentation, protocol encryption, and authorization only for the staff required to operate the infrastructure.

Disaster recovery procedures are in place (**section 12**).

10. Strategic Plan

The dados.IPB has a strategic safeguard plan that includes the following aspects:

- Dados.IPB related backups are performed daily. Daily backups are full system backups with captures of all virtual servers. In this way, the state of the virtual machine (VM) can be restored from the last backup;
- The backups are stored in a secure location and are only available to authorized employees;

- The repository is classified as a higher-priority service by SI/CCOM, the maximum recovery time will be within 4 working hours;
- SI/CCOM takes all necessary precautions to ensure the stability and security of the repository's administration and operation. This includes vulnerability management, patch management, updates, and backup and restore backup procedures;
- Bandwidth, connectivity, and storage space sufficiently satisfy user needs. Storage space is self-monitored, with regular alerts and reports sent to IT staff;
- Hardware and software inventories and configuration information are recorded in an application managed by SI/CCOM, which takes inventory as there are changes to the systems;
- If any malicious or illegal content is detected in the repository, the SI/CCOM will immediately inform the SDBIPB about it in order to take the necessary measures (e.g. disassociation/deleting datasets, among others);

In addition, the IPB seeks and strives to:
- Maintain the "trusted" repository (maintain the integrity and quality of the datasets, collections, metadata, and associated documents);
- Ensure that digital resources are preserved in the long term;
- Ensure that all data published on Dados.IPB is protected;
- Implement good preservation practices;
- Improve the speed and efficiency with which information is preserved and retrieved;
- Develop, improve and maintain cost-effective, appropriately located and regularly reviewed storage systems;
- Optimize the use of repository space and repository operation for robust, reliable, and secure preservation purposes;
- Install and maintain all the fundamental preservation functions for the Dataverse application;
- Update the software related to the repository and preservation of data published in it when necessary.

11. Costs and e Financing

It is important to analyze, control and manage the necessary costs for long-term preservation. The SDBIPB and the SI/CCOM should estimate the costs involved in preservation-related activities for the dataset, collections, metadata, and associated documents and record the findings in their Preservation Plan. The focus for costs should include:

- costs related to technological infrastructure, including maintenance and upgrades of preservation system, physical space, equipment (servers, storage and backup resources) and software (operating systems, antivirus, preservation tools, etc.);
- human resource related costs, based on qualification and workload requirements, including recruitment and ongoing training;
- operational costs of preservation actions, such as digitization, quality control, conversion, and format migration, according to the chosen preservation strategies;

- costs related to the management of intellectual property rights, as well as carrying out an inventory of rights;
- other related costs, including indirect factors such as the periodicity of preservation and monitoring actions, integration with other systems, growth forecast of the Dados.IPB volume, among others.

To date, the costs have not been estimated. However, due to the limited amount of data currently published in the repository, the repository management has concluded that Dados.IPB and the IPB have sufficient resources to carry out the current Preservation Plan. During the course of the repository's development, the cost estimate will be updated according to the institution' and users' needs.

The financing of the strategic actions for the digital preservation of the datasets published in the Dados.IPB, collections, metadata and associated documents are under the responsibility of the administrative bodies of the Polytechnic Institute of Bragança and the SDBIPB, and there are also other sources of resources related to national and international projects.

12. Contingency and disaster recovery plan

In the event of an incident, the SI/CCOM is responsible for restoring the Dados.IPB, datasets, collections, documents, and associated metadata already published, following the IPB incident policy. As the repository is a critical IPB service, the SI/CCOM will resolve the incidents related to Dados.IPB with the highest priority within 4 working hours.

The SI/CCOM is responsible for compliance with the Incident Policy and for security improvements to Dados.IPB. The IT staff will focus on reducing the number of security incidents and the potential damage caused to final users for security issues. This includes any incident that affects information security at IPB, incidents that compromise the confidentiality and integrity of data, as well as unwanted incidents that affect data availability.

13. Succession Plan

A repository that claims to be reliable guarantees continued access to the data entrusted to it. Termination of activity by any of the institutions involved in the creation, implementation, and maintenance of the Dados.IPB puts access to the preserved data at risk. The termination of activity may be caused by different reasons, such as, for example, lack of financial allocation to ensure the necessary investments for the implementation of actions to safeguard datasets, and collections, the allocation of technical staff to support the operations of the repository, the renewal of vital components of the system, the change of strategy by the IPB, by legal imposition, by a natural disaster, among others.

In the context of the cessation of activities by some of the entities involved in the preservation of the data of Dados.IPB, actions will be implemented to ensure continuity of service with the least possible interruption.

One of the scenarios to ensure continued access to the data is to transfer files from Dados.IPB to POLEN - the national repository created by FCT and FCCN.

14. Support

The IPB may revise this preservation policy at its sole discretion without notice. Please check the Dados.IPB information page regularly for our current practices. The approved preservation policy must be reviewed within a maximum of three years. The formulation and review of this preservation policy are essential steps in fulfilling its strategic goals and responsibilities: it provides strategic guidance both for initiating any measures that are necessary for the protection of its collections and dataset, and for the meeting, or extending, nationally and internationally agreed standards for the preservation of digital material, including CoreTrustSeal certification.

The preservation policy helps the IPB comply with legislation and accountability requirements and the expectations of its user communities. The IPB ensures that it is at the forefront of technical advances by adopting a long-term strategic approach to digital preservation, which includes monitoring hardware and software developments, obsolescence, and migration of its collections and datasets (when necessary). The SDBIPB and SI/CCOM aim to continuously improve all aspects of the preservation workflow by embedding quality awareness in all processes and by promoting training for users.

The SDBIPB and the SI/CCOM undertake to ensure a support service for the resolution of problems and questions about the repository and the issues related to data preservation. The technical support services operate on working days from 9 am to 5 pm, by email: dados@ipb.pt, with a response and intervention start-up time of one hour for critical requests, four hours for urgent requests and 24 hours for other requests. Please also check out the existing guides, which are constantly updated.

The dispositions of this document will come into force the day after its approval.

15. References

This policy is developed using the following resources:

1. Preservation Policy of the Digital Library of the Polytechnic Institute of Bragança: https://portal3.ipb.pt/uploads/bibliotecas/Poli%CC%81tica%20Preservacao%20Biblioteca%20Digital%20v1.pdf
2. Preservation Policy of Dados.IPB: https://docs.google.com/document/d/1cElg53PInMXm1_hMPx9pavoWK2Ek2OU0aYt0j_VubfA/edit#
3. Digital Preservation Policies and Procedures, Texas Digital Library: https://texasdigitallibrary.atlassian.net/wiki/spaces/TDRUD/pages/291635428/Digital+Preservation+Policies+and+Procedures
4. The Dataverse Project, "Harvard Dataverse Preservation Policy": https://support.dataverse.harvard.edu/harvard-dataverse-preservation-policy
5. Purdue University Research Repository (PURR), "PURR Digital Preservation Policy": https://purr.purdue.edu/legal/digitalpreservation
6. DataverseNO Preservation Plan: https://site.uit.no/dataverseno/about/policy-framework/preservation-policy/preservation-plan/
7. Fiocruz's Digital Preservation of Collections Program: https://www.arca.fiocruz.br/bitstream/handle/icict/44220/Programa_Preservacao_Digital_Portugues?sequence=4&isAllowed=y