Final Report

GLIM-BioData
Gateway for Living Data Management

Within the scope of **PNCADAI - Programa Nacional de Ciência Aberta e Dados Abertos de Investigação**

Part of **Medida RE-C05-i08 - Ciência Mais Digital**

do **PRR - Programa de Recuperação e Resiliência**

9 February 2025 to 31 December 2025

Date: 20/01/2026
Prepared by: Luciana Peixoto, Jorge Oliveira and Bruno Louro
Revised: Sérgio Matos

Index

1. Progress Monitoring

The GLIM-BioData project was conceived to establish a robust and sustainable Research Data Management Centre, consolidating BioData.pt as the national reference for the management and valorisation of life and health sciences data in Portugal. Throughout the project, activities were guided by the overarching objective of fostering data sharing and reuse through the implementation of best practices, policies, and technical solutions aligned with FAIR principles, Open Science, and national and European strategies for data-driven research.

The work plan was structured around complementary objectives addressing both non-sensitive and sensitive data domains. These included the development of tailored processes for biodiversity and Human Federated data, the establishment of a national FEGA Portugal helpdesk, the definition of legal and ethical frameworks ensuring compliance with the General Data Protection Regulation, and the deployment of interoperable tools and services supporting data curation, publication, and reuse. In parallel, capacity building constituted a core pillar of the project, with extensive training programmes in research data management, and data stewardship delivered by ELIXIR-certified trainers, reinforcing national expertise and fostering a sustainable community of data professionals.

Long-term sustainability was addressed through systematic impact monitoring, using quantitative and qualitative Key Performance Indicators and the RI-PATHS methodology to assess outcomes and inform continuous improvement. This approach was complemented by strengthened collaboration at the national level and active engagement with key European infrastructures and initiatives, including the GDI Consortium, ELIXIR, FEGA, and EOSC. Together, these efforts ensure interoperability, strategic alignment, and the integration of GLIM-BioData within the European research data management ecosystem.

This section provides an overview of the progress achieved across all Work Packages by listing the milestones and deliverables completed during the project. Together, these outputs demonstrate the successful implementation of the GLIM-BioData work plan, the effective coordination of activities at national and European levels, and the achievement of the project's objectives in terms of impact, interoperability, and long-term sustainability within the life and health research data ecosystem.

1.1. Deliverables and Milestones

Below is an overview of all deliverables and milestones, as well as the implementation concerning the contractual timetable (Table 1).

Table 1 - Application Timetable vs Execution – Deliverables and Milestones

Activity	Deliverable /Milestone	Estimated schedule	% Execution as of the report date	
Detailed Work Plan	**D1.1** - GLIM-BioData Centre Work Plan	9 May 2025	Expected	100
			1st quarter progress report	100
			2nd quarter progress report	100
			Actual	100
Implementation and adoption of an institutional policy for research data management and sharing by the centre's institution(s)	**D1.5** - Definition of the institutional policy for RDM and sharing	Up to 9 months after the contract start date (October 2025)	Expected	100
			1st quarter progress report	20
			2nd quarter progress report	70
			Actual	100
	D1.6 - Implementation and adoption of the institutional policy for RDM and sharing by the GLIM-BioData participant institution(s)	By the end of the contract	Expected	100
			1st quarter progress report	0
			2nd quarter progress report	0
			Actual	100
	M1.1 - Roadshow about the GLIM-BioData Centre and the integration in the Consortium-GDI	By the end of the contract	Expected	100
			1st quarter progress report	20
			2nd quarter progress report	40
			Actual	100
	M1.2 - Roadshow about human-sensitive data and alignment with FEGA	By the end of the contract	Expected	100
			1st quarter progress report	0
			2nd quarter progress report	40
			Actual	100
	M3.1 - Report on outreach activities with FEGA stakeholders	By the end of the contract	Expected	100
			1st quarter progress report	0
			2nd quarter progress report	0
			Actual	100
	M3.3 - Guidelines to address Legal and Ethics Issues in RI	By the end of the contract	Expected	100
			1st quarter progress report	0
			2nd quarter progress report	10
			Actual	100

Activity	Deliverable /Milestone	Estimated schedule	% Execution as of the report date	
Availability of data sets with reuse potential	D2.1 - Curation and datasets availability plan for Non-sensitive Life Data	Up to 9 months after the contract start date (October 2025)	Expected	100
			1st quarter progress report	0
			2nd quarter progress report	50
			Actual	100
	D3.1 - Curation and datasets availability plan for Sensitive Life and Health Data	Up to 9 months after the contract start date (September 2025)	Expected	100
			1st quarter progress report	0
			2nd quarter progress report	40
			Actual	100
	D2.2 - Publication of datasets for Non-sensitive Life Data	By the end of the contract	Expected	100
			1st quarter progress report	0
			2nd quarter progress report	0
			Actual	100
	D3.2 - Publication of datasets for Sensitive Life and Health Data	December 2025	Expected	100
			1st quarter progress report	0
			2nd quarter progress report	0
			Actual	75
	M2.2 - DMPortal integration in POLEN	July 2025	Expected	100
			1st quarter progress report	10
			2nd quarter progress report	50
			Actual	100
	M2.1 - Open-source electronic lab notebook (eLabFTW) implementation	November 2025	Expected	100
			1st quarter progress report	10
			2nd quarter progress report	70
			Actual	100
	M3.2 - Standard Operating Procedures for FEGA Portugal Helpdesk	December 2025	Expected	100
			1st quarter progress report	0
			2nd quarter progress report	50
			Actual	100

Activity	Deliverable /Milestone	Estimated schedule	% Execution as of the report date	
Training sessions on good research data management practices according to the plan proposed by the proposing entities	**D4.1** - Training sessions on best practices in RDM for Life Data - "Ready4Intensive" Course - GLIM 1st Edition	3 and 4 July 2025 in CiiMAR	Expected	100
			1st quarter progress report	30
			2nd quarter progress report	100
			Actual	100
	D4.2 - Training session on best practices in RDM for Life Data - "Ready4Intensive" Course - GLIM 2nd edition	9 and 10 October 2025 at UCoimbra	Expected	100
			1st quarter progress report	0
			2nd quarter progress report	20
			Actual	100
	D4.3 - Training Session for Data Stewards - 1st session	7 July 2025 Online	Expected	100
			1st quarter progress report	10
			2nd quarter progress report	100
			Actual	100
	D4.4 - Training Session for Data Stewards - 2nd and 3rd session	rescheduled 2 October 2025 Online 19 November 2025 UCIBIO	Expected	100
			1st quarter progress report	0
			2nd quarter progress report	20
			Actual	100
	M4.1 - Training on accessing, processing, and curating datasets	rescheduled 3 November 2025 Online 17 November 2025 UCIBIO	Expected	100
			1st quarter progress report	0
			2nd quarter progress report	20
			Actual	100
	M4.2 - Training in eLABFTW	28 April 2025 in CiiMAR in collaboration with Centre GDI FAIRway	Expected	100
			1st quarter progress report	100
			2nd quarter progress report	100
			Actual	100
	M4.3 - Training in eLABFTW	22 September in CCMAR	Expected	100
			1st quarter progress report	10
			2nd quarter progress report	50
			Actual	100

Activity	Deliverable /Milestone	Estimated schedule	% Execution as of the report date	
	Completion of 100% of the sessions in the proposed plan	By the end of the contract	Expected	100
			1st quarter progress report	14,3
			2nd quarter progress report	40
			Actual	100
Integration and collaboration in a national network of data stewards	**D5.1** - Integration into the national data stewards network to be operated by the Consortium	By the end of the contract	Expected	100
			1st quarter progress report	100
			2nd quarter progress report	100
			Actual	100
	M5.3 -Establishment of a core group of data stewards for sensitive life and health data	November 2025	Expected	100
			1st quarter progress report	0
			2nd quarter progress report	80
			Actual	100
Participation in the Consortium's coordination activities	**D1.7** - Participation in the Consortium-GDI coordination activities (Participation in at least 80% of Consortium meetings)	By the end of the contract	Expected	100
			1st quarter progress report	10
			2nd quarter progress report	60
			Actual	100
Providing a public version of the centre's quarterly progress reports	**D1.2** - 1st Progress report	13 May 2025	Expected	100
			1st quarter progress report	100
			2nd quarter progress report	100
			Actual	100
	D1.3 - 2nd Progress report	10 September 2025	Expected	100
			1st quarter progress report	0
			2nd quarter progress report	100
			Actual	100
	D1.4 - Final report	By the end of the contract	Expected	100
			1st quarter progress report	0
			2nd quarter progress report	0
			Actual	100

Activity	Deliverable /Milestone	Estimated schedule	% Execution as of the report date	
	Information made available every four months	By the end of the contract	Expected	100
			1st quarter progress report	33,3
			2nd quarter progress report	67
			Actual	100
Centre GDI Impact	M5.1 - Assessment of BioData.pt node and communities services	June 2025	Expected	100
			1st quarter progress report	80
			2nd quarter progress report	100
			Actual	100
	M5.2 - Participation in projects promoted by European infrastructures (ESFRI) and/or funded by European funds	By the end of the contract	Expected	100
			1st quarter progress report	25
			2nd quarter progress report	90
			Actual	100
	M5.4 - Key Performance Indicators definition and monitorization	By the end of the contract	Expected	100
			1st quarter progress report	10
			2nd quarter progress report	50
			Actual	100

As part of the activities carried out during **the entire project period**, 16 deliverables and 14 milestones were completed in alignment with the established timeline and validated (except for this final report).

1.2. Explanation of Progress

During the first reporting period (9 February - April), **three** key **deliverables** were completed in alignment with the established timeline. These included deliverable **D1.1** - GLIM-BioData Centre Work Plan, which defines the strategic and operational framework for the project; **D1.2** - First GLIM-BioData progress report, which provides an overview of the work developed during the first months of implementation; and **D5.1** - Integration and collaboration in the Portuguese network of data stewards. Additionally, **one Milestone**, **M4.2,** was achieved with the successful delivery of a training session on eLabFTW, held on April 28, 2025, at CIIMAR, in partnership with the FAIRway RDM Centre.

As part of the activities carried out during the second reporting period (May - August), **three deliverables** and **two milestones** were completed in alignment with the established timeline. These include deliverables **D1.3** - 2nd Progress report, **D4.1** - Training sessions on best practices in RDM for Life Data - "Ready4Intensive" Courses - GLIM 1st edition 2025, at CiiMAR, **D4.3** - Training Sessions for Data Stewards - 1st session (Online), and two milestones, **M2.1** - Open-source electronic lab notebook (eLabFTW) implementation and **M5.1** - Assessment of BioData.pt node and communities services.

Since September and until the end of the project **ten deliverables** and e**leven milestones** have been reached, namely **D1.4** - Final report, **D1.5** - Definition of the institutional policy for RDM and sharing with the consequent **D1.6** - Implementation and adoption of the institutional policy for RDM and sharing by the GLIM-BioData participant institution(s), **D1.7** - Participation in the Consortium-GDI coordination activities (without evidence report, but running minutes signature, **D2.1** - Curation and datasets availability plan for Non-sensitive Life Data, **D2.2** - Publication of datasets for Non-sensitive Life Data, **D3.1** - Curation and datasets availability plan for Sensitive Life and Health Data, **D3.2** - Publication of datasets for Sensitive Life and Health Data **D.4.2** - Training sessions on best practices in RDM for Life Data - "Ready4Intensive" Courses - 2 nd edition, and the **D4.4** - Training Sessions for Data Stewards - 2nd session and 3rd session. Milestones reached in this period are: **M1.1** - Roadshow about the GLIM-BioData Centre and the integration in the Consortium-GDI and simultaneously **M1.2** - Roadshow about human-sensitive data and alignment with FEGA, **M2.2** - DMPortal integration in POLEN (delayed from the second reporting period), **M3.1** - Report on outreach activities with FEGA stakeholders,**M3.2** - Standard Operating Procedures for FEGA Portugal Helpdesk, **M3.3** - Guidelines to address Legal and Ethics Issues in RI, **M4.1** - Training on accessing, processing, and curating datasets, **M4.3** - Training in eLABFTW II, **M5.2** - Participation in projects promoted by European infrastructures (ESFRI) and/or funded by European funds, **M5.3** - Establishment of a core group of data stewards for sensitive life and health data.

Throughout the entire project period, significant progress was made with the completion of **16** key **deliverables** and **14 milestones** in line with the defined schedule. A detailed account of project execution, including small deviations from the initial plan, is provided in the next section, organised by work packages (WP).

2. Overall Execution

During the current reporting period (9 February- 31December 2025), the GLIM-BioData project maintained a steady pace of implementation, achieving 16 deliverables and 11 milestones, all in line with the established timeline. This corresponds to a cumulative execution of 30 deliverables and milestones out of 30 (100 %), which exceeds the expected progress at this stage.

Progress was distributed across all Work Packages (WP).

WP1 - Coordination and Management

This WP aimed to coordinate and manage the GLIM-BioData centre and to articulate all the WPs and comprised four tasks.

Task 1.1 Preparing and monitoring the work plan for the GLIM-BioData

D1.1 - GLIM-BioData Centre Work Plan - **9 May 2025**

The preparation of a comprehensive and exhaustive work plan for the GLIM-BioData (D1.1) was completed by 9 May 2025. After validation, it was made available at https://biodata.pt/glim#wp and https://bit.ly/4bbLvV5.

D1.2 - 1st Progress report - **13 May 2025** and **D1.3** - 2nd Progress report - **10 September 2025**

The GLIM-BioData work plan implementation was monitored, and four-monthly progress reports were prepared, sent and shared publicly in BioData.pt Website https://biodata.pt/glim#wp, 1st report here https://bit.ly/4sPzbA3 and 2nd report here https://bit.ly/3LLjI3x.

D1.4 - Final report - **End of Project**

The deliverable D1.4 - Final Report will be achieved through the submission of the present document in January, thus ensuring the timely delivery of the project's mandatory reporting obligations. This deliverable provides a comprehensive overview of the all progress made, challenges encountered, and mitigation measures adopted, contributing to transparent monitoring of the project's implementation.

Task 1.2 Define, implement, and adopt a policy for management and sharing research data

A policy for the management and sharing of biological research data, specifically Life and Health Data, by all GLIM-BioData members was developed. This data management policy was aligned with the National Strategy for Open Data under development by INCoDe 2030 (D1.5), and its adoption was promoted, with implementation monitored (D1.6). This policy was expected to be a requirement for funding, support, or certification of BioData.pt's services.

D1.5 - Definition of the institutional policy for RDM and sharing - **October 2025**

This deliverable established the organisational and technical framework for the responsible management and sharing of life and health sciences data, aligned with FAIR principles and the principle of "as open as possible, as closed as necessary". It defined clear approaches for handling non-sensitive and sensitive data, including open dissemination with rich metadata and controlled access mechanisms for sensitive data through trusted repositories such as FEGA Portugal. The policy set out access

categories and key technical and organisational measures to ensure compliance with the GDPR, support data reuse, and reinforce trust in data management practices.
This deliverable was successfully achieved within the expected timeline.

D1.6 - Implementation and adoption of the institutional policy for RDM and sharing by the GLIM-BioData participant institution(s) - **End of Project**

This deliverable documented the implementation and adoption of the institutional Research Data Management and sharing policy across GLIM-BioData participant institutions, building on the policy framework established in D1.5. The policy was refined through structured consultation with the BioData.pt research community, ensuring that monitoring and evaluation mechanisms were practical, relevant, and aligned with institutional needs.
The final policy version incorporated a completed monitoring and evaluation framework, including defined KPIs and reporting mechanisms, and was made widely accessible through institutional communication channels. Measures to support policy adoption, including integration with BioData.pt services and clear acknowledgement procedures, were defined to promote compliance, transparency, and sustainable implementation.

The document outputted from D1.5 and D1.6 was deposited on Zenodo with the following permanent identifier: https://zenodo.org/records/17988033 and the current version (1.1) of this policy is also available at https://biodata.pt/pt/documents/biodatapt-research-data-management-rdm-policy

Task 1.3 Articulate and coordinate GLIM-BioData's actions with the Consortium-GDI

D1.7 - Participation in the Consortium-GDI coordination activities - **End of Project**

Participation in the Consortium-GDI meetings was carried out, namely in the 1st Re.Data General Assembly Meeting (26 February 2025) with GDI Centres and subsequent meetings, and the GLIM-BioData actions were coordinated in line with the consortium's principles (including the invitation to participate in the Kick-Off Meeting, as well as the round table "Research Data Management Centres for Living Data: Driving Open Science and Innovation" organised by GLIM-BioData during Bioinformatics Open Days, 26 March 2025).
The integration of members into the networks developed by the Consortium was actively promoted.
Inês Chaves, as WP1 Leader, joined the National Interest Group on Open Science and Research Data Management.
Materials and resources of the Re.Data Consortium—such as the Characterisation of Research Data Management Support Professionals survey—were disseminated to GLIM-BioData members.

Task 1.4 Dissemination and outreach

The project's results were disseminated as described in the Communication Plan: Visual Identity (logo, website, social networks, newsletter); Internal Communication (mailing lists, Teams channels, and bi-monthly team meetings via video conference); External Communication and Knowledge Transfer (dissemination of GLIM-BioData information through BioData.pt media channels, presentations at conferences (ELIXIR All Hands 2025 - Poster), and dissemination of training activities); and Disclosure of Impact (KPI monitoring, including newsletter subscribers, web page views, social media followers and shares, number of published scientific articles, oral presentations, and participation in training activities).

Roadshows promoting the GLIM-BioData principles were conducted (M1.1), alongside Roadshows focused on human-sensitive data and alignment with FEGA Portugal (M1.2).

M1.1 - Roadshow about the GLIM-BioData Centre and the integration in the Consortium-GDI and **M1.2** - Roadshow about human-sensitive data and alignment with FEGA - **End of Project**

This milestone aims the promotion of Roadshows about the GLIM-BioData principles (M1.1 - End of project). Roadshows about human-sensitive data and alignment with FEGA Portugal (M1.2 - End of project). The roadshows included the visit to the associates: UCIBIO Lisboa - January 8, UCIBIO Porto - January 21, CiiMAR - April 28, UCoimbra - June 30, CCMAR - September 22, Nova Medical Schoo - October 8, iBET/ITQB - November 4. Roadshows in GiMM, CEBAL and FCUL could not be carried out due to scheduling constraints; however, they are already planned for the first quarter of 2026. Other associates or possible future associates were visited namely ISA - April 23, Politécnico de Leiria - May 12, Politécnico de Setúbal - May 14 and RISE-Health - September 15. During the roadshows, the data management and sharing policy efforts were presented, along with the consortium's activities, established networks, and efforts to promote the integration of members into these networks. In these Roadshows were moment to engaging researchers and institutional representatives in discussions about research data management practices and the role of GLIM-BioData in the national and European RDM landscape.

WP2 - Management of Non-sensitive Life Data for Sharing and Reuse

The objective of this WP was to reuse currently available non-sensitive data from biodiversity omics and to enhance these data by making their respective transformed data products interoperable with other data streams within Earth sciences. This WP included three tasks. Collaboration with the Centre GDI FAIRway, GBIF, and iRe:RESEARCH was fostered.

Task 2.1 Hosting and planning of omics diversity datasets availability

A data management plan was provided by the data steward to the researchers, to expedite the process of data publication in open repositories, ensuring that data quality-assured formats and outputs complied with current research standards, with particular emphasis on those applicable to the field of biological diversity (DarwinCore) (D2.1). The non-sensitive data and respective metadata produced by GLIM-BioData in biodiversity and microbiota research were deposited for public access in DMPortal (D2.2). To ensure the maintenance and sustainability of DMPortal, migration to POLEN was initiated (M2.1 – July 2025). Training on accessing and reusing publicly available datasets was provided to promote data reusability in line with the EOSC interoperability framework, covered in Task 4.3.

D2.1 - Curation and datasets availability plan for Non-sensitive Life Data - **October 2025**

This deliverable presents the Curation and Datasets Availability Plan for non-sensitive life and health data under WP 2 of GLIM-BioData RDM Center. The plan outlines how datasets will be curated, governed, and made available through BioData.pt's DMPortal, a Dataverse-based repository for open and non-sensitive research data.

The document defines processes for dataset submission, curation, metadata management, access authorization, and long-term sustainability. It also specifies how legal and ethical compliance will be ensured and how outreach and support structures, such as the DMPortal Helpdesk, will support data providers and users working with biodiversity omics data. The document outputted from this deliverable was deposited on Zenodo with the following permanent identifier:

https://zenodo.org/records/17573973.

M2.2 - DMPortal integration in POLEN - **July 2025**

Within the scope of GLIM-BioData research in biodiversity and microbiota, datasets were annotated using state-of-the-art metadata standards and deposited for public access in DMPortal. Given the critical importance of long-term persistence for annotated datasets submitted to DMPortal, it is essential to identify a robust and sustainable infrastructure for data storage. Accordingly, the objective of this milestone was to ensure the maintenance and sustainability of the Dataverse instance currently hosted by BioData.pt through DMPortal (https://dmportal.biodata.pt/).

During the course of the GLIM-BioData project, several initiatives were undertaken to establish a protocol with FCCN for the migration of DMPortal to the POLEN infrastructure. In addition to multiple informal discussions, a dedicated meeting between POLEN and BioData.pt was held on September 12 to address the benefits and urgency of this migration.

Furthermore, during the FCCN Forum (May, Coimbra), an oral presentation entitled *"DMPortal and GLIM-BioData"* was delivered, highlighting the migration process and its associated benefits. These include exploring synergies with the POLEN Repository; assessing full or partial integration between DMPortal and the POLEN Repository;

addressing the specific metadata and structured data needs of the Life and Health Sciences; actively contributing to data curation through community engagement; and facilitating the import of annotated datasets through the development of APIs and metadata standards.

The final outcome of this milestone was the inclusion of DMPortal as part of the second pilot group to be integrated into the POLEN DataHub.

Task 2.2 EMO BON and EOSC FAIR-EASE biodiversity data reuse case

Aligned with the One Health approach, this project used EMO-BON and FAIR-EASE datasets to monitor the spread of virulent bacterial clones in aquatic environments and to develop predictive models for human health risks. Within GLIM-BioData, machine learning algorithms were applied to identify genes and metabolic functions as biomarkers for clinically important bacteria, through the continuous integration of omics datasets from European marine ecosystems with environmental data.

One of the identified targets of the One Health framework is to understand connections between anthropogenic activity and environmental microbiome taxonomy and function, which can trigger pathogenic risks for humans, as well as all the other organisms in the given ecosystem. Conducted exploratory analysis assessed a model to predict biological ontology categories (GO terms, KEGG pathways) from environmental data collected during EMO-BON sampling. To predict the ontology term distributions, geographic and environmental features (temperature, salinity, chlorophyll) with anthropogenic pressure indexes filtered from momics-lite repository (https://github.com/GenMicroLab/momics-lite) served as data inputs. The goal was to perturb anthropogenic features in the trained models and observe shifts in predicted functional categories, potentially revealing novel associations between anthropogenic pressure and pathogenic functions.

There are fundamental limitations that currently prevent reaching meaningful conclusions. With only 181 samples available at the moment and substantial percentage of missing environmental data values, the dataset is too small to handle the dimensionality mismatch. Even when restricting to a single GO aspect such as molecular function, the output space contains ~1,200 terms to predict. Models including MLPs, XGBoost, and linear regression inevitably overfit to this sparse data and lack extrapolation capability. This preliminary work demonstrates the need for more data integration from (a) newly analysed EMO-BON samples by metaGOflow and (b) similar sampling campaigns analysed by MGnify workflow. As the training data grows, this ready-to-deploy modeling strategy available from GitHub (https://github.com/GenMicroLab/glim_ml) can yield invaluable insights into anthropogenic impacts on microbial functional profiles.

D2.2 - Publication of datasets for Non-sensitive Life Data - **End of Project**

Deliverable D2.2 aims to enable and demonstrate the publication of datasets containing Non-Sensitive Life Data through DMPortal, ensuring that all operational

requirements are met. This includes validating deposition workflows in real-world conditions.

Deliverable D2.2 achieved its strategic objective by demonstrating the operational capacity of DMPortal for non-sensitive life data publication. **Thirteen** datasets were successfully deposited and published until the end of 2025, validating the workflows and procedures established in D2.1. In the beginning of 2026, one more dataset was deposited (MEGABLAST searches were run for each of the 181 EMO-BON samples (ENA project PRJEB51688) Two alignments were performed, (1) on assembled contigs and (2) on merged and filtered reads. The dataset contains the best BLAST matched sequences against the Virulence Factor Database (VFDB_setA) and the MEGARes 3.0.0 microbial resistance database.). All datasets are publicly available at https://dmportal.biodata.pt

Task 2.3 Electronic laboratory notebooks

An instance of the open-source electronic laboratory notebook eLabFTW was installed and made available to all GLIM-BioData members (M2.1). Users were able to deposit data directly into BioData.pt instance or via eLabFTW interface.

M2.1 - Open-source electronic lab notebook (eLabFTW) implementation - **November 2025**

This deliverable corresponded to the installation and availability to all members of GLIM-BioData of an instance of the open-source electronic laboratory notebook, eLabFTW. This instance supports training practices in Ready For BioData Management Programme Couses, namely eLabFTW and Ready4Intensive courses.

The BioData.pt eLabFTW instance is available at elab.biodata.pt, and in-house closed eLabFTW instances were installed and available at CCMAR and CIMAR partners.

Following its increasing use by BioData.pt members and its further promotion in BioData.pt courses, we recently joined the informal group of eLabFTW users promoted by the Re.Data consortium.

WP3 - Management of Sensitive Life and Health Data for Sharing and Reuse

The objective of WP3 was the development of policies for the management of Sensitive Life and Health Data in compliance with the GDPR. The development of Standard Operating Procedures for FEGA Portugal Helpdesk (M3.2) and the publication of datasets following curation and in accordance with these policies (D3.1, D3.2). Also included the reporting of outreach activities with FEGA Portugal stakeholders (M3.1).

D3.1 - Curation and datasets availability plan for Sensitive Life and Health Data - **September 2025**

This deliverable documented the curation workflows, governance, legal/ethical considerations and operational roles required to curate, preserve and make available sensitive life and health datasets under FEGA Portugal. It provides a stepwise curation

plan (pre-submission, submission, ingestion & validation, curation, access control setup, controlled distribution, monitoring) and defines the dataset availability policy and phased rollout.

The document outputted from this deliverable was deposited on Zenodo with the following permanent identifier: https://doi.org/10.5281/zenodo.17237390

M3.1 - Report on outreach activities with FEGA stakeholders - **End of Project**

Milestone M3.1 aimed to report on outreach activities with Federated European Genome-phenome Archive (FEGA) Portugal stakeholders. Sharing national operations with FEGA and Central EGA (CEGA) practices. This milestone aligned with all the tasks of WP3. These activities aimed to engage different stakeholder groups in the use, access and submission of data, while promoting good practices in data management and sharing, including sensitive data. held during the BioData.pt Technical Meeting (https://biodata.pt/events/6th-biodatapt-elixir-pt-technical-meeting) and the BioData.pt All Hands event (https://events.biodata.pt/event/12/timetable/#20251117.detailed).

Another important activities included a session within the National Data Stewards Network focused on sensitive data, promoted by RE.DATA on 17 June and reported in M5.3; the FEGA Roadshows reported in M1.2 and a session delivered as part of the Training Data Stewards for Life Sciences - Session II https://events.biodata.pt/event/8/overview. Collectively, these initiatives strengthen data literacy, support compliance with legal and ethical requirements, and foster coordination between technical and scientific communities.

Task 3.1 Create FEGA Portugal Helpdesk

A helpdesk for FEGA Portugal was established through the deployment of a Request Tracker system, ensuring that all user queries were organised and that Standard Operating Procedures (SOPs) for the operation of the Helpdesk were defined (M3.2). These SOPs were aligned with those proposed by the Central EGA Operational Committee.
A Data Steward was allocated to the Helpdesk and trained to gain a comprehensive understanding of the FEGA submission system. The Helpdesk supported and empowered users from national institutions throughout the data deposition process, as well as data reuse, by providing access to datasets involving sensitive life and health data (D3.2).

D3.2 - Publication of datasets for Sensitive Life and Health Data - **December 2025**

Deliverable D3.2 aimed to enable and demonstrate the publication of datasets containing Sensitive Life and Health Data through FEGA Portugal, ensuring that all legal, ethical, and operational requirements are met. This includes preparing institutional stakeholders, aligning legal frameworks, and validating deposition workflows in real-world conditions.

To address these challenges, WP3 focused D3.2 activities on initiating and sustaining a concrete, real-world deposition process with a national research institution. Extensive discussions were carried out with the GiMM institute, acting as a pilot depositor of sensitive data into FEGA Portugal.

This engagement spanned several months and involved close collaboration between technical staff, data stewards, legal experts, and institutional representatives. The central activity was the **co-development of the first Data Processing Agreement (DPA)** specifically tailored for the deposition of sensitive life and health data into FEGA Portugal.

Although no dataset was ultimately deposited within the reporting period, several critical advances were accomplished:

- **Elaboration of a full Data Processing Agreement (DPA):**
  The DPA defines roles and responsibilities between the data controller and FEGA Portugal as data processor, clarifying issues such as data access control, security measures, retention policies, and liability.
- **Clarification of legal and operational responsibilities:**
  The process enabled a detailed mapping of obligations related to GDPR compliance, institutional governance, and FEGA operational procedures, significantly reducing ambiguity for future depositors.
- **Validation of deposition workflows:**
  Through iterative discussions, the full legal and procedural workflow required for sensitive data deposition was effectively tested and refined, identifying bottlenecks and points requiring additional guidance or documentation.
- **Capacity building and trust establishment:**
  The sustained interaction with GiMM contributed to building institutional trust in FEGA Portugal as a secure and compliant infrastructure for sensitive data sharing.

M3.2 - Standard Operating Procedures for FEGA Portugal Helpdesk - **December 2025**

Milestone M3.2 aimed to define, formalise, and make operational the **Standard Operating Procedures (SOPs)** governing the Federated European Genome-phenome Archive (FEGA) Portugal Helpdesk. These SOPs ensure consistent, auditable, and high-quality user support while aligning national operations with FEGA and Central EGA (CEGA) practices.

The milestone was fully achieved through the **development and publication of a comprehensive set of SOPs**, covering the full lifecycle of helpdesk operations and sensitive data submission support. All SOPs were implemented and made available through a dedicated, version-controlled repository hosted on GitHub, ensuring transparency, sustainability, and long-term maintainability.

The SOPs were designed in alignment with guidelines and operational expectations defined at the federated level by the Central EGA Operational Committee, while addressing node-specific responsibilities of FEGA Portugal.

The SOP portfolio demonstrates a high level of operational maturity:

- **11 SOPs** formally defined and published
- **Clear identifiers and versioning**, enabling traceability
- **Documented procedural steps** (3–9 steps per SOP), ensuring reproducibility

Task 3.2 Legal and Ethics Issues

The objective of this task was to develop a set of guidelines and templates to address legal and ethical issues commonly encountered in the stewardship of sensitive data within research infrastructures (M3.3).

Given the complexity of managing sensitive life and health data, particularly in the context of Federated European Genome Archive (FEGA) Portugal and the broader requirements of the General Data Protection Regulation (GDPR), it is essential to provide clear guidance to data providers, service providers, and data users. Accordingly, the objective of this milestone is to establish comprehensive guidelines that support responsible data handling practices across BioData.pt | ELIXIR PT Services.

M3.3 - Guidelines to address Legal and Ethics Issues in RI - **End of Project**

This milestone consisted of transposing and adapting the ELIXIR Europe's ELSI guidelines to the Portuguese context, ensuring compliance with national legislation while maintaining alignment with European standards and best practices. The policy incorporates specific guidance on compliance with Portuguese Act no. 58/2019 and the GDPR, as well as alignment with international instruments such as the Nagoya Protocol, the EU AI Act, and the European Convention on Human Rights.

The resulting document, BioData.pt ELSI Policy Version 4.0, was finalized on 19 December 2025. The final outcome of this milestone is the approval and implementation of the BioData.pt ELSI Policy v4.0, which serves as the authoritative reference for addressing legal and ethical issues in research infrastructures handling sensitive life and health data.

The document outputted from this deliverable was deposited on Zenodo with the following permanent identifier: https://doi.org/10.5281/zenodo.18085715

Task 3.3 Creating Awareness

Many institutions were still unaware of the available solutions covering the complete life cycle of sensitive data, including its collection, storage, processing, sharing, and re-use. In coordination with WP1, a roadshow was carried out in close engagement with stakeholders and major institutions handling sensitive data (M1.2), with the aim of raising awareness of existing resources such as FEGA and Beacon, in an articulated manner. This milestone was achieved simultaneously with M1.1 Roadshow about the

GLIM-BioData principles (M1.1 - End of project) and also related to the Task 1.4 Dissemination and outreach.

WP4 - Training and Capacity Building

This WP was dedicated to the planning, structuring, development, preparation, implementation, evaluation, and certification of training activities in the areas of bioinformatics, data management, and data stewardship. Synergies with the Re.Data consortium were identified and fostered.

Task 4.1 "Ready for BioData Management?" Pool of Trainers

Training needs in the fields of bioinformatics, data management, and data stewardship across all associated institutions were identified during the BioData.pt Technical Meeting. The BioData.pt Training Platform met to identify potential trainers. Trainers from the GLIM-BioData Centre were certified by ELIXIR through a "Train-the-Trainers" course and employed active learning methodologies, particularly project-based learning.

Task 4.2 "Ready for BioData Management?" Training

Training in the management of Life and Health Data was offered to empower researchers and technicians nationwide, in alignment with the ELIXIR Training Programme.

The Ready for BioData Management? intensive course was designed to provide life sciences researchers with a comprehensive introduction to Research Data Management (RDM) across the entire data lifecycle. Through a combination of theoretical content, group discussions, and practical exercises, participants engaged with tools such as Galaxy, Dataverse, and the Data Stewardship Wizard (DSW). The course targeted researchers, PhD students, postdoctoral researchers, and laboratory managers working with biomedical or life sciences data. No prior experience in RDM was required, although participants were expected to have some familiarity with handling research data. Training was delivered by experts from BioData.pt using resources from the ELIXIR RDMkit.

Two "Ready4Intensive" training sessions were delivered: one focused on Life Data and Health Data at CiiMAR on 3–4 July 2025 (D4.1), and a second session in 9-10 October 2025 at the University of Coimbra (D4.2).

D.4.1 - Training sessions on best practices in RDM for Life Data - "Ready4Intensive" Courses - 1st edition - **3 and 4 July 2025 in CiiMAR**

The deliverable D4.1, corresponding to the delivery of training sessions on best practices in Research Data Management (RDM) for life data, was successfully completed at CIIMAR on 3 and 4 July 2025, through the implementation of the *Ready*

for BioData Management: Intensive Course. This two-day, hands-on training brought together a diverse group of researchers, students, and data professionals from various BioData.pt Associated Members, focusing on key aspects of RDM throughout the data lifecycle, from planning and documentation to sharing and reuse (Table 2). Participants were introduced to practical tools and workflows such as Galaxy, Dataverse, and the Data Stewardship Wizard, gaining applied skills in line with FAIR data principles. The course was developed and delivered by members of the GLIM-BioData RDM Centre. A total of 22 participants attended the training (1 from FCUL, 2 from UCIBIO, 2 from BioData.pt, 5 from Universidade do Minho, and 12 from CIIMAR), demonstrating strong interest and active engagement with RDM best practices in the life sciences domain. This deliverable directly supports the capacity-building objectives of the GLIM-BioData project by strengthening national expertise and promoting alignment in open science and research data management practices.

Table 2 - Training sessions on best practices in RDM for Life Data - "Ready4Intensive" Courses - 1st edition - 3 and 4 July 2025 in CiiMAR (https://events.biodata.pt/event/9/)

Time	Session	Type	Summary	Trainer(s)
Day 1 - 3 July				
13:30	Introduction to BioData.pt and GLIM		BioData.pt - Services and Tools	Luciana Peixoto
13:45	Introduction to RDM	T	Interactive presentation motivating and introducing RDM	Daniel Faria
15:15	Data Collection Overview	T	Discussion of the key issues at the stage of data collection	Daniel Faria
15:45	Coffee Break*			
16:00	Data Processing & Analysis Overview	T	Discussion of the key issues at the stages of data processing and analysis	Daniel Faria
16:30	Data Processing & Analysis Hands-On	P	Group exercise on the workflow platform Galaxy	Gil Oliveira
17:45	Wrap Up	T	Quick discussion of the day's lessons	Trainers & Participants
Day 2 - 4 July				
9:30	Quick Recap / Intro	T	Short summary of day 1 and introduction to day 2	Daniel Faria
9:45	Data Preservation Overview	T	Discussion of the key issues at the stage of data preservation	Daniel Faria
10:15	Data Storage Hands-On	P	Quick group exercise on the storage-cost estimation tool of the DSW	Jorge Oliveira
10:45	Coffee Break*			
11:00	Data Sharing & Reuse Overview	T	Discussion of the key issues at the stages of data sharing and reuse	Daniel Faria
11:30	Data Sharing & Reuse Hands-On	P	Group exercise on the DMPortal, an instance of DataVerse	Gil Oliveira
13:00	Lunch Break*			
14:00	Data Management Planning Overview	T	Discussion of the key issues at the stage of planning	Daniel Faria
14:30	Data Management Planning Hands-On	P	Group exercise on the Data Stewardship Wizard (DSW)	Jorge Oliveira
16:00	Wrap Up	T	Quick discussion of the course's lessons	Tr& Participants

D.4.2 - Training sessions on best practices in RDM for Life Data - "Ready4Intensive" Courses - 2 nd edition - **9 and 10 October 2025 in UCoimbra**

The deliverable D4.2, corresponding to the delivery of training sessions on best practices in Research Data Management (RDM) for life data, was successfully completed at University of Coimbra on 9 and 10 October 2025, through the implementation of the *Ready for BioData Management: Intensive Course*. This two-day, hands-on training brought together a diverse group of researchers, students, and data professionals from various BioData.pt Associated Members, focusing on key aspects of RDM throughout the data lifecycle, from planning and documentation to sharing and reuse. Participants were introduced to practical tools and workflows such as Galaxy, Dataverse, and the Data Stewardship Wizard, gaining applied skills in line with FAIR data principles. The course was developed and delivered by members of the GLIM-BioData RDM Centre. A total of 27 participants attended the training (25 from UCoimbra, 1 from UCIBIO, 1 from UMinho), demonstrating strong interest and active engagement with RDM best practices in the life sciences domain. This deliverable directly supports the capacity-building objectives of the GLIM-BioData project by strengthening national expertise and promoting alignment in open science and research data management practices.

Table 3 - Training sessions on best practices in RDM for Life Data - "Ready4Intensive" Courses - 2 nd edition - 9 and 10 October 2025 in UCoimbra (https://events.biodata.pt/event/11/)

Time	Session	Type	Summary	Trainer(s)
9 July			
11:00	Introduction to BioData.pt and GLIM		BioData.pt - Services and Tools	Luciana Peixoto
11:30	Introduction to RDM	T	Interactive presentation motivating and introducing RDM	Daniel Faria
12:30	Data Collection Overview	T	Discussion of the key issues at the stage of data collection	Daniel Faria
12:50	Lunch Break*			
14:00	Data Collection Hands-On	P	Group exercise on the electronic lab notebook eLabFTW	Gil Oliveira
15:30	Data Processing & Analysis Overview	T	Discussion of the key issues at the stages of data processing and analysis	Daniel Faria
15:50	Data Processing & Analysis Hands-On	P	Group exercise on the workflow platform Galaxy	Jorge Oliveira
17:10	Wrap Up	T	Quick discussion of the day's lessons	Trainers & Participants
10 July			
9:30	Quick Recap / Intro	T	Short summary of day 1 and introduction to day 2	Daniel Faria
9:40	Data Preservation Overview	T	Discussion of the key issues at the stage of data preservation	Daniel Faria
10:00	Data Storage Hands-On	P	Quick group exercise on the storage-cost estimation tool of the DSW	Jorge Oliveira

10:45	Coffee Break*			
11:00	Data Sharing & Reuse Overview	T	Discussion of the key issues at the stages of data sharing and reuse	Daniel Faria
11:20	Data Sharing & Reuse Hands-On	P	Group exercise on the DMPortal, an instance of DataVerse	Gil Oliveira
12:50	Lunch Break*			
14:00	Data Management Planning Overview	T	Discussion of the key issues at the stage of planning	Daniel Faria
14:20	Data Management Planning Hands-On	P	Group exercise on the Data Stewardship Wizard (DSW)	Jorge Oliveira
16:00	Wrap Up	T	Quick discussion of the course's lessons	Trainers & Participants

Task 4.3 Training in Management of Biodiversity/Life Data

To encourage data reuse, training was developed on accessing, processing, and curating the EMO-BON and FAIR-EASE datasets (M4.1). This supported new queries on biodiversity data with improved interoperability through Research Data Management services. Using the FAIR-EASE Interdisciplinary Data Discovery and Access Service, data were used to query existing aggregated biodiversity results alongside a range of environmental and climatic datasets.

M4.1 - Training on accessing, processing, and curating datasets - **22 September 2025**

The milestone M4.1, corresponds to the successful completion of the first and second sessions of the Training on accessing, processing, and curating datasets, held online on 3rd November 2025 and in person on 17th November 2025. These sessions were organised by BioData.pt within the scope of the GLIM-BioData Research Data Management Centre and conducted by David Palecek from CCMAR, in a Challenge Based Learning approach. EMO-BON establishes a long-term omics observatory for marine biodiversity through bimonthly sampling of coastal waters and both soft and hard sediments across more than 20 stations in Europe. Standard operating procedures are applied for sample collection, sequencing, and workflow analysis to ensure consistent taxonomic and functional annotation. FAIR principles are implemented using appropriate ontologies, RO-Crates, and a Python-based data analysis toolkit prepared for deployment in Virtual Research Environments (VREs). On Table 4 is the course schedule and the program. The training materials are also available in TeSS.

https://tess.elixir-europe.org/materials/emo-bon-metagenomics-from-backend-integration-to-frontend-processing.

Table 4 - Training on accessing, processing, and curating datasets schedule and program ([Training on accessing, processing, and curating datasets 2025](#))

3rd November	**Session 1 - Accessing and Processing EMO-BON Metagenomic Data** - Online
14:00-14:15	Welcome and Course Introduction - Luciana Peixoto
14:15-14:30	Introduction to the EMO-BON campaign
14:30-15:00	Demonstration: Exploring the Data
15:00-15:30	Discussion and Questions
17th November	**Session 2 - 17th Nov: EMO-BON Metagenomics: From Backend Integration to Frontend Processing** - Faculty of Pharmacy of the University of Porto (UCIBIO Porto)
10:30-11:00	Session 2 overview
11:00-11:30	Backend Demonstration (RO-Crates and SPARQL Endpoint)
11:30-12:00	Using and Extending Analysis Dashboards
12:00-13:00	Challenge: Integration with Other MGnify Datasets

Session 1, conducted online, was attended by 12 participants: 1 each from UAveiro, UMinho, CiiMAR, and BioData.pt; 2 from CCMAR; and 3 from FCUL and UCIBIO, reflecting broad institutional interest and engagement.

Session 2, conducted in person as a hands-on session, was attended by 6 participants: 1 each from CiiMAR, UAveiro, and ELIXIR-NL, and 3 from FCUL.

This milestone contributes to the capacity-building goals of the GLIM-BioData project by supporting the adoption of digital tools for research data management across associated institutions.

The milestone was successfully achieved, although **not on the originally scheduled date** due to the division of the course into two sessions and its co-location with the All Hands meeting.

Task 4.4 Training in eLABFTW

Two training sessions were held at CCMAR and CiiMAR for new users of eLabFTW, an open-source electronic laboratory notebook, to promote FAIR principles from the start of research data collection (M4.2 and M4.3). Aimed at researchers, laboratory technicians, and data managers, these hands-on sessions explored the key features of eLabFTW, including experiment documentation, inventory management, and team collaboration tools, supporting participants in adopting best practices for data management and reproducibility.

M4.2 - Training in eLABFTW - **28 April 2025 in CiiMAR in collaboration with Centre GDI FAIRway**

The milestone M4.2, which consisted of delivering a hands-on training session in eLabFTW at CIIMAR, was successfully completed on 28 April 2025. This training was organised in collaboration with the FAIRway RDM Centre, and conducted by Luciana Peixoto from GLIM-BioData, Maria Paola Tomasino and Carolina Ventura Costa from the GLIM-BioData and also the FAIRway RDM Centres. Gil Poiares-Oliveira and Bruno Louro also supported the overall coordination of the session and ensured its alignment with other GLIM project tasks. On Table 5 is the course schedule and program.

Table 5 - eLabFTW course at CIIMAR schedule and program (https://events.biodata.pt/event/7/)

	eLabFTW course at CIIMAR program
14:00-14:15	Introduction to BioData.pt and the GLIM-BioData project - Gateway for living data management. Luciana Peixoto and Gil Poiares Oliveira (BioData.pt)
14:15-14:30	Introduction to FairWay project - A pathway to promote open science through research data life cycle management. Maria Paola Tomasino (CIIMAR)
14:30-15:00	Data collection in marine monitoring: EMO BON. Bruno Louro (CCMAR)
15:00-16:30	eLabFTW - Hands on exercise: open source laboratory notebook. Carolina Ventura Costa (CIIMAR)

The session was attended by 11 participants from different BioData.pt Associated Members (1 from FCUL, 1 from BioData.pt, 1 from CCMAR, and 8 from CIIMAR), reflecting broad institutional interest and engagement. This milestone contributes to the capacity-building goals of the GLIM-BioData project by supporting the adoption of digital tools for research data management across associated institutions. It also reflects the growing collaboration between the GLIM-BioData RDM Centre and other RDM centres.

M4.3 - Training in eLABFTW - **22 September in CCMAR**

The milestone M4.3, which consisted of delivering a hands-on training session in eLabFTW at CCMAR, was successfully completed on 22 September 2025. This training was organised in collaboration with the CCMAR-LA, and conducted by Luciana Peixoto from GLIM-BioData, Maria Paola Tomasino and Bruno Louro. On Table 6 is the course schedule and program.

Table 6 - eLabFTW course at CCMAR schedule and program (https://events.biodata.pt/event/13/)

	eLabFTW course at CCMAR program
14:00-14:45	Introduction to BioData.pt and the GLIM-BioData project - Gateway for living data management.
14:45-15:15	Introduction to open electronic laboratory notebook manager, eLabFTW.
15:15-17:00	Hands on exercise on eLabFTW - Document Experiments and resources

The session was attended by 20 participants, 17 from CCMAR, and 3 from University of Algarve, reflecting broad institutional interest and engagement. This milestone contributes to the capacity-building goals of the GLIM-BioData project by supporting the adoption of digital tools for research data management across associated institutions. It also reflects the growing collaboration between the GLIM-BioData RDM Centre and other RDM centres.

Task 4.5 Training Data Stewards for Life Sciences

The Training Data Stewards for Life Sciences course was organised in three sessions instead of the originally planned two sessions. So, although only one session had initially been planned for 17 November, two sessions were ultimately delivered, as reflected in this description. Deliverable D4.4 corresponds to the successful completion of the second and third sessions of the Training Data Stewards for Life Sciences course, held online on 2 October 2025 and in person on 19 November. These sessions were organised by BioData.pt within the scope of the GLIM-BioData Research Data Management Centre as described in figure 1 and in the link https://events.biodata.pt/event/8/.

Figure 1 - Schematic representation of all sessions organisation of the Training Data Stewards for Life Sciences course.

D4.3 - Training Sessions for Data Stewards - 1st session - **7 July 2025 (Online)**

The deliverable D4.3 corresponds to the successful completion of the first of three sessions of the *Training Data Stewards for Life Sciences* course, held online on 7 July 2025. Organised by BioData.pt within the framework of the GLIM-BioData Research Data Management Centre, this fully hands-on session introduced participants to essential tools and practices for supporting research data management (RDM) from the earliest stages of the data lifecycle.

The programme covered a broad range of key topics relevant to data stewardship in the life sciences. Inês Chaves and Gil Poiares Oliveira provided an overview of the data lifecycle and the role of data stewards; Pedro Barros (ITQB) introduced controlled vocabularies and the Ontology Lookup Service; Carolina Ventura Costa (CIIMAR) demonstrated the use of electronic lab notebooks through eLabFTW; Ricardo Leite led a session on version control with Git and GitHub; and Francisco Pina Martins (FCUL/IPS) presented workflow management using Snakemake.

This first session also marked the launch of the course's *project-based learning* component, in which participants work on real-life inspired case studies across all stages of the data lifecycle. This methodology, based on active learning principles, promotes deeper engagement and practical skill development by encouraging participants to apply concepts in collaborative, problem-solving contexts. The first project stage involved identifying a case study, selecting relevant vocabularies, and designing a data processing pipeline.

A total of 12 participants attended the training (3 from FCUL, 2 from CIIMAR, 2 from BioData.pt, 2 from CCMAR, 1 from Universidade do Algarve, 1 from GIMM and 1 from Universidade de Aveiro). The diversity of institutional affiliations and scientific backgrounds among participants highlights the broad relevance and appeal of the course, confirming the growing interest in data stewardship across different domains.

The deliverable D4.3, which comprised the Training Sessions for Data Stewards 2025, was completed through the successful delivery of the first session, which took place online on July 7, 2025 (https://shorturl.at/wCln4)

D4.4 - Training Sessions for Data Stewards - 2nd session - **17 November 2025 (in person).**

The second session began with a brief catch-up and a presentation of the students' completed Stage A of the project, part of the course's *project-based learning* component. It then moved on to key principles and national services related to preserving, sharing, and reusing research outputs. The programme covered nonsensitive data sharing through Dataverse (DMPortal), presented by Gil Oliveira, followed by an overview of the Research Data Management strategy at FCT|FCCN delivered by Filipa Pereira and Pedro Sobral. They also presented the Polen Repository, a national service supporting data preservation. The session concluded with Jorge Oliveira (INESC-ID | IST), who discussed the sharing of sensitive data using FEGA and Beacon.

The third session took place face-to-face during the All Hands meeting in Porto, beginning in the Grand Hall of Building EA in the morning and continuing in Room 03.P1.E3 in the afternoon. It opened with a presentation by Mijke Jetten (ELIXIR NL) on the professionalisation of data stewards. The session then addressed several core aspects of research data management, including the development of Data Management Plans, the evaluation of FAIRness, and the provision of feedback on DMPs. Participants also examined funder open science policies and relevant institutional policies. The session concluded with the students' project presentations.

A total of 9 participants attended the session 2 (3 from FCUL, 2 from CIIMAR, 1 from BioData.pt, 2 from CCMAR, 1 from Universidade do Algarve). A total of 13 participants attended the session 3 (3 from FCUL, 1 from CIIMAR, 2 from BioData.pt, 1 from CCMAR, 1 from Universidade do Algarve, 2 from GiMM, 3 from INESC-ID/IST). The diversity of institutional affiliations and scientific backgrounds among participants highlights the broad relevance and appeal of the course, confirming the growing interest in data stewardship across different domains.

Together, these sessions offer a structured and practical path towards professionalising data stewards in life sciences research, contributing to national capacity building in open science and FAIR data management.

Task 4.6 Capacity Building

Anchored in training courses and other science dissemination events such as technical meetings, community and platform meetings, and general meetings with members of all associates, and conferences, the GLIM-BioData promoted national and international collaboration and create links between academia and industry (Microbiome Community Facilitator that came from industry). The organisation of the 3rd Microbiome Summit, that will happen in April 2026, organised by BioData.pt Microbiome Community was started.

WP5 - Impact and Sustainability

Strategies for the sustainable growth of GLIM-BioData were developed and implemented, including financial planning, resource allocation, and the enhancement of data management infrastructure, in collaboration with WP1. Opportunities for skills development, career advancement, and training for GLIM-BioData members were promoted in collaboration with WP4.

Task 5.1 Facilitation of collaborations within communities and Portuguese Data Steward Network

GLIM-BioData members were encouraged to participate in communities of researchers from different scientific fields in order to foster collaboration among national research groups, using mailing lists, Slack channels, and shared folders on cloud-based file-sharing infrastructure.

Active dialogue with the scientific community was also promoted through the creation of dedicated Open Science Working Groups (Non-sensitive Life Data and Sensitive Life and Health Data) within the associated institutions, aimed at sharing experiences and best practices in Open Science and fostering collaboration and integration with the Portuguese Data Steward Network (D5.1). Within this network, GLIM-BioData members attended monthly meetings, shared useful links and resources on data stewardship via the network's mailing list, and contributed to the definition of the Research Data Management (RDM) Data Professional profile by providing insights into life sciences and biological data handling requirements.

D5.1 - Integration and collaboration in the national network of data stewards - **March 2025**

The deliverable D5.1, which comprised GLIM-BioData RDM Centre integration and collaboration in the Portuguese Network of Data Stewards (PNDS), was fulfilled through the participation of BioData.pt's Data Steward, Miguel Cisneiros, in this national network coordinated by Re.Data. His involvement in the network was formalised by his attendance at the welcome meeting for new members, organised by Re.Data. Additionally, his subscription to the official mailing list (datastewards [at] lists.redata.pt) provides further evidence of active engagement in the network. These actions mark the effective inclusion of the GLIM-BioData RDM Centre in the national data stewardship landscape and reinforce its collaborative role within the broader Portuguese RDM ecosystem. Other BioData.pt associate members should join this network, as we will promote the network on our communication channels as well as in the communities.

Although this deliverable was initially scheduled for completion only at the end of the project (see Table 1 of the 1st quarter of 2025 GLIM-BioData progress report), we anticipated its execution and successfully completed it by the end of March 2025. The active participation in the PNDS continued until the end of the project.

Task 5.2 Identification and Development of Resources

Work was carried out towards the identification and development of essential national computational and software resources required for cutting-edge research in the life and health sciences, ensuring that these resources complied with European standards and protocols, thereby enhancing their visibility and impact (M5.1).

M5.1 - Assessment of BioData.pt node and communities services - **June 2025**

The assessment of BioData.pt node and communities services, which corresponds to milestone M5.1, was successfully achieved within the planned timeline (Evidence Report Milestone M5.1 - Assessment of BioData.pt node and communities services). This milestone encompassed the evaluation of BioData.pt | ELIXIR Portugal Node and Community Services through the implementation of a two-tier badge system, distinguishing "Node Services" and "Community Services" according to criteria such as maturity, quality, and availability.

The process involved the following steps:
- Establishment of an Evaluation Board;
- Draft of the Service Evaluation Form based on guidelines used internationally, namely by other ELIXIR Nodes;
- Publishing a description of the process in the BioData.pt Website, including an open call for expressions of interest via the ELIXIR Portugal Node Coordinator;
- Contacting maintainers of current services as to appeal to their participation on the expression of interest;
- Convene a meeting of the Evaluation Board as to analyse each of the filled-out forms;
- Update the Service pages on the BioData.pt website with the selected services.

For each service, a decision was reached to either:
1) Add it to the catalogue;
2) Remove it from the catalogue;
3) Promote or demote it.

Task 5.3 Participation in projects promoted by European infrastructures (ESFRI) and projects funded by European funds

This task promoted the active participation in European projects and in initiatives promoted by ESFRI (e.g. the ELIXIR Europe infrastructure) (M5.2), which had the potential to transform national strategies in the areas of health and life sciences through the sharing of knowledge, resources, and collaborations.

M5.2 - Participation in projects promoted by European infrastructures (ESFRI) and/or funded by European funds - **End of Project**

The milestone M5.2, which consisted of actively participating in European projects and projects promoted by ESFRI (e.g., the ELIXIR Europe infrastructure), which have the potential to transform national strategies in the areas of health and life sciences, through knowledge, resources, and collaborations.

The portfolio of grants for 2025, is composed of 7 EC-funded grants (4 of them dedicated to infrastructures) and 6 ELIXIR Commissioned Service grants. Some of these projects are being managed by BioData.pt, and for that, the participants have a cost center.

Five EC-grants proposals are currently under review (3 of them dedicated to infrastructures), including one coordinated by the ELIXIR- Europe for which results are expected to be released in Jun 2026:

Furthermore, we are planning the submission of two ELIXIR projects in 2026, in the Node and People tier, related to Node development, and others related to Training. These projects would start in 2027.

Task 5.4 Visibility and Advocacy

The visibility of the activities, developments, and contributions of the GLIM-BioData Centre was promoted at both national and international levels. Advocacy was undertaken to highlight the importance of national, data-driven research in bioinformatics for advancing health technologies and the life sciences.

The integration in the Portuguese Data Stewards Network (D5.1), within which GLIM-BioData established and maintained a biological data community sub-group and set up a core group of data stewards for sensitive life and health data (M5.3). In addition, active contact and interaction with other working groups, networks, and initiatives at both national and international levels were promoted. GLIM-BioData also aligned closely with the Portuguese Data Stewards Network to disseminate key information, events, training opportunities, and resources related to data stewardship and Research Data Management (RDM).

M5.3 - Establishment of a core group of data stewards for sensitive life and health data - **November 2025**

The milestone M5.3, which comprises the establishment of a core group of data stewards for sensitive life and health data, was initiated during the National Data Stewards Network meeting organised by Re.Data, on June 17th 2025 (https://shorturl.at/mpTGI). On this occasion, BioData.pt formally presented the creation of a national Core Group of Data Stewards for Life and Health, supported by the GLIM-BioData project. The initiative aims to connect data stewards, bioinformaticians, and professionals across Portugal, fostering collaboration, disseminating best practices, promoting Open Science and FAIR principles, and supporting the visibility of relevant training opportunities, events, and career opportunities. 16 members are already registered, and the next step will be to schedule its first meeting.

Task 5.5 KPIs Definition and Impact Monitorisation

The quality and impact of the Research Infrastructure's activities were monitored qualitatively and quantitatively through the use of impact pathways (P) from the RI-PATHS methodology, which had already been implemented at BioData.pt, in addition to the commitments established by CoARA. Key Performance Indicators (KPIs) were used for impact monitoring (M5.4).

M5.4 - Key Performance Indicators definition and monitorisation - **End of Project**

The milestone M5.4, which consisted in the definition and monitorisation of GLIM-BioData Key Performance Indicators using the RI-Path Framework.

The RI-Path Framework provides a structured methodology for planning, monitoring, and assessing impact in Research Infrastructures (RIs) and RI-related projects. It is grounded in a pathway-based logic, which conceptualises impact as a progressive process moving from enabling activities and outputs towards medium- and long-term outcomes.

Rather than treating impact as a single end result, RI-Path recognises that **sustainable impact emerges through interconnected pathways**, each representing a sequence of actions, outputs, and changes involving different stakeholder groups.

Within this framework, **Strategic Objectives** play a central role. Each Strategic Objective is explicitly linked to one or more **RI-Path impact pathways** (e.g. collaboration, capacity building, resources, governance, or visibility). This linkage ensures that project monitoring goes beyond output counting and instead focuses on **how and why change is expected to occur**. The first step of this methodology is the identification of Strategic Objectives followed the identification of the Key Performance Indicators (KPI) for each objective. The monitorisation of the KPI followed the reporting periods.

Strategic Objective – Facilitate Collaboration

RI-Path Pathway: P11 – Collaboration and Co-creation
Primary RI-Path Stage: Medium-term outcomes
Secondary Contribution: Short-term outcomes

This strategic objective primarily contributes to **medium-term impact**, as it focuses on the consolidation of networks, co-creation practices, and sustained interactions among stakeholders (Table 7).

Table 7 - Facilitate Collaboration KPI list

Number of Meetings among GLIM community
Number of events hosted/organised by GLIM
Number of collaborations with other stackeholders (REDATA, MIRRI, Instruct, EuroBioimaging, GenomePT)
Number of collaborations with other GDI centres (GBIF, FAIRWAY, CeSDHR, ...)

The KPIs associated with collaboration pathways demonstrate **consistent and cumulative progress** throughout the reporting period.

Strategic Objective – Resource Identification and Development

RI-Path Pathway: P4 – Resources and Infrastructure
Primary RI-Path Stage: Short-term outcomes
Secondary Contribution: Medium-term outcomes

This objective is predominantly positioned within the **short-term RI-Path stage**, as it focuses on the provision, and dissemination of Data Management resources and Tools, including the number of courses (Table 8).

Table 8 - Resource Identification and Development KPI list

Number of resources/Tools
Number of training Courses
Page views (cumulative)

Strategic Objective – Capacity Building

RI-Path Pathway: P6 – Capacity Building and Skills Development
Primary RI-Path Stage: Medium-term outcomes
Secondary Contribution: Long-term impact

Capacity building is anchored in **medium-term impact**, as it reflects changes in skills, competences, and organisational capabilities (Table 9).

Table 9 - Capacity Building KPI list

Trainers list
Number of trainings sessions in universities or research institutions (Ready for BioData Management, Curating Datasets, eLabFTW)
Number of scientific users/trainees (cumulative)

Capacity building activities display a progressive and accelerating trajectory.

Strategic Objective – Visibility and Advocacy

RI-Path Pathway: P13 – Communication, Dissemination and Advocacy
Primary RI-Path Stage: Short-term outcomes
Secondary Contribution: Medium- and long-term impact

Visibility and advocacy actions primarily generate **short-term outcomes** by increasing awareness, reach, and stakeholder engagement (Table 10).

Table 10 - Visibility and Advocacy KPI list

Number of Posts in Social Media
Visitors on Website (GLIM)
Roadshows
Number of promotional events, exhibitions, fairs
Visits to (high-level) scientific events

Visibility and advocacy indicators show high and sustained performance.

Strategic Objective – Sustainable Node Development

RI-Path Pathway: P2 – Governance and Sustainability
Primary RI-Path Stage: Long-term impact
Secondary Contribution: Medium-term outcomes

This strategic objective is directly aligned with **long-term RI-Path impact**, as it focuses on sustainability, governance structures, and the durability of project outcomes beyond the funding period (Table 11).

Table 11 - Sustainable Node Development KPI list

Number of ELIXIR Project submissions
Number of ELIXIR Project financed
Number of European Projects Submissions

Although smaller in absolute numbers, this strategic objective shows steady development.

The milestone M5.4 - Key Performance Indicators definition and monitorization has reached 100% completion, as initially planned for this reporting period.

Financial Execution

	Total budget until the end of the Project (Values without IVA)	Execution until 1st report	Execution until 2nd report	Execution until the end of the Project
Total Budget for activities	**220 200 €**	**15 685 €** (7.12%)	**74 585 €** (34%)	**220 200 € €** (100%)
Human resources and service provision - includes BioData.pt HR, hiring a Data Steward, hiring a SysAdmin and service provision of a DPO	139 000 €	9 000 € (6.5%)	61 000 € (44%)	139 000 € (100%)
Training actions - Trainers, fee for Train-the-Trainer, and travel and accommodation	15 200 €	600 € (4%)	5 500 € (35%)	15 200 € (100%)
Events - Technical meeting, All hands, Roadshow, travel and accommodation	18 000 €	3 000 € (16.6%)	5 000 € (28%)	18 000 € (100%)
Technical Resources - server, workstation and data center maintenance	48 000 €	3085 € (6.4%)	3085 € (6.4%)	48 000 € (100%)

3. Additional Relevant Considerations

The GLIM-BioData project carried out a wide range of coordination, communication, training, and integration activities during this period.

The GLIM-BioData consortium actively engaged in several national and international initiatives that reinforced its visibility and integration within the broader research data ecosystem. BioData.pt participated in the GBIF training course "Publicação de dados de biodiversidade através do GBIF - 2025" (25-27 June, ISA, Lisbon), where Data Steward Miguel Cisneiros represented the project and strengthened expertise in FAIR data practices (https://shorturl.at/GRDcu). BioData.pt team also participated in major national science events, including the Jornadas de Computação Científica 2025 (7 May, Aveiro) (https://shorturl.at/pYqfP) and Encontro Ciência 2025 (9-11 July, Lisbon), promoting the project's role in advancing Open Science and Research Data Management (https://shorturl.at/F7DrT).

At the European level, BioData.pt took part in the ELIXIR Heads of Nodes Retreat 2025 (https://shorturl.at/X6Enp), represented by ELIXIR Portugal Head of Node Professor Ana Teresa Freitas, and was also represented by Jorge Oliveira at the Genome of Europe Consortium Meeting (https://shorturl.at/gp3fh), further consolidating the project's presence in strategic international infrastructures. In parallel, GLIM-BioData actively contributed to the 2025 ELIXIR All Hands Meeting in Thessaloniki (3-7 June), where the BioData.pt team presented their work and exchanged knowledge with European counterparts.

Capacity-building activities were reinforced through participation in the Bioinformatics Workshop hosted by the School of Tourism and Maritime Technology of the Polytechnic Institute of Leiria (https://shorturl.at/00QRc), as well as through the ELITMa Module training on Node Data Management Strategy (https://shorturl.at/IZQ2u).

These activities highlight the consortium's strong commitment to community building, international collaboration, and the promotion of best practices in research data management across the life and health sciences.

4. Sustainability Plan

The GLIM-BioData Centre, integrated in BioData.pt, was established to strengthen national capabilities in the management, curation, and sharing of life and health data, with a strong alignment to FAIR principles and the One Health approach. Over the course of the project, significant progress was achieved in establishing robust data management infrastructure, developing training programmes, fostering national and international collaborations, and contributing to European Research Infrastructures (RIs) such as ELIXIR.

BioData.pt is presently coordinating the implementation of the Portuguese node of the European "Genomic Data Infrastructure" network, which it will manage when deployed. The infrastructure will host the genomes of Portuguese individuals as part of the European 1+ Million Genomes initiative. In this context, BioData.pt, together with

SPMS (Serviços Partilhados do Ministério da Saúde), is also the entity responsible for the data management and computational architecture proposal for the national strategy to implement genomic medicine in Portugal. Additionally, BioData.pt is responsible for implementing the National node of the European Genome Archive (EGA) network, which enables the management of sensitive data obtained by scientific projects in Portugal, giving European and international visibility to research conducted in Portugal and participating in European efforts in the field of Big Data analysis.

BioData.pt is responsible for managing the Portuguese node of the European ELIXIR network (European Life Sciences Infrastructure for Biological Information), a key component of the ESFRI (European Strategy Forum on Research Infrastructures) network.

As a member of the National Roadmap for RI (RNIE), BioData.pt maintains a close partnership with FCT and its digital services unit FCCN, with whom a cooperation protocol has been signed.

BioData.pt is currently a member of the RCTS Network, maintained by FCCN. As a distributed digital infrastructure, the support of FCCN has provided access and support to RCTS services such as software licences, hosted web platforms, connectivity solutions and access to RNCA National Network for Advanced Computing. BioData.pt | ELIXIR Portugal web services are already running on RNCA infrastructure, and reaching a worldwide audience, especially with respect to the popular Beacon and FEGA Portugal federated human data services.

Strategic Sustainability Objectives

This report highlight the sustainability strategy of the Centre, detailing ongoing activities, funding streams, collaborations, and planned initiatives to ensure the long-term impact and viability of GLIM-BioData beyond the official project end date.

The sustainability of GLIM-BioData focuses on five key objectives:

1. **Integration into National and European Research Infrastructures:** Active participation in European projects and ESFRI initiatives, including ELIXIR Europe, to leverage knowledge, resources, and collaborations, thereby influencing national strategies in health and life sciences.
   a. Collaborations with ELIXIR Europe and ELIXIR nodes, have positioned GLIM-BioData as a key national node in life and health data management and strengthened its role in European infrastructures.
   b. Engagement in ELIXIR Commissioned Services has facilitated knowledge exchange and capacity building, creating a stable platform for future collaborations.
2. **Financial Sustainability:** Management of a diverse portfolio of grants, ensuring that infrastructure development, personnel, and operational costs are supported through a combination of EC-funded projects, ELIXIR Commissioned Services, and co-funded initiatives.

a. BioData.pt is a non-profit association currently formed by 15 members, which are entities dedicated to Research, Development and Innovation, covering the entire national territory (Figure 1). Its members include Universidade do Minho (UMinho), Centro Interdisciplinar de Investigação Marinha e Ambiental (CIIMAR), Applied Molecular Biosciences Unit (UCIBIO), Universidade de Aveiro (UA), Universidade de Coimbra (UC), Instituto de Tecnologia Química e Biológica (ITQB NOVA) e NOVA Medical School, Faculdade de Medicina da Universidade de Lisboa (FMUL), Instituto de Biologia Experimental e Tecnológica (iBET), Instituto Gulbenkian de Ciência (IGC), Instituto de Medicina Molecular (iMM), Faculdade de Ciências da Universidade de Lisboa (FCUL), Instituto de Engenharia de Sistemas e Computadores - Investigação e Desenvolvimento (INESC-ID), Instituto Superior Técnico (IST), Centro de Biotecnologia Agrícola e Agro-alimentar do Alto Alentejo (CEBAL), and Centro de Ciências do Mar (CCMAR). Despite its autonomy from its members, BioData.pt relies on the existing excellence in research at associated institutions. These synergies result in a huge potential for BioData.pt to participate in transnational research and provide medium-term national economic impact.

b. Milestone **M5.2**: Active participation in European projects and ESFRI-promoted initiatives, including ELIXIR Europe, prove this (M5.2 Evidence report).

c. The portfolio of grants for 2026, is composed of 7 EC-funded grants (4 of them dedicated to infrastructures) and 6 ELIXIR Commissioned Service grants. Some of these projects are being managed by BioData.pt, and for that, the participants have a cost centre.

d. Five EC-grants proposals are currently under review (3 of them dedicated to infrastructures), including one coordinated by the ELIXIR-Europe for which results are expected to be released in Jun 2026.

3. **Capacity Building and Training:** Continued development of training programmes for researchers, data stewards, and technicians, ensuring a sustainable pipeline of skilled personnel in bioinformatics, data management, and sensitive data stewardship.
   a. Development and delivery of intensive training courses in Research Data Management (RDM), FAIR principles, and the management of sensitive life and health data.
   b. Certification of trainers through ELIXIR's "Train-the-Trainers" programme. In 2026 a TtT will be delivered in Portugal organised by BioData.pt.
   c. Hands-on sessions with eLabFTW and FAIR-EASE datasets promoted adoption of best practices in data collection, curation, and reuse.
   d. Training activities targeted researchers, PhD students, postdoctoral researchers, lab managers, and data stewards, creating a sustainable community of skilled professionals.

4. **Community Engagement and Networking:** Maintenance and expansion of national and international networks, including the Portuguese Data Stewards

Network, Open Science Working Groups, and engagement with life sciences communities to facilitate knowledge sharing, collaboration, and the adoption of best practices.

   a. Maintenance of Open Science Working Groups for Non-sensitive Life Data and Sensitive Life and Health Data (BioData.pt Communities).
   b. Monthly meetings open to all members will be maintained for resource sharing, network promotion and funding opportunities

5. **Monitoring and Impact Assessment:** Implementation of KPIs and adherence to frameworks such as CoARA and RI-PATHS to ensure the quality, impact, and visibility of the Centre's activities.

   a. Regular reporting of outputs, publications, and training metrics ensured visibility of impact and alignment with European standards.

These GLIM-BioData achievements collectively ensure that BioData.pt remains a sustainable, high-impact research infrastructure in the national and European life and health sciences landscape.

5. Conclusion

The GLIM-BioData Centre has successfully achieved its planned milestones within the project timeline, establishing a strong foundation for sustainability. The combination of diversified funding, trained personnel, robust infrastructure, and active European engagement ensures that the Centre can continue to deliver high-quality services, support national and European research strategies, and foster a FAIR, open-science culture.

Future sustainability will be secured through:

- Continued participation in European and ELIXIR-led initiatives.
- Expansion of the training and capacity-building programme.
- Integration of new infrastructure and computational resources into national research frameworks.
- Ongoing monitoring of impact and engagement with both national and international communities.

GLIM-BioData is thus positioned to maintain its strategic relevance and operational excellence beyond the end of the initial project period.