MISSIONE 4
ISTRUZIONE
RICERCA

ITINERIS

Digitalized IBISBA-IT distributed platforms, protocols, and validated pipelines for biomolecule/microorganism discovery, characterization, and engineering for bioprocess development and a circular economy transition

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA

Deliverable number:	D6.23
Work package:	WP6 – Terrestrial Biosphere
Intermediate Objective:	IO7
Deliverable type:	X Document, report
	☐ Websites, patent filings, videos, etc.
	☐ Other: please specify .....
Dissemination level:	X Public
	☐ Restricted
Estimated delivery (bimester):	B14
Actual delivery date:	31/03/2025
Author(s) (Partner-OU):	Mauro Di Fenza, Nicola Curci, Federica De Lise, Angela Capaccio, Beatrice Cobucci-Ponzano
Reviewed by:	ITINERIS Executive Board
Note:	

IR0000032 – ITINERIS, Italian Integrated Environmental Research Infrastructures System - CUP B53C22002150006 (D.D. n. 130/2022)
Funded by EU - Next Generation EU
Mission 4 “Education and Research” - Component 2: “From research to business” -
Investment 3.1: “Fund for the realisation of an integrated system of research and innovation infrastructures”

Table of contents

1. INTRODUCTION

1.1 Purpose of the document

This document outlines the management aspects involved in developing the IBISBA-IT digital infrastructure, which encompasses digitalised distributed platforms, protocols, and validated pipelines for biomolecule/microorganism discovery, characterisation, and engineering to facilitate bioprocess development and a transition to a circular economy.

The development of the IBISBA-IT digital infrastructure is grounded in the architecture crafted by IBISBA partners University of Manchester, VTT, and Universitat Autonòma de Barcelona, as detailed in the PREP-IBISBA deliverable D8.2[1] (Preliminary architecture for the IBISBA digital infrastructure). The primary aim of the digital infrastructure is to establish a digital ecosystem that provides tools, resources, datasets, services, and data management capabilities. This ecosystem is designed to bolster research, innovation, collaboration, and knowledge exchange among various entities, including IBISBA-IT research facilities, IBISBA (comprising IBISBA Central and National Nodes), the 22 research infrastructures (RIs) participating in ITINERIS, and the broader user community.

In this document, digital infrastructure and e-infrastructure will be used interchangeably.
It is important to note that this document is intended as a dynamic entity, subject to continual updates and refinements in response to evolving needs and developments.

1.2 Definitions and Acronyms

API

API stands for Application Programming Interface. An API is a set of rules, protocols, and tools that allow different software applications to communicate with each other. It defines the methods and data formats developers can use to interact with a particular software component, service, or platform, abstracting the underlying implementation details.

Digital Objects

A digital object refers to any entity or resource in a digital format that can be accessed, manipulated, or transmitted electronically. It encompasses a wide range of digital content, including but not limited to documents, images, videos, datasets, software applications, and web pages. Digital objects are characterised by their ability to be stored, replicated, and distributed across computer networks, making them versatile and easily shareable. They often have associated metadata, which provides descriptive information about the object, such as its title, author, creation date, and format.

DOI

A DOI (Digital Object Identifier) is a unique, permanent identifier assigned to digital resources (such as research articles, datasets, software, or reports) to ensure they can be reliably found, accessed, and cited over time.

ELIXIR
ELIXIR is a European life sciences infrastructure enabling researchers to access and analyse life science data to improve the value and impact of life science research on public health, the environment and the economy.

EOSC
EOSC stands for the European Open Science Cloud. It is an ambitious initiative by the European Commission to create a trusted, open environment for sharing and managing research data and services across borders and disciplines. EOSC aims to provide researchers and innovators with seamless access to a wealth of resources, including data, tools, and computing services, to facilitate collaboration and accelerate scientific discovery.

FAIR
FAIR is an acronym that stands for *Findable, Accessible, Interoperable,* and *Reusable*. It represents a set of guiding principles designed to enhance the usability and effectiveness of data assets, particularly in the context of scientific research and data management. These principles were developed to address data sharing, integration, and reuse challenges in the digital age.

FAIR Digital Objects
FAIR data objects refer to digital entities that adhere to FAIR principles, such as datasets, collections, or individual files.

GitHub
GitHub is a web-based platform used for hosting and collaborating on software development projects. It provides version control using Git, a distributed version control system and offers various features and tools to facilitate collaboration among developers working on the same project.

JSON
JavaScript Object Notation (JSON) is a lightweight, human-readable data interchange format widely used in web development to transmit data between a client and a server.

JSON API
JSON API is a specification for building APIs that use JSON as the primary data format for requests and responses.

OpenAPI
OpenAPI is a specification for building and documenting APIs and provides a standardised way to describe the structure, functionality, and usage of APIs, making it easier for developers to understand and interact with them.

PID
A PID is a long-lasting, globally unique, Persistent Identifier assigned to a digital object, such as a dataset, document, or research publication. PIDs are used to ensure the persistent and reliable referencing of digital objects, even as they move across different systems and repositories.

SEEK
SEEK or FAIRDOM-SEEK is an open-source web-based cataloguing and commons platform for sharing heterogeneous scientific research datasets, models or simulations, processes and research outcomes. It preserves associations between them, along with information about the people and organisations.

SEEK-ID
A persistent, unique identifier assigned to digital objects within SEEK-based platforms, such as FAIRDOMHub or IBISBAKhub.

URT
Unità di Ricerca presso Terzi (Third-Party Research Unit). These units are collaborative research entities that operate in partnership with other institutions, such as universities, research centres, or public and private entities.

2. STATE OF THE ART

2.1 IBISBA-IT: The first Italian Research Infrastructure for Industrial Biotechnology

IBISBA-IT (Industrial Biotechnology Innovation and Synthetic Biology Accelerator - Italy) is the Italian Node of the European distributed research infrastructure IBISBA (www.ibisba.eu). IBISBA is a pan-European distributed research infrastructure dedicated to industrial biotechnology, federating European expertise to provide researchers from academia and industry across the globe with integrated services for all stages of bioprocess development projects. ESFRI classifies the RI under the Health & Food domain for its strong focus on developing innovative bioprocesses and biomanufacturing, making it suitable for the production of pharmaceuticals, functional molecules for food and health-related products. However, IBISBA's activities have a strong environmental dimension, particularly in the development of sustainable bioprocesses that contribute to the circular bioeconomy, waste valorisation and carbon-neutral production systems.

IBISBA-IT operates within four well-defined research domains: Synthetic Biology, Green Chemistry, Sustainable Bioenergy, and Functional Food. Within these areas, its mission is to pioneer the development of new molecules and processes through enzyme/protein discovery and engineering and advance novel biotransformations and bioprocesses. These objectives are pursued through modern, holistic approaches encompassing 'omics' sciences, bioinformatics, molecular engineering, and bioreactor technologies. Ultimately, IBISBA-IT plays a vital role in bolstering Italian scientific research in industrial biotechnology, offering support to national and local governments, and fostering training initiatives to nurture expertise in this field.

2.2 Partner Institutions involved

IBISBA-IT (www.ibisba.it) is distributed throughout the country under the leadership of the CNR Institute of Biosciences and BioResources based in Naples (IBBR-NA). Alongside IBBR-NA, IBISBA-IT includes three other CNR partners: the Institute of Agricultural Biology and

Biotechnology (IBBA), the Institute of Food Sciences (ISA), and the Institute for the Animal Production System in the Mediterranean Environment (ISPAAM). Additionally, it involves several prominent academic institutions such as the University of Naples' Federico II' (UNINA), the University of Milano Bicocca (UNIMIB), the University of Insubria (UNINSUBRIA), the Alma Mater Studiorum - University of Bologna (UNIBO), the University of Tuscia (UNITUS), and the Insubric Research Institute Foundation for Life (FIIRV).

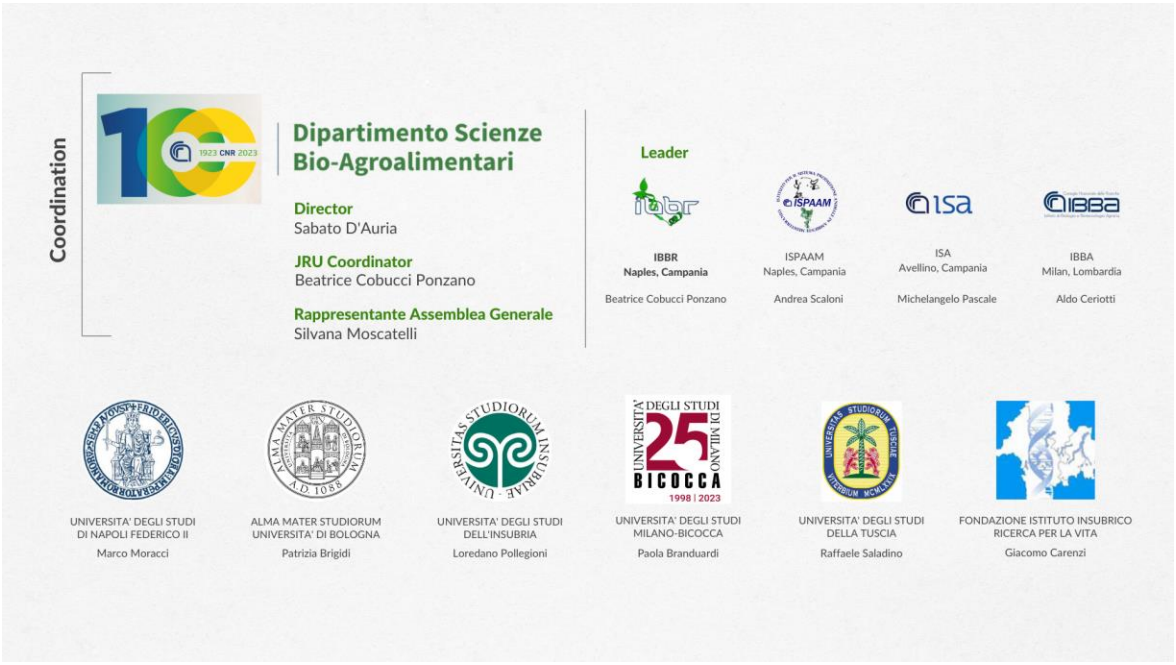

Figure 1. IBISBA-IT partner institutions.

IBBR is involved in the identification & characterization of novel CAZymes from extremophiles for their exploitation in lignocellulosic biomasses and waste products hydrolysis and valorisation.

IBBA focuses on the optimization of toxic polypeptides expression in eukaryotic hosts.

ISA is dedicated to optical protein biosensors for continuous monitoring of analytes.

ISPAAM is engaged in applying innovative omics approaches for proteomic monitoring changes during fermentation processes.

UNINA specialises in discovering and engineering enzymes for applications in bioremediation and waste treatments.

UNIMIB works on genetic improvement of microbial strains, metabolic engineering & control and validation of fermentation processes.

UNINSUBRIA is dedicated to *in silico* design and protein engineering for novel enzymatic activities.

UNIBO aims to identify novel sources of biomass, enzymes and bio-products.

UNITUS is involved in identifying and characterising biobased chemicals and natural bioactive substances.

FIIRV focuses on the isolation and characterisation of novel biomolecules.

2.3 Building a national digital infrastructure

IBISBA-IT is composed of a multidisciplinary partnership of autonomous facilities, each varying in their unique competencies and capabilities aimed at strengthening the overall Italian contribution within IBISBA. Similarly, the digital infrastructure must be federated to support the node's operations and deliver IBISBA-IT's research activities while adhering to the FAIR principles for data and other FAIR Digital Object management. Furthermore, the digital infrastructure must be configured to meet the specifications required by the IBISBA Knowledge Hub (IBISBAKHub) and the ITINERIS Central Hub.

2.4 IBISBA Knowledge Hub

The IBISBA Knowledge Hub (https://hub.ibisba.eu/), a customised version of the FAIRDOMHub repository[2], is a central, versatile online resource for storing, managing and sharing a wide range of scientific research data, models, processes, and outcomes. Members can oversee projects, log diverse data objects (e.g. experimental results, sample details, documents, and presentations) alongside relevant metadata, and control access to the content. Essentially, the Hub functions as an online repository and collaborative space for IBISBA's assets and project outcomes, offering a comprehensive suite of services.

Users can independently manage project assets, organising, sharing, and disseminating materials resulting from collaborative efforts and project-related events. The Hub catalogues project outcomes, standard operating procedures, workflows, template libraries, and small file repositories while also indexing large data files stored both internally and externally (Figure 2).

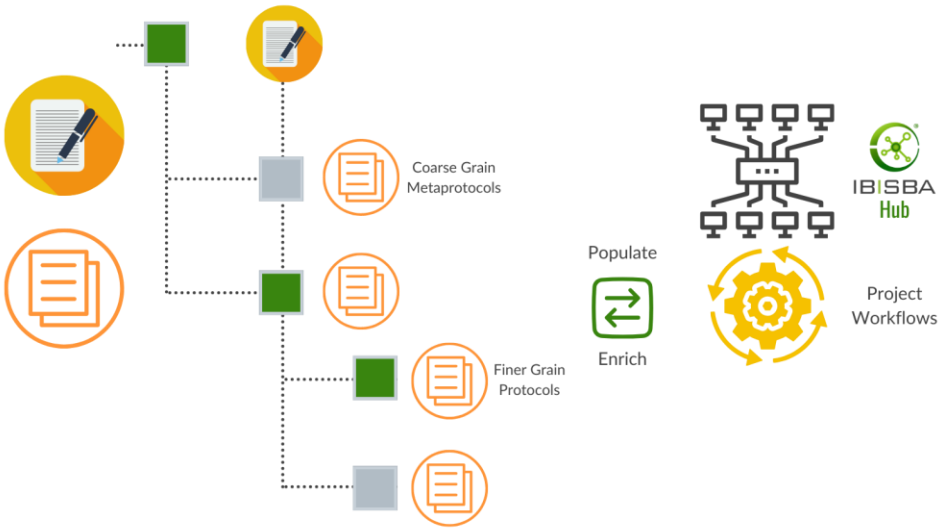

Figure 2. Modular workflows using standards and content managed by the IBISBA Knowledge Hub.

Precise access control mechanisms enable project managers to grant external users access as needed. Users can manage data resources using the Hub as the primary user management tool or in conjunction with local resources.

Security is paramount, with the Hub ensuring secure data storage while offering options for private storage with the flexibility to share or publish data when appropriate. It facilitates the storage, registration, sharing, and access of standardised protocols, shared computational workflows, and data format templates. Users can describe and establish relationships between data assets using rich metadata relevant to samples, organisms, models, protocols, data, computational workflows, and documents. Integration with third-party tools, such as modelling tools (e.g., JWS Online), metadata capture tools (e.g., Rightfield), and development repositories (e.g., GitHub), is seamless. The Hub offers an OpenAPI and JSON API conformant interface, enabling secured programmatic access for tools. One of its key features is its ability to establish connections between different assets, offering a comprehensive catalogue of IBISBA-specific knowledge assets, complete with information about the individuals and organisations involved.

The platform's architecture ensures that all research assets are meticulously catalogued with their associated metadata. This unified view of assets is organised according to the ISA framework (https://isa-tools.org/index.html), which stands for Investigation (the project context), Study (a unit of research), and Assay (analytical measurement). Additionally, the platform offers flexibility regarding content location, allowing small datasets to be uploaded directly to the Hub while larger datasets can be referenced externally, residing in third-party data repositories.

As a result, the IBISBA Knowledge Hub's catalogue encompasses a diverse array of data stores, spanning locally-installed e-infrastructures, nationally-provided e-infrastructures, and public archives such as the Core Data Resources and recommended Deposition Databases of ELIXIR. This versatility ensures that IBISBA researchers have seamless access to a wealth of valuable resources, fostering collaboration and innovation within the scientific community.

3. AIM OF THE WORK

Data management practices within IBISBA must meet the required standards for future operations. Many facilities lack centralised storage systems, leading to data being collected and maintained inconsistently and securely. Additionally, data sharing often relies on email or local network transfers, with inadequate data management policies and limited use of metadata collection software. Despite some awareness of the FAIR principles among facility operators, implementation still needs to be completed due to a lack of expertise and training. As data is generated and analysed, various tasks such as data collection, storage, transfer, processing, analysis, and sharing require the implementation of centralised services and resources.

IBISBA-IT provides its output in the form of experimental data, protocols, meta-protocols, and project workflows to streamline bioprocess development. These data streams necessitate dynamic and flexible support from the digital infrastructure, and all digital objects must be securely stored in trusted repositories for long-term access.

The aim of the work is to digitalise harmonised data (*from D6.13 - Harmonized environmental metagenomic and enzyme characterization datasets to build standardized and interoperable data files to be stored, shared, and used for data integration into process models*) and standardised protocols and pipeline for the bioconversion and valorisation of a natural resource (e.g. lignocellulosic biomass) to create a data repository for the steps of the bioprocess performed by the services of IBISBA-IT.

IBBR-NA will manage the collection of various digital assets, including facility protocols and meta-protocols, ensuring they are properly stored, indexed, and accessible. Compliance with data format standards (*D6.3 - Standards for data acquisition and storage developed and tested to facilitate digitalization of bioprocess analysis and accelerate bioprocess development*) is critical to guarantee that these digital objects are correctly maintained on the IBISBAkHub and ITINERIS Central Hub. In addition, all assets must be accompanied by metadata that aligns with EOSC and ESFRI standards and shared in portable formats recognised by the broader Life Science RI community. These assets must be indexed and made accessible according to the permissions defined by IBISBA data governance. Furthermore, IBBR-NA will oversee the storage and indexing of other key files, such as documents and publications, to ensure they can be easily accessed and retrieved.

Data storage will be carried out across two units: one located at IBBR-NA, and the other at the IBBR-NA URT within the Department of Biology at the University of Naples Federico II, Together, these units will form a digital infrastructure named IBISBA-IT Data Nexus, or simply Nexus.

Access to the Nexus requires authorisation and authentication of facility-specific user credentials. Digital security spanning all components of the digital infrastructure, including the Nexus, data repositories held by partner facilities, and the Central Data Hub must be guaranteed. Security includes various aspects, such as site identification, authentication, and authorisation. Site identification ensures that sites can be readily recognised as expected sites, while authentication verifies the identity of individuals or computers attempting to access information. Lastly, authorization protocols determine access permissions for authenticated users and specify the extent of their access rights.

3.1 Implementation and configuration of the Nexus

The foundation of IBISBA-IT digital infrastructure lies in developing and configuring advanced data storage solutions. The installation involved setting up two Network-Attached Storage (NAS) devices configured to meet the RI's specific data management needs. This setup ensures seamless integration

into the existing network, enabling efficient handling of large datasets. The configuration process included optimising storage settings, network parameters, and security protocols to maximise performance and safeguard data integrity. These measures are necessary to implement a resilient and high-performing storage system that efficiently supports the RI's research activities.

3.1.1 Installation of QNAP TS-1277XU-RP NAS Servers

The QNAP TS-1277XU-RP NAS are high-performance devices designed for data-intensive environments. They feature powerful processors, offering scalable, high-speed storage and uninterrupted operation, while the QuTS hero Operating System (OS) provides advanced protection, including features such as duplication, compression, and snapshots indispensable for data management in storage systems, as they ensure data integrity and provide mechanism for data protection and recovery (https://www.qnap.com/en/). Each NAS also supports virtual machines, enabling flexible deployment of lightweight applications and secure remote access through Virtual Private Networks (VPN) and firewalls. These servers, equipped with 48 hard drives, ensure optimal balance of high storage capacity, data integrity, and reliability for data-intensive tasks and long-term operations.

3.1.2 Network reliability and resilience

The inclusion of APC Smart-UPS units protects the NAS servers from power outages and fluctuation, minimising the risk of data loss and hardware damage. Additionally, a dedicated firewall provides robust security measures, managing and controlling network traffic to safeguard the infrastructure from external threats, ensuring the confidentiality and integrity of data.

3.1.3 Advanced NAS functionality with Linux-based architecture

The QuTS hero OS is built on a Linux-based architecture designed to provide reliable, high-performance, and secure data storage while supporting data-heavy applications. QuTS hero enhances system efficiency through its multitasking interface, simplifying the management of data, user accounts, and applications. The built-in App Centre expands the NAS's capabilities with tools like Container Station and Virtualization Station. These tools enable the deployment of data management applications, the IBISBA-IT website, and an instrument booking system within isolated, secure environments (containers). Additionally, they allow for the simultaneous running of multiple operating systems, enhancing flexibility to the NAS infrastructure, optimising hardware usage, and maintaining isolated environments for different applications.

3.1.4 Fortified network security and secure data transmission

To safeguard the infrastructure against threats and vulnerabilities, pfSense, a powerful, open-source firewall and routing platform was deployed. It provides a multi-layered defence system, which analyses network traffic to ensure that only legitimate traffic is allowed through, while potentially malicious data flow is blocked. The platform also incorporates an Application Layer Gateway (ALG) to manage application-specific protocols securely and seamlessly. Network Address Translation (NAT) adds another layer of protection by shielding private device IP addresses behind a public one, keeping internal systems safe and hard for external attackers to directly access them.

Secure communication is maintained through Virtual Private Networks (VPN) integration, with OpenVPN establishing encrypted data transmission for remote access and site-to-site connections, facilitating secure connections across the RI facilities. Intrusion Detection and Prevention Systems (IDS/IPS) add another layer of defence by monitoring network activity in real time and responding to potential threats through signature-based and anomaly detection techniques.

Granular firewall rules further enhance security, allowing precise control over network access by regulating data flows based on IP addresses and other parameters. This robust setup ensures integrity and reliability of the network while safeguarding data transmission.

3.2 Implementation and configuration of lab equipment scheduling

An instrument booking system has been developed to manage access to the RI, ensuring organised and efficient use of the facilities. The system allows users to create accounts, submit access requests, and comply with predefined access policies. Integrated into the RI's website (www.ibisba.it) hosted on the NAS housed at CNR-IBBR, the booking system relies on Microsoft Bookings for managing appointments. This system allows users to reserve instruments and related services, ensuring that all resources are appropriately allocated and utilised. Additionally, it enables data extraction and instrument usage monitoring, providing insights into equipment use patterns, supporting better resource planning, helping identify underused instruments, and ensuring maintenance schedules are aligned with actual usage, ultimately improving operational efficiency. The integration with Microsoft Bookings streamlines the booking process, offering a user-friendly interface while maintaining control over access based on availability and policy requirements.

4. PROTOCOL AND WORKLOW DIGITALISATION

4.1 Workflow digitalisation

Digitalisation aims to streamline laboratory workflows by reducing human intervention and implementing machine-readable digital ecosystems and automation. Although inherently complex, digitalisation is essential for scientific research, where a vast amount of experimental data must be collected, processed, and shared across institutions. This transformation involves responsible data management, meaning the ethical and systematic handling of data throughout its lifecycle (from its generation to its preservation and reusability) while also being compliant with FAIR principles, helping researchers adhere to standardised data stewardship practices [3,4]. Figure 3 illustrates the concept behind the data lifecycle.

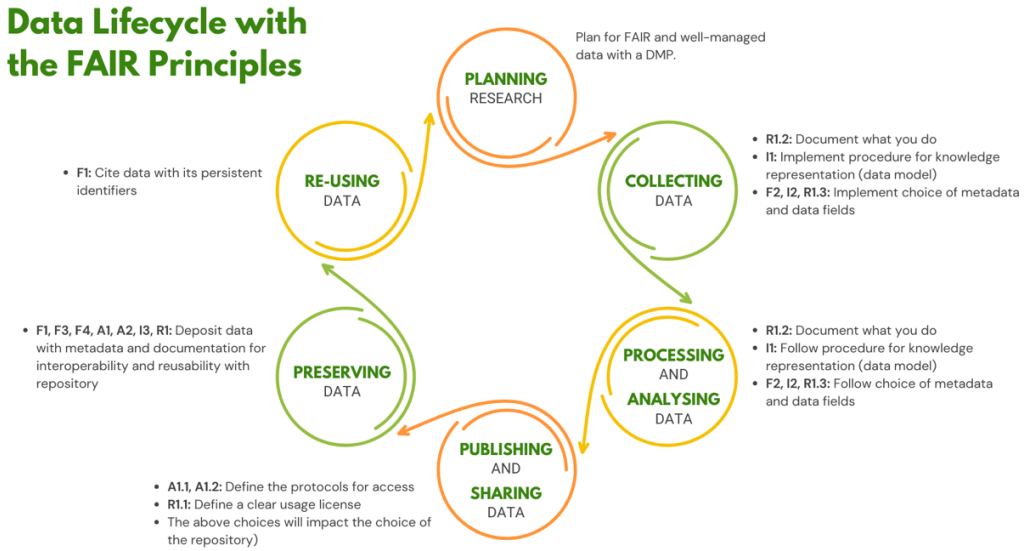

Figure 3. Data Lifecycle with the FAIR principles. Adapted from the original by Erik Schultes, Barbara Magagna, and Alessa Gambardella, FAIR Data Stewardship, 5 December 2024. Licenced under CC-BY-NC-SA 4.0.

By implementing automated workflows, standardised metadata structures, and long-term storage solutions, research infrastructures can enhance efficiency while ensuring compliance with Open Science frameworks[5]. As one might expect, transitioning to complete digitalisation is an incremental process requiring considerable effort.

For IBISBA-IT, digitalisation is essential in creating a federated research environment where experimental data, protocols and other assets are seamlessly integrated across multiple facilities. By implementing automated data workflows, IBISBA-IT aims to build a digital infrastructure that transitions from traditional, manual data management to an automated (and FAIR-compliant) digital ecosystem.

In many cases, IBISBA-IT laboratories still rely on fragmented systems for data management, often using standalone spreadsheets, paper-based notebooks, or local file storage solutions that make laboratory procedures laborious and inefficient. As a consequence, data collection and sharing may represent an issue, as they still rely on methods that increase the risk of data loss and limited accessibility. Additionally, metadata documentation, crucial for assuring reproducibility and data reuse, is often inconsistent or absent. These gaps slow research and make it inadequate to comply with international standards, particularly the FAIR principles. To address these issues, IBISBA-IT has outlined a digitalisation strategy that integrates automated data workflows, data and metadata management, and long-term storage solutions. This strategy is supported by an IBISBA-IT Data Management Plan (DMP), which provides a foundational framework for data handling within the infrastructure (https://ibisbahub.eu/documents/75?version=4). Derived from the DMPs developed in the IBISBA 1.0[6] and PREP-IBISBA[7] projects, the IBISBA-IT DMP was designed as an integrative component to emphasise laboratory processes rather than the broader project-focused approach previously adopted. The DMP was created using Data Stewardship (DS) Wizard, an open-source online tool part of ELIXIR and a recommended FAIR Supporting Resource (FSR) for the creation of DMPs in compliance with FAIR principles.

The IBISBA-IT DMP is a proposal and is intended to be a living document, evolving to become more detailed, comprehensive, and fully aligned with FAIR principles.

Given the complexity and magnitude of the task, full digitalisation will be a gradual process. To date, a certain level of automation has been implemented to optimise and streamline various laboratory operations. This digitalised workflow reduces the manual workload for researchers while improving and making data management more efficient. At the core of this digitalisation workflow is a set of interoperable tools, each playing a critical role in the data lifecycle, from collection and processing to archiving and sharing. The following paragraphs will illustrate the main steps of the workflow digitalisation. A detailed data lifecycle description is provided by the DMP.

4.2 Digitalisation of Bioprocess Protocols and Data Management

The digitalisation process revolves around laboratory activities. In the specific case of this deliverable and in relation to the research field in which IBISBA-IT operates, these activities are represented by bioprocesses. One of the aims was to produce digital Standard Operating Procedures (SOPs) for an entire bioprocess pipeline. These protocols were generated following the ISA metadata schema[8] and uploaded to the IBISBAKHub. As they are part of the same pipeline, the SOPs were stored as a unified set with the same SEEK-ID persistent identifier (https://ibisbahub.eu/sops/142?version=2). These versions may be updated after this document is published. However, the versioning feature provided by the IBISBAKHub will allow users to browse any newer versions that may become available. Ultimately, the protocols will be assigned a DOI through DataCite, according to the procedures established by the ITINERIS Consortium. In addition to the PDF-saved SOPs, machine-readable digital versions in JSON format were generated. Metadata descriptions describe their interdependencies in the pipeline, i.e., the execution order for correct (re)use. An example of an SOP is in the annex attached to this document.

Besides digital SOPs, another critical component of a digitalisation system is an Electronic Lab Notebook (ELN). Among the various ELNs evaluated, ElabFTW (https://www.elabftw.net) has been identified as a promising application due to its structured approach to protocol management, metadata documentation, and version control. ElabFTW is the most popular open-source electronic lab notebook for research labs. Created at the Curie Institute, it is used by Universities worldwide to create personal and shared notebooks. ElabFTW ensures data provenance, providing a clear and transparent record of how data was generated and processed across its lifecycle. The ELN acts as a central point where experiment metadata is recorded and linked to instrument-generated datasets and every dataset can be traced back to its origin and experimental setup[9].

The process begins with researchers logging experimental parameters directly into ElabFTW through custom forms. These entries register details such as sample information, instrument names, serial numbers, and experimental conditions, documenting every stage of the research. In parallel, instrument-generated raw data and metadata are saved to a dedicated local folder. At this point, an *ad hoc* Python script, monitoring the folder automatically extracts selected metadata (e.g. instrument settings and sample identifiers) and parses them into JSON and XML format files. The selected metadata is then ready to be manually imported into ElabFTW, where it is merged with the pre-existing experimental records. The script is developed to check for new files at regular intervals. By combining manual input from researchers with machine-readable metadata, ElabFTW provides a view of each experiment and its related datasets. A demo of the script designed for the Fast Protein Liquid Chromatography (FPLC) system (Cytiva ÄKTA Pure, controlled via Unicorn software) can be found on the IBISBAKHub following this link: https://ibisbahub.eu/data_files/125?version=1.

Following data collection, the next priority is long-term storage and access management, which is managed in two ways: using the Nexus for all produced data intended for internal use, and the IBISBAKHub for publishing scientifically relevant data.

4.3 Data preservation and management using the NAS

The QNAP NAS servers have been deployed to provide high-capacity, fault-tolerant storage, ensuring that all research data is archived. As previously described, these servers operate using QuTS hero OS, which incorporates deduplication, compression, and snapshot technologies to maintain data integrity while optimising storage efficiency. Currently, manual intervention is required using USB storage devices to transfer raw experimental data (and metadata) to its destination storage location on the Nexus. However, automating this step through direct API integration between the instrument and the storage system is desirable, eliminating the need for manual file transfers.

4.4 Data preservation and management using the IBISBAKHub and ITINERIS Hub

Selected datasets on the Nexus intended for sharing are transferred to the IBISBAKHub, where they are assigned a SEEK-ID persistent identifier.

The ElabFTW record ID is linked to the corresponding dataset stored in IBISBAKHub through its SEEK-ID, ensuring its long-term preservation. If the dataset is stored only on the Nexus, the ElabFTW record is associated with the dataset folder by linking them through their shared name, corresponding to the ElabFTW ID. This connection ensures data provenance because every dataset can be traced back to the experimental conditions under which it was produced. When datasets are ready for public dissemination, they are transferred to the ITINERIS Hub and assigned a DOI, making them public and citable.

Figure 4 illustrates how the digitalised workflow underpinning the IBISBA-IT e-infrastructure aligns with the data lifecycle.

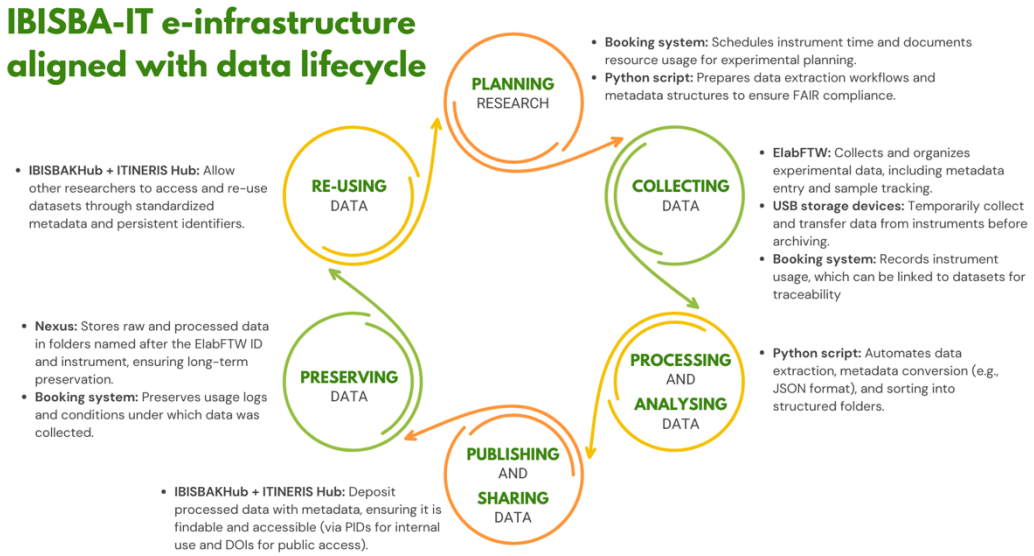

Figure 4. IBISBA-IT e-infrastructure aligned with data lifecycle

Looking forward, automation via the ElabFTW API will further streamline the process by allowing metadata extraction and import to occur automatically, strengthening the link between ElabFTW and IBISBAKHub, and enabling real-time metadata transfer directly from laboratory instruments. Through these processes, ElabFTW ensures that data and metadata remain accurate, traceable, and aligned with FAIR principles, supporting both internal workflows and external collaboration.

5. FUTURE ENHANCEMENTS AND OPEN CHALLENGES

The current system provides a foundation for digitalisation, which will require further enhancements, including:

- Ontology integration to enhance metadata standardisation and interoperability.
- Implementing automated data validation checks to improve metadata accuracy.
- Strengthening cross-system integration between ElabFTW, Nexus, the IBISBAKHub, and the ITINERIS Hub for a seamless research data ecosystem.

By addressing these challenges, IBISBA-IT aims to set a new standard for digital research infrastructure, ensuring that its workflows remain efficient, scalable, and fully aligned with IBISBA's European best practices.

REFERENCES

[1] Admin, IBISBA, & Goble, C. (2022). D8.2 Preliminary architecture for the IBISBA digital infrastructure. IBISBA. https://doi.org/10.34701/IBISBA.1.DOCUMENT.39.1

[2] FAIRsharing.org: FAIRDOMHub; FAIRDOMHub. https://doi.org/10.25504/FAIRsharing.nnvcr9

[3] Schultes, Erik. 'Data Stewardship Plan Templates Designed to Support the FAIR Principles'. 1 Jan. 2023 : 1 – 3. https://doi.org/10.3233/FC-221508

[4] Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., et al. (2016). *The FAIR Guiding Principles for scientific data management and stewardship*. *Scientific Data*, 3, 160018. https://doi.org/10.1038/sdata.2016.18

[5] Guellec, D. and C. Paunov (2018), "Innovation policies in the digital age", *OECD Science, Technology and Industry Policy Papers*, No. 59, OECD Publishing, Paris, https://doi.org/10.1787/eadd1094-en

[6] O'Donohue, M. (2020). *Data mamangement plan – IBISBA 1.0*. https://doi.org/10.34701/IBISBA.1.DOCUMENT.8.1

[7] O'Donohue M. (2020). PREP-IBISBA Data Management Plan. PREP-IBISBA. European Commision, CORDIS: https://cordis.europa.eu/project/id/871118/results

[8] Sansone, S.-A., Rocca-Serra, P., Gonzalez-Beltran, A., Johnson, D., & ISA Community. (2016). ISA Model and Serialization Specifications 1.0. Zenodo. https://doi.org/10.5281/zenodo.163640

[9] Carpi, Nicolas: eLabFTW – the open source lab notebook. FOSDEM 2021, FOSDEM VZW, 2021. https://doi.org/10.5446/53327