Grant Agreement No.: 101080718
Call: HORIZON-HLTH-2022-STAYHLTH-01-two-stage
Topic: HORIZON-HLTH-2022-STAYHLTH-01-05-two-stage
Type of action: HORIZON-RIA

BIO-STREAMS

D5.3 Bio-Streams Platform Beta

BIO-STREAMS Platform – Beta Version Demonstrator

Revision: v.1.0

Work package	WP5
Task	Task 5.6
Due date	30/04/2026
Submission date	30/04/2026
Deliverable lead	Netcompany – Intrasoft
Version	1.0
Authors	Marios Logothetis, Anastasios Gogos, Petros-Sozon Dimitrakopoulos (INTRA), Dimitrios Eleftheriou (STS), Thelma Androutsou (CSCY)
Reviewers	Svetlin Hansov & Blagoy Palev (BLOCKS) | Leipuviene Kristina (SIA)

Abstract	The BIO-STREAMS Platform Beta presents the first integrated demonstrator of the platform, validating its distributed architecture and deployment approach. It enables secure access to clinical data across multiple healthcare institutions through central services and hospital-side components. Key elements include the Node Gateway for connectivity and the Information Management System (IMS) for data querying and aggregation. Deployed using containerized services and automated CI/CD pipelines, the platform demonstrates initial integration

	and operational readiness in a development environment, establishing the foundation for further development and final validation. The Dashboard serves as the primary user-facing interface of the platform, enabling stakeholders to access data, manage studies, and monitor system behaviour across integrated components.
Keywords	Bio-streams platform, distributed data platform, biobank, healthcare data, integration, clinical data access, hospital pilot, containerized architecture, digital health application

DOCUMENT REVISION HISTORY

Version	Date	Description of change	List of contributor(s)
1.0	30/04/2026	Final preparation / curation before submission	M. Logothetis, A. Gogos, P. Dimitrakopoulos
0.9	29/04/2026	Addressing review comments, Consolidation of final version	M. Logothetis, A. Gogos, P. Dimitrakopoulos
0.8	28/04/2026	Review from BLOCKS	Svetlin Hansov, Blagoy Palev
0.7	26/04/2026	Review from SIA	Leipuviene Kristina
0.6	24/04/2026	Review + contributions from CSCY	Thelma Androutsou
0.5	20/04/2026	Review from STS, curation	Dimitrios Eleftheriou, A. Gogos
0.4	17/04/2026	References, Abbreviations, curation	Anastasios Gogos
0.3	16/04/2026	Curation, new content, refinements	Anastasios Gogos
0.2	07/04/2026	1st draft content added	Anastasios Gogos, Petros-Sozon Dimitrakopoulos
0.1	02/04/2026	1st version of the document	Marios Logothetis, Anastasios Gogos

Disclaimer

Co-funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the other granting authorities. Neither the European Union nor the granting authority can be held responsible for them.

Copyright notice

Project co-funded by the European Commission in the Horizon Europe Programme		
Nature of the deliverable:	DEM	
Dissemination Level		
PU	*Public, fully open, e.g. web*	**X**
SEN	*Sensitive, limited under the conditions of the Grant Agreement*	
Classified R-UE/ EU-R	*EU RESTRICTED under the Commission Decision No2015/ 444*	
Classified C-UE/ EU-C	*EU CONFIDENTIAL under the Commission Decision No2015/ 444*	
Classified S-UE/ EU-S	*EU SECRET under the Commission Decision No2015/ 444*	

* *R: Document, report (excluding the periodic and final reports)*
DEM: Demonstrator, pilot, prototype, plan designs
DEC: Websites, patents filing, press & media actions, videos, etc.
DATA: Data sets, microdata, etc
DMP: Data management plan
ETHICS: Deliverables related to ethics issues.
SECURITY: Deliverables related to security issues
OTHER: Software, technical diagram, algorithms, models, etc.

Executive summary

As part of Work Package 5, the BIO-STREAMS Platform Beta represents the first integrated demonstrator of the BIO-STREAMS platform, providing a functional prototype that validates the overall system architecture, deployment approach, and initial component integration.

The platform enables secure and controlled access to distributed clinical data across multiple European healthcare institutions. Its architecture combines central platform services with hospital-side deployments, allowing data to remain within institutional boundaries while being accessed through a unified querying mechanism.

The current beta version includes the deployment and integration of core infrastructure services, such as reverse proxy and authentication mechanisms, as well as key application components including the Dashboard, Active Health application, Node Gateway, and the Information Management System (IMS). Through this setup, the platform supports end-to-end data flow, where data requests are processed centrally and executed across distributed hospital environments, with backend querying capabilities validated and user-facing interaction mechanisms under ongoing development. In parallel, the platform also exposes user-facing services and engagement tools, including the Serious Games suite, supporting the execution of pilot activities across different stakeholder groups.

A key aspect of the platform is the use of the Node Gateway component to establish secure connectivity with hospital Node Bundles, overcoming typical network restrictions such as firewalls and NAT configurations. On top of this connectivity layer, the IMS provides data aggregation and query-processing capabilities, enabling unified access to clinical data across multiple sites.

The platform is deployed using a container-based architecture supported by automated CI/CD pipelines, ensuring reproducibility and consistent deployment across environments. Initial validation confirms the successful integration of core components and the operational readiness of the platform in a development setting.

This beta release establishes a solid foundation for the BIO-STREAMS platform, demonstrating the feasibility of its distributed approach and preparing the ground for further integration, extended functionality, and full system validation in the final release.

Table of contents

List of figures

Co-funded by the European Union

List of tables

Abbreviations

ACT	Active Health Application
ANA	Analytics Service (Matomo)
API	Application Programming Interface
AUTH	Authentication Service (Keycloak)
BNB	BIO-STREAMS Node Bundle
CI/CD	Continuous Integration / Continuous Deployment
DASH	Dashboard
DevOps	Development & Operations
IMS	Information Management System
IP	Internet Protocol
LOGS	Distributed Log Service
METRICS	Metrics Monitoring Service
NAT	Network Address Translation
NGW	Node Gateway
SERG	Serious Games Suite
SSH	Secure Shell
TCP	Transmission Control Protocol
UI	User Interface
URL	Uniform Resource Locator

Co-funded by the European Union

1 Introduction

The present deliverable, D5.3 - BIO-STREAMS Platform Beta, provides a concise overview of the current state of the BIO-STREAMS platform. As a DEM (Demonstrator) type deliverable, its objective is to document the availability of a functional platform prototype, highlighting its main components, deployment approach, and integration status.

The BIO-STREAMS platform aims to enable secure and controlled access to distributed clinical and synthetic data across multiple European healthcare institutions. At this stage, a beta version of the platform has been established, integrating core infrastructure services, application-layer components, and initial data-access mechanisms.

This document focuses on presenting the current operational setup of the platform, including its deployment workflow and the interaction between its main components. Detailed technical descriptions, extended testing results, and full system validation will be provided in the final version of the platform deliverables.

Co-funded by the European Union

2 Platform Overview

The BIO-STREAMS platform follows a distributed architecture designed to support secure data access and processing across multiple stakeholders, including healthcare institutions, researchers, and application services. The overall system integrates central platform services with distributed hospital-side components, enabling a unified data access layer while respecting institutional boundaries.

The high-level architecture of the platform is illustrated in Figure 1, where the main system components and their interactions are presented.

The platform consists of the following key components:

- **Dashboard and Application Services:**

  User-facing components such as the Dashboard and the Active Health application provide access to platform functionalities.

- **Node Bundles (Hospital-side components):**

  Deployed within hospital environments, Node Bundles are responsible for storing and managing clinical data locally. These components operate within protected institutional networks and are not directly exposed to external systems.

- **Node Gateway:**

  The Node Gateway acts as the communication bridge between the central platform and the distributed Node Bundles. It enables controlled and secure connectivity to hospital environments, overcoming network restrictions such as firewalls and NAT configurations.

- **Information Management System (IMS):**

  The IMS is a central microservice responsible for processing and aggregating data queries. It distributes requests to the appropriate hospital nodes via the Node Gateway and returns unified results to the requester. This component plays a key role in enabling cross-site data access within the platform.

- **AI Tools and Supporting Services:**

  The platform includes additional components such as synthetic data generators and analytical tools, supporting advanced data processing and decision-making capabilities.

- **Core Infrastructure Services:**

  Supporting services include authentication with Keycloak [1], reverse-proxying and routing with Traefik [2], observability with Prometheus [3] and Grafana [4], centralized logging with Seq [5] and analytics with Matomo [6]. These components ensure secure access, monitoring, and operational reliability.

Overall, the platform implements a modular and containerized architecture, where individual components are deployed as independent services and integrated through well-defined interfaces.

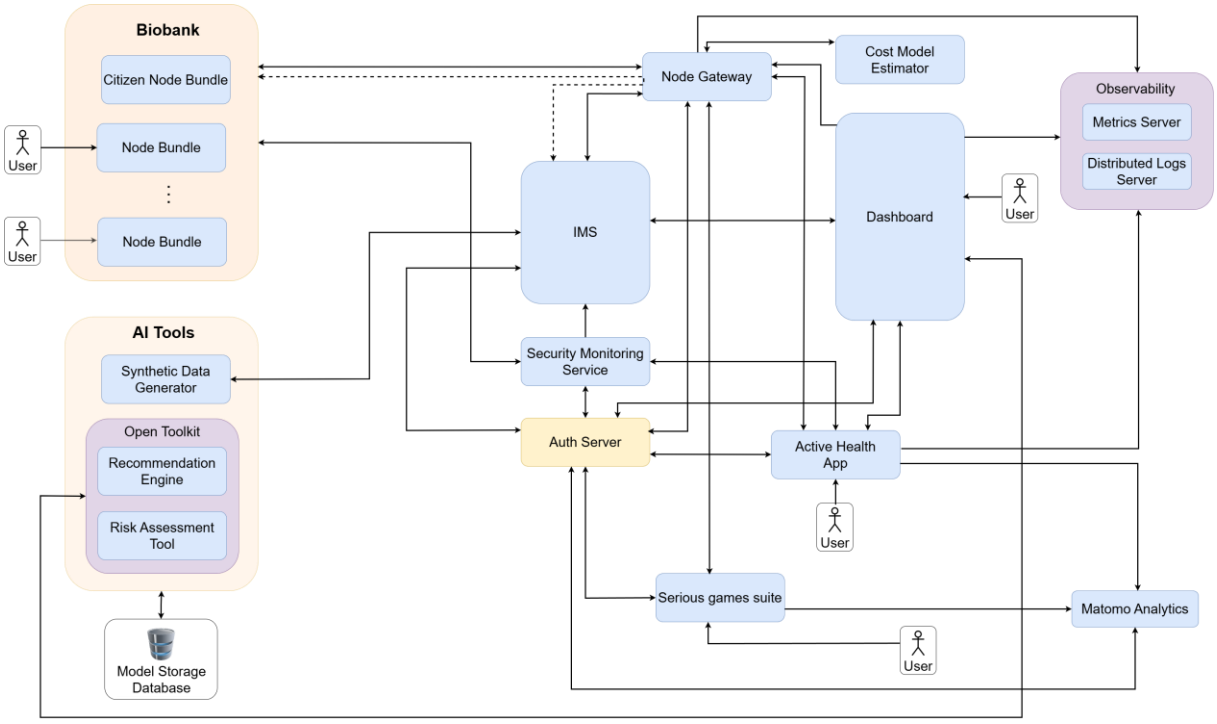

Figure 1: BIO-STREAMS architecture diagram from D5.1

3 Demonstrator Scope

The BIO-STREAMS Platform Beta represents an initial integrated version of the platform, focusing on the availability and interaction of its core components. As a demonstrator, this version aims to validate the overall architecture, deployment approach, and basic data flow across the system.

At the current stage, the following elements are included in the platform beta. Most core components have been implemented and integrated, while selected functionalities remain under development and planned for subsequent releases.

- **Core infrastructure services**, including reverse proxy, secure access, and authentication mechanisms, enabling controlled exposure of platform services.
- **Central platform services**, such as the Dashboard, Active Health application, Node Gateway, and supporting backend services, deployed and accessible through the platform environment.
- **Querying capabilities**, enabled through the integration of the Information Management System (IMS) with multiple hospital Node Bundles.
- **Distributed connectivity with hospital environments**, achieved through the Node Gateway component, allowing communication with Node Bundles deployed within protected institutional networks.
- **Monitoring and observability tools**, providing basic metrics collection, logging, and system visibility.

Within this context, the Dashboard acts as the central user-facing entry point of the platform, providing stakeholders with access to data, study management functionalities, and system monitoring capabilities across the integrated services. The current beta version demonstrates multiple end-to-end data flows across the platform components. In the case of biobank data requests, these are processed through the IMS and routed to distributed hospital nodes. Results are then aggregated and returned, validating the underlying backend functionality. The data access and integration mechanisms have been successfully established, forming a solid foundation for further extension, while user-facing querying functionalities will be enhanced in subsequent platform iterations.

At the same time, the platform remains under active development. The following aspects are either partially implemented or planned for future releases:

- **Extended integration of all project components**, as some components are still in the process of being onboarded or fully integrated.
- **Advanced data processing capabilities**, including full support for synthetic data generation and more complex query handling mechanisms.
- **Enhanced performance and scalability features**, particularly for long-running or high-volume queries.
- **Comprehensive testing and validation coverage**, including large-scale and production-level scenarios.

Overall, the BIO-STREAMS Platform Beta serves as a functional prototype, demonstrating the feasibility of the platform architecture and its ability to support distributed data access across multiple stakeholders, while setting the foundation for the final system release.

In addition to data access and processing capabilities, the BIO-STREAMS Platform Beta supports user-facing services that are essential for the execution of the project's pilot activities. These include the Dashboard, which serves as the main interface for healthcare professionals and researchers, the Active Health application targeting end-users, and the Serious Games

suite accessible through the platform marketplace. Together, these components demonstrate the platform's ability to support multiple stakeholder interactions, extending beyond data retrieval to real-world usage scenarios in clinical and pilot environments.

4 Deployment & Demonstration Setup

The deployment of the BIO-STREAMS Platform Beta follows a structured and automated approach based on containerized services and CI/CD [7] pipelines. The platform is deployed across dedicated environments (e.g., development and production), ensuring flexibility and reproducibility of the setup.

The deployment process is organized in a sequence of stages, reflecting dependencies between infrastructure and application components.

4.1 Deployment Strategy

The platform relies on automated pipelines to deploy and manage its services with Docker [8] containers. Each major group of components is deployed through dedicated pipelines, ensuring consistency and controlled rollout of the platform.

Container images are retrieved from both public repositories (e.g., Docker Hub [9]) and the project's private container registry (a Harbor [10] instance) and are deployed as part of multi-service compositions with Docker Compose [11].

The deployment process is executed through a set of automated pipelines corresponding to infrastructure, platform services, and integration components.

4.2 Core Infrastructure Deployment

The initial step in the deployment process involves the setup of core infrastructure services, which provide the foundation for the rest of the platform.

This includes:

- **Reverse Proxy and Routing (Traefik):**
  A reverse proxy service is deployed to manage incoming traffic, route requests to the appropriate services, and handle HTTPS termination. Service exposure is centrally controlled through this component.
- **Authentication and Access Control (Keycloak):**
  A centralized authentication service is deployed to manage user identities and secure access to platform components. This service is integrated with user-facing applications and backend services.

These components also establish the necessary container networking configuration, enabling communication between services and defining the platform's exposure model. An overview of the infrastructure deployment pipelines is presented in Figure 2.

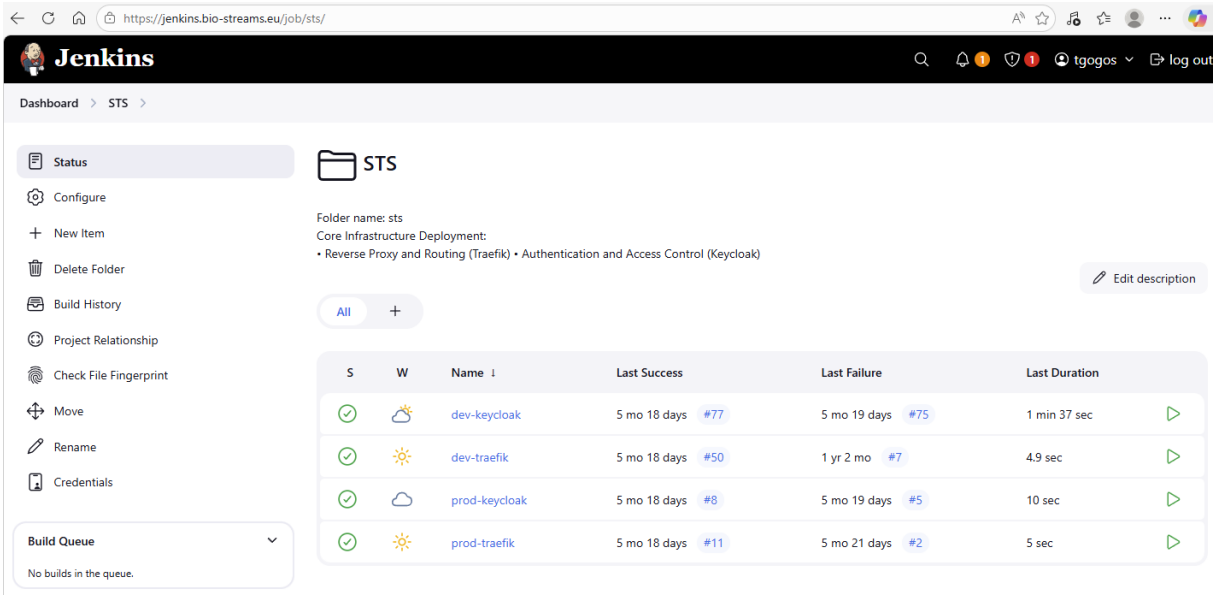

Figure 2: Jenkins - view of the core infrastructure services deployment-pipelines

4.3 Platform Services Deployment

Following the infrastructure setup, the main platform services are deployed. These include both application-layer components and supporting services.

Key deployed services include:

- Dashboard and Active Health applications, providing user-facing interfaces.
- Node Gateway, enabling connectivity with distributed hospital Node Bundles.
- Database services, supporting platform data storage.
- Analytics and monitoring tools, such as metrics collection and logging services.
- Additional supporting services, including connectivity and system management components.

All services are deployed as containerized applications and integrated through the shared networking layer, with external access managed through the reverse proxy. The deployment pipelines for the main platform services are illustrated in Figure 3.

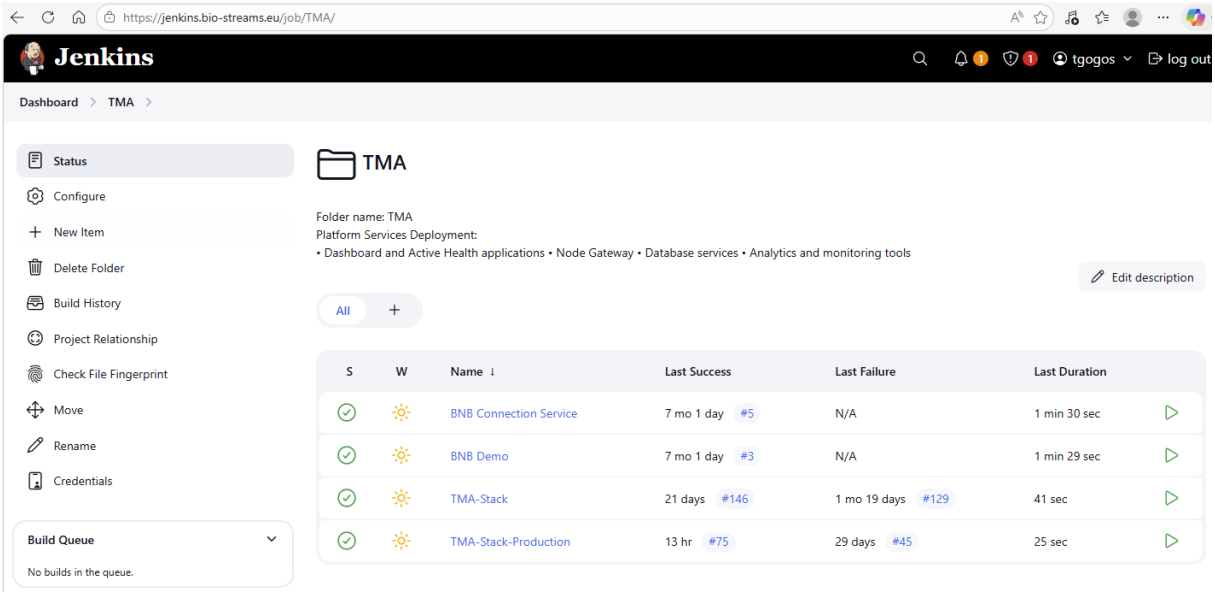

Figure 3: Jenkins - view of the main platform services deployment-pipelines

4.4 IMS Deployment and Data Integration

The Information Management System (IMS) is deployed as a containerized microservice within the platform environment and plays a central role in enabling data access across distributed hospital nodes.

The IMS integrates with:

- **Node Gateway**, which provides the communication channel to hospital Node Bundles.
- **User-facing applications**, such as the Dashboard, which is expected to integrate querying functionalities and interact with the IMS. At the current stage, the IMS has been validated through backend-level integration, with user-facing query interaction planned for subsequent platform iterations.

Due to the constraints of hospital environments (e.g., NAT and firewall restrictions), direct access to Node Bundles is not possible. The Node Gateway establishes and maintains the required connectivity, enabling the IMS to route queries and retrieve data from distributed nodes.

Through this setup, the platform supports a distributed data access model, where queries are processed centrally but executed across multiple hospital environments. The IMS has been successfully integrated with multiple hospital Node Bundles through the Node Gateway, enabling validated end-to-end data flow across distributed environments.

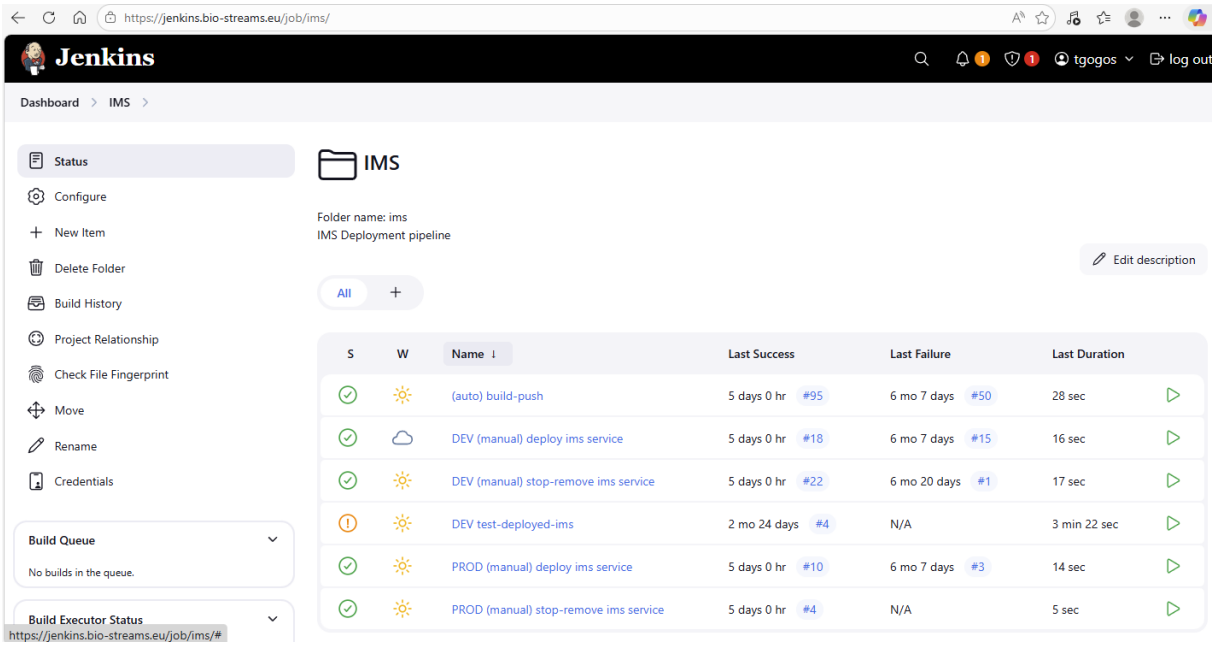

Figure 4: Jenkins - view of the IMS deployment-pipelines

4.5 Environments and Configuration

The current CI/CD setup supports 2 deployment environments, one for development and one production, with environment-specific configurations. The 3rd one appearing in Figure 5 and Figure 6 is the environment that hosts the actual CI/CD services and is accessible only by administrators.

Service configuration is managed through environment variables and deployment pipelines, allowing services to adapt to different environments without requiring modifications to the underlying container images.

This approach ensures:

- consistent deployment across environments
- flexibility in configuration management
- separation between development and production setups

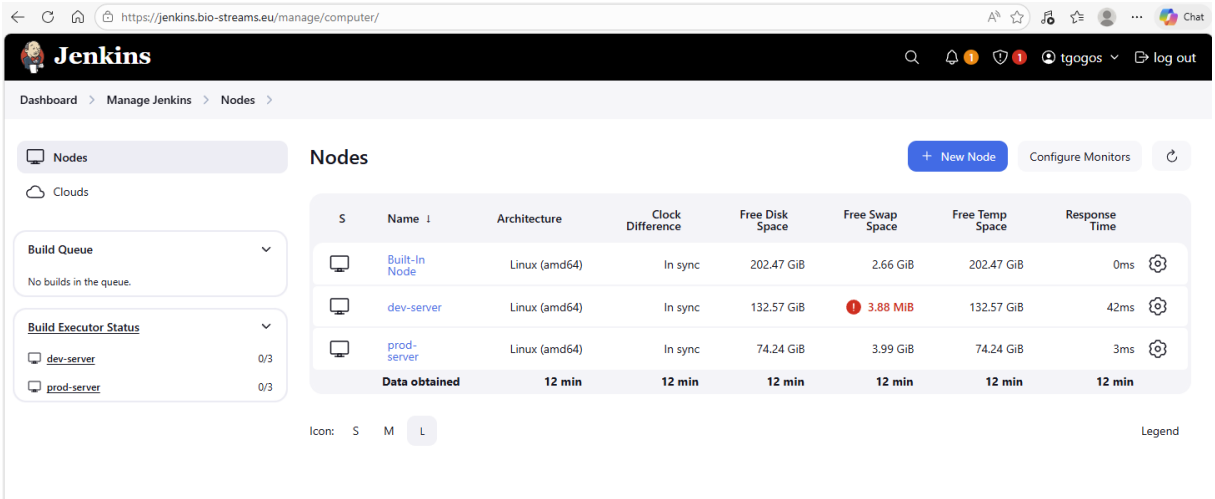

Figure 5: Jenkins view - environments (Nodes) used

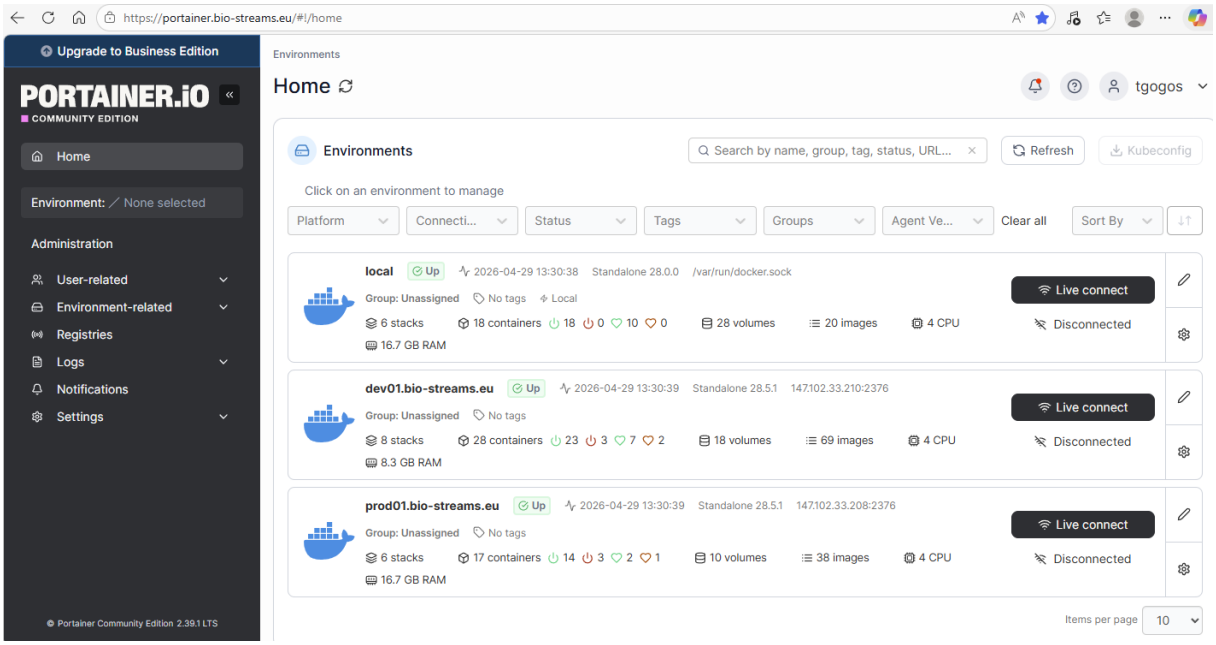

Figure 6: Portainer view - environments used

5 Source Code & Platform Access

The BIO-STREAMS Platform Beta is supported by a set of repositories and deployed services that enable both development and demonstration of the platform components.

5.1 Source Code Repositories

The source code of the main platform components is maintained under the BIO-STREAMS GitHub [12] organization. The repositories listed in this section include both source code and deployment-related artefacts of the BIO-STREAMS platform. Certain components provide full implementation source code, while other repositories focus on deployment, configuration, and orchestration aspects (e.g., container definitions, environment variables, configuration details, CI/CD pipelines). Together, these repositories enable the integration, deployment, and operation of the platform across environments.

Component / Scope	Repository	Description
Core infrastructure services	https://github.com/bio-streams-eu-project/sphynx-services	Contains configuration and deployment artifacts related to core infrastructure components, including reverse proxy and authentication services
Platform services	https://github.com/bio-streams-eu-project/tma-services	Includes the main platform services such as Dashboard, Active Health application, Node Gateway, and supporting backend components
Information Management System services	https://github.com/bio-streams-eu-project/ims	Contains the implementation of the IMS component, responsible for distributed query processing and data aggregation across hospital nodes

Table 1 – Main platform component repositories

These repositories follow standard version control practices and support collaborative development across the consortium partners.

5.2 Platform Access Points

The BIO-STREAMS Platform Beta is accessible through a set of web-based entry points corresponding to its main services and components. These access points can be grouped into user-facing applications and supporting platform services.

User-facing services:

- Dashboard: https://dashboard.bio-streams.eu/
- Active Health application: https://activehealth.bio-streams.eu/
- Serious Games (Marketplace): https://marketplace.bio-streams.eu/

The marketplace provides access to multiple serious games developed within the project, targeting user engagement and behavioral interventions.

Core platform services:

- Authentication (Keycloak): https://auth.bio-streams.eu/

Operational and monitoring services:

- Analytics (Matomo): https://analytics.bio-streams.eu/
- Metrics monitoring (Grafana): https://metrics.bio-streams.eu/
- Logging service: https://logs.bio-streams.eu/

Supporting services (optional / under evaluation):

- Localization platform (Weblate [13]): https://localization.bio-streams.eu/

Some internal components, such as the Node Gateway, are not publicly exposed, as they are used for secure communication with hospital-side deployments.

5.3 CI/CD & Infrastructure Services

The BIO-STREAMS Platform Beta is supported by a set of infrastructure and DevOps services that enable automated deployment, secure access, and operational monitoring of the platform.

The platform follows a container-based deployment approach, where individual components are packaged as container images and deployed through CI/CD pipelines. As illustrated in Figures 2, 3 and 4, the deployment and operation of the platform services are performed through a controlled and automated process with Jenkins. Pipelines are used to orchestrate the build and deployment of the application components, providing clear visibility into execution status, success/failure states and deployment history across environments. In parallel, Portainer offers an operational view of the deployed services (see Figure 7), allowing monitoring of running containers, environments, and system status. Together, these tools support a transparent and manageable deployment lifecycle, ensuring consistency across environments and facilitating maintenance and troubleshooting activities.

This CI/CD workflow supports the full lifecycle of platform services, including build processes, container image management, automated deployment, and validation steps. This approach reduces manual intervention and enables the reliable and continuous integration of platform components.

In addition to deployment automation, the platform integrates operational services that support its day-to-day functioning. These include authentication and access control mechanisms, analytics services for user interaction monitoring, as well as observability tools for metrics collection and log aggregation. Together, these services provide the necessary foundation for maintaining platform reliability, monitoring performance, and supporting debugging and maintenance activities.

This combination of CI/CD practices and operational tooling ensures that the BIO-STREAMS platform can evolve continuously while maintaining stability across its distributed components.

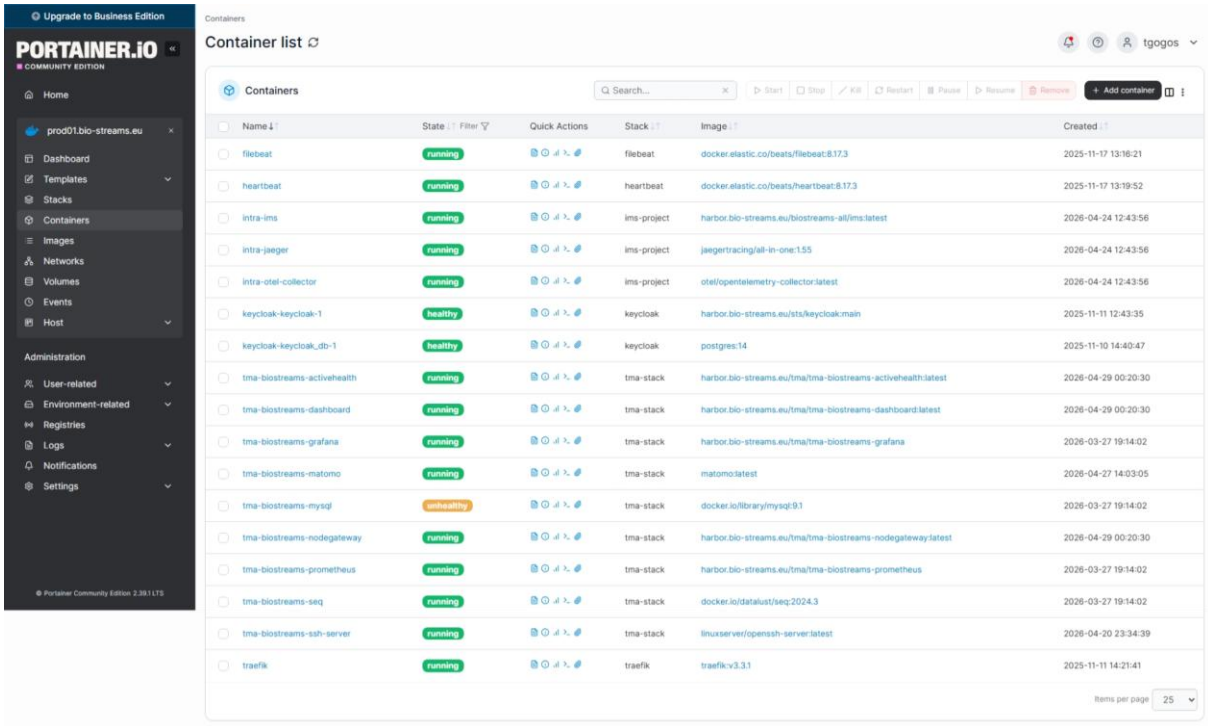

Figure 7: Portainer snapshot of platform services running on the Production environment

6 Conclusions & Next Steps

The BIO-STREAMS Platform Beta demonstrates the successful deployment and initial integration of the platform's core components, providing a functional prototype that validates the overall system architecture and deployment approach.

At this stage, the platform supports:

- deployment of core infrastructure services
- integration of central platform components
- connectivity with distributed hospital Node Bundles
- initial data querying and aggregation through the IMS component

The current implementation confirms the feasibility of the platform's distributed architecture, particularly in enabling secure communication with hospital environments through the Node Gateway and supporting centralized query processing via the IMS.

While the beta version establishes a solid foundation, further work is required to reach the final platform release. Future efforts will focus on:

- completing the integration of all project components
- introduction of user-facing interfaces for advanced querying and data exploration through the IMS
- enhancing data processing capabilities, including full support for synthetic data
- improving performance and scalability
- extending testing and validation activities under more realistic operational conditions
- refining user-facing functionalities and overall system usability

The final version of the platform will build upon the current beta implementation, providing a more complete, robust, and fully validated system aligned with the project objectives.

7 References

[1] “Keycloak: Open Source Identity and Access Management | Add authentication to applications and secure services with minimum effort. Keycloak provides user federation, strong authentication, user management, fine-grained authorization, and more.,” [Online]. Available: https://www.keycloak.org/.

[2] “Traefik is the leading open-source reverse proxy and load balancer for HTTP and TCP-based applications that is easy, dynamic and full-featured.,” [Online]. Available: https://traefik.io/traefik .

[3] “Prometheus: From metrics to insight. Power your metrics and alerting with the leading open-source monitoring solution.,” [Online]. Available: https://prometheus.io/.

[4] “Grafana: Query, visualize, alert on, and understand your data no matter where it’s stored. With Grafana you can create, explore, and share all of your data through beautiful, flexible dashboards.,” [Online]. Available: https://grafana.com/.

[5] “Seq - The self-hosted search, analysis, and alerting server,” [Online]. Available: https://datalust.co/seq.

[6] “Matomo: Google Analytics alternative that protects your data and your customers' privacy,” [Online]. Available: https://matomo.org/.

[7] “CI/CD Documentation - Basecamp folder that provides the official documentation for the BIO-STREAMS CI/CD environment. (Access provided upon request),” [Online]. Available: https://3.basecamp.com/5618330/buckets/32813804/uploads/8019226872.

[8] “Docker: Accelerated Container Application Development | Docker is a platform designed to help developers build, share, and run container applications. We handle the tedious setup, so you can focus on the code.,” [Online]. Available: https://www.docker.com/.

[9] “Docker Hub simplifies development with a powerful container registry for storing, managing, and sharing Docker images. By integrating seamlessly with your tools, it enhances productivity and ensures reliable deployment, distribution, and acces,” [Online]. Available: https://hub.docker.com/.

[10] “Harbor - an open source registry that secures artifacts with policies and role-based access control, ensures images are scanned and free from vulnerabilities.,” [Online]. Available: https://goharbor.io/.

[11] “Docker Compose is a tool for defining and running multi-container applications. It is the key to unlocking a streamlined and efficient development and deployment experience.,” [Online]. Available: https://docs.docker.com/compose/.

[12] “GitHub,” [Online]. Available: https://github.com/.

[13] “WEBLATE: Web-based continuous localization - Copylefted libre software, used by over 2,500 libre software projects and companies in over 165 countries.,” [Online]. Available: https://weblate.org/el/.