



**Implementation of the new,  
extended LIFEWATCH VRE with the  
new functionalities for an easy access  
and use**



<b>Deliverable number:</b>	D6.9
<b>Work package:</b>	WP6
<b>Intermediate Objective:</b>	
<b>Deliverable type:</b>	<input checked="" type="checkbox"/> Document, report
	<input type="checkbox"/> Websites, patent filings, videos, etc.
	<input type="checkbox"/> Other: please specify .....
<b>Dissemination level:</b>	<input checked="" type="checkbox"/> Public
	<input type="checkbox"/> Restricted
<b>Estimated delivery (bimester):</b>	B13
<b>Actual delivery date:</b>	30/04/2025
<b>Author(s) (Partner-OU):</b>	Giacinto Donvito (INFN-Ba)
<b>Reviewed by:</b>	
<b>Note:</b>	

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

### Sommario

1. INTRODUCTION.....	4
2. WP6.14 INITIAL OBJECTIVES.....	4
3. ACTIVITY STATUS.....	4
4. ACHIEVEMENTS TO DATE.....	6
<b>PRODUCTION ROLLOUT FOR LIFEWATCH.....</b>	<b>6</b>
<b>SECURITY AND COMPLIANCE.....</b>	<b>6</b>
<b>PERFORMANCE AND SCALABILITY.....</b>	<b>7</b>
5. FUTURE PLANS.....	7
<b>AI &amp; DATA ANALYTICS.....</b>	<b>7</b>
<b>ENHANCED INTEROPERABILITY.....</b>	<b>7</b>
<b>COMMUNITY EXPANSION &amp; SUSTAINABILITY.....</b>	<b>7</b>
6. CONCLUSION.....	8
USE CASE: NAAVRE DEPLOYMENT.....	9

## 1. INTRODUCTION

This deliverable presents the completed rollout of the new LifeWatch VRE instance. Building on the groundwork laid by previous initiatives, the updated VRE enhances usability and extends functionality to meet the specific use cases defined by the ITINERIS project. By integrating containerized Jupyter Notebook and RStudio services into a unified platform, we have created a seamless environment for data discovery, analysis, and workflow orchestration. The upgraded VRE not only supports today's requirements—such as optimized access to distributed environmental datasets and interactive AI-driven analytics—but also provides a scalable, modular foundation that can evolve to accommodate future tools, services, and research scenarios championed by the LifeWatch community.

## 2. WP6.14 INITIAL OBJECTIVES

The WP6.14 activity of the ITINERIS project has two specific aims:

1. **Implement and extend a Virtual Research Environment (VRE)** for the LifeWatch community, in line with European requirements and project goals. This VRE will provide seamless, optimized access and management of heterogeneous data, addressing data locality, data diversity, and metadata handling. Work will be carried out by INFN personnel and external experts to leverage diverse expertise for a robust, flexible, and high-level VRE platform.
2. **Upgrade and expand the existing hardware infrastructure** built under the LifeWatchPLUS PON project. This upgrade will support new project requirements and accommodate additional environmental data being collected by the Italian and European LifeWatch networks.

To achieve the above objectives, it was necessary to acquire new hardware capable of delivering advanced cloud computing and storage services in support of the LifeWatch community. This investment also enabled full-stack testing of different software solutions that can be offered as Virtual Research Environments (VREs) for LifeWatch users. In particular, the following platforms have been deployed and validated: **Galaxy, Jupyter Notebook, RStudio, and NaaVRE.**

This document therefore describes the activities carried out for the acquisition, installation, and configuration of the hardware, as well as the tests and validation activities performed to evaluate the aforementioned services deployed on the cloud platform, enhanced with the resources acquired within the ITINERIS project. These activities – both those specifically supporting the hardware infrastructure and those related to the software tools made available to the LifeWatch community – have been conducted by the technologist staff recruited under the project.

## 3. ACTIVITY STATUS

In order to reach this objective, we acquired the needed hardware that is intended to improve the infrastructure that is already hosted at INFN-Bari data center and already connected to Lifewatch Italy services.

In particular, we have acquired three different categories of hardware:

1. Fat CPU Nodes

High-density compute units optimized for general-purpose, compute-intensive workloads. These nodes commonly offer configurations such as approximately 192 CPU cores per node. Their role is to provide scalable and flexible CPU cycles in a distributed cloud-HPC architecture.

## 2. CPU + GPU Nodes

Specialized hybrid nodes designed for accelerated computing and AI workloads. Each GPU node combines CPUs with 4 × NVIDIA H100 GPUs, offering powerful AI/machine-learning capability at the edge of the distributed infrastructure. These nodes are also enabling researchers to run demanding workloads using GPU acceleration near data sources.

## 3. Storage Nodes

Dedicated storage infrastructure that supports both mass storage and high-performance operations. Providing the persistence and throughput needed for data-intensive applications and serving as the backbone for the distributed HPC-Cloud environment built under TeRABIT.

Node Type	Configuration & Purpose
Fat CPU Nodes	~192 CPU cores; general compute services across multiple national sites
GPU Nodes	CPU + 4 × NVIDIA H100 GPUs; tailored for AI/ML and accelerated computing needs
Storage Nodes	Mass & high-performance storage ~360TB per host

At the end of the project, all core resources for the LifeWatch community are now in production. We have deployed fully configured all the possible solutions: Galaxy, Jupyter Notebook, RStudio server instances and NaaVRE, providing researchers with versatile, on-demand computational environments. These services leverage our flexible OpenStack and Kubernetes infrastructure, allowing dynamic scaling of CPU, GPU, and storage resources based on real-time workload requirements. The deployment has been carried out by the two Technologist with an ICT cloud computing expertise specializing in scientific applications, ensuring best practices in performance, security, and resource optimization.

During the initial phase of the project, we conducted an in-depth analysis of the LifeWatch community’s requirements, exploring a range of technical architectures for the VRE. Based on this groundwork, we selected a hybrid deployment model—combining an OpenStack instance for stable, long-term compute and storage with a Kubernetes cluster for dynamic, containerized workloads. This approach ensures transparent, optimized access to heterogeneous datasets while providing researchers with the flexibility to scale resources on demand. By aligning our infrastructure choices with both user needs and European interoperability standards, we have laid a robust foundation for the LifeWatch VRE Service.

By combining an IaaS foundation with container orchestration, these two solutions deliver a fully virtualized research environment that addresses a wide spectrum of scientific workloads. The OpenStack layer guarantees persistent, long-running services—ideal for traditional pipelines and data curation—while the Kubernetes-managed Docker platform enables rapid, on-demand provisioning of compute-intensive tasks. Researchers can seamlessly burst into GPU- and FPGA-equipped nodes for AI-driven analyses or high-throughput processing of massive datasets, all with intelligent data-locality optimization to minimize transfer times. This hybrid architecture ensures both the reliability required for continuous services and the agility demanded by cutting-edge, large-scale data processing.

By leveraging these virtualization technologies, we not only address the immediate requirements of the ITINERIS project—such as seamless deployment of containerized services and dynamic resource allocation—but also lay the groundwork for future expansion tailored to the LifeWatch community’s evolving needs. The chosen architecture’s inherent flexibility allows us to introduce new applications or use cases without major infrastructure changes: whether it’s integrating specialized analytics tools, onboarding novel data-services, or scaling out GPU-driven AI pipelines, the same underlying OpenStack-and-Kubernetes platform will support them all. In this way, the VRE we’ve built today becomes a living ecosystem, ready to adapt and grow alongside the scientific ambitions of LifeWatch Italy.

## 4. ACHIEVEMENTS TO DATE

### PRODUCTION ROLLOUT FOR LIFEWATCH

The production rollout has delivered a fully operational cloud-based environment tailored to the needs of the LifeWatch research community. Core analytical services – **Galaxy, Jupyter Notebook, RStudio, and NaaVRE** – have been deployed as **containerized instances orchestrated via Kubernetes** and running on **OpenStack-based virtual machines**. This setup provides flexible allocation of **CPU, GPU, and FPGA quotas on demand**, enabling researchers to run diverse workloads ranging from data-intensive analyses to AI-driven workflows. The platform currently serves **over 300 active users within the Italian LifeWatch community**, supporting daily interactive analysis and fostering reproducible science.

A key milestone in this rollout is the implementation of **resource federation**. By interconnecting **three INFN data centers** with **two regional cloud providers**, the platform enables transparent access to approximately **1 PB of environmental and biodiversity datasets**. This federation model ensures that data can be discovered, accessed, and processed seamlessly, regardless of where it is physically hosted, and lays the groundwork for scalable integration with broader European infrastructures such as EOSC.

Together, these achievements demonstrate the maturity of the infrastructure in production and its capability to support advanced Virtual Research Environments (VREs) for the LifeWatch community.

### SECURITY AND COMPLIANCE

The security and compliance framework of the infrastructure has been designed to ensure trust, interoperability, and adherence to European regulations. At the core of the solution lies the **INDIGO Identity and Access Management (IAM)** service, which provides federated authentication and authorization across multiple domains, including **eduGAIN** and the **LifeWatch SSO**. Multi-Factor Authentication (MFA) is enforced for all user accounts, significantly enhancing protection against unauthorized access and ensuring that only verified users can interact with sensitive services and data.

In parallel, the infrastructure adopts the **AARC Blueprint Architecture (BPA)**, a widely recognized set of guidelines for building interoperable Authentication and Authorization Infrastructures (AAI) within the European Open Science Cloud (EOSC) ecosystem. By aligning with the AARC BPA, the platform ensures seamless federation with other research infrastructures, while at the same time embedding mechanisms for **GDPR-compliant data handling**. This guarantees that all user data, credentials, and research outputs are managed according to best practices in privacy, accountability, and transparency.

Together, these measures establish a robust **security-by-design approach**, balancing the need for accessibility and federation with stringent requirements for compliance and user data protection.

## PERFORMANCE AND SCALABILITY

The infrastructure has demonstrated a high degree of reliability and elasticity in production. The **Notebook and RStudio services achieved an uptime of 99.2%**, confirming the robustness of the underlying architecture and the effectiveness of monitoring and operational procedures. This level of availability ensures that researchers can rely on the platform for continuous access to their analysis environments, minimizing interruptions in scientific workflows.

In terms of scalability, the system has been configured with **dynamic auto-scaling policies** implemented on Kubernetes clusters. During peak load, the infrastructure successfully sustained **up to 600 concurrent sessions**, automatically provisioning additional resources to meet demand. This elastic behavior guarantees that the platform can respond efficiently to usage spikes, while optimizing resource utilization under normal operating conditions.

Overall, these results validate the platform's ability to combine **service reliability with flexible scaling**, establishing a production-grade environment that can adapt to the evolving needs of the LifeWatch community.

## 5. FUTURE PLANS

### AI & DATA ANALYTICS

Future activities will focus on strengthening the platform's capability for advanced data analytics and artificial intelligence. In particular, the infrastructure will support the deployment of **distributed machine learning pipelines**, based on frameworks such as **TensorFlow** and **PyTorch**, strategically positioned close to data sources to minimize latency and optimize performance. In parallel, a dedicated **data movement and replication module** will be integrated to handle large-scale biodiversity datasets, ensuring efficient distribution and high availability across federated nodes.

### ENHANCED INTEROPERABILITY

Another key objective is to extend the **federation model** beyond the current infrastructure partners, establishing operational links with **EURO-ARGO** and **EMBRC**. These integrations will strengthen the marine and environmental domains and foster cross-disciplinary collaboration. At the same time, the infrastructure will adopt and expose **FAIR data services** through the **EOSC Portal**, ensuring compliance with European Open Science Cloud standards and maximizing data discoverability and reuse.

### COMMUNITY EXPANSION & SUSTAINABILITY

The sustainability of the infrastructure is also a priority. Plans include the **onboarding of three additional Italian research institutions by Q4 2025**, thereby broadening the user community and strengthening national coverage. To secure long-term viability, the project will define **cost-recovery and co-funding models** under the ITINERIS governance framework, balancing public

investment with community-driven contributions. This approach ensures both financial sustainability and continuous innovation.

## 6. CONCLUSION

This deliverable provides evidence of the project's achievements in building and operating a **robust, scalable, and secure cloud-based environment** to serve the needs of the LifeWatch community. Through the acquisition of new hardware, the deployment of containerized services, and the implementation of federated access across multiple national and regional data centers, the project has successfully enabled researchers to access advanced computing, storage, and analytical resources in a seamless and reliable manner.

The integration of tools such as **Galaxy, Jupyter Notebook, RStudio, and NaaVRE** has already demonstrated the capacity of the infrastructure to support diverse scientific workflows, while the adoption of **state-of-the-art security and compliance frameworks** ensures that services are interoperable, trustworthy, and aligned with European standards such as EOSC and GDPR. Performance indicators, including high service uptime and elastic scalability under peak demand, further confirm the maturity of the platform in production.

Looking forward, the project's **next steps will consolidate support for artificial intelligence and machine learning workloads**, expand the federation to include additional national and European research infrastructures, and strengthen the sustainability model through new institutional onboarding and governance-driven cost-recovery strategies. These developments will not only broaden the user base but also guarantee the **long-term scientific and societal impact of ITINERIS**, positioning it as a cornerstone for environmental and biodiversity research in Italy and its integration within the European Open Science Cloud ecosystem.

## LIFEWATCH USE CASE: NAAVRE DEPLOYMENT

The most important use case that has been supported in the DataCenter empowered using ITINERIS is an implementation of NaaVRE ([https://naavre.net/docs/NaaVRE\\_Interface/](https://naavre.net/docs/NaaVRE_Interface/)) dedicated to LifeWatch (<https://itineris.lifewatchitaly.eu/vreapp>), so the following section presents a comprehensive account of the activities undertaken to enable and support the NaaVRE Virtual Research Environment within the ITINERIS infrastructure. These activities encompassed the deployment of containerized services on Kubernetes clusters, the integration with federated authentication and authorization mechanisms, and the optimization of data access and processing pipelines. Particular emphasis is placed on demonstrating how NaaVRE extends the LifeWatch service portfolio, providing researchers with advanced, reproducible, and scalable data analysis workflows in alignment with EOSC interoperability and FAIR data principles.

NaaVRE Documentation

GitHub



Follow the tutorial

Try NaaVRE

Get in touch



### Turn Jupyter notebooks into reusable components

NaaVRE lets you containerize cells of Jupyter notebooks. Use these reusable components to build workflows.



### Compose workflows and run them in the cloud

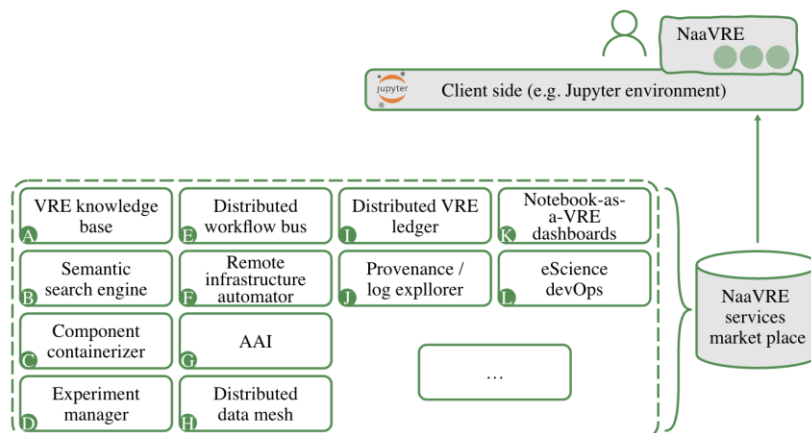
NaaVRE lets you compose workflows and run them into a cloud environment, with different parameters or input data.



### Find, reuse and share research assets

Find notebooks, containerized cells, workflows and datasets to run your own experiments in NaaVRE.

NaaVRE (Notebook as a Virtual Research Environment) provide user-centric support in the lifecycle of research activities, for example, discovering and accessing research assets or composing and executing application workflows. The Notebook as a Virtual Research Environment (NaaVRE) platform is a VRE platform developed by LifeWatch VLIC; it allows users to create customizable Collaborative Virtual Labs to conduct domain research activities [1].



1. Zhao, Zhiming, Spiros Koulouzis, Riccardo Bianchi, Siamak Farshidi, Zeshun Shi, Ruyue Xin, Yuandou Wang et al. "Notebook-as-a-VRE (NaaVRE): From private notebooks to a collaborative cloud virtual research environment." Software: Practice and Experience 52, no. 9 (2022): 1947-1966. <https://doi.org/10.1002/spe.3098>

The deployment of **NaaVRE (Narrative Virtual Research Environment)** within the distributed data center operated by LifeWatch Italy, with resources located in **Lecce and Bari**, represented a

major milestone in demonstrating the capability of the ITINERIS infrastructure to support advanced Virtual Research Environments. This activity required the design and configuration of a Kubernetes-based environment on **bare metal machines**, enabling a production-grade platform for container orchestration and scientific services. The deployment was technically coordinated and carried out by **Antonio Marra** and **Salvatore Tataro** (University of Salento), who ensured the correct setup of the cluster and its alignment with LifeWatch operational requirements.

During the installation phase, a number of technical challenges and best practices were identified, which were consolidated into revised requirements for future deployments. For instance, the initial configuration highlighted the **critical importance of having a persistent volume provisioner** capable of dynamically allocating storage volumes to users and services. In the absence of such a mechanism, the cluster would have required manual intervention by system administrators each time a new volume was requested, severely limiting scalability and usability. Similarly, the team observed that several aspects of the infrastructure — including **domain name assignment, public IP availability, reverse proxy configuration, and SSL certificate management** — need to be fully prepared in advance, prior to initiating the deployment, in order to avoid misconfigurations and service interruptions. These lessons learned have been documented and are now part of the reference deployment guidelines.

Once the cluster was fully operational, the deployment of NaaVRE itself was carried out using **Helm**, the Kubernetes package manager, and the publicly available Helm chart hosted on GitHub. This activity was led by **Stefano Rosato** (EquiUP), with dedicated support from **Spiros Koulouzis** and **Gabriel Pelouze** (LifeWatch ERIC). The process followed a well-structured methodology: the repository was first cloned, and the template configuration file (`values/values-example-basic.yaml`) was adapted into a specific configuration file (`values/values-itineris.yaml`) tailored to the ITINERIS infrastructure. This customization step required the inclusion of key parameters such as the fully qualified domain name, randomly generated secrets and credentials, API keys, and the configuration of the virtual labs to be enabled within NaaVRE. The ownership and maintenance of this configuration file remain under the responsibility of **Stefano Rosato (EquiUP)**, ensuring accountability and continuity in the lifecycle of the deployment.

The customized configuration was then applied to the Kubernetes cluster using Helm commands, effectively launching NaaVRE and its associated services in the production environment. Throughout this process, several issues were encountered, ranging from minor misconfigurations to interoperability problems. However, each of these was systematically analyzed, resolved, and the corresponding fixes were subsequently integrated into the public repository, thus strengthening the maturity of the deployment artifacts.

Overall, this deployment not only resulted in a fully functional instance of NaaVRE within the LifeWatch distributed data center, but also produced a set of **validated operational guidelines** and **deployment best practices**. These outcomes will significantly simplify and accelerate future installations, while also contributing to the broader reproducibility and sustainability of Virtual Research Environments across the LifeWatch and ITINERIS federated infrastructure.

In the next picture you can see the NaaVRE portal installed in production for the Lifewatch community.

