D4.2 - Infrastructure & Services Definition (b-final)

Work Package	WP 4, EO4EU Data Marketplace Ecosystem
Lead Author (Org)	Marios Sophocleous (EBOS)
Contributing Author(s) (Org)	Name (organisation) NKUA, ECMWF, NVCR, SIS, CINECA, HES-SO, DANAOS, KEMEA, ENG, IES, MEEO, IVI
Due Date	31.05.2024
Date	3.06.2024
Version	V0.1

Dissemination Level

X	PU: Public
	PP: Restricted to other programme participants (including the Commission)
	RE: Restricted to a group specified by the consortium (including the Commission)
	CO: Confidential, only for members of the consortium (including the Commission)

Versioning and contribution history

Version	Date	Author	Notes
0.1	11.03.2024	EBOS	TOC
0.2	15.04.2024	NVCR, SIS, CINECA, HES-SO, DANAOS, KEMEA, ENG, IES, MEEO, IVI	Contributions to the deliverable
0.3	30.05.2024	EBOS	1st complete version for peer review
0.4	31.05.2024	NKUA & ECMWF	Peer Review
1.0	03.06.2024	EBOS	Submission-ready version after peer review

Terminology

Terminology/Acronym	Description
AI	Artificial Intelligence
AR	Augmented Reality
API	Application Programming Interface
AUX	Auxiliary
CI/CD	Continuous Integration/Continuous Development
CPU	Central Processing Unit
DSL	Domain Specific Language
EO	Earth Observation
EU	European Union
GDL	Graph Description Language
GPU	Graphics Processing Unit
HPC	High-Performance Computing
IaaS	Infrastructure as a Service
KG	Knowledge Graph
ML	Machine Learning
PaaS	Platform as a Service
RAM	Random Access Memory
SSO	Single Sign-On
UMM	User Management Module
VR	Virtual Reality
WFE	Workflow Editor
WP	Work Package
XR	Extended Reality

Table of Contents

List of Figures

List of Tables

Executive Summary

Deliverable “D4.2 - Infrastructure & Services Definition- Final” focuses on updating the infrastructure requirements for each component of the EO4EU platform, presenting an updated architecture and detailing the consolidated infrastructure needs. This deliverable builds upon the foundations laid in D4.1 “D4.1 - Infrastructure & Services Definition (a)”, which defined the initial specifications of the cloud and High-Performance Computer (HPC) infrastructure and established guidelines for data storage, connectivity, and networking with external data sources.

In D4.2, we detail the infrastructure updates necessary to support the evolving requirements of the EO4EU platform. The updated architecture reflects the current state of the platform, ensuring it remains modular and scalable to accommodate new services and components. While the HPC infrastructure will not be actively utilised at this stage, the training and running of machine learning algorithms are efficiently managed on the CINECA cloud infrastructure. However, provisions have been made to seamlessly integrate HPC resources if future needs arise, ensuring the platform's adaptability and robustness.

Critical elements of D4.2 include:

- **Updated Infrastructure Requirements:** Detailed specifications for each component highlight the changes and enhancements needed to support the EO4EU platform's current and future needs.
- **Revised Architecture:** An updated architectural blueprint incorporating new services and components while ensuring compatibility with existing systems. This architecture supports a flexible and modular approach, allowing for efficient integration and scalability.
- **Consolidated Infrastructure Needs:** This is a comprehensive overview of the consolidated infrastructure requirements for the entire EO4EU platform. This includes computational and storage needs and networking and connectivity considerations, ensuring a cohesive and efficient operational environment.
- **Provisions for HPC Integration:** Although HPC resources are not currently required, the document outlines the provisions to integrate HPC with the existing cloud infrastructure. This ensures the platform's preparedness to handle increased computational demands and maintain high performance.

By providing these updates and detailed requirements, D4.2 ensures that the EO4EU platform remains robust, scalable, and ready to meet the diverse needs of its users, both now and in the future. The deliverable ensures that the platform's infrastructure can support the ongoing development and deployment of services, maintaining high performance and reliability standards.

1 Introduction

This document represents the second release in a series of deliverables to fulfil the requirements outlined in Task T4.1, specifically focusing on 'Integration Planning & Service Provision Specification.' Deliverable D4.2's primary purpose is to provide updated infrastructure requirements for each component of the EO4EU platform and present a revised architecture that aligns with the platform's current needs. Building upon the foundations established in D4.1 [1], this deliverable incorporates new insights and developments to ensure the platform remains robust and scalable.

The key objectives of this report include detailing the updated specifications for the cloud infrastructure (WEkEO[1] & CINECA) and ensuring it supports the latest service components and processing methods. It presents an updated architectural blueprint that integrates new components and services while maintaining modularity and scalability. The report offers a comprehensive overview of the consolidated infrastructure requirements for the entire EO4EU platform, encompassing computational, storage, and networking needs. Additionally, it outlines provisions for the future integration of HPC resources, ensuring the platform is prepared to handle increased computational demands if necessary. These objectives are achieved by incorporating insights from the structural and technical components identified in Work Packages (WPs) 2 & 3, along with a thorough analysis and assessment of the evolving user requirements and system specifications. Although the current focus is on utilising the CINECA cloud infrastructure for training and running machine learning algorithms, this document ensures that provisions are in place for seamless integration of HPC resources if future needs arise.

By detailing these updates and revisions, Deliverable D4.2 ensures that the EO4EU platform remains adaptable, accessible, and capable of supporting diverse data processing methods. This aligns with the project's overarching vision of enhancing the usability of environmental observation information.

1.1 Purpose of the Document

Deliverable D4.2 provides an updated technical analysis of the infrastructure and services required for the EO4EU Platform. Building on D4.1, it updates the requirements, specifications, and technical components needed to deploy and integrate software components within cloud and local networks. This document refines the infrastructure to support a comprehensive ecosystem that enhances data access and processing.

Acknowledging the contributions of WP2 and WP3, which analysed user requirements and developed components, D4.2 incorporates the latest insights to update the platform's infrastructure and architecture. The primary objective is to present an updated system architecture detailing interfaces and integration demands based on evolving requirements. It also establishes updated performance metrics to ensure the effective operation of the EO4EU framework and addresses provisions for potential HPC integration. Linked to WP5, which focuses on deploying and integrating use cases with the EO4EU Platform, D4.2 defines specific infrastructure requirements and service provisions for each use case, leading to overall system enhancements and optimisations. This document ensures that the EO4EU platform remains adaptable, accessible, and capable of supporting diverse data processing methods per the project's vision.

1.2 Relation to Other Activities and Deliverables

Deliverable 4.2 is essential in updating the infrastructure and services of the EO4EU Platform, connecting directly with 'WP2 – Requirements Elicitation and Conceptual Framework Specification-Task T2.3 Technical Specifications, Interoperability Requirements, and Scalability Analysis.' This work

[1] https://www.wekeo.eu/

package and task define and analyse the System Architecture of the Software Components, including their functional and technical requirements.

As WP2 continues to analyse end-user requirements and system features, a concurrent evaluation of technical requirements and system specifications of the EO4EU platform is conducted. This dual analysis aligns with the operational and technical necessities outlined in D4.2, ensuring seamless integration with ongoing project developments. Additionally, 'WP3 – Data Orchestration & Machine Learning - T3.2 Systems and Services Orchestration' remains crucial in supporting the needs and requirements of all Software Components designated for deployment on cloud infrastructure. Tasks T3.1, T3.2, T3.3, and T3.4 and their corresponding Deliverables D3.1-D3.6 in WP3 provide significant technical requirements and provisions for services, aiding in the appropriate design, development, and deployment of the required infrastructure.

D4.2 continues the deployment of EO4EU Software Components, evaluating operations, functionalities, applications, and outcomes. The process involves technologies such as the Kubernetes[2] container orchestration system for automating deployment and management and the Apache KAFKA,[3] a distributed event streaming platform. Performance testing and evaluation of software responses to end-user requests lead to an in-depth investigation of the necessary infrastructure and software services for the EO4EU platform. Moreover, D4.2 includes an updated analysis of parameters such as access and streaming of multiple Earth Observation (EO) Data Sources, data multi-processing, computing power assessment, HPC architecture and infrastructure, cluster/cloud infrastructure, communication framework requirements, and data storage and organisational needs.

This deliverable is closely linked with 'WP5 – EO Data Uptake Demonstration of the EO4EU frameworks,' which is dedicated to deploying and integrating various use cases within the EO4EU Platform. As each use case is deployed and tested, a deeper understanding of the infrastructure requirements and service provisions is developed, contributing to new technical insights. These insights enhance and optimise the overall system architecture and infrastructure of the EO4EU Platform.

1.3 Deliverable Overview and Report Structure

The following provides a summary of the Deliverable in Chapters and gives the corresponding Report Structure:

Chapter 1: Introduction

- This section outlines the document's primary objectives and scope, explains its interactions with other activities and deliverables within the EO4EU project, and provides an overview and roadmap of the entire deliverable for reader navigation.

Chapter 2: Updated Services Definitions & Requirements

- This section is central to the document, detailing the updated definitions and requirements of the services within the EO4EU platform. It includes the Data Tier, Platform as a Service (PaaS), Machine Learning (ML), and Front-End Tier.

Chapter 3: Summary of Updated Infrastructure & Services Specifications

- This section consolidates and summarises the key updated specifications related to infrastructure and services, providing a concise reference point for readers.

Chapter 4: Infrastructure as a Service (IaaS) Tier

- This section provides the specifications of the maximum resources allocated from the cloud providers for the EO4EU platform.

Chapter 5: Future Infrastructure Integration Planning

[2] https://kubernetes.io/

[3] https://kafka.apache.org/

- This section provides the provisions made for future infrastructure needs and integrations.

Chapter 6: Conclusion

- The chapter summarises the key findings, insights, and implications derived from exploring updated service definitions, requirements, and infrastructure specifications.

2 Updated Services Definitions & Requirements

This chapter provides an updated high-level description of the EO4EU platform components and services based on the architecture established in D2.4 [1] and the latest developments. It details each component's and service's revised infrastructural needs to plan their deployment on the available infrastructure, ensuring maximum performance and scalability. This updated analysis ensures that the EO4EU platform remains robust, efficient, and adaptable to evolving requirements.

2.1 EO4EU Platform Architecture

The EO4EU architecture (Figure 1) comprises five distinct tiers and their subsequent components, depicted in Table 1. In this section, however, the minimum resource requirements for each element will be defined for IaaS and the OpenStack component because those resources depend on the components' requirements running and on the actual users and their queries.

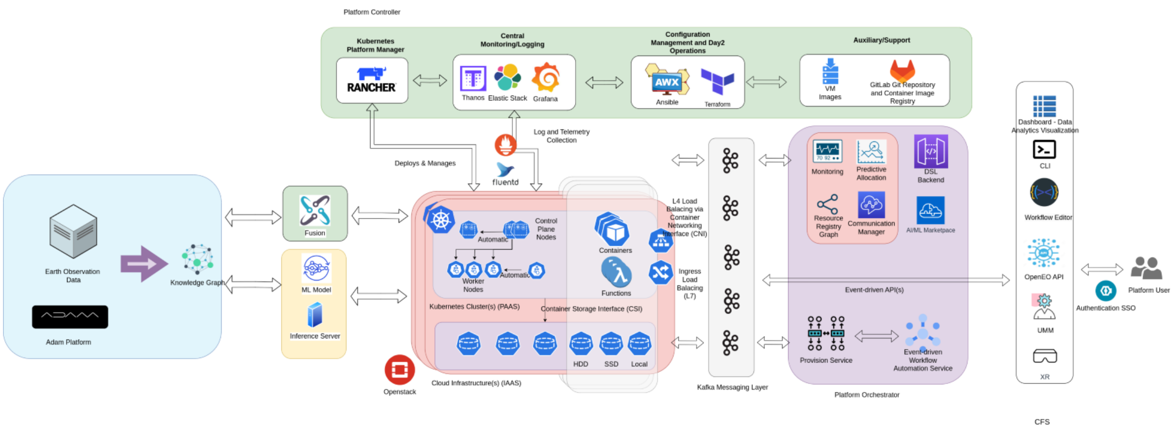

Figure 1: EO4EU Platform Architecture.

Table 1. Components & Tiers of EO4EU Platform.

Components/Tiers	Data Tier	IaaS Tier	PaaS Tier	ML Tier	Frontend Tier
Knowledge Graph	✓				
OpenStack		✓			
Platform Controller					
Kubernetes Platform Manager			✓		
Auxiliary (AUX)/support-GitLab Git Repository and Continuous Integration/Continuous Development (CI/CD)			✓		
AUX/support-GitLab Container Image Registry			✓		
AUX/support-Config Management & Day2 Operations			✓		
Authentication Single-Sign-On (SSO)/User Management Module (UMM)			✓		
Platform Orchestrator			✓		
Provision Service			✓		
Pre/Post-Processor			✓		
Resource Registry			✓		
Artificial Intelligence (AI)/ML Marketplace			✓		
KAFKA Message Bus			✓		
Domain Specific Language (DSL) Engine			✓		
Data Fusion Engine			✓		

Machine Learning Model				✓	
Machine Learning Inference Server				✓	
Machine Learning Controller				✓	
Machine Learning Wrapper				✓	
Online Portal/Data Analytics Visualization					✓
Workflow Editor					✓
Extended Reality (XR)/Virtual Reality (VR) System					✓
Extended OpenEO Application Programming Interface (API)					✓

2.2 Data Tier

In the updated Data Tier (Figure 2) for D4.2, the platform now integrates a Knowledge Graph (KG) to manage the diverse data obtained from various sources. This component plays a central role in enhancing and standardising the incoming data, streamlining the preprocessing tasks previously handled by multiple components in D4.1.

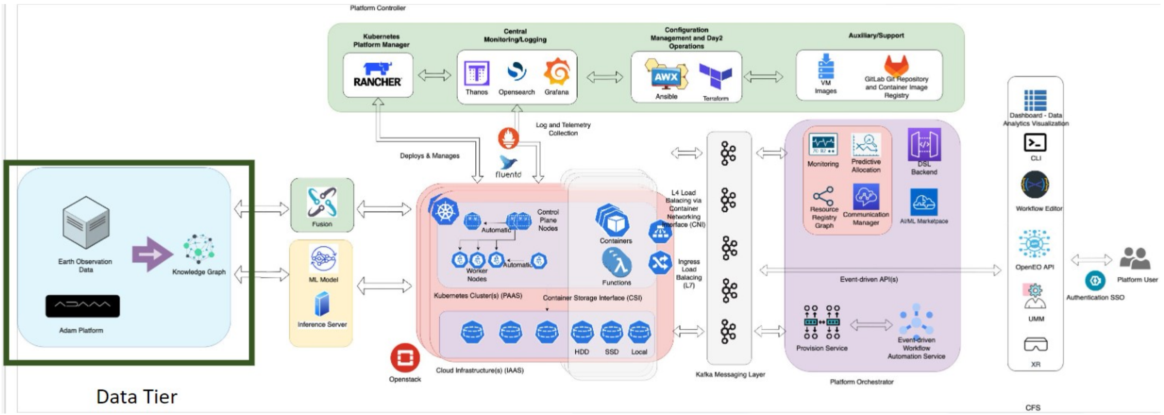

Figure 2. Data Tier part of the EO4EU Architecture.

2.2.1 Knowledge Graph

The KG in the Data Tier enhances user search capabilities, enabling access to Earth Observation (EO) data through semantic search and free-text queries. It processes and organises datasets from sources like Copernicus Services and third-party platforms like ADAM, using metadata embeddings for a structured, unified approach to diverse EO data.

Designed for user-friendliness, the KG makes EO data accessible to non-experts, improving search accuracy and relevance through semantic metadata processing. By converting user queries into internal representations, the KG efficiently matches them with dataset vectors, providing targeted and relevant results.

The minimum infrastructure requirements to run the KG are outlined in Table 2 below.

Table 2. Minimum Infrastructure Needs for KG.

Specification	Minimum Values	Infrastructure
Computational Needs	4 Central Processing Units (CPUs)	CINECA Cloud
Memory Needs	5 GB Random Access Memory (RAM)	
Storage Needs	1 TB	
Graphics Processing Unit (GPU) Needs	Med (depending on user queries)	

2.3 Platform as a Service (PaaS) Tier

PaaS (Figure 3) delivers a higher-level cloud computing service than IaaS. Building upon the foundational infrastructure provided by IaaS

, PaaS offers additional services and tools to simplify application development, deployment, and management.

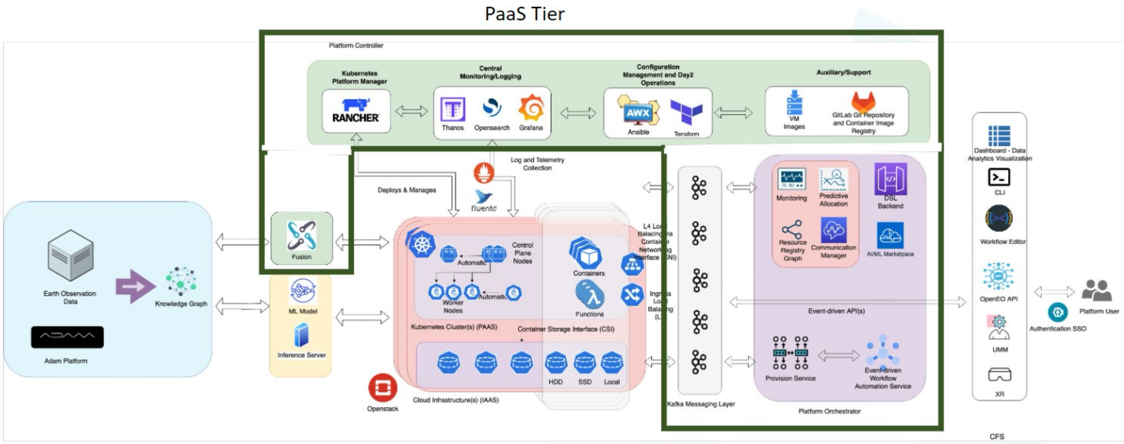

Figure 3: PaaS Tier part of the EO4EU Architecture.

2.3.1 Platform Controller

The Platform Controller includes all necessary components and services to manage and operate the entire platform. This encompasses the Kubernetes Platform Manager, the AUX/support-Gitlab Git Repository and CI/CD, the AUX/support-Gitlab Container Image Registry, the AUX/support-Configuration Management & Day 2 Operations and the Authentication SSO components responsible for resource monitoring, and simplifying application deployment on the platform.

The minimum infrastructure requirements for the Platform Controller are detailed in Table 3 below.

Table 3. Minimum Infrastructure Requirements for Platform Controller.

Sub-Component	Specification	Minimum Values	Infrastructure
Kubernetes Platform Manager (Rancher)	Computational Needs	24 CPUs	WEkEO
	Memory Needs	90 GB RAM	
	Storage Needs	385 GB	
	GPU Needs	None	
AUX/support Gitlab (Git Repository & Container Image Registry	Computational Needs	On-demand	
	Memory Needs	8 GB RAM	
	Storage Needs	766 GB	
	GPU Needs	None	
AUX/support-Configuration Management & Day 2 Operations	Computational Needs	12 CPUs	
	Memory Needs	38 GB RAM	
	Storage Needs	160 GB	
	GPU Needs	None	
Authentication SSO/UMM	Computational Needs	On-demand	
	Memory Needs	1 GB RAM	
	Storage Needs	300 GB	
	GPU Needs	None	

2.3.2 Platform Orchestrator

The Platform Orchestrator, central to the system, comprises five sub-components: the Provision Service, the Pre-Processor, the Post-Processor, the Resource Registry/Resource Registry Handler and the AI/ML Marketplace. The AI/ML Marketplace database includes all the ML models available to users. It is currently accessible through the Workflow Editor (WFE) while defining their preferred workflow. In contrast to the Inference Server Service, where the ML models are trained and the ML components, where the ML models take data as input and generate output to be visualised, the AI/ML Marketplace aims to be used in the future as an online store and database where users can upload their models and buy any models that other users have licensed.

The minimum infrastructure requirements for the Platform Orchestrator are detailed in Table 4 below.

Table 4. Minimum Infrastructure Requirements for Platform Orchestrator.

Sub-Component	Specification	Minimum Values	Infrastructure
Provision Service*	Computational Needs	2 CPUs	CINECA Cloud
	Memory Needs	4 GB RAM	
	Storage Needs	100 GB	
	GPU Needs	None	
Pre-Processor	Computational Needs	2 CPUs	
	Memory Needs	4 GB RAM	
	Storage Needs	100 GB	
	GPU Needs	None	
Post-Processor	Computational Needs	2 CPUs	
	Memory Needs	4 GB RAM	
	Storage Needs	100 GB	
	GPU Needs	None	
Resource Registry	Computational Needs	4 CPUs	
	Memory Needs	6 GB RAM	
	Storage Needs	110 GB	
	GPU Needs	None	
AI/ML Marketplace	Computational Needs	8 CPU	
	Memory Needs	32 GB	
	Storage Needs	25 GB	
	GPU Needs	None	

*In the case of autoscaling, the resources may be multiplied by the number of replicas.

2.3.3 KAFKA Message Bus

The Apache Kafka message bus manages real-time data streams and ensures reliable message delivery within the EO4EU platform. It supports event streaming and log aggregation, enabling seamless data flow between components. Kafka producers send messages to specific topics, which are then consumed by platform components, ensuring real-time data processing. Kafka's architecture supports high throughput and low latency, ideal for handling large data volumes. Operating within the Dockerized environment, Kafka integrates with the EO4EU message bus framework, ensuring continuous data availability for processing and analysis.

The minimum infrastructure requirements for the Apache Kafka message bus are detailed in Table 5 below.

Table 5. Minimum Infrastructure Requirements for KAFKA Message Bus.

Specification	Minimum Values	Infrastructure
Computational Needs	16 CPUs	CINECA Cloud
Memory Needs	8 GB RAM	
Storage Needs	50 GB	
GPU Needs	None	

2.3.4 DSL Engine

The DSL Engine validates and controls System workflow and supports the development of the Graph Description Language (GDL) for the WFE. GDL defines system nodes, attributes, metadata, and relationships. The DSL Engine, integrating with the WFE via the WFE AUX Service, validates and compiles workflows for deployment. Valid workflows are converted to YAML and sent to the AUX Service, while invalid ones generate error reports for resolution.

The minimum infrastructure requirements for the DSL Engine are detailed in Table 6 below.

Table 6. Minimum Infrastructure Requirements for DSL Engine.

Specification	Minimum Values	Infrastructure
Computational Needs	4 CPUs	CINECA Cloud
Memory Needs	8 GB RAM	
Storage Needs	5 GB	
GPU Needs	None	

2.3.5 Data Fusion Engine

The Data Fusion Engine enhances situational awareness by merging data readings. It performs two main functions: creating fusion models and pipelines for spatiotemporal data aggregation and dynamically executing workflows in parallel based on user requests. These pipelines are available on the AI/Marketplace and used within the Workflow Execution Environment.

When a new workflow starts, the Fusion Proxy sets it up in the Kubeflow environment, handles execution, initialises the environment, runs the pipeline, and publishes results. Notifications are sent via Kafka. Computation tasks are managed for high productivity, using HPC or GPU environments as needed. Pipelines run in the Dockerized Kubeflow environment and communicate results through the EO4EU message bus.

The minimum infrastructure requirements for the Data Fusion Engine are shown in Table 7 below.

Table 7. Minimum Infrastructure Requirements for Data Fusion.

Specification	Minimum Values	Infrastructure
Computational Needs	0.2 CPU s	CINECA Cloud
Memory Needs	1 MB RAM	
Storage Needs	1 TB	
GPU Needs	Med (depending on user queries)	

2.4 Machine Learning Tier

The ML Tier (Figure 4) of the EO4EU platform includes several essential ML components: an ML models' repository, an inference server, an ML controller, and ML wrapper components.

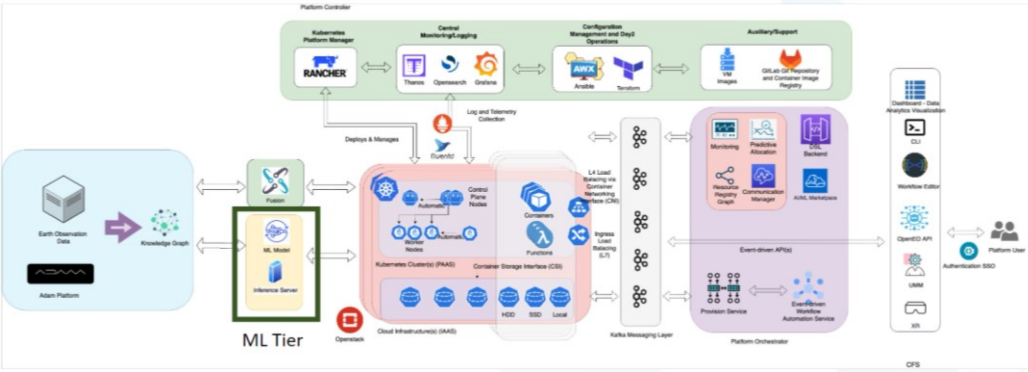

Figure 4: ML Tier part of the EO4EU Architecture.

The ML models are trained on diverse datasets to provide accurate predictions and insights. The inference server trains these models, whilst the ML models' repository stores all trained models with version control and metadata for efficient management. The ML controller is used to create and delete a workflow. The ML wrapper's role is to parse the data and query the inference server to complete the workflow task and store the results. These components enable the platform to efficiently process and analyse large data volumes, delivering valuable insights. Although initially, there was a challenge in terms of network storage for the outputs of the ML models (GeoTIFFs), the challenge has been overcome by using different Cloud optimised files and also by optimising the details and information stored in the GeoTIFF files so that they are not over-engineered. However, they can still provide the information the user is looking for.

The minimum infrastructure requirements for the ML Tier are detailed in Table 8 below.

Table 8. Minimum Infrastructure Requirements for ML Tier.

Sub-Component	Specification	Minimum Values	Infrastructure
ML Models	Computational Needs	-	CINECA Cloud
	Memory Needs	-	
	Storage Needs	10 GB*	
	GPU Needs	None	
ML Inference Server**	Computational Needs	1 CPUs	
	Memory Needs	4 GB RAM	
	Storage Needs	-	
	GPU Needs	High	
ML Controller	Computational Needs	1 CPUs	
	Memory Needs	1 GB RAM	
	Storage Needs	-	
	GPU Needs	None	
ML Wrapper	Computational Needs	1 CPU/per workflow	
	Memory Needs	2 GB	
	Storage Needs	-	
	GPU Needs	None	

*Depending on the number and size of models.

**Number of CPUs at which clock speed is needed: Autoscalable depending on need. Memory needed (RAM): Autoscalable depending on need. Storage required (disk space): No persistent storage; download models from ML Models Repository. GPU needs (Low/med/high): Great speed up from

GPUs and other ML accelerators. CPUs can be used instead. There is currently one inference server by use cases. If one inference server is insufficient, a second will be created.

2.5 Front-End Tier

The Front-End Tier (Figure 5) delivers a multi-dimensional User Interface (UI) encompassing the Online Portal/Data Analytics Visualisation, Workflow Editor, XR/VR and the Extended OpenEO API, allowing users to interact with and control the platform effectively.

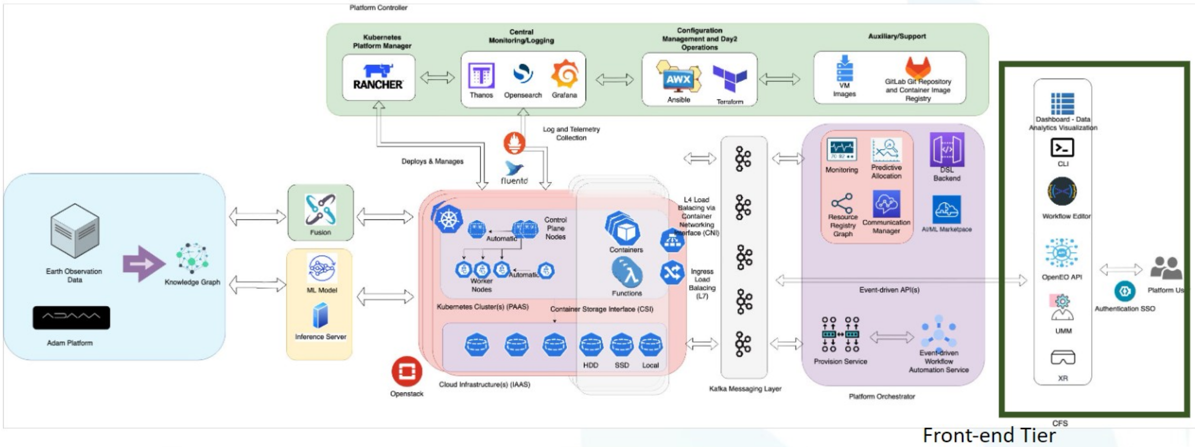

Figure 5: Front-End Tier part of the EO4EU Architecture.

2.5.1 Online Portal/Data Analytics Visualisation

The Data Analytics Visualization tool helps with decision-making and policy development by offering advanced visualisations and analytics. It makes understanding data more accessible through interactions with AI/ML modules and uses augmented reality for an immersive experience with Earth Observation (EO) data. Designed for scientists and European Union (EU civilian users, it supports real-time analytics and helps understand research data better. The tool combines different visualisation methods to highlight important events related to weather, climate, and environmental changes. It is an easy-to-use platform for exploring various datasets, especially environmental observations and EO data.

The minimum infrastructure requirements for the Data Analytics Visualization tool are shown in Table 8 below.

Table 9. Minimum Infrastructure Requirements for Online Portal/Data Analytics Visualisation.

Specification	Minimum Values	Infrastructure
Computational Needs	2 CPUs	CINECA Cloud
Memory Needs	16 GB RAM	
Storage Needs	80 GB	
GPU Needs	None	

2.5.2 Workflow Editor

The Workflow Editor is essentially the Front End of the DSL Engine, accessible through the EO4EU Online Portal. Through the Workflow Editor, users can visually create workflows from the data sources to data fusion to specific ML processing using the available ML models and eventually visualise the output data through the Data Analytics Visualisation component.

The minimum infrastructure requirements for the Workflow Editor tool are shown in Table 10 below.

Table 10. Minimum Infrastructure Requirements for Workflow Editor.

Specification	Minimum Values	Infrastructure
Computational Needs	2 CPUs	CINECA Cloud
Memory Needs	2 GB RAM	
Storage Needs	5 GB	
GPU Needs	None	

2.5.3 Extended/Virtual Reality

The XR system elevates the visualisation and exploration of Earth Observation (EO) data via a web-based interface. It incorporates Augmented Reality (AR) and VR components, enabling real-world data analysis and 3D visualisation on compatible hardware. Designed explicitly for EO data, these interfaces offer interactive tools to enhance comprehension of environmental observations. Leveraging the Web XR device API, the XR system seamlessly communicates with other EO4EU applications to access processed EO data. Its objective is to deliver a brief and immersive experience for users engaging with EO data in augmented and virtual realities.

The minimum infrastructure requirements for the XR/VR component are shown in Table 11 below.

Table 11. Minimum Infrastructure Requirements for XR/VR component.

Specification	Minimum Values	Infrastructure
Computational Needs	2 CPUs/instance	CINECA Cloud
Memory Needs	2 GB RAM/instance	
Storage Needs	-	
GPU Needs	None	

2.5.4 Extended OpenEO API

The OpenEO API manages control interfaces and facilitates communication between users and various EO4EU software components. Its primary function is establishing intelligent data interfaces and enabling seamless communication among software modules. The API defines user requirements and requests services and aligns with system component specifications to generate the requested user data. Acting as a conduit for intelligent communication between EO4EU applications and software functions, it functions as a remote communication server, receiving user requests and delivering EO data in the requested format. Additionally, the API interacts with the Kubernetes Platform Software Management and other Control Software modules to retrieve real-time and continuous data from the requested cluster or available servers and software resources for processing EO data.

The minimum infrastructure requirements for the Extended OpenEO API are shown in Table 12 below.

Table 12. Minimum Infrastructure Requirements for Extended OpenEO API.

Specification	Minimum Values	Infrastructure
Computational Needs	2 CPUs	CINECA Cloud
Memory Needs	32 GB RAM	
Storage Needs	80 GB	
GPU Needs	None	

3 Summary of Updated Infrastructure & Services Specifications

Table 13. Summary of updated minimum infrastructure & services specifications.

<table>
<tr><th>Service/Component</th><th>Tier</th><th>Requirements</th><th>Infrastructure</th><th>Total Infrastructure Requirements*</th></tr>
<tr><td>KG</td><td>Data</td><td>Computational Power: 4 CPUs
Memory: 5 GB RAM
Storage: 1 TB
GPU Demand: Med (depending on user queries)</td><td>CINECA Cloud</td><td rowspan="6">WEkEO
Computational Power: 36 CPUs
Memory: 137 GB RAM
Storage: 1.6 TB
GPU Demand: None

CINECA Cloud
Computational Power: 50 CPUs
Memory: 130 GB RAM
Storage: 2.7 TB
GPU Demand: High when needed</td></tr>
<tr><td>Platform Controller</td><td>PaaS</td><td>Computational Power: 36 CPUs + on demand for some sub-components
Memory: 137 GB RAM
Storage: 1.6 TB
GPU Demand: No need</td><td>WEkEO</td></tr>
<tr><td>Platform Orchestrator</td><td>PaaS</td><td>Computational Power: 18 CPUs
Memory: 50 GB RAM
Storage: 435 GB
GPU Demand: No need</td><td rowspan="4">CINECA Cloud</td></tr>
<tr><td>KAFKA Message Bus</td><td>PaaS</td><td>Computational Power: 16 CPUs
Memory: 8 GB RAM
Storage: 50 GB
GPU Demand: No need</td></tr>
<tr><td>DSL Engine</td><td>PaaS</td><td>Computational Power: 4 CPUs
Memory: 8 GB RAM
Storage: 5 GB
GPU Demand: No need</td></tr>
<tr><td>Data Fusion Engine</td><td>PaaS</td><td>Computational Power: 0.2 CPUs
Memory: 1 MB RAM
Storage: 1 TB
GPU Demand: Med (depending on user queries)</td></tr>
</table>

ML Components	ML	Computational Power: 2 CPUs (auto-scalable on needs) Memory: 7 GB RAM (auto-scalable on needs) Storage: 10 GB, depending on the number and size of models GPU Demand: High		
Online Portal/Data Analytics Visualisation	Front -End	Computational Power: 2 CPUs Memory: 16 GB RAM Storage: 80 GB GPU Demand: No need		
Workflow Editor	Front -End	Computational Power: 2 CPUs Memory: 2 GB RAM Storage: 5 GB GPU Demand: No need		
XR/VR System	Front -End	Computational Power: 2 CPUs Memory: 2 GB RAM Storage: - GPU Demand: None		
Extended OpenEO API	Front -End	Computational Power: 2 CPUs Memory: 32 GB RAM Storage: 80 GB GPU Demand: No need		

*Some resources are on-demand; hence, depending on the needs at that specific time and considering the number of simultaneous users/use cases, the resources are dynamically distributed to allow the best user experience for all users whilst maintaining an optimal resource allocation.

4 Infrastructure as a Service (IaaS) Tier

IaaS (Figure 6) abstracts the infrastructure, enabling the PaaS components to dynamically and programmatically use compute, storage, and networking resources through the OpenStack REST APIs.

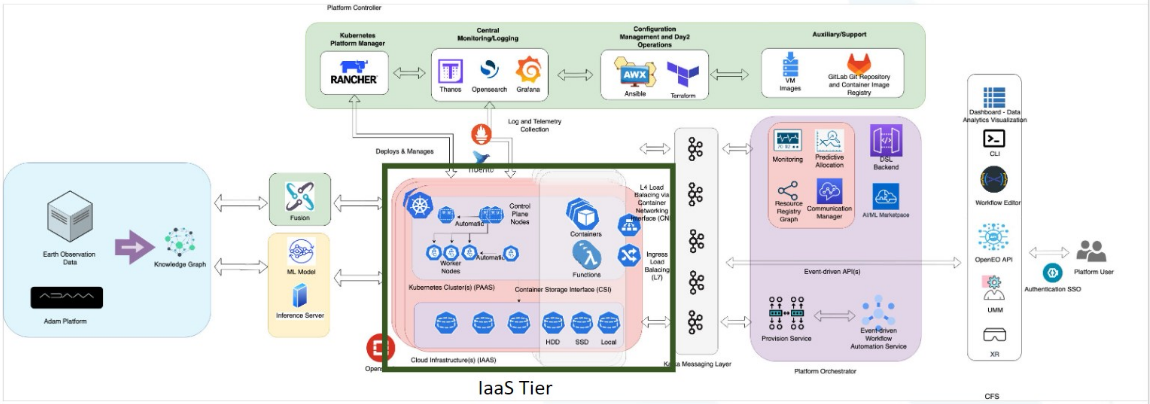

Figure 6: IaaS Tier part of the EO4EU Architecture.

4.1 OpenStack

OpenStack, an open-source cloud computing platform, offers robust Infrastructure as a Service (IaaS) capabilities. It provides essential cloud infrastructure services, including computing, storage (both block and object), and networking. As a foundational layer for higher-level services like Kubernetes and the PaaS tier, OpenStack remains a flexible and versatile solution for building and managing cloud environments.

OpenStack infrastructure requirements depend entirely on the resources needed by the rest of the platform components; hence, a different approach was followed to plan this component. Table 14 below shows the maximum infrastructure resources dedicated to the EO4EU platform (quotas). It is important to note that as the number of applications running on Kubernetes increases, the infrastructure needs will correspondingly rise. Eventually, OpenStack infrastructure specifications will encompass all the resources allocated by cloud providers for the whole platform.

Table 14. Maximum Infrastructure resources dedicated to the EO4EU Platform (quotas).

Infrastructure	Infrastructure Specifications Quotas
WEkEO	Computational Power: 175 CPUs Memory: 3.9 TB RAM Storage: 62.5 TB GPU Demand: No need
CINECA Cloud	Computational Power: 250 CPUs Memory: 1.8 TB RAM Storage: 9 TB GPU Demand: High need for some components

5 Future Infrastructure Integration Planning

The chapter on "Future Infrastructure Integration Planning" elaborates on the core principles guiding the design and implementation of the EO4EU platform. The updated Infrastructure Integration Planning also maintains the following attributes:

- Interoperability is paramount, ensuring seamless communication and data exchange among all software components. Standard protocols and data formats, such as JSON in the EO4EU Project, facilitate smooth communication.
- Scalability is a crucial consideration, with the platform designed to accommodate growing data volumes and user bases. Horizontal and vertical scaling capabilities enable the system to expand as needed.
- Security measures are robust, safeguarding sensitive data, user credentials, and platform functionalities. Encryption, authentication, and authorisation mechanisms, including the AAA framework, intelligently control EO4EU software components and data access.
- Efficient data processing workflows are prioritised to handle big data effectively. Distributed computing techniques and compression algorithms optimise the processing of multiple EO data streams while selectively processing desired data.
- Fault tolerance is ensured through resilient hardware and software resources that sustain data overloads and concurrent user activity. Redundancy and fault-tolerant strategies guarantee continuous operation.
- High performance is pursued through code optimisation, caching mechanisms, and efficient algorithms to enhance system responsiveness and efficiency.
- Comprehensive documentation detailing integration architecture, API specifications, data formats, and data flows is maintained. This documentation aids development and future maintenance, ensuring the platform remains well-documented and accessible to stakeholders.

While the initial plan included integrating HPC for the ML Inference service, its immediate deployment is optional. The scalable and adaptable infrastructure of the EO4EU platform obviates the need for HPC. However, the platform is architected with flexibility, allowing for seamless integration of HPC in the future should the demand arise. This strategic approach underscores the platform's readiness to evolve in response to changing requirements. It ensures that resources are allocated efficiently to meet the needs of data processing and user activity.

It is essential to mention that the post-project plan is for CINECA to keep the resources available for the EO4EU platform for two years after the project completion, while for WEkEO, if the required resources are minimised, the same 2-year post-project period will be ensured.

6 Conclusion

This document denoted as "D4.2 – Infrastructure & Services Requirements - Final", marks a critical milestone as the second deliverable from the efforts of Work Package 4 within the EO4EU project. It provides an in-depth update on multiple facets crucial to the continued successful establishment and functioning of the EO4EU platform.

This document addresses an essential aspect: the detailed update of the services and components constituting the EO4EU platform. Through meticulous analysis, the document offers a comprehensive understanding of the purposes each element serves within the broader framework of the project. Additionally, this deliverable presents the updated infrastructure requirements essential for the ongoing deployment of the platform's components and services. This foresight ensures a refined understanding of the technological prerequisites, setting the stage for subsequent phases of development.

The document presents a thorough overview of the available infrastructure earmarked for hosting the EO4EU platform, with particular attention to WEkEO and CINECA Cloud infrastructures, given that HPC will not be used at this stage. However, provisions for potential HPC integration are discussed, ensuring future readiness. The strategic allocation of these infrastructures influences the hosting strategy for various components or groups (Tiers) of components and services. The document further elucidates a well-defined methodology for achieving a multi-cloud infrastructure, encapsulating all available resources to ensure flexibility and efficiency in deploying EO4EU platform components and services.

Regarding hosting strategy, the document provides explicit details on the infrastructure chosen for each component or group, accompanied by the rationale behind these decisions. The Platform Controller and most other components are hosted on the CINECA Cloud, utilising a Kubernetes Multi-Cluster. The document also discusses the readiness to integrate HPC resources, ensuring optimal performance for critical Machine Learning/Inference Server services. To encapsulate the collective needs of all components, the document furnishes a summarised version of the updated infrastructure requirements, ensuring a balanced and efficient allocation of resources and setting the stage for seamless integration of services within the EO4EU platform.

Conclusively, the document not only updates the intricacies of infrastructure and service requirements but also strengthens the foundation for the forthcoming phases of the EO4EU project. The detailed insights into computational, memory, storage, and GPU capabilities presented in the concluding section serve as a valuable reference for evaluating and optimising the platform's performance.

7 References

[1] C. EO4EU, “D4.1 - Infrastructure & Services Definition (a),” EO4EU, 2023.

[2] C. EO4EU, “D2.4 - Technical, Operational and Interoperability, specifications and Architecture,” EO4EU, 2023.