Horizon 2020 Project LETHE
"A personalized prediction and intervention model for early detection and reduction of risk factors causing dementia, based on AI and distributed Machine Learning."

Research and Innovation Action
H2020-SC1-DTH-2020-1
GA 101017405

Duration: 48 months from 01/01/2021
Coordinator: Sten Hanke, FH JOANNEUM GESELLSCHAFT MBH

Deliverable ID.:	**D2.3**	
Deliverable title:	**LETHE architecture update**	
Planned delivery date:	30/06/2023 (M30)	
Actual delivery date:	03/07/2023 (M31)	
Editor:	Ville Tenhunen, EGI Foundation	
Contributing partners:	EGI, ER, FHJ, FORTH, INFO, MU	
Internal reviewer:	Nico Kaartinen, KAASA	
Checked and released by:	Sten Hanke (FHJ)	
Dissemination Level:	X	PU = Public;
		CO = Confidential
Type:	Report	

This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No. 101017405.

This deliverable reflects only the authors' view and the Commission is not responsible for any use that may be made of the information it contains.

Document Information and History

Deliverable description (from DoA)
This deliverable provides the first version of technical architecture defining all modules (as responsibilities) as well as interfaces (and responsibilities) for the validation trials.

Please refer to the Project Quality Handbook for guidance on the review process and the release numbering scheme to be used in the project.

Release number	Release date	Author [Person and Organisation]	Milestone*	Release description /changes made
0.1	5.4.2023	Ville Tenhunen (EGI)	Reviewed ToC	
0.2	13.6.2023	Ville Tenhunen (EGI)	Version for review	
0.9	29.6.2023	Ville Tenhunen (EGI)	Final version	

** The project uses a multi-stage internal review and release process, with defined milestones. Milestone names include abbreviations/terms as follows:*

- *TOC = "Table of Contents" (describes planned contents of different sections);*
- *Intermediate: Document is approximately 50% complete – review checkpoint;*
- *ER = "External Release" (i.e. to commission and reviewers);*
- *Proposed: document authors submit for internal review;*
- *Revised: document authors produce new version in response to internal reviewer comments*
- *Approved: Internal project reviewers accept the document.*

Table of Contents

Table of Figures

List of Tables

1 Executive Summary

The LETHE project will produce the platform to manage data for a personalised prediction and intervention model for early detection and reduction of risk factors causing dementia, based on AI and distributed Machine Learning.

The platform has a modular architecture which contains several components on raw data sources, data ingestion, unstructured data storage, data processing, structured data storage, AI/ML data model design, AI/ML ops and data presentation.

In the LETHE project the platform must follow strict data security and privacy procedures and practices because of the sensitive nature of the data. The LETHE Platform will be built on the EGI Federated Cloud solution which offers centralised services for data transfer, storage and processing including federated authentication and authorization methods. Deployment has designed the way where it is possible to move to another environment.

The platform architecture contains two middleware components. One for the sensor data and one for other data. Both will be integrated with a data lake, data processing services and finally as structured data in a data warehouse. The AI/ML core modules use data from the data warehouse for the AI/ML solutions which offer information to the risk factor identification and intervention solutions as dashboards, reports and other data products.

2 About This Document

2.1. Role of Deliverable

This deliverable is the final description of the LETHE architecture. This is part of the LETHE system documentation together with other related deliverables described in the chapter 2.2.

The deliverable describes the key components of the LETHE architecture. It gives a higher-level picture about LETHE applications and components and their integrations between each other.

2.2. Relationship to other LETHE deliverables

This deliverable is related to a few LETHE deliverables. The most important ones are:

- D2.2 LETHE architecture and scenarios: This deliverable is the previous version of the LETHE architecture description and deployment scenarios.
- D2.4 LETHE deployment setup and handbook: This deliverable describes the deployment activities following the proposed architecture and address differences in the LETHE architecture since D2.2.
- D4.3 LETHE digital interventional model: This deliverable refers to the implementation scheme regarding the LETHE Sensing Ecosystem.
- D5.3 Deployment of the LETHE infrastructure: In this deliverable, the LETHE project infrastructure and its components has been documented.

2.3. Major changes comparing the previous version of the deliverable

This deliverable focuses on the update of the most functional parts of the architecture of the LETHE i.e. the deliverable describes information architecture, reference architecture and system architecture. Instead, some chapters of the D2.2 have been skipped with summaries, mentions or references to the previous version because the content of those chapters are mostly unchanged.

These chapters are as follows:

In Section 3 “Introduction”

- 3.1. Conceptual Architecture Overview
- 3.2. Architecture Principles
- 3.3. Logical Architecture

In the section 4 “Rules, regulations and standards”

- 4.1 Regulatory Requirements
- 4.2. Security Policy
- 4.3. Standards

In the section 5 “Business Architecture”

- 5.1. Stakeholders
- 5.2. Use Cases

Application deployment issues are described in the deliverable D2.4 and therefore section 9 "Application Deployment Scenarios" of the D2.2 has skipped here totally.

2.4. Structure of the Document

This document focuses on the LETHE information and system architecture.

In the chapter 4 "Use cases and requirements" the deliverable summarizes use cases and present requirements of the LETHE system.

Chapter 5 "Information architecture" describes information architecture.

The chapter 6 "System architecture" draws a picture about the LETHE system by describing the LETHE system context, applications and their main components.

Chapter 7 "Infrastructure architecture" summarizes infrastructure elements of the LETHE system based on the deliverable D5.3.

2.5. Abbreviation

Abbreviations i.e. acronyms used in this document.

Table 1. Abbreviation

Abbreviation	Explanation
AAI	Authentication and Authorization Infrastructure
AI	Artificial intelligence
APA	The Activity Profiling Algorithm
CD	Continuous Deployment
CI	Continuous Integration
CPU	Central processing unit a.k.a. central processor
CTMS	Clinical Trial Management System
DNS	Domain Name System
DWH	Data warehouse
EOSC	European Open Science Cloud
EIF	European Interoperability Framework
ELK	Elasticsearch, Logstash, Kibana
FINGER	The Finnish Geriatric Intervention Study to Prevent Cognitive Impairment and Disability
GB	Giga Byte
GDPR	General Data Protection Regulation
GPU	Graphics processing unit

HA	High Availability
IaaS	Infrastructure as a Service
ICT	Information and Communication Technology
IdP	Identity Provider
kW	Kilowatt
mIoT	Mobile Internet of Things, here essentially mobile health devices
QoS	Quality of Service
RAM	Random Access Memory
RBAC	Role-Based Access Control
SLA	Service Level Agreement
SP	Service Provider
SSL	Secure Sockets Layer
UI	User Interface
UPS	Uninterruptible Power Supply
VM	Virtual Machine
VO	Virtual Organisation

2.6. Glossary

Glossary of terms used in this document in alphabetical order.

Table 2. Glossary

Term	Explanation
Data lake	A data lake is a system or repository of data stored in its natural/raw format, usually object blobs or files. It is possible to store data as-is and run different types of analytics: from dashboards and visualizations to big data processing, real-time analytics, and machine learning.
Metadata	Additional information related to single record data or to a data set that enables data discoverability and improves interoperability.
Data catalogue	Inventory of the available data to enable data discoverability.
Data warehouse	A data warehouse (DW) is a system used for reporting and data analysis and is considered a core component of business intelligence. DWs are central repositories of integrated data from one or more disparate sources.[1]
ELK Stack	“ELK” is the acronym for three open-source projects: Elasticsearch, Logstash, and Kibana. Elasticsearch is a search and analytics engine.

	Logstash is a server-side data processing pipeline that ingests data from multiple sources simultaneously, transforms it, and then sends it to a "stash" like Elasticsearch. Kibana lets users visualize data with charts and graphs in Elasticsearch[2]. The combination of these three tools is commonly referred to as “ELK Stack”.
Prospective data	User input data via wearables, mobile applications, mobile health devices (mIoT), surveys etc.
Retrospective data	Data from previous research projects and databases
Structured data store	Repository of structured data
Use case	A use case is a written description of how users will perform tasks on the IT systems.
Virtual organisation	A group of people (e.g. scientists, researchers) with common interests and requirements, who need to work collaboratively and/or share resources (e.g. data, software, expertise, CPU, storage space) regardless of geographical location. They join a VO in order to access resources to meet these needs, after agreeing to a set of rules and Policies that govern their access and security rights (to users, resources and data). [3]

Additional glossary is possible to find from the EGI pages: http://go.egi.eu/glossary

3 Introduction

The aim of the LETHE project is to establish a digital-enabled intervention for cognitive decline prevention based on the evolution of a successful protocol evolving into an ICT based preventive lifestyle intervention through individualized profiling, personalized recommendations, feedback and support, well targeted on a population stratified by cost-effective biological biomarkers. The LETHE solution will be tested in a feasibility study validating the achieved improvements.

LETHE will go beyond and provide a data-driven risk factor prediction model for older individuals at risk of cognitive decline, building upon big data analysis of cross-sectional observational and longitudinal intervention datasets from 4 clinical centres in Europe including the 11 years analysis of The Finnish Geriatric Intervention Study to Prevent Cognitive Impairment and Disability (FINGER).

The three major achievements expected in LETHE are:

1. A data-driven risk factor prediction model for older individuals at risk of cognitive decline building up on big data analysis of cross-sectional observational and longitudinal intervention datasets
2. Novel digital biomarkers, for early detection of risk factors, based on unobtrusive ICT-based passive and active monitoring
3. A digital enabled intervention for cognitive decline prevention based on the evolution of a successful protocol (FINGER) evolving into an ICT based preventive lifestyle intervention through individualized profiling, personalized recommendations, feedback and support (FINGER 2.0), well targeted on a population stratified by cost-effective biological biomarkers

The LETHE Platform has two main data sources:

- Data from previous research projects and databases (retrospective data, phase I)
- User input data via wearables, mobile applications, mobile health devices (mIoT), surveys etc. (prospective data, phase II)

The LETHE platform includes applications, services and integrations to fulfil goals of the LETHE project. In this deliverable these elements of the platform have described.

4 Use cases and requirements

This chapter summarises use cases and requirements described in the D2.2.

4.1. Use cases

As described in D2.2 LETHE stakeholders are:

- patients
- patients with dementia and caregivers
- healthy persons
- health care professionals
- scientists
- clinicians
- politicians and health care policy makers
- researchers on AI and machine learning
- researchers on Big Data frameworks and processing

Previously identified use cases are as follows:

1. User collects and delivers prospective data to the RADAR-base
2. User collects and delivers prospective data to the LETHE Middleware
3. Scientist stores retrospective data
4. Scientist receives reports from the dashboard
5. Developer develops an AI core module
6. The expert previews the results from the AI modules and adjusts the plan
7. The participant receives alerts and notifications from the system

These are presented in detail in the D2.2.

4.2. Functional requirements

Functional requirements specify what the system should do.

4.2.1 Licenses

Here licenses mean licenses of the LETHE Platform components

R1.1 Licenses
Licenses have to be included in the service or resource information as clearly as possible. Information has to contain a link to the exact license description. *essential*

4.2.2 Authentication and Authorization

On-demand platforms contain multiple authentication and authorization issues to be solved based on users' needs and requirements.

According to the type of resources, various authentication methods will be made available to the systems. This is used to access cloud resources; it is possible to use federated identity management systems based on standards such as OpenID Connect tokens.

R1.2 Authentication and authorization
For service provider authentication, EGI Check-In should be used. *essential*

R1.3 Local accounts
Local accounts (i.e username/password) to access resources will also be needed. *essential*

4.2.3 LETHE Platform

Following functionalities must be defined during the project with the best practices and based on agile service development methods i.e., detailed use cases and user stories must be described with product owner(s) to backlogs.

R1.4 The LETHE Platform
The platform covers at least following areas to develop: • Retrospective data storage and sharing • Data transferring services • Prospective data ingestion from user equipment and other sources • Data harmonisation services • Analysing and computing platform for developing the AI core module, model training and deployment • Platform for the visualizing and reporting solutions • Portal functionalities including discoverability and search functionalities • Collection and labelling of data • Solution integrations *essential*

4.2.4 Deployment Mechanism

The deployment mechanism is the action used to put a built application or resource into a platform where users can find and use it.

R1.5 Deployment mechanism
The LETHE Platform deployment mechanism must be well described.

essential

4.3. Non-Functional Requirements

Non-functional requirements specify how the system performs a certain function.

4.3.1 Usability

Standardization organisation ISO defines usability as "The extent to which a product can be used by specified users to achieve specific goals with effectiveness, efficiency, and satisfaction in a specified context of use."[4] In the case of web-based services this means merely “a small learning curve, easy content exploration, findability, task efficiency, user satisfaction, and automation”.[5]

R2.1 Usability
Usability of upcoming services must follow W3C Guidelines[6] *essential*

4.3.2 Accessibility

Web accessibility allows everyone, including people with disabilities, to perceive, understand, navigate and interact with the Internet.

R2.2 Accessibility
The LETHE Platform and vertical matchmaking user interfaces must consider regulations like The Web Accessibility Directive (Directive (EU) 2016/2102)[7] *essential*

4.3.3 Performance and QoS

Quality of service (QoS) is the description or measurement of the overall performance of a service, such as response time, throughput, utilization.

R2.3 Performance and QoS
During the project, based on users’ need and requirements, target level of performance and QoS have to be defined and measures which are needed taken. *essential*

4.3.4 Scalability

Scalability in network-available services, which might be defined as the ability of an application to handle growth efficiently, is typically achieved by making them available on multiple devices.

R2.4Scalability
The LETHE Platform components must be scalable enough to support uptake resources and services users' needs. *essential*

4.3.5 Availability and Capacity

Capacity means in this context that there are sufficient resources available to fulfil users' requests.

R2.5 SLA
Service level agreements (SLAs) will be negotiated with resource and service providers. *essential*

R2.6 Capacity and availability
The project must identify service performance requirements based on users' needs, plan the resources required to fulfil the requirements and ensure performance monitoring. *essential*

4.3.6 Virtual machine environments

R2.7 Deployment of virtual machine environments
Each VM should be created as a new server that can be used, either standalone or as part of a larger, cloud-based infrastructure. *essential*

4.3.7 Security and Data Protection

Data security and protection are regulated with multiple laws and other regulations.

R2.8 Security and data protection
Services must consider applicable regulations like General Data Protection Regulation (GDPR)[8]. Additionally, guidelines from the D6.1 *Data Management plan and ORDP I*[9] and T1.3 *Risk management* must be implemented also in these activities. *essential*

R2.9Secure transfer of sensitive data
All sensitive personal data which will be transferred to the cloud environment have to be pseudoanonymized or anonymized before transferring. All connections to the cloud environment must be strongly encrypted and secured. *essential*

R2.10 Processing the sensitive data
Sensitive data processing will take place only in the cloud environment. LETHE data processing in laptops, desktops or other portable devices is not allowed. *essential*

R2.11 Backups
Backups are automatically created disk images of VMs. Enabling backups for VMs enables system-level backups at weekly intervals, which provides a way to revert to an older state or create new VMs. *essential*

R2.12Firewalls
Cloud Firewalls are a network-based, stateful firewall service provided for the VM (Environment) access. Cloud firewalls block all traffic that isn't expressly permitted by rules. *essential*

R2.13 Data centre physical security requirements
Facility: • Tier 2. 99% availability. • 200 kW of power maximum in data centre of high density. • electric ring with UPS and group generator Security: • Specialized security personnel • Intelligent indoor and outdoor video surveillance system • Smoke detection system • Gas extinguishing system Operation: • Customized support backed by technical team of experts with presence onsite IT; no outsourcing.

• Following ITIL guidelines (not certification) *essential*

R2.14 Data security during testing
▪ The use of production databases in testing is not allowed. When using actual data that contain personal information, this is removed before using them, especially for systems that belong to the category of the outmost criticality. In addition, all necessary measures should be taken to ensure the confidentiality of commercially sensitive user data and market operators. ▪ Copy of actual data in application testing system must be based on specific process. ▪ The actual data are deleted from the test system applications immediately after completion of the tests. ▪ Actual data used from the production systems on test systems are recorded for possible future audits. *essential*

R2.15 Software approval
The following should be checked before setting a software to production: ▪ The software is not infected with viruses. ▪ It is compatible with other products used by the project. ▪ Can work in the environment for which it is intended and identifies the parameters that must be configured. ▪ Delivered and accompanied by relevant manuals. ▪ Fulfils the required functionality. ▪ Meets the security standards set by security policy. ▪ Checks and tests of the above, to detect and correct serious mistakes. The results are considered negative in terms of security, if the following applies: ▪ Detect serious errors in software. ▪ Cases arise where data do not correspond to the calculated estimated results (audit). ▪ The manuals or operating instructions are not available or are inadequate. ▪ The software documentation is not available or is insufficient. *essential*

4.3.8 Interoperability

R2.16Interoperability
Interoperability between different services and components must be defined. Interoperability principles must follow New EIF i.e., European Interoperability Framework by ISA2 programme principles.[10] *essential*

4.3.9 Standards and Architecture Framework

Architecture definitions have been described in this document.

R2.17 Architecture compliance
The LETHE Platform must follow standards and architecture definitions described in this document. *essential*

R2.18 Configuration management
Configuration management processes and procedures must be defined to provide and maintain a logical model of all configuration items and their relationships and dependencies. *essential*

R2.19 Change management
Change management processes and procedures must be defined to ensure changes to configuration items are planned, approved, implemented and reviewed in a controlled manner to avoid adverse impact of changes to services or the customer's receiving services. *essential*

5 Information architecture

5.1. Data Sources and Datasets

5.1.1 Retrospective Data

The retrospective knowledge base, which will be used to generate the initial prediction model, will be based on 4 different data sets provided by the clinical partners of the LETHE consortium.

These data sets will be harmonized and combined to a single, structured source of information.

The data originate from clinical studies, clinical observations and insurance companies, and consist of demographics, clinical markers, cognition-test-results and -markers, physical and functional status, health status, lifestyle information and information about mood and quality of life.

A detailed description of the data records and the harmonization process can be found in D2.1.

5.1.2 Prospective Data

Prospective data in LETHE will be collected via variable apps, tools and wearables. The sources of data can be separated as follows:

- Data from Smartphone Sensory system (passive data)
- Data from Wearables (passive data) + Cardio data (active and passive data)
- Meditation data is collected from the LETHE app
- Nutrition data is collected via questionnaires in LETHE app
- Mental health data collected from the LETHE app (active data)
- Additional behavioural modules as well as smart glasses and social robots (active and passive data)

Additionally, the LETHE app contains a calendar which is used for events related to the LETHE study (study visits, online meetings).

A detailed overview of all collected data from the sensing ecosystem is described in D4.3.

The retrospective data model and variables defined in the deliverable D2.1 “Structured data repository”. An initial description of all used standards within the project as well as the interoperability framework are described in D2.5. During the project these models and interoperability aspects evolve because requirements of the project.

6 System architecture

6.1. Overview and method

The LETHE platform architecture contains two separate systems; Phase I for developers and data providers data sharing and Phase II for trials with participants, clinicians and applications produced within the project.

In this deliverable the C4 model is used to describe both phases.

The C4 model is an "abstraction-first" approach to diagramming software architecture, based upon abstractions that reflect how software architects and developers think about and build software[11].

In the C4 model viewpoints are organized to hierarchical levels:

1. Context diagrams: show the system in scope and its relationship with users and other systems;
2. Container diagrams: decompose a system into interrelated containers. A container represents an application or a data store;
3. Component diagrams: decompose containers into interrelated components, and relate the components to other containers or other systems;
4. Code diagrams: provide additional details about the design of the architectural elements that can be mapped to code.

Because level 2 term “container” can be misunderstood, this document uses the term “High-level technical building blocks”.

For level 1 to 3, the C4 model uses 5 basic diagramming elements: persons, software systems, containers, components and relationships[12].

Levels 3 and 4 are not used in this deliverable because it is recommended to use only those diagrams which add value[13]. Level 3 information is described with the container specifications where detailed explanation of LETHE Platform applications are given. Level 4 descriptions are useful for developers and those change numerous times during the project and maintenance phase.

6.2. Phase I: System context (level 1)

Following context picture describe the LETHE Phase I services. Basic idea of the phase is sharing data between researchers and developers who are using FINGER data to the upcoming AI models.

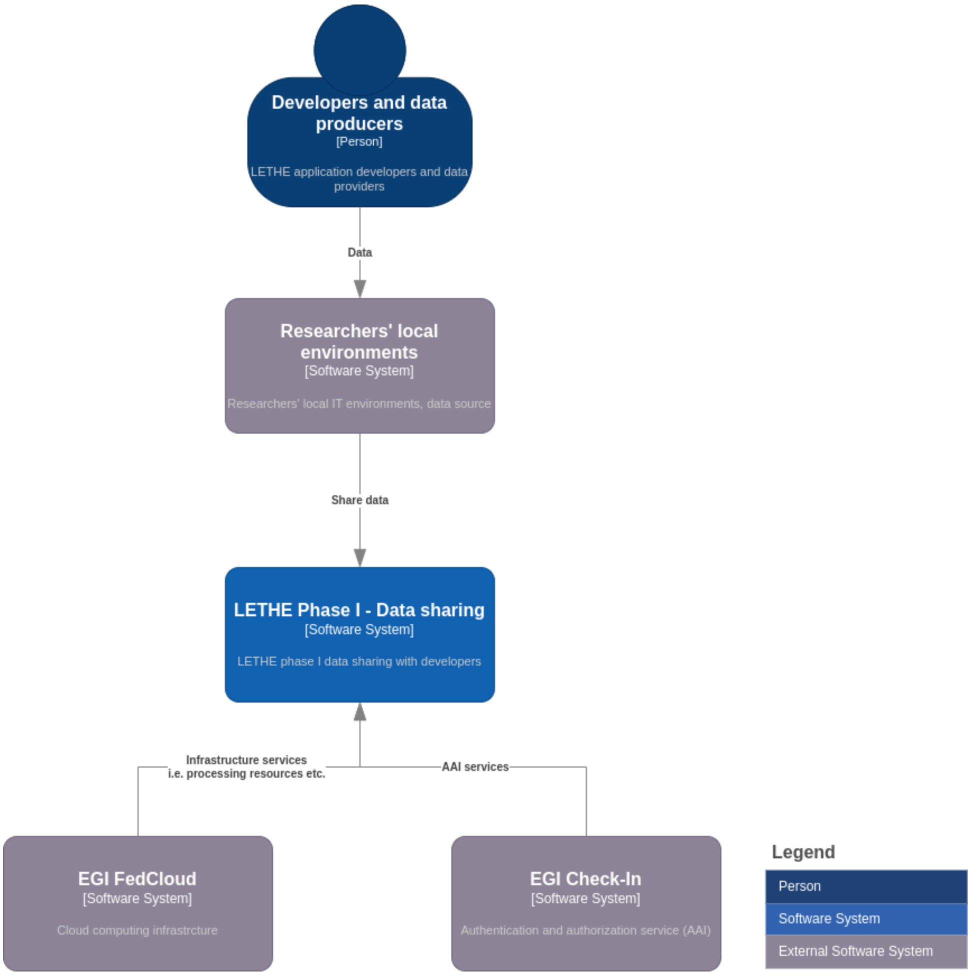

Figure 1. LETHE Phase I - Context

Given the sensitive nature of the datasets manipulated, an encrypted central storage is provided by the EGI Federated Cloud. The secure transfer of the pseudonymised datasets is ensured by SSL transfers provided by EGI Datahub service using web interfaces or 'lite' clients at sources (Data providers).

Virtual Machines (VM) are deployed via the EGI Cloud Compute service to access and process the data in the storage.

All individuals involved in depositing, accessing and processing the data are registered within the EGI Check-in service (based on their IdPs) to access and use the above mentioned EGI services in uniform and easy way.

6.2.1 Local environments

Before data pseudonymization, data locates researchers' local environments. Data is processed on local research environment based on principles described in chapter "Information architecture" of this deliverable. Architecture of these environments follows local regulations and deployments.

All network connections from local environments to the LETHE Platform are encrypted and secured with local and cloud service measures.

6.2.2 EGI FedCloud

The EGI Federated Cloud Platform (FedCloud)[14] is a federated cloud infrastructure. It is an IaaS-type cloud consisting of academic, private clouds and virtualised resources built around open standards. The requirements of the scientific community drive its development.[15]

For the LETHE Phase I the EGI FedCloud offers all needed cloud infrastructure services.

More detailed description will be included in the chapter "Infrastructure architecture" of this deliverable.

6.2.3 EGI Check-In

EGI Check-in[16] is a proxy service that operates as a central hub to connect federated Identity Providers (IdPs) with EGI service providers. Check-in allows users to select their preferred IdP so that they can access and use EGI services in a uniform and easy way.

This means in LETHE Phase I that clinical partners and developers who are sharing or using data in the Phase I can access to the platform with their own institution's credentials.

More detailed description will be included in the chapter "Infrastructure architecture" of this deliverable.

6.3. Phase I: High-level technical building blocks (level 2)

High-level technical building-blocks figure opens the software system "LETHE Phase I – Data sharing" block of the Level 1 figure above. In practice, this specified which EGI FedCloud services are in use in the LETHE Project Phase I.

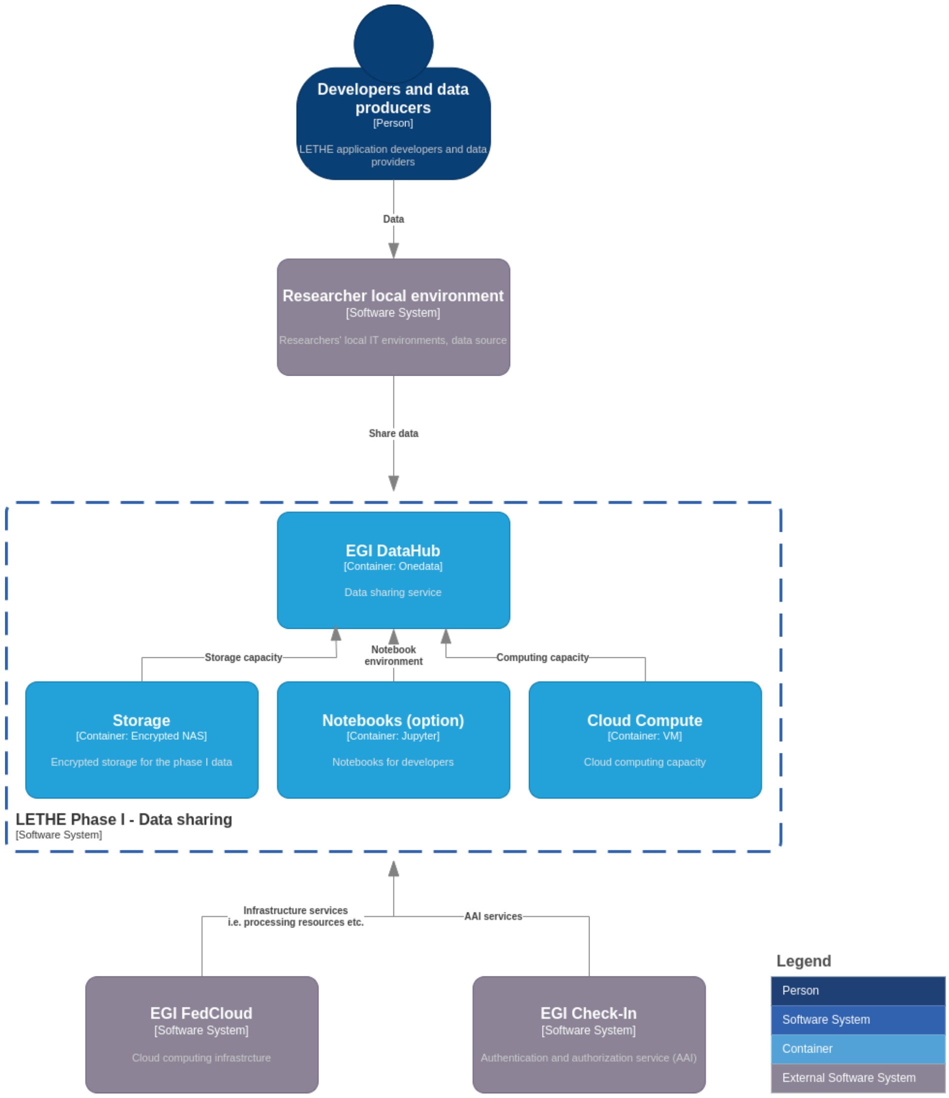

Figure 2. LETHE Phase I - Building bloks

6.3.1 EGI DataHub

EGI DataHub[17] is the service which brings data close to computing to exploit it efficiently, and to publish a dataset and make it available to a specific community, or worldwide, across federated sites. EGI DataHub is based on the Onedata[18] technology.

Security functionalities of the EGI DataHub has tailored for the LETHE project because the nature of the data used in the project phase II. This includes also encrypted storage volumes in the EGI's service provider facilities. The secure transfer of the pseudonymised datasets is ensured by SSL transfers.

6.3.2 Storage

Storage attached to the EGI DataHub services is an encrypted NAS storage from the EGI Online Storage[19] service.

6.3.3 Notebooks

Also, EGI Notebooks are available for specific tasks. Notebooks is a browser-based tool for interactive analysis of data. Notebooks is based on the JupyterHub technology and supports multiple programming languages including Julia[20], Python[21], R[22], GNU Octave[23] and MATLAB[24]. In the LETHE project Python is mainly used.

6.3.4 Cloud Compute

Cloud computing resources the project get from SCIGNE Cloud Compute[25] which is federated with EGI FedCloud i.e. FedCloud delivers SCIGNE Cloud Computing service for the project users. This arrangement gives opportunity to deploy compute servers with the operating system.

6.4. Phase II: System context (level 1)

Following figure describes a LETHE Platform and its context including relation with users and other systems.

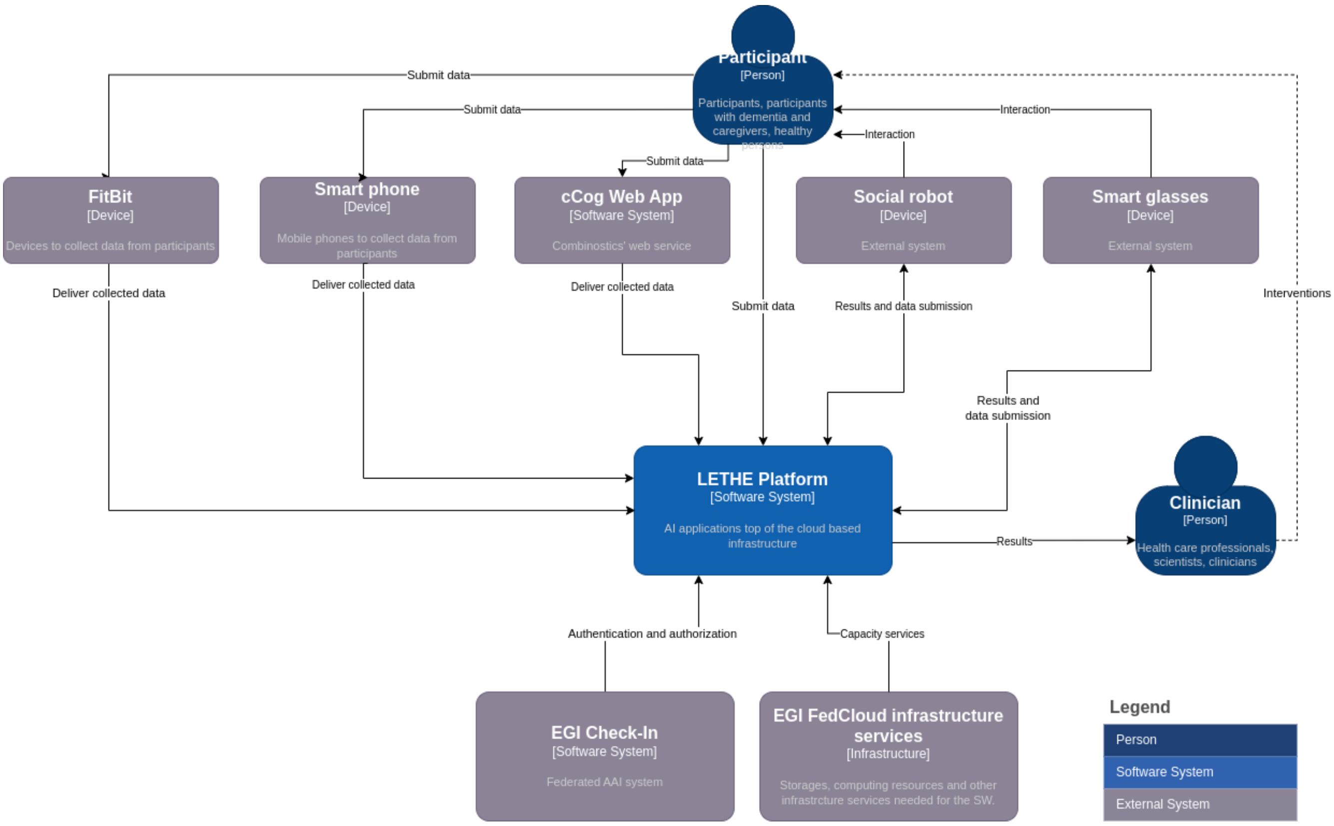

Figure 3. LETHE Phase II - Context

LETHE Platform has two major user groups; trial participants and clinicians.

In this diagram “participant” covers stakeholder groups participants, participants with dementia and caregivers and healthy persons. Respectively “clinician” covers groups health care professionals, scientists, clinicians.

The LETHE Platform has relation following external systems or applications:

6.4.1 FitBit Smartwatches and Fitbit App

Fitbit Sense Smartwatches[26] are used to collect physical activity data, cardio data and sleep data. Normally, Fitbit data is generated from the participant, transmitted to the Fitbit Data Warehouse and subsequently collected from RADAR-base. The data, upon acquisition is sent from the RADAR-base local storage to the LETHE Data Lake.

6.4.2 Smart Phone

The pRMT app[27] (RADAR-base Passive Remote Monitoring App) is a complementary phone application that is installed on the participants' phones, with the explicit purpose of periodically harvesting phone sensor data. The same procedure is followed, upon collection of phone sensor data, they are stored in the RADAR-base local storage, and subsequently submitted to the LETHE Data Lake.

6.4.3 Smart Glasses

In the LETHE project the voice app and Bose FRAMES audio glasses has used. These are audio glasses with open-ear technology specifically designed for exercise use. They allow handsfree interaction with the LETHE app without cancelling noise, thus allowing for safe outdoor use.

Bluetooth Glasses Device is a receiver and recorder of audio from the companion Glasses App (Android). Glasses App receives audio from the glasses and calls our own API for text analysis. Once the analysis is performed, or in any case a message must be delivered to the user, calls another API to generate speech from text.

Glasses use cloud infrastructure developed in Google Cloud, as Google Cloud Functions (serverless). The API is designed to query also the LETHE DWH and receives event data to engage actions.

When data analysis is done, before sending back analysis results along with synthetized text to the Glasses App, data is stored in structured form in the LETHE DWH. Data stored in a structured form can be considered equivalent to "answers" to questionnaires, so the same level of data protection established in the LETHE protocol for questionnaires is applied.

6.4.4 cCog Web App

cCog[28] is a web-based cognitive test tool for detecting neurodegenerative disorders by Combinostics. cCog web app is used in the LETHE project to collect participant data and deliver it to LETHE System for next level analysis.

6.4.5 Social robot

The Temi robot is a fully autonomous personal robot with advanced capabilities such as voice recognition, natural language processing, and sensors. It is customized for the LETHE project to suit the needs and goals of individual users and can provide personalized support and motivation to help individuals make sustainable positive changes in their lives. The complementary LETHE Robot App is used together with the Temi Robot to interact with a selection of the features.

The robot will remind the participant of daily events, will ask them to give feedback when they are on the “red path” of involvement or if you want to complete the cognitive training exercises. The robot will do this on its own or the user may do it at any given time.

In addition to that the robot will assist the user while doing their exercises by showing them the according training video on the robot’s screen while they are doing their exercises.

The robot application reads data from the Data warehouse (e.g. daily event data, being on the red path etc.) and writes data into the Data Lake (robot specific error and status data).

6.4.6 EGI Check-In

EGI Check-In is an authentication and authorization solution for the LETHE Platform applications. It is described in the details next chapter of the deliverable (Infrastructure architecture).

6.4.7 EGI FedCloud infrastructure

The EGI Federated Cloud Platform (FedCloud) is a federated cloud infrastructure. It is an IaaS-type cloud consisting of academic, private clouds and virtualised resources built around open standards. EGI FedCloud is describe in details next chapter of the deliverable (Infrastructure architecture).

6.5. Phase II: High-level technical building blocks (level 2)

Next level description is container diagram which represents an application and data storages in the LETHE Platform.

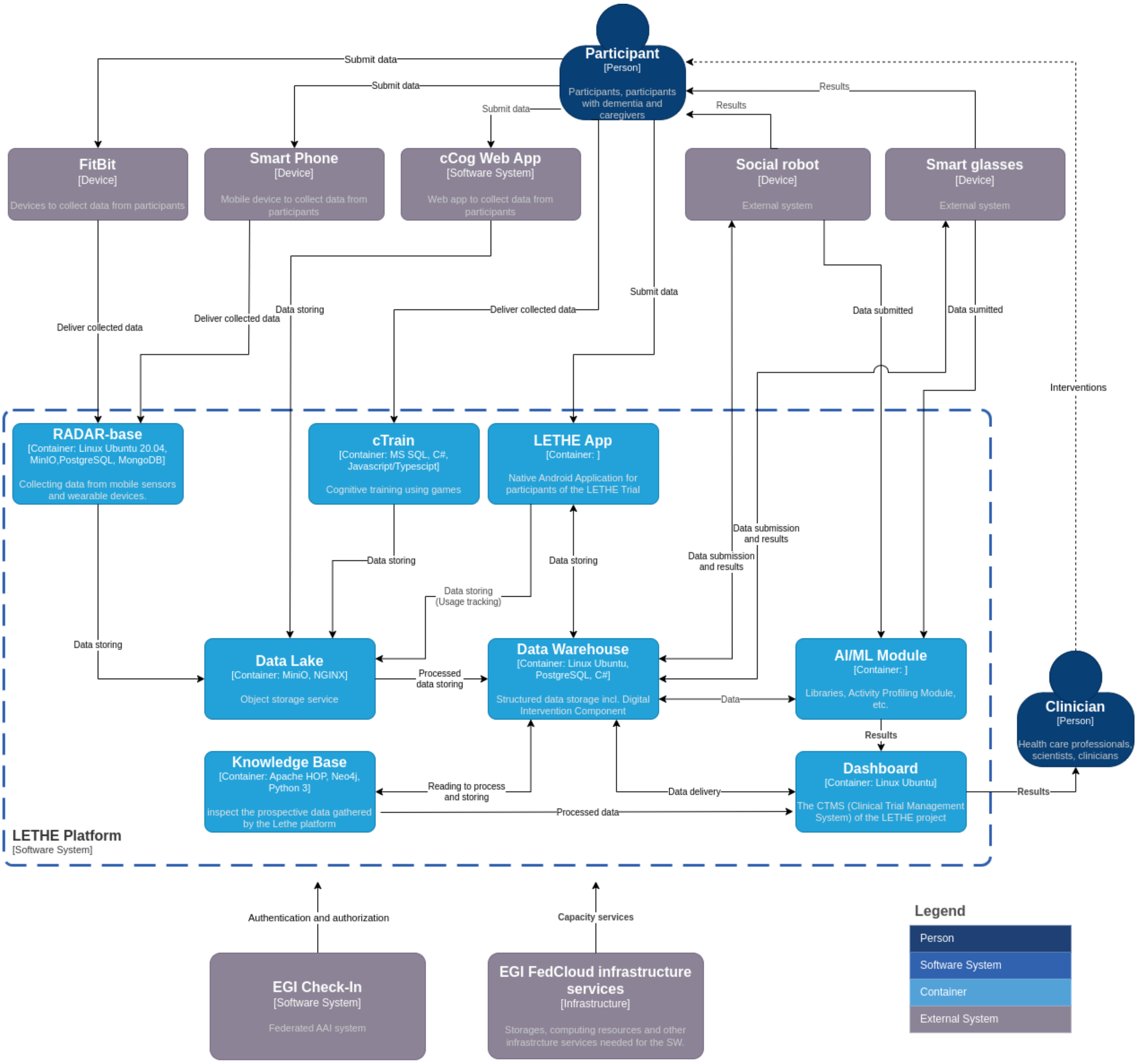

Figure 4 . LETHE Phase II - Building blocks

6.5.1 RADAR-base

RADAR-base is an end-to-end platform that allows for the enrolment and monitoring of participants in remote clinical trials by collecting data from mobile sensors and wearable devices such as smartwatches. LETHE project uses Kubernetes version of the RADAR-base.

Core components:

- Storage node
- Storage resource manager
- Kubernetes cluster

- DNS Server
- Object storage

In the LETHE project RADAR-base get data from FitBit devices and smart phones. Data is stored to the Data Lake.

6.5.2 cTrain

In cTRAIN, users can do cognitive training using six different games. These games contain multiple difficulty levels and after training the users can track their progress through score chart.

Core components are:

- Azure web service
- Azure blob storage
- Azure MS SQL database storage
- Data Lake

cTrain gets it data from users of the LETHE platform. cTrain deliver the data to the Data Lake.

6.5.3 LETHE App

The LETHE App offers a range of essential features and serves as the companion application for the LETHE trial. Once the onboarding process is completed, participants gain access to initial questionnaires, which are conveniently accessible within the app. Following randomization in the trial, additional sections are made available to participants, including emotion tracking, an integrated calendar, a task list featuring a dedicated Tiny Habits section, weekly reports, intervention-specific videos, and a diary section.

LETHE App is a native Android Application for participants of the LETHE Trial.

LETHE App submit to Data Warehouse and Data Lake. LETHE App retrieve data from the Data Warehouse.

6.5.4 Data Lake

Data Lake provide the object storage service for the LETHE platform. Accept data in CSV and JSON format. To send data to the Data Lake, it is mandatory to use the appropriate methods provided by the Data Lake API service.

Core components of the Data Lake:

- MinIO Server
- MinIO mc
- MinIO Console
- MinIO SDK Java
- NGINX

Data Lake receives data from RADAR-base, cCog Web App, cTrain and LETHE App through the Data Lake API. Data Lake delivers data to the Data Warehouse via *Data Lake to DWH* component.

Data Lake API provides a secure interface to send data to the Data Lake. Any received request is validated for sender authentication and authorization. Data Lake API is RESTful API and its components are:

- Spring Boot
- Java 11 framework
- Apache Camel

Data Lake to DWH gathers data from Data Lake, prepares it and send it to the Data Warehouse using the interface provided by the Data Warehouse itself. Its components are:

- Spring Boot
- Java 11 framework
- Apache Camel
- ELK Stack

6.5.5 Data Warehouse

Data Warehouse provides structured data storage for the LETHE project through a modern, distributed, load balanced, and secure API surface that is connected to a highly available relational database system. The DWH services the LETHE Dashboard and by extension their respective users.

In addition, it provides data to the Digital Intervention Component required for said service calculations and a storage platform to save the results. Periodically, Digital Intervention Component checks DWH data, aggregates, compares, and writes data right back to the DWH.

Lastly, it provides a structured data storage location for all the services such a RADAR-base that use the Data Lake as a proxy to save their respective data for the LETHE project.

Core components of the Data Warehouse are:

- Kubernetes Cluster - Load Balancer Hosting 4 API nodes
- 2 Virtual Machines - Primary and Secondary HA Databases
- Persisted storage volumes - Databases and backups
- GitLab Container Registry
- Azure Monitor - API Monitoring
- Azure Dev Ops - Source Control

Data Warehouse serves several LETHE applications; Data Lake, LETHE App, Notifications, AI/ML Module, Knowledge Base and Social Robot, Smart Glasses and Dashboard.

6.5.6 Knowledge Base

Knowledge Base provides the ability to inspect the prospective data gathered by the LETHE platform, related to the participants of the pilot. Knowledge Base provides two different UI: a more complex one, that offers the most dynamic way possible to analyse the collected data; a more friendly one, less dynamic but able to show data very effectively.

Core components of the Knowledge Base are

- Apache HOP
- Neo4J
- NeoDash

Knowledge Base is part of the feature engineering process. The extracted feature will be stored in the Data Warehouse.

6.5.7 AI/ML Module

AI/ML Module contain several services. AI core modules will be made with Python. TensorFlow[29] and Keras[30] libraries will be also used. MLflow[31], Jupyter notebooks[32] and other solutions will be in the EGI FedCloud[33] and CPU/GPU VMs from the computing cloud.

Prediction Service is one of the components of the AI/ML Module. It provides the result returned by the application of any AI prediction model registered in the service. The models are the result of the AI research activities carried out within the Lethe Project itself.

Explainability Service provides information related to the AI models explainability. As for the Prediction Service, any model must be registered in the service.

The core components of both Prediction Service and Explainability Service are:

- Scikit-learn 1.0.2, Flask 2.2.2, Pandas 1.5.1 as the AI framework
- Python 3
- MLServer

The Activity Profiling Algorithm (APA) will be utilised to analyse the target group daily living activities. Through behavioural profiling of the study participants, and by evaluating them at various time intervals, the algorithm will detect and reveal abnormal structures and patterns of the study participants behaviour.

APA Module is a Python application that reads and writes data from/to the DWH. APA module will be made with Python, TensorFlow, Keras and PyTorch libraries will be also used. MLflow, Jupyter notebooks and other solutions will be in the EGI FedCloud and CPU/GPU VMs from the computing cloud.

Its core components are:

- Python 3
- Pandas, TensorFlow, PyTorch, Keras, Scikit learn as the AI frameworks / libraries
- Uvicorn FastAPI[34] – Model Serving

The unification of AI module with the APA will be decided upon performance analysis and workload balance of the models. In case all models can be served at the already allocated resources and the chosen technology stack aligns, then merging of the models will effectively alleviate overhead and will be integrated.

In any other case, that it is deemed that further effort will be required without the respective overhead/performance gain, the APA will be deployed as a different component.

6.5.8 Dashboard

The CTMS (Clinical Trial Management System) of the LETHE project, short LETHE Dashboard, provides an overview of participants who have completed the onboarding process with the LETHE app and participants who have dropped out of the study. Other pages include the detail of each page as well as the eCRF where the visits are documented. Users of the Dashboard are clinicians as well as researchers.

Core components of the Dashboard are

- Java for the Android Application
- Google Playstore (app has published in the store)

Other technical components used for the application are

- React
- Docker

Dashboard uses data from Data Warehouse.

7 Infrastructure architecture

7.1. Infrastructure Architecture Overview

As described in the D2.4 “LETHE deployment setup and handbook“ and D5.2 "Deployment of the LETHE infrastructure" the LETHE infrastructure is based on two main concepts: Cloud Computing and Infrastructure-as-Code. This means that the LETHE platform is implemented in a federated cloud computing environment, EGI FedCloud, and the platform uses a virtual infrastructure that can be defined in software instructions.

In the LETHE platform infrastructure, two management systems are in use:

- The Infrastructure Manager (IM), which is a free and open-source application to assist users with the creation of infrastructure in a cloud environment and the deployment of software.
- Kubernetes, which is an open-source container management system that assists developers in deploying software applications.

The infrastructure of LETHE is deployed in one of the providers of the EGI Federation, which uses OpenStack as its cloud technology. Among other things, there are three main resources that are managed through the OpenStack platform:

- Computing power
- Storage
- Networking

The LETHE project is organised in two phases:

- Phase I is the model generation, where preparatory work is carried out to harmonise and optimise existing data and design a prediction mode.
- Phase II is the intervention implementation and includes the validation and feasibility trials.

Both phases have different needs in terms of infrastructure.

Phase I requires two virtual machines with 16 CPUs, 16GB RAM and 700GB storage in total.

Phase II needs a much bigger computing environment with 122 vCPUs, 488GB RAM and approximately 16TB storage. Additionally, Phase II environment can offer also GPU resources to developers. Capacity requirements might change during the project when applications will be updated.

7.2. Authentication and Authorisation

The authentication and authorisation need of the LETHE platform are fulfilled by the EGI Check-in service. EGI Check-in is an identity management system that provides a full AAI solution, it operates as an authentication proxy and an authorisation management system that combines and releases user attributes from different sources. Check-in enables access to EGI services and resources using federated authentication mechanisms. Specifically, the proxy service is operated as a central hub between Identity

Providers (IdPs), such as Google, ORCID, LinkedIn and those adhered to eduGAIN[35] , amongst others, and Service Providers (SPs) establish and maintain a technical and trust relation only to a single entity, the EGI Check-in, instead of managing many-to-many relationships. In this context, the proxy acts as a service provider towards the identity providers and as an identity provider towards the service providers.

Through the EGI Check-in, users can authenticate with the credentials provided by the IdP of their home organisation (e.g. via eduGAIN), as well as using social identity providers, or other selected external identity providers (support for eGOV IDs is also foreseen). To achieve this, the EGI Check-in supports a wide range of standards and open technologies, including OpenID Connect, OAuth 2.0, SAML 2.0 and X.509, which facilitates interoperability and integration with existing AAI services. In addition to serving as an authentication proxy, the EGI Check-in provides a central Discovery Service (Where Are You From – WAYF) for users to select their preferred IdP.

Authorisation in the LETHE platform is implemented by a Virtual Organisation (VO). In simple terms, a Virtual Organisation is just a group of users that represent a community. EGI VOs are created to group researchers who aim to share resources across the EGI Federation to achieve a common goal as part of a scientific collaboration. VOs are also the mechanism established to grant users access to the resources of the EGI Federation. Users must follow enrolment workflows to join a VO, whose request must be approved or rejected by a VO Manager. Inside a VO, users can be part of different groups and take different roles, according to their permissions in the different applications of the platform.

For prospective data gathering and ingestion via RADAR-base and LETHE Middleware, EGI Check-in is also used to end user authentication and authorisation.

7.3. Deployment

The deployment plan has been described in the D2.4 "LETHE deployment setup and handbook"[36]. As described in the document, the deployment process has been divided into two parts:

- The infrastructure. Instead of using traditional physical server computers, LETHE is deployed on a virtual infrastructure, using cloud technology. Herein, the hardware components are allocated on demand and its deployment is designed, managed and documented in software templates.
- The LETHE software components. Each module of the LETHE platform installs specific software for its functioning. Like the hardware allocation, the required installation of software is fully automated and managed via software templates.

The LETHE components deployed on the cloud of the EGI Federation (EGI FedCloud) fully support the CI/CD framework principles throughout the lifecycle of the project development until its final availability for use.

The infrastructure of the LETHE project has been built and deployed with a set of services and tools based on the European Open Science Cloud (EOSC), such as the Infrastructure Manager[37]. These tools facilitate the rapid adaptation needed in the LETHE infrastructure to provide the necessary computing resources as the project evolves. For LETHE platform developers, LETHE's infrastructure basically follows an Infrastructure-as-a-Service (IaaS) model, where they have on-demand access to cloud-hosted servers and storage.

8 Conclusion

This document describes the basic architecture of the LETHE Platform. Objective of the architecture is to specify a platform for applications and processes. This deliverable describes the architecture produced by several project WPs for LETHE trials.

This architecture is based on interoperable modules and applications. Architecture described in the deliverable has been produced for production mode i.e., even if architecture changes are possible, they are not expected during trials. Based on the ICT best practices and information security requirements continuous maintenance operations will take their place.

LETHE architecture contains modular applications with their own technology stack. Interoperable APIs make integrations between applications possible.

Because of the sensitive nature of the LETHE data information security measures and requirements have to be on high level.

In addition to the applications presented in the architecture being installed for the project trials, this can be used as a basis for continuous integration and LETHE application development in the future.

9 References

[1] https://en.wikipedia.org/wiki/Data_warehouse

[2] https://www.elastic.co/what-is/elk-stack

[3] EGI Glossary https://ims.egi.eu/display/EGIG/Virtual+organisation

[4] https://en.wikipedia.org/wiki/Usability

[5] https://en.wikipedia.org/wiki/Web_usability

[6] W3C Guidelines, https://www.w3.org/WAI/test-evaluate/

[7] https://digital-strategy.ec.europa.eu/en/policies/web-accessibility

[8] https://gdpr.eu/

[9] D6.1 Data Management plan and ORDP I

https://teams.microsoft.com/l/file/769644B8-3B97-437F-8FA0-D3D3DF359178?tenantId=714ce493-9d00-45e9-9d52-89dfa568262b&fileType=pdf&objectUrl=https%3A%2F%2Ffhjoanneum.sharepoint.com%2Fsites%2FGEBLETHE%2FFreigegebene%20Dokumente%2FWP6%20-%20Data%20Protection%20and%20Ethical%20Impact%2FLETHE%20D6.1.pdf&baseUrl=https%3A%2F%2Ffhjoanneum.sharepoint.com%2Fsites%2FGEBLETHE&serviceName=teams&threadId=19:19ff423f14ca423e9b13c56d7153b3fc@thread.tacv2&groupId=aed7df51-4a0b-456e-b10e-7dccce589409

[10] https://ec.europa.eu/isa2/eif_en

[11] C4 model: https://c4model.com

[12] https://en.wikipedia.org/wiki/C4_model

[13] https://c4model.com

[14] https://www.egi.eu/tab/federated-cloud/

[15] https://www.egi.eu/egi-infrastructure/

[16] https://www.egi.eu/service/check-in/

[17] https://www.egi.eu/service/datahub/

[18] https://onedata.org/

[19] https://www.egi.eu/service/online-storage/

[20] https://julialang.org/

[21] https://www.python.org/

[22] https://www.r-project.org/

[23] https://octave.org/

[24] https://www.mathworks.com/products/matlab.html

[25] https://grand-est.fr/en/services/on-demand-servers-and-containers/

[26] https://www.fitbit.com/global/us/products/smartwatches/sense

[27] https://radar-base.org/docs/prmt/

[28] https://www.combinostics.com/2020/08/01/ccog-a-web-based-cognitive-test-tool/

[29] https://www.tensorflow.org/

[30] https://keras.io/

[31] https://mlflow.org/

[32] https://www.egi.eu/service/notebooks/

[33] https://www.egi.eu/egi-infrastructure/

[34] https://fastapi.tiangolo.com/

[35] https://edugain.org/

[36] D2.4 "LETHE deployment setup and handbook"

[37] M. Caballer, I. Blanquer, G. Moltó, and C. de Alfonso, “Dynamic Management of Virtual Infrastructures,” *J. Grid Comput.*, vol. 13, no. 1, pp. 53–70, Mar. 2015, doi: 10.1007/s10723-014-9296-5.