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.:	**D2.2**	
Deliverable title:	**LETHE architecture and scenarios**	
Planned delivery date:	31/12/2021 (M12)	
Actual delivery date:	31/12/2021 (M12)	
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.

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	10.11.2021	Ville Tenhunen, EGI Foundation	Proposed	
0.2	13.11.2021	Nico Kaartinen, KAASA	Reviewed	
0.3	30.11.2021	Ville Tenhunen, EGI Foundation	Revised	
0.4	12.12.2021	Nico Kaartinen, KAASA	Reviewed	

** 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 personalized 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.

This LETHE 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 FedCloud solution which offer centralized services for data transfer, storage and processing including federated authentication and authorization methods.

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.

Deployment of the LETHE Platform must be done in iterative processes in several WPs. Decisions about middleware component roles and responsibilities are needed. This architecture and deployment scenarios have to be updated during the project.

2 About This Document

This deliverable provides the first version of the technical architecture defining all modules (as responsibilities) as well as interfaces (and responsibilities) for the validation trials for the LETHE platform.

2.1. Role of Deliverable

This document defines and specifies an architecture of the LETHE Platform which will serve this solution and produces ICT services for the whole lifecycle of the LETHE.

This deliverable is the basis for the LETHE Platform development in terms of the architecture and deployments scenarios. The deliverable defines all necessary components of the architecture and specifies detailed roles of components.

This deliverable is the first version of the architecture definition and will be updated with D2.3.

2.2. Structure of the Document

The structure of this document follows common enterprise architecture definition structures which has architecture principles, user requirements, business architecture, information architecture, service and application architecture, system and technology architecture and finally deployment and maintenance architecture described. All these aspects of the enterprise architecture have not been fulfilled perfectly but the document has selected those parts which are essential for the LETHE Platform development.

Section 3 is an introduction to this document and the LETHE Platform architecture including the overall concept of the platform. Following sections contain detailed descriptions for rules, regulations and standards (Section 4), business architecture (Section 5), information architecture (Section 6), service architecture (Section 7, infrastructure architecture (Section 8) and application deployment scenarios (Section 9).

For the technical development of the LETHE Platform the most important sections are 7-9 which describe also applications and their running environments but also deployment scenarios. For example, Section 5 and 6 support this purpose and contain mainly references to the resources which has already been described somewhere else.

2.3. Abbreviation and Glossary

2.3.1 Abbreviation

Abbreviations i.e. acronyms used in this document.

Table 1. Abbreviation

Abbreviation	Explanation
AAI	Authentication and Authorization Infrastructure
AI	Artificial intelligence
CMDB	Configuration Management Data Base
COU	Collaborative Organisation Unit

DOI	Data Object Identifier
DPIA	Data Protection Impact Assessment
DW	Data warehouse
EIF	European Interoperability Framework
FINGER	The Finnish Geriatric Intervention Study to Prevent Cognitive Impairment and Disability
GDPR	General Data Protection Regulation
IdM	Identity Management
IdP	Identity Provider
mIoT	Mobile Internet of Things, here essentially mobile health devices
PIA	Privacy Impact Assessment
PITR	Point-In-Time Recovery
POC	Person of Concern
QoS	Quality of Service
RBAC	Role-Based Access Control
SOP	System Operating Procedure
SOPs	Standard Operating Procedures
SP	Service Provider
URI	Uniform Resource Identifier
VM	Virtual Machine

2.3.2 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. A data lake is usually a single store of data including raw copies of source system data, sensor data, social data etc., and transformed data used for tasks such as reporting, visualization, advanced analytics and machine learning.[1]
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 mart	Subset of data from data warehouse provided to satisfy a specify business analysis requirement.

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.[2]
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]
Cold storage	A storage solution where data is accessed less frequently or rarely and can provide even slower accesses.

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

3 Introduction

3.1. Conceptual Architecture Overview

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

A simplified concept of the LETHE Platform is presented in the following picture.

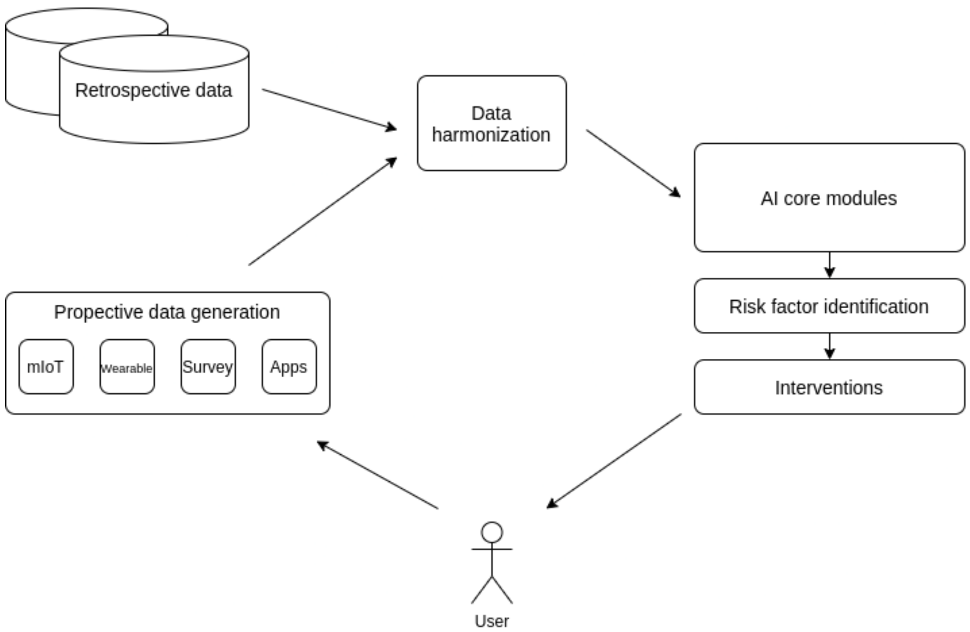

Figure 1. Simplified concept of the LETHE

Generally, the LETHE Platform has two main data sources:

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

This data from different sources has to be harmonized for the AI Core Modules. Feature selection applies for regression ML modules, otherwise deep learning models will be used. These create reports, visualizations etc. which are used for the risk factor identification. Interventions or other measures with users are followed by these processes based on data managed on the LETHE Platform.

Within the LETHE project framework architecture components and their relations with the project work packages are presented in figure 2.

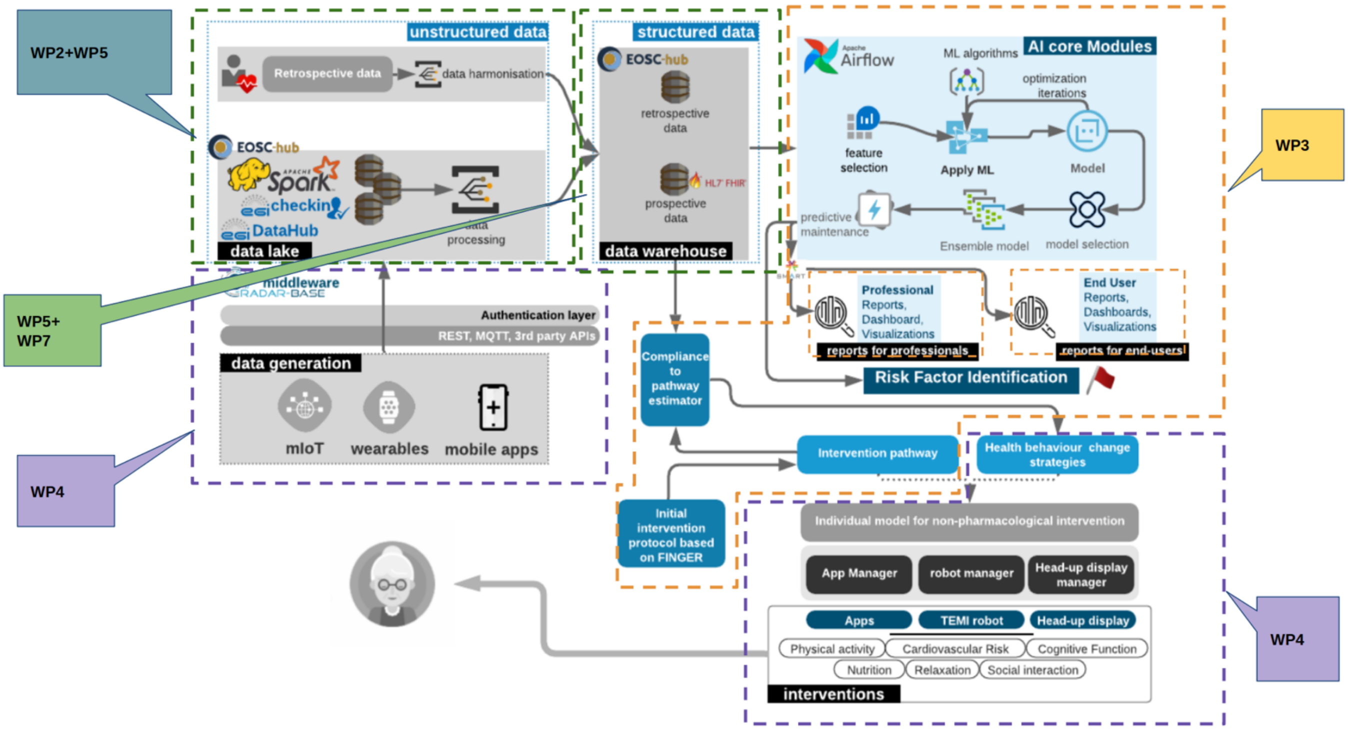

Figure 2. The conceptual architecture of the LETHE platform

3.2. Architecture Principles

The principles governing the LETHE logical architecture are:

- Scalability: All components of the LETHE system will scale in order to support growing user demand and business requirements;
- Availability: Increased system availability with fully redundant network and server infrastructure eliminating single points of failure;
- Accessibility: Systems and component have to follow regulations, guidelines and best practises as described later;

- Security: Software and hardware infrastructure will provide an end-to-end security model that protects data from malicious attacks or theft;
- Manageability: Ease of configuration, ongoing health monitoring, and failure detection are vital to the goals of availability, scalability, and security. The ability to manage the solution will also match future growth of the environment;
- Agility: The infrastructure will have the ability to quickly adapt to changes based on business needs. Agility is an essential aspect of maximising the benefit of IT to the business and it is considered of outmost importance for the system environments;
- Alignment with business needs: The infrastructure will meet the customer requirements for the system. Thus, certain technical choices will be driven by specific project needs and business practices.

3.3. Logical Architecture

The LETHE logical architecture defines and presents how the LETHE system and its components will be logically deployed and setup at different tiers, the off the shelves' tools to be used for design and implementation at each tier and in turn the applications to be built to enable the integrated LETHE system.

The deployment and operational model that will be applied will be based on a multi-tier architecture. The architecture of the system will be split into individual logical tiers. Through this multi-tier architecture scheme, the system's hardware and software is organized into individual subsystems (modular organization). In this way, the system is flexible and can adjust to dynamic changes of the operational environment as well as future technological changes of software/hardware.

The proposed system will be deployed in a 3-tier architecture, which will include:

- The Presentation Tier which includes the interaction environment with both the internal and external system users and the way the data will be presented. User access to available services will be through a single technological platform, which will provide identification possibilities and authorized access to the users.
- The Application Tier which comprises:
    - The business logic of applications that relate to the functionality of the offered applications and services, which are offered by the system.
    - The central infrastructure of identities' management, authentication, authorization and security of applications and systems and the related infrastructure that is required.
- The Data Tier is composed of the LETHE databases (both SQL and Non-SQL) which are characterized by high efficiency, availability, security and manageability. The databases to be deployed assure the high efficiency of the system and relate to storage systems and to information management regardless of whether this relates to transactional data, master data, or data of the offered system databases. The subsystems of the Application Tier will be able to share the common data models and the common data infrastructure.

The architecture platform is complemented with the following Tiers:

- Enterprise Security: It is related to the security infrastructure which protects the offered system. This security will be common for the entire architecture and will treat all issues regarding user access, automated provision of user rights, data encryption, data security and system security in a common way.

- Enterprise Management: It is related to the functionality that allows administrators to supervise the operation of all architecture tiers under the common environment and proceed to administrative changes and tasks of problem detections.

The system will be implemented using several different technologies and frameworks.

Below is the high-level diagram of the LETHE logical architecture with layers and modules stated. An in-depth analysis of these modules follows in chapter *7. Service Architecture*.

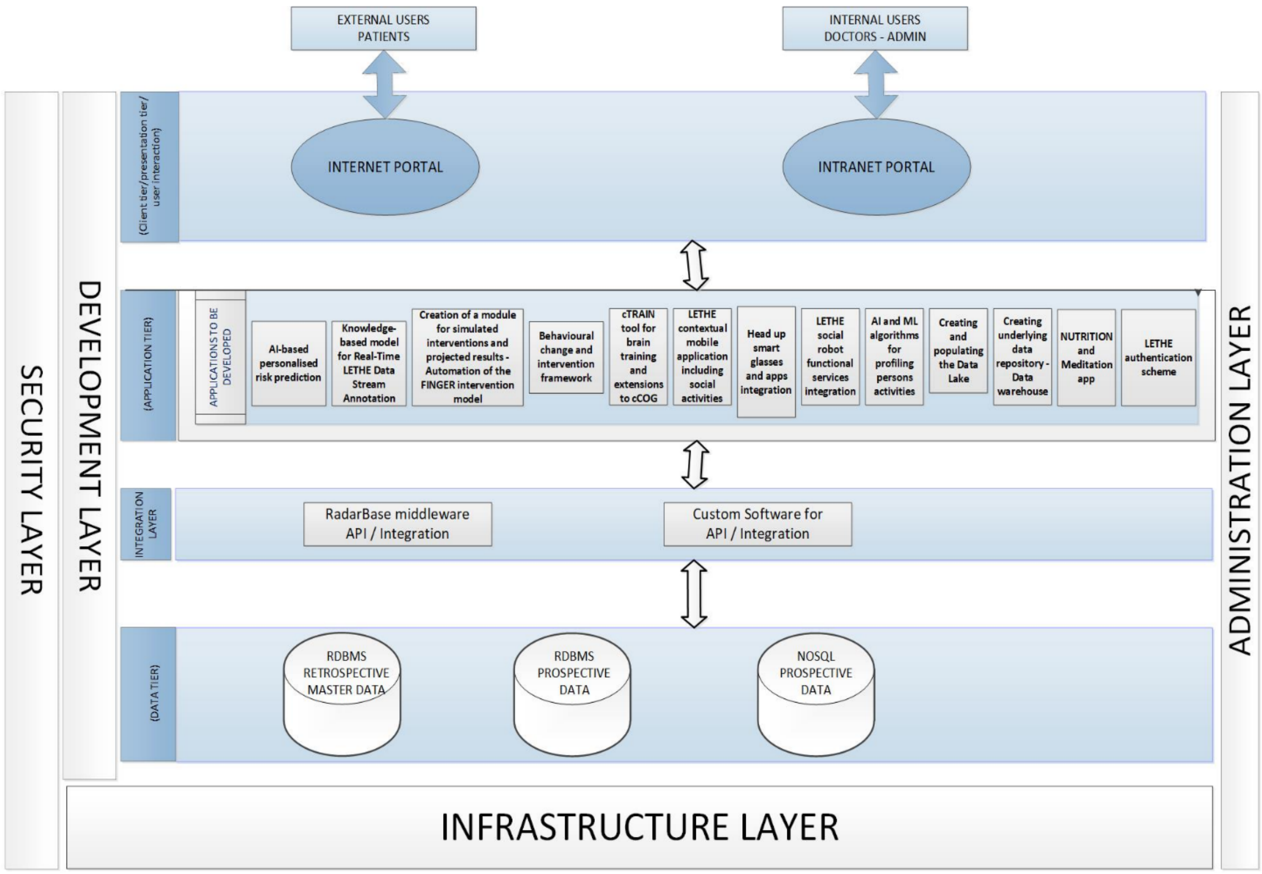

Figure 3. High level diagram of the LETHE logical architecture

4 Rules, regulations and standards

4.1. Regulatory Requirements

4.1.1 The General Data Protection Regulation

The key regulatory requirement LETHE will follow is The General Data Protection Regulation (GDPR)[4]. GDPR was enacted by the European Union (The European Parliament and of the Council of the EU, 2016) to deepen and harmonize personal data protection regulations.

Among other articles of the GDPR for the LETHE Platform the article 32[5] is essential. It states a framework for sensitive data management including:

"Taking into account the state of the art, the costs of implementation and the nature, scope, context and purposes of processing as well as the risk of varying likelihood and severity for the rights and freedoms of natural persons, the controller and the processor shall implement appropriate technical and organisational measures to ensure a level of security appropriate to the risk, including inter alia as appropriate:

a. *the pseudonymisation and encryption of personal data;*
b. *the ability to ensure the ongoing confidentiality, integrity, availability and resilience of processing systems and services;*
c. *the ability to restore the availability and access to personal data in a timely manner in the event of a physical or technical incident;*
d. *a process for regularly testing, assessing and evaluating the effectiveness of technical and organisational measures for ensuring the security of the processing."*

In the LETHE project GDPR issues have been managed in the deliverable *D10.2 POPD requirements*[6] and in the *D6.1 Data management Plan and ORDP I*[7]

4.1.2 Web Accessibility Directive

Because the LETHE Platform uses and produces also web interfaces it is important to take into account requirements of the Web accessibility directive[8]

The Directive obliges websites and apps of public sector bodies to meet specific technical accessibility standards. There are a limited number of exceptions that include broadcasters and live streaming.

The Directive requires:

- an accessibility statement for each website and mobile app;
- a feedback mechanism so users can flag accessibility problems or request information published in a non-accessible content;
- regular monitoring of public sector websites and apps by Member States, and reporting on the results.
- The Directive complements the European Accessibility Act which covers a wide range of products and services also in the private sector. Further European legislation supports people with disabilities in other areas including electronic communications, audio-visual media services, eBooks, eCommerce and ICT equipment. Highlights of how these EU policies affect digital accessibility are in the infographic on 'Digital Economy and Society' legislation.

4.2. Security Policy

The purpose of this section is to describe information security policy to be used in LETHE project. The following general rules for systems customization will be followed for all operating systems, databases and applications of LETHE:

- Services and applications which are not used will be deleted or disabled.
- The latest updates and patches will be installed in the system after they become available (time of application will depend on criticality of issues involved), if this procedure does not affect the smooth operation of the system and does not violate the operational requirements. Before installing security versions, they will be checked based on software manufacturer instructions for compatibility issues with the systems used by the organization.
- Accounts with high rights (root or admin) will not be used for actions that may be carried out and with low account lower allowances.
- Use of administrator accounts and users with high access rights (root or admin) will be limited.

LETHE maintains the list which describes LETHE Platform applications, business operations, technical characteristics (OS, databases, network, and communication links), business managers, managers of information technology and the physical location of the IT infrastructure.

The hosting organization is experienced in running high-security hosting services on behalf of its customers and has developed an Information Security Management System (ISMS) for its secure hosting services. As such, the hosting organization is already fully equipped with the policies and procedural capabilities that are required in order to host high security development and test environments and will use its experience in order to ensure the highest possible levels of security for both environments.

4.2.1 Procedural Measures

The procedural elements required in order to maintain a secure testing, development and production environment include:

- Security Policies: Staff (administrators and developers) will be granted access only if they understand and comply with their security obligations as outlined in the environment's security policies;
- Security Monitoring: Comprehensive monitoring and alerting system for the data centre services;
- Audit Trails: Regular reviews of audit logs conducted by an administrative person in the data center.
- Account Administration and Password Management: Conducted according to the established policy, including account auditing and provisioning, access control and Single Sign-On;
- Incident Management: The hosting organization will maintain an incident management policy and associated incident response procedures (categorisation, triggers and associated actions) as part of its hosting solution;
- Server Management: System audits to determine if bug fixes and/or security patches are needed.
- Business Continuity Plan: Regularly maintained and tested by the hosting organization to ensure its validity in the event of a real disaster. It details how the organisation recovers from a disaster or service interruption;
- Backup/Restore and Storage Management: The backup operation will include testing of security policy for backup procedures, reviewing of backup policy and compliance with this backup policy;

- Service Desk: Registers, classifies, prioritises and forwards unsolved requests to the respective units for solving. Restores a disturbed or impaired service performance while reducing negative business impact;
- Configuration Management: CMDB records the installed equipment, the software modules and their configuration. Replacement/configuration of installed equipment, HW/SW upgrades and maintenance;

4.2.2 Physical Measures

The measures for the physical protection of the IT applications and infrastructure are summarised below.

- Ensure and document the protection of the IT applications and infrastructure;
- Identify all access points through each physical security perimeter and measures to control entry;
- Identify processes, tools and procedures to monitor physical access to the perimeter(s);
- Describe procedures for the use of physical access controls including: (i) visitors access management; (ii) recovery procedures; (iii) prohibition of inappropriate use of physical access controls.
- Define procedures for reviewing access authorisation requests and revocation;
- Identify procedures for escorted access within the security perimeter of non-authorised personnel

4.2.3 System Maintenance Security

The purpose of this section is to identify security requirements on operating systems, databases and applications used in the LETHE project and ensure that they are properly parameterized in terms of security, managing and protecting from unauthorized access. The term 'system' means the operating systems, databases and applications of LETHE.

The following general rules for system customization should be followed for all operating systems, databases and applications of LETHE:

- Services and applications which are not used should be deleted or disabled.
- The latest updates and patches must be installed in the system after they become available (time of application will depend on criticality of issues involved), if this procedure does not affect the smooth operation of the system and does not violate the operational requirements. Before installing security versions, they will be checked with the software manufacturer for compatibility issues with the systems used by the organization.
- Accounts with high rights (root or admin) must not be used for actions that may be carried out with lower privileges.
- Use of administrator accounts and users with high access rights (root or admin) should be limited.

4.2.4 Supply, Development and Maintenance of Information Systems

Any new system implementation and/or technology infrastructure is assessed and adequately controlled for security vulnerabilities which may endanger LETHE's functions.

The introduction of control and security mechanisms in the design and development of new systems and applications is considered necessary and imperative to meet the security requirements in such a way as to ensure compliance with the existing legislative framework, customer confidence and continuity of operations.

Requirements for the development, installation and maintenance of information apply to all LETHE units and suppliers responsible for the development, installation and maintenance of information resources.

4.2.5 Third Parties and Outsourcing Security Management

The interconnection of LETHE with any third company and/or external partner should be implemented only if there is official documentation of the business necessity.

The minimum requirements and security standards in regard to network links of third parties (including those of suppliers, cooperating institutions and companies) to the LETHE Platform and their access to it are similar than requirements to project partners.

LETHE for each contract with third parties should name a responsible manager to control the contract and monitor and evaluate the provided services.

The manager's responsibility is to contact the third party and resolve any issues that may arise. In addition, he/she will supervise third-party services, with a view to ensure the quality of these services as well as compliance with this security policy.

A list of employees responsible to manage third party contracts should exist and be maintained.

4.3. Standards

4.3.1 Information Security Management

ISO/IEC 27001. The LETHE project partners and service providers should also care about overall security of its services. The project including cloud computing and data storage providers should follow Information Security Management standard ISO/IEC 27001:2013 guidelines.

4.3.2 Repository Certifications (Optional)

CoreTrustSeal. The CoreTrustSeal Trustworthy Data Repositories Requirements reflect the characteristics of trustworthy repositories. As such, all requirements are mandatory and are equally weighted, standalone items. Although some overlap is unavoidable, duplication of evidence sought among requirements has been kept to a minimum where possible.

5 Business Architecture

The Business architecture represents holistic, multidimensional business views of: capabilities, end-to-end value delivery, information, and organizational structure; and the relationships among these business views and strategies, products, policies, initiatives, and stakeholders[9].

In this document a whole set of the business architecture approaches are not described because the focus is merely on system level architecture and deployment scenarios.

This section presents overall views to the LETHE project stakeholders and processes including use cases and requirements. Rules, regulations and standards are described in the section 4 of this document and services in the section 7.

5.1. Stakeholders

By the LETHE deliverable D9.2 *Dissemination and Communication Plan*[10] stakeholders that will be involved in the project are the following:

- European Institutions, Agencies, Joint Undertakings.
- Ministries and government authorities
- Healthcare and Welfare Sector
- eHealth Industry
- European Network of citizens' associations
- EOSC Scientific Community
- EU citizens at risk of cognitive decline
- General scientific community
- General public

Other words used in the LETHE proposal stakeholders of the LETHE platform 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

5.2. Use Cases

A use case is a written description of how users will perform tasks on the IT systems.

In this document we use following simplified specifications for the use case:

Id: <id of the case>

Use Case: <Title of the use case>

Actor(s): <anyone or anything who is using the system>

Scope: <The system, sub-system or component that will provide the interactions>

Brief: <The body of the use case is simply a paragraph which informally describes what happens. In this case this is equivalent with a simple user story "As a <role> I can <capability>, so that <receive benefit>">

Id	1
Use case	User collects and delivers prospective data to the RADAR-base
Actor(s)	Patient, healthy persons
Scope (modules)	Mobile phones, mIoT, wearables, mobile devices
Brief	As a user, I want to use my wearables to collect data and deliver it to the platform.

Id	2
Use case	User collects and delivers prospective data to the LETHE Middleware
Actor(s)	Patient, healthy persons
Scope	Surveys, calendar, cTrain, cCoach, FitBit nutrition app, meditation app
Brief	As a user, I want to use my wearables to collect data and deliver it to the platform.

Id	3
Use case	Scientist store retrospective data
Actor(s)	Scientists, clinicians
Scope	Sensitive data storage and EGI DataHub
Brief	As a scientist, I want to store retrospective data sets over secured connections to the encrypted storage.

Id	4
Use case	Scientist receive reports from the dashboard
Actor(s)	Scientists, health care professionals, clinicians
Scope	CCoach, cMRI & cDSI
Brief	As a scientist, I want to receive reports from the LETHE Platform dashboard

Id	5
Use case	AI core module development
Actor(s)	Researchers on AI and machine learning, AI/ML developers
Scope	AI code modules, LETHE Data Warehouse
Brief	As AI/ML developer/researcher, I want to have structured data from the API for AI/ML Core Module development

Id	6
Use case	The expert previews the results from the AI modules and makes adjustments to the plan
Actor(s)	Clinicians, scientists, health care professionals
Scope	AI core modules, dashboards, reports
Brief	As an expert, I want to receive results from the AI core modules to my desktop and adjust plans for patients based on risk factor identification

Id	7
Use case	The patient receives alerts and notifications from the system
Actor(s)	Health care professionals, patients, healthy persons
Scope	Applications, robots, head-up displays
Brief	As a patient, I want to receive alerts from the intervention systems via applications or other equipment I'm using

User stories and use cases will be updated based on results of the User Requirements TF of the LETHE project. More detailed requirements are defined in the TF's working document the Agenda and Results of the Professional-User Requirements Workshop[11] (see an Annex 1) and from the LETHE Advisory Board Meetings (see Annex 2).

5.3. Functional Requirements

In this document each functional or non-functional requirement is flagged as either essential, desired, suspended or unstable.

Functional requirements specify what the system should do, such as:

- Business Rules
- Authentication (which methods users can use)
- Authorization (which levels users need)
- Deployment mechanism
- Possible licenses
- Ordering

5.3.1 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*

5.3.2 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. In particular this is used to access cloud resources; it is possible to use federated identity management systems based on standards such as OpenIDConnect 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*

5.3.3 LETHE Platform

Following functionalities have to be defined during the project with the best practices and based on agile service development methods i.e., detailed use cases and user stories have to 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 the AI core module developing, 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*

5.3.4 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 has to be well described *essential*

R1.6 Virtual machines
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*

5.4. Non-Functional Requirements

Non-functional requirements specify how the system performs a certain function such as:

- Usability (regulations and guidelines)
- Performance and QoS (for example response time, throughput, utilization)
- Scalability (component's scalability)
- Capacity (capacity specification on vertical matchmaking)
- Interoperability
- Other regulatory requirements

5.4.1 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."[12] In the case of web-based services this means merely "a small learning curve, easy content exploration, findability, task efficiency, user satisfaction, and automation".[13]

R1.7 Usability
Usability of upcoming services have to follow W3C Guidelines[14] *essential*

5.4.2 Accessibility

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

R1.8 Accessibility
The LETHE Platform and vertical matchmaking user interfaces have to take into account regulations like The Web Accessibility Directive (Directive (EU) 2016/2102)[15] *essential*

5.4.3 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.

1.9 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*

5.4.4 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.[16]

R1.10 Scalability
The LETHE Platform components have to be scalable enough to support uptake resources and services users' needs. *essential*

5.4.5 Availability and Capacity

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

R1.11 SLA
Service level agreements (SLAs) will be negotiated with resource and service providers. *essential*

R1.12 Capacity and availability
The project has to identify service performance requirements based on users' needs, plan the resources required to fulfil the requirements and ensure performance monitoring. *essential*

5.4.6 Security and Data Protection

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

R1.13 Security and data protection
Services have to take into account applicable regulations like General Data Protection Regulation (GDPR)[17]. Additionally, guidelines from the D6.1 *Data Management plan and ORDP I*[18] and T1.3 *Risk management* have to be implemented also in these activities. *essential*

R1.14 Secure 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 have to be strongly encrypted and secured. *essential*

R1.15 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*

R1.16 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*

R1.17 Firewalls
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*

R1.18 Data center physical security requirements
Facility: • Tier 2. 99% availability. • 200 kW of power maximum in data center 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*

R1.19 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*

R1.20 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, in order 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*

5.4.7 Interoperability

R1.21 Interoperability
Interoperability between different services and components have to be defined. Interoperability principles have to follow New EIF i.e. European Interoperability Framework by ISA2 programme principles.[19] *essential*

5.4.8 Standards and Architecture Framework

Architecture definitions have been described in this document.

R1.22 Architecture compliance
The LETHE Platform has to follow standards and architecture definitions described in this document. *essential*

R1.23 Configuration management
Configuration management processes and procedures have to be defined to provide and maintain a logical model of all configuration items and their relationships and dependencies.

R1.24 Change management
Change management processes and procedures have to 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.

6 Information architecture

6.1. Data Sources and Datasets

6.1.1 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.[20]

6.1.2 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 (active and passive data)
- Data from Wearables (passive data)
- Data from Meditation App (passive data)
- Data from Nutrition App (active data)
- Cardio data (active and passive data)
- Calendar integration (active data)
- Mental health data (active data)
- Additional behavioural modules as well as smart glasses and social robots (active and passive data)

A detailed overview of all collected data from the sensing ecosystem is described in D4.3.[21]

6.2. Data Models

The retrospective data model defined in the deliverable D2.1[22] and prospective data models will be defined in the deliverable which will be defined later

Sensing systems will be described by WP4 and system APIs will be described by WP5.

6.3. Semantic Interoperability

Interoperability plays a significant role within the project. Therefore, syntactic as well as semantic operability is needed and common standards in health care should be used. A detailed description of all used standards within the project as well as the interoperability framework are described in D2.5.[23]

7 Service Architecture

This section describes the services which are included into the LETHE Platform in terms of the technical structure, constraints, and characteristics of software components and the interfaces between them.

7.1. Service Architecture Overview

In order to fulfil the requirements specified in section 5, the solution for the LETHE system is implemented through a service-oriented architecture.

The architecture involves different groups of services:

- Data generation
- Data collection, data storing and data processing
- Data analysing and AI/ML modules
- Reporting, visualizing and end user interactions
- Governance and other support services.

Below is a representation of these service groups:

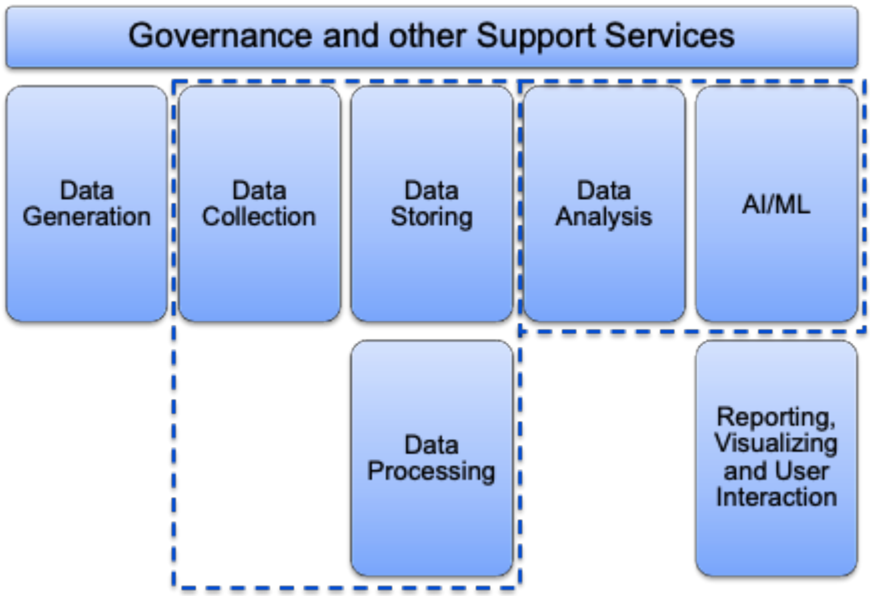

Figure 4. Service groups

The architecture solution foresees that the platform can be located in a cloud infrastructure following different approaches, depending on the available resources, described into section 9.

The LETHE project is divided in two phases which reflect its project plan.

Phase 1 of the project has already started and will end August 2022. This phase will provide an initial prediction model, developed to predict the progression of dementia and the related risk factors.

The model is based on timeseries and is trained using different data sets provided by four clinical research centers in Europe, partners of the LETHE project. When tested and validated, the model will be deployed as REST API for remote model invocation.

The next image shows the components of the system services for this first phase of the project.

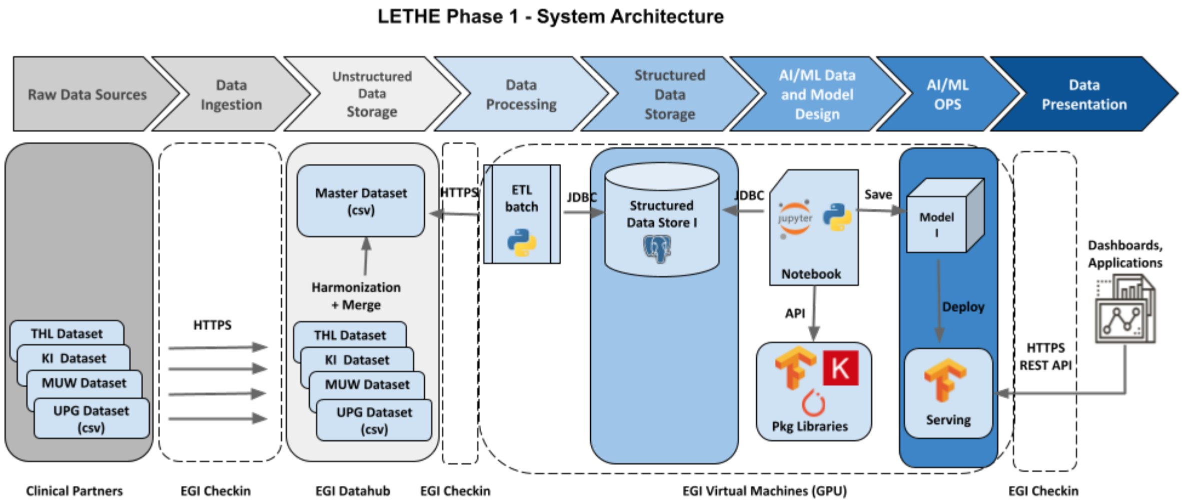

Figure 5. LETHE Phase 1

Within LETHE phase 2, the initial prediction model will be extended and validated. This will be done using both regression and deep learning. The required features will be selected from:

- long-term additional data provided by the four clinical centers
- active and passive data collected through the LETHE sensors and apps ecosystem that will represent the new digital biomarkers.

This new model will be tested and optimized iteratively. Each version of the model will be deployed as a service with a REST API.

To reach this goal, the system architecture of LETHE phase 2 is more complex: the next image shows the components of the system services for this phase of the project.

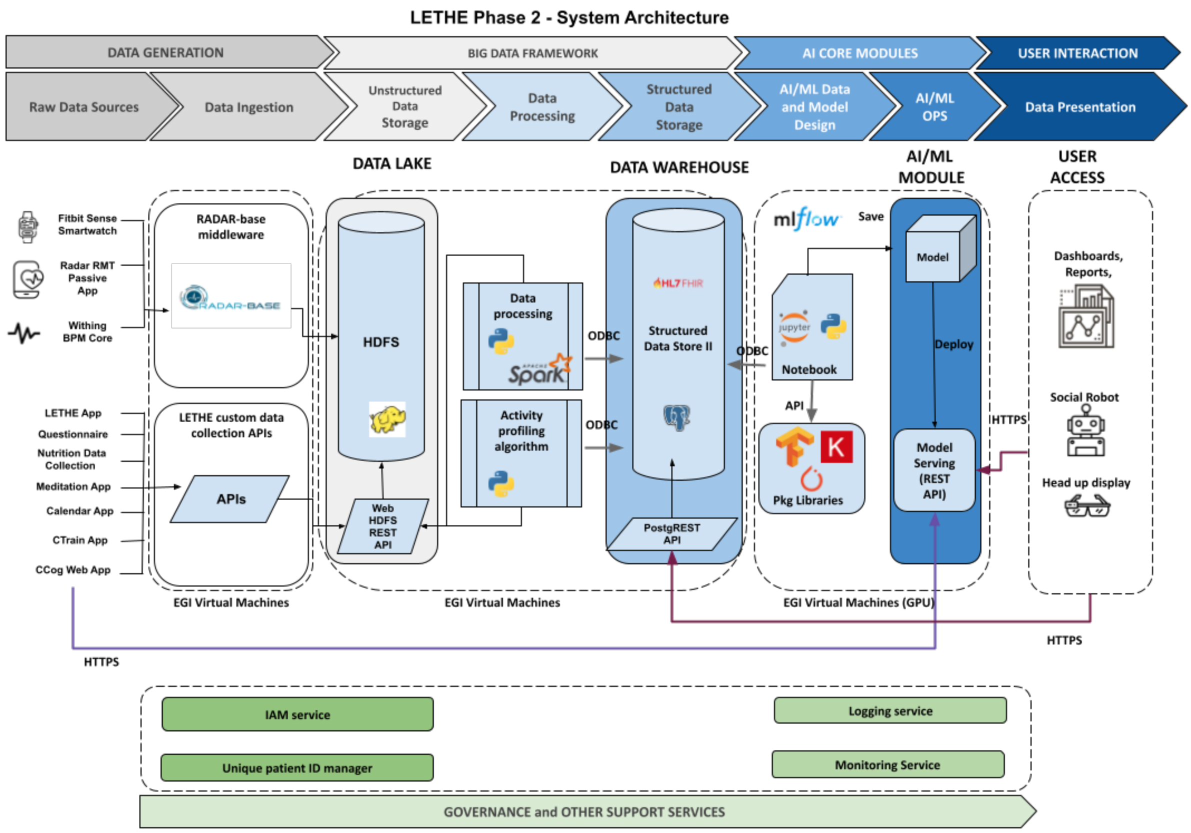

Figure 6. LETHE Phase 2

The following sections describe each group of services. For each service, these three components are provided:

- the interface, which defines how a service provider will execute requests from a service consumer.
- the contract, which defines how the service provider and service consumer should interact.
- the implementation, which is a custom code or a software component.

7.2. Data Generation Services

The data generation services are the devices and the applications producing the raw data and the software components that allow to collect the raw data for the LETHE system.

The following picture shows the different data sources and the software components of the data generation services.

The data sources can produce two different types of data:

- *passive data*: generated by IoT sensors of wearable devices and mobile devices where user interactions are not required. The data is produced as a real-time stream of events. RADAR-base is used to collect this type of data.
- *active data*: generated by mobile and web applications requiring user interactions. The data is produced as a file that records the activity of the user. A set of LETHE data collection APIs are used to collect this type of data.

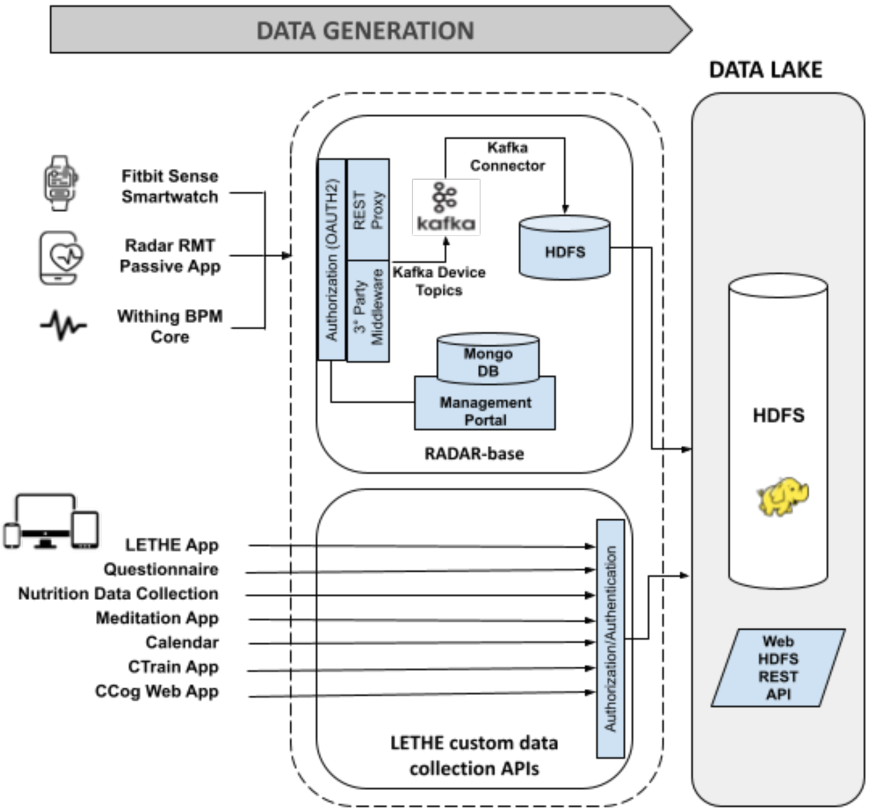

Figure 7. Data sources

7.2.1 RADAR-base

RADAR-base is an open source data collection and management platform based on Apache Kafka for remote assessment of diseases using various wearable devices and mobile application technologies.[24]

Considering that the RADAR-base Docker distribution will be deprecated by the end of 2021 and that the RADAR-base Kubernetes solution is not yet stable, the contingency plan would be to develop a custom middleware platform to serve as a means for sensor data collection.

7.2.2 Passive Data Sources

The passive data is produced by these physical activity and cardio devices. Some possible examples about these devices are:

- Fitbit Sense Smartwatch
- Radar RMT Passive App
- Whitings BPM Core.

It is collected using RADAR-base.

The RADAR-base involved components depend on the device:

- Physical activity smartwatch:
    - Data is stored into the mobile device connected to the smartwatch;
    - Data is sent to RADAR-base into two possible ways, depending on the solution provided by the smartwatch vendor:
        - Data is sent from mobile device to the smartwatch vendor cloud storage service, RADAR-base, through third-party API, reads data from this cloud storage service and sends it to a dedicated device Apache Kafka topic;
        - Data is sent from the mobile device middleware to RADAR-base through a REST proxy and that routes it to a dedicated device Apache Kafka topic;
    - The data is stored to the internal cold storage (HDFS storage) using a Kafka connector.
- Radar RMT Passive App:
    - Data is temporarily stored in the mobile device;
    - The mobile device middleware sends data to RADAR-base through a REST proxy and sends it to a dedicated device Apache Kafka topic
    - The data is stored into the internal cold storage using an Apache Kafka connector.
- Withing BPM Core: TBD

The data is sent from the RADAR-base local HDFS storage to the LETHE Data Lake, possibly using the Web HDFS REST API interface provided by the Data Lake.

This synchronization task between RADAR-base cold storage and the Data Lake will be a scheduled batch. Another possible solution could be RADAR-base cold storage hosted directly into the Data Lake. This will also avoid data security constraints to be applied even to RADAR-base cold storage.

Other RADAR-base modules that will be used with LETHE to manage passive data collection are the Authentication service and the Management Portal.

The Management Portal is a web application providing a user interface to manage projects, enrolling participants and managing the association of participants with corresponding data sources (devices and apps). It controls the authentication and authorization of data operations and de-Identification of all the data identifiers.

7.2.3 LETHE Data Collection APIs

The LETHE collection APIs represent a custom layer that is used by active data sources to collect data and send it to the LETHE Data Lake.

For each of the foreseen active data sources, a dedicated custom API will be implemented to perform at least these tasks:

- It will collect data locally (into the device) and group new data (not yet sent) into a file, including the patient identifier. For the same patient, data is collected by different data sources. It is important that these different data sets can be associated to the same patient. For this reason, the management of a unique identifier associated to each patient user is required.
- It will use LETHE authentication and authorization. Authentication and authorization module can be managed using EGI Check-In[25] and OAuth2 based solutions as described in the chapter 4.6. Details of the authentication and authorisation applications will be defined and decided in WP5 (T5.4).
- It will send the data file to LETHE Data Lake using REST API interface.

The LETHE collection APIs will be defined later, when all active data sources and their interfaces will be known.

7.2.4 Active Data Sources

The active data is produced by these mobile and web applications:

- Questionnaire
- Nutrition Data Collection
- Meditation App
- Calendar
- CTrain App
- CCog Web App
- LETHE App

These applications have their own APIs and data models. This data will be collected in to the Data Lake layer and it will be harmonized in the data processing layer.

7.3. Data Storing, Processing and Interoperability Services

7.3.1 Data Lake

The Data Lake will store raw data coming from data sources, where transformation and other processing procedures are not yet applied.

The Apache Hadoop Distributed File System (also HDFS[26]) is used for the Data Lake storage.

HDFS directly stores data as files: it allows to be compliant with any type of raw data coming from data sources. It has different types of processing interface. The WebHDFS REST API interface[27] allows to access HDFS storage using Spark framework.

It also offers strong and easy scalability, with the possibility to distribute the data storage over different nodes, depending on the estimated amount of data coming from data sources.

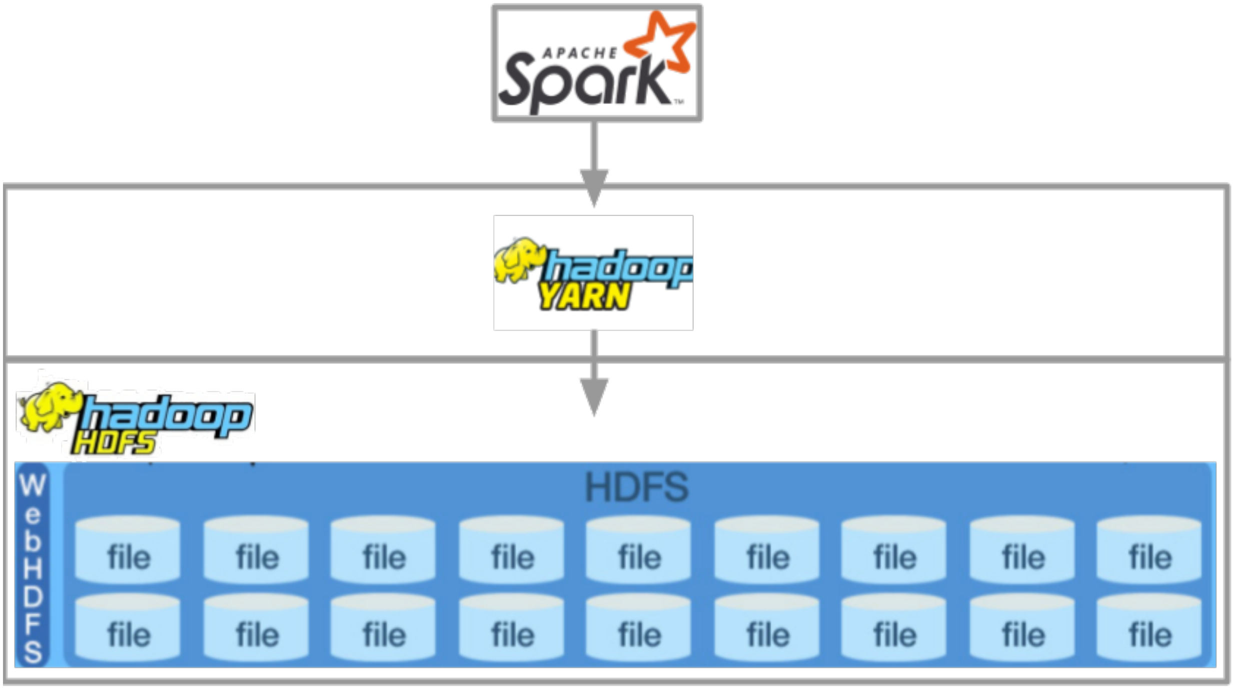

Figure 8. HDFS

7.3.2 Data Processing

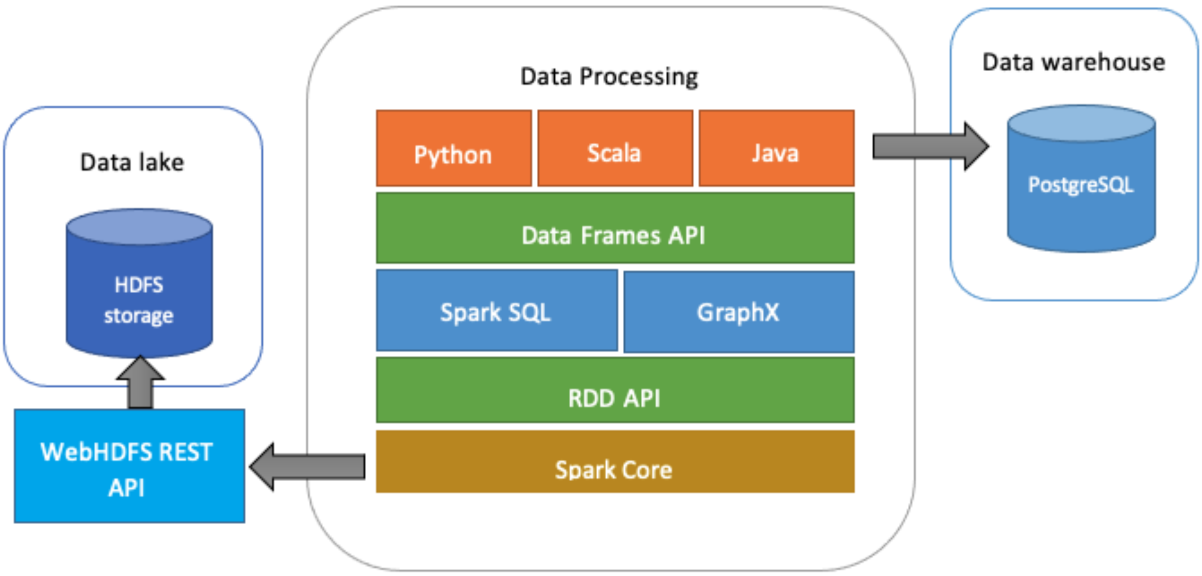

Figure 9. Data processing

The LETHE Data Lake stores raw data produced by the different data sources and it is not yet ready to be used by the AI/ML module. For example: the Data Lake manages only data insert, so it can store incomplete, wrong, or duplicated data.

The success of any AI/ML model is as good as its input data. A High-quality training data set is extremely critical in healthcare and a deciding factor for its end result. For these reasons, the raw data stored in the Data Lake has to be processed.

The processing of raw data will include at least these tasks:

- Validation: incomplete, duplicated, syntactic and semantic wrong data has to be intercepted and marked as *wrong*
- Fixing: when possible, the wrong data has to be fixed and marked as valid;
- Cleaning: when fixing is not possible, the wrong data has to be removed from the next processing steps;
- Aggregating: by time (time granularity of passive data can be too), by patient (data for the same patient comes from different data sources);
- Annotate: it encompasses giving labels and metadata tags that form the base for any algorithm by establishing the grounds to create machine learning models;
- Transform: the data is transformed to be compliant with the LETHE Data Warehouse data format;
- Consistency check: it establishes when data is considered complete and consistent and can be used for AI/ML;
- Enrichment: it adds metadata to enable the catalogue and the data discoverability necessary to satisfy the interoperability requirements.

The data processing component is a batch which reads data from the LETHE Data Lake, using the WebHDFS REST API interface, processes data using the defined rules and writes the result data into the LETHE Data Warehouse, using ODBC interface.

The batch is a custom Python code that uses Spark framework[28], that allows to quickly process large amount of data.

Depending on the final LETHE infrastructure (different hosts for the Data Lake, the Processing Module and the Data Warehouse), the processing batch needs authorization methods to access the Data Lake and the Data Warehouse.

7.3.3 Data Warehouse

The Data Warehouse is the data store for the structured data ready to be used by the AI Core Module.

The Data Warehouse is implemented using instances of PostgreSQL database service.

During the first phase of the project (see Figure 5), an ETL process will load the Master Data sets into the Datawarehouse without any type of data transformation, into a set of tables of a dedicated PostgreSQL instance. This loading operation will be a one-shot procedure, after that this database instance will become immutable, without further additional inserts.

For this reason, the ETL process is not considered a software component of the LETHE system, but an operational procedure as the retrospective data harmonization.

During the second phase of the project (see figure 6), the raw data produced by the Data Generation service, will be processed and then loaded into the Datawarehouse. A dedicated PostgreSQL instance is used during the second phase of the project. The database schema will reflect the data schema described into section 6.

To manage the access to these instances:

- for internal modules, ODBC interface is used;
- the PostgREST[29] service will be provided and used by the external modules.

7.3.4 Interoperability

The interoperability functionality allows to share data with external applications and cloud service. The basic mode of interoperability is through APIs.

A third-party client, in order to access the data, must be properly registered within the LETHE system and consequently possess the necessary credentials to log in. Once in possession of the credentials, the third-party client can, through appropriate API calls, request a list of all the "tables" (data tables, in a Data Warehouse perspective) present in the system to which it is enabled, as well as perform data extractions by any of the aforementioned tables.

The API allows the client to page the data of any table based on time intervals. In this way the client is free to acquire the data of interest in pull mode.

The Data Warehouse interoperability depends on the used data format.

In case of HL7FHIR[30] data format, SMART[31] framework could be introduced into LETHE to enable third-party applications to extract HL7FHIT data from the Data Warehouse.

7.3.5 Metadata and Data Catalogue

The Metadata and Data Catalogue provides a discovery functionality that can facilitate the interoperability for data stored into the Data Warehouse.

The Data Warehouse shall offer controlled and harmonized data for the AI Core Modules which will be organised with catalogue functionalities. This will be defined in detail in WP5.

7.4. Analysis and AI/ML Services

During the second phase of the project, the development of the advanced models will apply different approaches including feature selection for regression model and deep learning models.

The models will be tested and optimized iteratively. After validation, the models will be deployed and REST APIs will be provided for the model serving.

The figure shows the AI pipeline as the life cycle of the models.

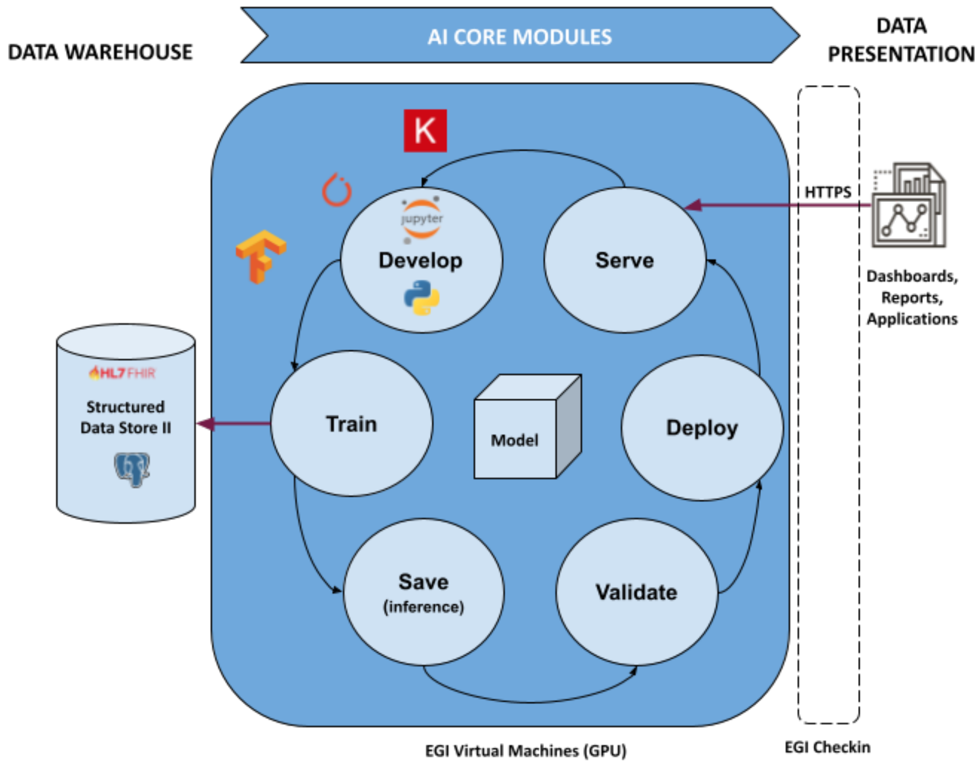

Figure 10. AI Pipeline

All the software components required for the AI Core Modules will be part of the LETHE Platform because data for develop, train, test and validate cannot be exported outside the LETHE Platform to fulfil security and privacy requirements.

The software components of the AI Core Modules will access data stored into the Data Warehouse after using authorization policies depending on the used protocol.

The source code for the model algorithm will be developed using:

- Python programming language,
- Notebook environment:
  - EGI Notebook[32],
  - Jupyter notebook service as part of AI Core Modules.
- required AI/ML framework and libraries such as TensorFlow[33] and PyTorch[34].

To manage the AI pipeline required by the LETHE project phase two, the MLflow framework[35] will be used.

7.4.1 MLFlow

MLflow[36] is as an open source platform for the complete machine learning lifecycle. A complete machine learning lifecycle includes data analysis and preparation, model training, model evaluation, model deployment, and model maintenance.

MLflow supports models load, save, and deployment with scikit-learn, TensorFlow and Spark platforms.

MLflow is built as a Python package and provides open REST APIs and commands to:

- Log important parameters, metrics, and other data that is important to the machine learning model
- Track the environment a model is run on

- Run any machine learning codes on that environment
- Deploy and export models to various platforms with multiple packaging formats.

MLflow is implemented as several modules, where each module supports a specific function.

The MLflow modules used into LETHE architecture are:

- *Tracking*: it implements REST APIs and the UI for parameters, metrics, artifacts, and source logging and viewing;
- *Projects*: it defines the specification on how to run the model training code. It includes the platform configuration, the dependencies, the source code, and the data that allow the model training to be executed through MLflow;
- *Models*: it defines the general model format in the MLmodel file.

8 Infrastructure architecture

In this section LETHE platform's infrastructure components such as storages, computing resources etc are specified.

8.1. Infrastructure Architecture Overview

The LETHE platform will be built on the top of the EGI FedCloud[37] services where main components are:

- EGI DataHub[38]
- Encrypted storage
- Encrypted data transferring
- Virtual machines offering the platform for the applications middleware solutions (like databases etc.) and computing (incl. GPUs and CPUs)
- Notebooks services
- Authentication and authorisation services

Services will be deployed and maintained by EGI federation service provider.

8.2. Infrastructure Component Descriptions

8.2.1 Data Generation and Collection

8.2.1.1 Retrospective

Each clinical partner will upload their retrospective data set to the LETHE cloud, where it will be stored as an unstructured data source. The data harmonization concept will be developed in Task 2.1 to merge the retrospective data accordingly into the structured retrospective master-dataset (D2.1).[39]

8.2.1.2 Prospective data

The collection of prospective data will be an ongoing process. Data will be collected via mobile apps, wearable sensors, questionnaires and tests, which the participants in the study will be asked to perform.

These data will be collected in the middleware layer and stored in the LETHE Data Lake (Figure7).

8.2.2 Data Storage and Computing Infrastructure

A storage and computing infrastructure had to be made available starting with data sharing phase 1 of the project. A series of options have been discussed in the early stages of the project. The solution agreed and adopted was based on a centralised architecture (Figure 11).

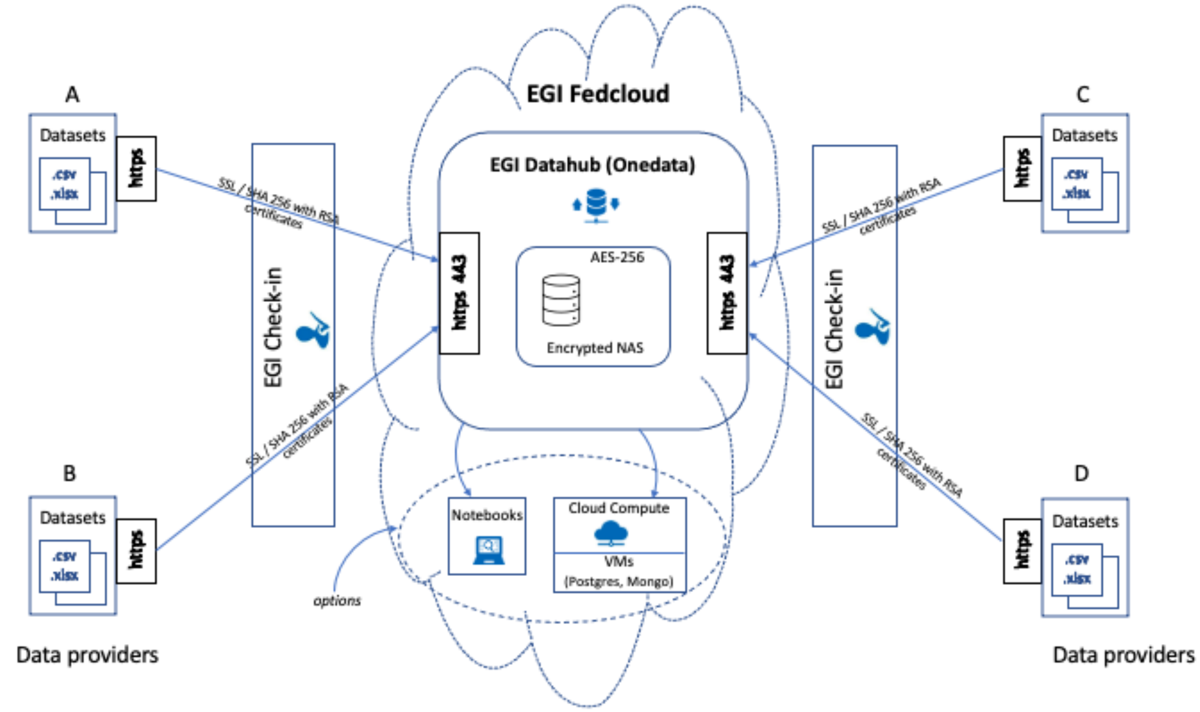

Figure 11. LETHE Data Storage and Computing Infrastructure - Data Sharing Phase 1

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).

Several Virtual Machines (VM) are deployed via the EGI Cloud Compute service to access and process the data in the storage. The minimal VM specifications agreed following a call to support the project are:

- 8 vCPU, 16GB RAM, 40GB disk (for storage management)
- 8 vCPU, 16GB RAM, 500GB disk (for merging the datasets hosted in the storage)
- 5 Intel Core i9-9820X or similar, 1 GPU NVIDIA 1081 TI, 32GB RAM (for Machine Learning tasks)

Also, EGI Notebooks are available for specific tasks.

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

The entire infrastructure described above will be further used during the data sharing phase 2 of the project when it is expected to be expanded with more resources.

8.2.3 Data Processing and Harmonization

Various datasets which are planned to be combined into a single, structured data source build the input for the processing and harmonization steps.

A major part of the harmonization process takes place in a distributed manner:

The structure and format of the joined master data set is defined in Task T2.1. Therefore, a map of all variables and information given in each data set is created: semantically similar and overlapping variables across all datasets are detected, and a common syntax is defined (D2.1)[40].

Based on the metadata and variable descriptions, a harmonization concept was established. In bi-weekly meetings, the progress and following steps were discussed among the clinical partners as well as with the programming team, which is responsible for the development of the ml-models. From these meetings, recoding principles were derived and summarized in a document D2.1[41].

Each clinical partner is responsible for the application of the principles and for the upload of the data records. Each data set will be stored twice (as raw data and as recoded / ready-to-be-harmonized data) as unstructured data source (i.e. .csv file) in the defined environment. (See previous chapter Data Storage and Computing Infrastructure)

Then, each recoded record will be cleaned and validated separately before all records will be stored as structured data all together. For this purpose, missing values, outliers, inconsistencies and inaccuracies in the data set need to be identified and corrected.

The data processing and harmonization process depends on the availability of the data and the availability of the Data Warehouse and the environment where the data can be stored and processed.

8.2.4 AI Core Modules

AI core modules will be made with Python. TensorFlow, Keras and Pkg libraries will be also used. ML Flow, Jupyter notebooks and other solutions will be in the EGI FedCloud and CPU/GPU VMs from the computing cloud.

8.2.5 Reports and Visualisations

The knowledge base has no specific requirements by the architecture point of view. This means that this document does not describe knowledge base models etc. In this document integrations and interoperability to the knowledge base are only described when relevant, but no knowledge base internal properties.

Deliverable D3.3 - Knowledge-base Model will contain the implemented model as well as the description of the knowledge base model for data stream annotation.

8.2.6 Applications for Interventions

The aim of this module is to analyse the participants' compliance to the pathways defined for them on the input side and to provide feedback to maintain or enhance compliance on the output side.

8.2.6.1 Compliance to Pathway

The input for the intervention module consists of the (compliance to the) *pathway estimator* module in combination with the *Intervention pathways*. The pathway estimator is determined by the automated FINGER protocol module and provides compliance to the default pathway as feedback.

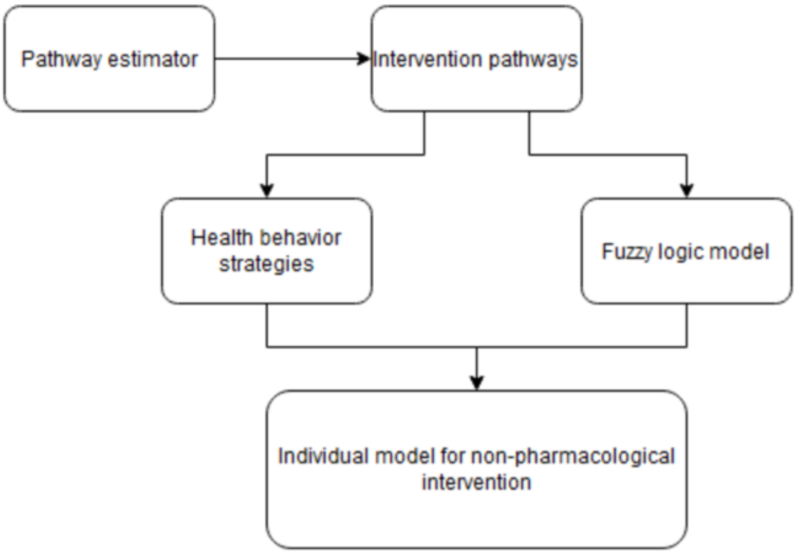

Figure 12. Intervention pathway

The intervention module needs to adjust its output to the patient's behavioural changes. While the pathway estimator module provides a rather fixed input by reporting the patient's compliance to the default protocol, the health behaviour strategies module will have to implement several strategies to adapt to behavioural changes. These strategies need to optimize the compliance to the protocol given the current situation, which again comprises the actual compliance to the protocol but also the potential to meet the requirements.

This potential will, for example, depend on the determined remaining time for the execution as well as the patient's current setting and location and also on different pathways. Therefore, in parallel to the standard pathway of the health behaviour strategies an additional *Fuzzy logic* module can provide a different suggestion based on this additional input.

8.2.6.2 Recommendations for Intervention Managers

The output of the intervention module must be provided via an interface that can be accessed by all managing modules used for the intervention (App manager, robot manager, Head-up display manager). The model needs to take the input data it gathered from the intervention pathways and strategies and dynamically decide whether to intervene and inform the applications. For this, the model will have to weigh the provided attributes.

For the output concept, the model needs to merge the information it got from the intervention strategy and the fuzzy logic model and create a readable output for the device managers. The payload will have a JSON format, as it is one of the most common formats used in data transfer environments. The communication between the involved modules will be based on the MQTT protocol, as it deals best with IoT devices and the architecture fits for the output concept. Using MQTT, the intervention model as well as the managers could act as clients in the structure and a MQTT broker could be established to handle the topics. This way clients can also act as publishers and send out a topic, for example for informing other devices that a message has been read and delete the respective notification.

The intervention module only depends directly on the output of the automated FINGER protocol and on the structured data in the LETHE Data Warehouse via the estimator for pathway compliance.

8.3. Authentication and Authorisation

The EGI Check-in service (also called EGI AAI proxy) enables access to EGI services and resources using federated authentication mechanisms. Specifically, the proxy service is operated as a central hub between federated Identity Providers (IdPs) residing 'outside' of the EGI ecosystem, and Service Providers (SPs) that are part of EGI. The main advantage of this design principle is that all entities need to establish and maintain 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 are able to 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 has built-in support for SAML, OpenID Connect and OAuth2[ix] providers and already enables user logins through Facebook, Google, LinkedIn, and ORCID. 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.

In the LETHE Platform EGI Check-in and Virtual Organisations (VO) are used as user authentication and authorisation mechanisms. This is needed for the platform because computing and storage resources are delivered from the EGI Federation.

In simple terms a Virtual Organisation is just a group of users. In 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 in Check-in are represented as Collaborative Organisation Units (COUs). A COU is more than just a group. It is the concept of groups combined with membership management and advanced enrolment workflows. COUs can also be organised in a hierarchical structure for creating groups or subgroups within a VO.[x]

For prospective data gathering and ingestion via RADAR-base and LETHE Middleware OAuth2 is also used to end user authentication and authorization.

RADAR-base provides a centralized study management system called Management Portal to manage multiple studies, users and their roles to guarantee security and privacy. This web application is the main user interface for planning and managing multiple studies, enrolling participants and managing the association of participants with corresponding data sources (devices and apps). This system controls the authentication and authorization of data operations and de-Identification of all the data identifiers.[xi][xii]

To make combined responses possible from the user both, RADAR-base and the LETHE Middleware have to deliver basic user information (i.e. uid, email, name).

The LETHE Middleware authentication and authorization scheme will be defined later, but it will be designed also to use OAuth2.

8.4. Integration Architecture

Integration architecture is a description that facilitates the communication between multiple systems in the LETHE platform i.e. integration architecture is comprised of structures which allow for the interoperability of different components of the platform.

In this context integration architecture description includes relations between

- applications used in the LETHE platform,
- possible APIs of the applications or components,
- data interchange formats or other file formats and
- their relations.

9 Application Deployment Scenarios

9.1. Module Responsibilities by WP

In this chapter has presented preliminary division of the LETHE Platform module development and deployment responsibilities. Definition based on phases I and II on 7.1 Service Architecture Overview.

9.1.1 Phase I

Table 3. Module responsibilities on phase I

Id	Module/System	WP
1	**Data ingestion (retrospective data)**	
1.1	EGI Check-In	5
2	**Unstructured data storage**	
2.1	EGI DataHub	5
3	**Data processing**	
3.1	EGI Virtual Machine (GPU)	5
4	**AI/ML Data and Model Design**	
4.1	Notebook	5
4.2	Pkg Libraries	5
5	**AI/ML Ops**	
5.1	AI/ML Module	3
6	**Data presentation**	
6.1	Dashboards, reports and applications	3

9.1.2 Phase II

Table 4. Module responsibilities on phase II

Id	Module/System	WP
7	**Data ingestion**	
7.1	EGI FedCloud VMs for prospective data applications	5
7.2	Middleware; RADAR-base	4
7.3	Middleware; LETHE	TBD
8	**Unstructured data Storage**	
8.1	LETHE Data Lake	5

9	**Data processing**	
9.1	Data processing	5
10	**Structured data storage**	
10.1	LETHE Data Warehouse	5
11	**AI/ML Data and Model Design**	
11.1	Notebooks	5
11.2	Pkg Libraries	5
12	**AI/ML Ops**	
12.1	AI/ML Module	3
13	**Data presentation**	
13.1	Dashboards, reports and applications	3
13.2	Intervention apps	4

9.2. Constraints and Schedules

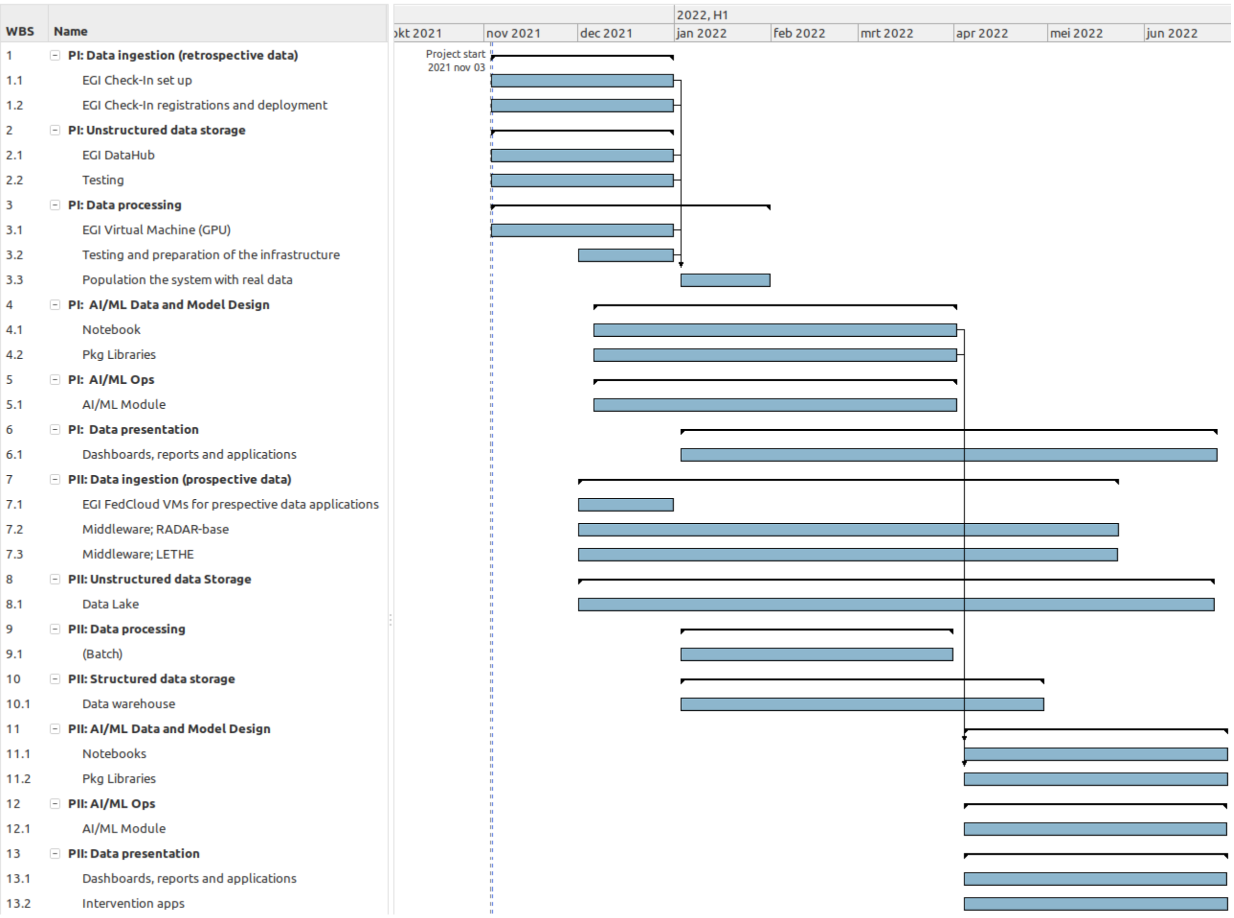

Figure 13. Preliminary GANTT for the deployments

Previous GANTT describes a preliminary order and timing for deployment of the LETHE platform phases I and II. It describes the first iterations of the system deployments until M16. Schedules will be updated by respective WP.

10 Conclusion

This document describes basic architecture of the LETHE Platform. Objective of the architecture is to specify platform to applications and processes for WPs which develop and deploy solutions.

This architecture based on interoperable modules and applications. Iterative development of the modules and other platform components are needed i.e. requirements and module architecture have to be updated by WPs which will develop solutions.

The LETHE Architecture contain following layers:

- Data sources
- Data Ingestion
- Unstructured data storage
- Data processing
- Structured data storage
- AI/ML data and model design
- AI/ML Ops and
- Data presentation

Maturity of some layer descriptions are higher than some other i.e. some specification information will be available later.

Additionally, this architecture specification includes a new component which has been defined in the earlier conceptual architecture specifications. Decisions about owners of this new LETHE platform middleware layer are required.

Before the next version of the architecture document essential parts of the architecture documentation have to update for other WP purposes. These are for example: use cases when the comprehensive data from the appropriate task force is available, application and deployment order when decisions about LETHE middleware has been made and phase I and phase II data flow descriptions.

Also, authentication and authorisation architecture have to be defined in details after the decisions about the custom the LETHE Middleware. Authentication and authorisation architecture will be managed by WP5. Deliverable D2.3 *LETHE architecture update* (M30) will be an update of this document based on developments within the different WPs.

11 Annex 1

Results from Professional Requirements Workshop, 08.11.2021

General requirements:

Nr	Requirement	Functional (f) non-functional (n)	Rating (0-5)
1.01	Alerts for categories of the progression of the disease	f	3.5
1.02	Alerts for categories of compliance-to-plan	f	4.3
1.03	Setup should be done with the GP / Neurologist together	n	3.8
1.04	Simple way of communication and data sharing between GP, specialist and patient	f/ n	4.3
1.05	Give GP feedback on what is going on and how patient is improving	f	3.5
1.06	Generation of an automatic report every 6 months	f	3.8
1.07	Information about local activities and persons (I.e. local FINGER activities)	f	2.8
1.08	Highlight lifestyle factors which are at risk or where the patient could improve	f	4.8
1.09	System should provide follow-ups for GP	f	3
1.10	Assign/forward a person of risk to another professional (I.e. prevention-nurse or specialist) via the tool	f	3.2
1.11	Provide evidence and information about the disease and therapies (FINGER)	f	4
1.12	LETHE as medical device	n	3
1.13	Alert if an update on the prevention model (new guidelines etc) is available	f	3.7
1.14	Show the availability of data for clinical studies	f	2.3
1.15	Risk-factor-diagram personalized for each person individually reordered (related to 1.08)	f	4.7
1.16	Efficiancy, help with workload instead of generating more work	n	5
1.17	Physician should be able to open a dashboard of a patient in preparation of a meet-up and see how they perform now compared to 6 months earlier	f	4.5

1.18	LETHE tool suggests by itself after the assessment if the patient could benefit from a LETHE guided intervention	f	4.7

Id	
Use case	Highlight lifestyle factors which are at risk or where the patient could improve
Actor(s)	professionals
Scope	Backend, Webview of LETHE system
Brief	GPs, professionals and specialists that use the LETHE System can view the (most progressing) Lifestyle-risk-factors of the patient in real-time and adapt the treatment according to that.

Graphical Design / Media Requirements:

Nr	Requirement	Functional (f) non-functional (n)	Rating (0-5)
2.01	Run as a local application on most computers	n	3.7
2.02	Run as a web application in most browsers	n	4.5
2.03	Offline version of the tool	n	2.7
2.04	Easily printable views and dashboards	n/F	4.3
2.05	Themed and easily printable UI	n	4.8
2.06	Excel / csv export of data	f	3.3
2.07	Adaptable dashboards (user can arrange Dashboards and information by themselves)	f	3.3
2.08	Tooltip-help by hovering over items	n	3.3
2.09	Application for professionals should run on mobile devices	n	3.2

Requirements on Medical History Entries, Surveys and Questionnaires/scales/tests:

Nr	Requirement	Functional (f) non-functional (n)	Rating (0-5)
3.01	Option for sending predefined questionnaires, scales and tests (like QoL)	f	4.5
3.02	Option of creating own questionnaires, scales, tests	f	3.2
3.03	Option of asking open questions to the patient over the LETHE tool	f	2.7
3.04	Specialists should be able to insert examination results (like cognitive tests) at visits	f	4.3

Requirements on Knowledge discovery, scientific usage and dashboards:

Nr	Requirement	Functional (f) non-functional (n)	Rating (0-5)
4.01	New risk stratifications and tools for specific target populations	f	4.2
4.02	Provide evidence on the intervention (minimal dose etc)	f	4
4.03	Positive messaging for patients (I.e. microlifes / micromorts)	f	4.3
4.04	Awareness assessment done through the app	f	3.8
4.05	Flexible adding and removing intervention areas and tools for measuring certain parameters	f	3.7
4.06	Color-coded information-matrix, visualization where the patient is placed compared to other persons	n	3.7
4.07	Temporal evolution of digital biomarkers	f	4
4.08	Temporal evolution of onset and progression of clinical symptoms	f	4
4.09	Change in risk factors over time, visualization of how it sums to a certain risk class	f	4.5

12 Annex 2

Summary of requirements for the Lethe app from the perspective of the Advisory Board

Material has collected in 3 occasions i.e. meeting 1 on 23 November 2021 with 3 members, meeting 2 on 26 November 2021 with 3 members and interview with 1 member on 26 November 2021. The same materials were addressed in the 3 occasions.

The information summarized in this document is based on feedback provided by members of the Lethe Advisory Board (AB). The AB includes six people who are at a higher risk of developing dementia and one person with mild dementia. To promote the meaningful participation of all members three separate meetings were organized.

General requirements

- Very user friendly and easy to interact with, easy to use, intuitive,
- Engaging: interactive not static.
- Fun, not boring.
- Easy to use, that doesn't frustrate the person, if it doesn't work, I stop using it.
- Something that you can speak to (like SIRI or Alexa)
- Something that looks nice when you look at it, attractive.
- It should not take too much of my personal time. It should help me, not add stress and work.
- Important that I can use the app at "my own time", not that "I have to". Flexibility and personalization are very important.
- It is important that participants can tailor the frequency and type of the reminders.
- Fits with the frame no matter where you are accessing it.
- The language should be easy to understand, images should be included. It should be adaptive. Friendly pictures
- The timeliness of the alerts is more important than the frequency. If I am going to forget, the reminder should be close to the event (to when I need to do/use it).

Specific requirements/features

Access, size and visualization:

- The Lethe app should be accessible from several devices (phone, tablet, computer) so that the person could decide how to access depending on the task or where the person is. Access from the phone is important as you can have it with you all the time.
- Size matters. Smaller if it is something that you need to carry or bring with you most of the time, but a keyboard and bigger screen helps if you need to complete information or spend more time.
- Tasks which include writing or reading (and for some for watching a video) are easier in a tablet or computer.
- Visuals: It would be good to have numbers and graphs, something that it is easy to interpret and possible to compare your performance overtime.

Content:

- TRACKING ACTIVITY: Help the person to monitor/track different areas relevant to brain health: e.g. how much exercise/workouts the person does, sleeping and waking time during the night, what/how much I eat, how much time I spend in relaxing activities vs how much time I spend working/in demanding activities. How much you have done of different activities (e.g. how many hours in a day you have read, how many hours you have exercised etc.).
- MONITORIG PROGRESS: Information about how you are performing and being able to monitor how you progress overtime is very important. Continuous information and reporting about how you progress in different areas.
- PROVIDE RECOMMENDATIONS, TIPS AND SUGGESTIONS:
  - Recommendations and information should not be general but specific for brain health (e.g. in nutrition not just about a healthy diet but about what to eat or avoid which is good for your brain). Specific and clear information about what to eat and what to avoid, not just general advice. The information should be accurate, up to date, something they can trust (e.g. there is a lot of contradictory information about food in internet, as for example the food that you should eat or you should avoid, this is confusing).
  - Advise about things related to your everyday life. Updates on relevant areas, agenda and reminders, suggestions for example about cooking
- ALERTS/ REMIDERS: To do things but also alerts when you are working too much/need to take time off or relax
- EVERYDAY LIFE (NOT CONNECTED TO BRAIN HEALTH): Access to music, newspaper, or other things that are of interest to the person
- THINGS THAT PROMOTE SOCIAL INTERACTION WITH OTHER PEOPLE

What would help in the long term?

- To be attractive the app should be very easy to use, should not ask often things i.e. not adding extra work.
- Something that “goes along with me” (adapts to my future circumstances and needs). Possibility to adapt the contents and advice as your situation changes (e.g. when the person gets older, or if the cognition deteriorates)
- Contents and services which are useful and interesting not just because they are related to brain health. It helps me not just with health problems but also with other interests (books, music). It is not all about the brain.
- Good brain games, something that keeps you fit and it is fun. Not boring.
- For people with cognitive impairment to have appropriate support would be very important, as they may feel they are not able to do everything on their own.
- The advice and suggestions should be tailored, “what is relevant to me”
- To have results/ feedback of how you are progressing during the 2 years. I am already trying to have a healthier lifestyle, without knowing exactly which impact this has on my cognitive function.

13 References

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

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

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

[4] https://gdpr.eu/

[5] https://gdpr.eu/article-32-security-of-processing/

[6] https://fhjoanneum.sharepoint.com/sites/GEBLETHE/Freigegebene%20Dokumente/WP10%20-%20Ethics%20requirements/D%2010.2/LETHE%20D10.2.pdf

[7] https://fhjoanneum.sharepoint.com/:b:/r/sites/GEBLETHE/Freigegebene%20Dokumente/WP6%20-%20Data%20Protection%20and%20Ethical%20Impact/LETHE%20D6.1.pdf?csf=1&web=1&e=ZlERss

[8] https://eur-lex.europa.eu/eli/dir/2016/2102/oj

[9] Business architecture, https://en.wikipedia.org/wiki/Business_architecture

[10] LETHE deliverable D9.2 Dissemination and Communication Plan
https://fhjoanneum.sharepoint.com/:w:/r/sites/GEBLETHE/Freigegebene%20Dokumente/WP9%20-%20Innovation%20Diss%20Expl/WP9%20Documents%20and%20KPIs%20for%20the%20PO/LETHE%20D9.2%20v0.9.docx?d=w3a74859d16bd48f09eac20862cd16060&csf=1&web=1&e=yqjQBL

[11] https://fhjoanneum.sharepoint.com/:w:/r/sites/GEBLETHE/Freigegebene%20Dokumente/WP2%20-%20Retrospective%20data%20harmonisation%20and%20design/Requirements/Prof.%20Req.%20Workshop/AgendaAndResults_LETHE_Prof_Req_Workshop.docx?d=wc800501243ec42a297dfd91a1edcfc8b&csf=1&web=1&e=3BgDwL

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

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

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

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

[16] https://www.w3.org/2001/03/WSWS-popa/paper33

[17] https://gdpr.eu/

[18] 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

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

[20] https://teams.microsoft.com/_#/school/files/WP2%20-%20Retrospective%20data%20harmonisation%20and%20design?threadId=19%3A1394a61aa8784f3586a6c8342450d30f%40thread.tacv2&ctx=channel&context=D2.1%2520Structured%2520data%2520repository&rootfolder=%252Fsites%252FGEBLETHE%252FFreigegebene%2520Dokumente%252FWP2%2520-%2520Retrospective%2520data%2520harmonisation%2520and%2520design%252FD2.1%2520Structured%2520data%2520repository

[21] D4.3 LETHE Sensing ecosystem (M16)

[22] https://teams.microsoft.com/_#/school/files/WP2%20-%20Retrospective%20data%20harmonisation%20and%20design?threadId=19%3A1394a61aa8784f3586a6c8342450d30f%40thread.tacv2&ctx=channel&context=D2.1%2520Structured%2520data%2520repository&rootfolder=%252Fsites%252FGEBLETHE%252FFreigegebene%2520Dokumente%252FWP2%2520-%2520Retrospective%2520data%2520harmonisation%2520and%2520design%252FD2.1%2520Structured%2520data%2520repository

[23] https://fhjoanneum.sharepoint.com/:w:/s/GEBLETHE/Eb3r4gZwrrBIsgslrF8k5lEBCOBIIOrnZxDjZaUUDCdD5A?e=nXLhGY

[24] https://radar-base.org/index.php/documentation/introduction/concepts-and-components-of-radar-base/

[25] https://docs.egi.eu/users/check-in/

[26] https://hadoop.apache.org/

[27] https://hadoop.apache.org/docs/r1.0.4/webhdfs.html

[28] http://sparkjava.com/

[29] https://postgrest.org/

[30] https://www.hl7.org/fhir/

[31] http://hl7.org/fhir/smart-app-launch/

[32] https://docs.egi.eu/users/notebooks/

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

[34] https://pytorch.org/

[35] https://mlflow.org/

[36] https://mlflow.org/

[37] https://www.egi.eu/federation/egi-federated-cloud/the-egi-federated-cloud-architecture/

[38] https://docs.egi.eu/users/datahub/

[39] https://teams.microsoft.com/_#/school/files/WP2%20-%20Retrospective%20data%20harmonisation%20and%20design?threadId=19%3A1394a61aa8784f3586a6c8342450d30f%40thread.tacv2&ctx=channel&context=D2.1%2520Structured%2520data%2520repository&rootfolder=%252Fsites%252FGEBLETHE%252FFreigegebene%2520Dokumente%252FWP2%2520-%2520Retrospective%2520data%2520harmonisation%2520and%2520design%252FD2.1%2520Structured%2520data%2520repository

[40] https://teams.microsoft.com/_#/school/files/WP2%20-%20Retrospective%20data%20harmonisation%20and%20design?threadId=19%3A1394a61aa8784f3586a6c8342450d30f%40thread.tacv2&ctx=channel&context=D2.1%2520Structured%2520data%2520repository&rootfolder=%252Fsites%252FGEBLETHE%252FFreigegebene%2520Dokumente%252FWP2%2520-%2520Retrospective%2520data%2520harmonisation%2520and%2520design%252FD2.1%2520Structured%2520data%2520repository

[41] https://teams.microsoft.com/_#/school/files/WP2%20-%20Retrospective%20data%20harmonisation%20and%20design?threadId=19%3A1394a61aa8784f3586a6c8342450d30f%40thread.tacv2&ctx=channel&context=D2.1%2520Structured%2520data%2520repository&rootfolder=%252Fsites%252FGEBLETHE%252FFreigegebene%2520Dokumente%252FWP2%2520-%2520Retrospective%2520data%2520harmonisation%2520and%2520design%252FD2.1%2520Structured%2520data%2520repository