A Demonstration of 4D Digital Avatar Infrastructure for Access of Complete Patient Information

Project acronym: MyHealthAvatar

Deliverable No. D3.8
Architecture Design v2.0

Grant agreement no: 600929

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

COVER AND CONTROL PAGE OF DOCUMENT	
Project Acronym:	MyHealthAvatar
Project Full Name:	A Demonstration of 4D Digital Avatar Infrastructure for Access of Complete Patient Information
Deliverable No.:	D3.8
Document name:	Architecture Design v 2.0
Nature (R, P, D, O)[1]	R
Dissemination Level (PU, PP, RE, CO)[2]	PU
Version:	5.0
Actual Submission Date:	15/03/2015
Editor: Institution: E-Mail:	Manolis Tsiknakis TEI-C tsiknaki@epp.teicrete.gr

[1] **R**=Report, **P**=Prototype, **D**=Demonstrator, **O**=Other

[2] **PU**=Public, **PP**=Restricted to other programme participants (including the Commission Services), **RE**=Restricted to a group specified by the consortium (including the Commission Services), **CO**=Confidential, only for members of the consortium (including the Commission Services)

ABSTRACT:

This deliverable describes MyHealthAvatar architectural blueprints and implementation activities. The emphasis, is on the definition and design of current MyHealthAvatar's architectural scheme based on the final user and functional requirements for implementation based on the selected use cases and scenarios. We followed the IEEE 1471 recommendation that defines an architecture as the fundamental organization of a system embodied in its components, their relationships to each other and to the environment and the principles guiding its design and evolution. The current version of the Deliverable also documents the steps were on capturing stakeholder needs; the process of making a series of architectural design decisions that resulted in a solution meeting those needs, its assessment against the stakeholder needs, and the refining of the solution until it is adequate and captures the architectural design decisions in a complete "Architectural Description". These iterative activities formed the core of the architecture definition process, and are reported in detail.

KEYWORD LIST:

User Requirements, Standards, Guidelines, Protocols, Formats, IT, State-Of-the-Art, Review, Semantics, Security, Epsos, Cloud

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement n^{o} 600929.

The author is solely responsible for its content, it does not represent the opinion of the European Community and the Community is not responsible for any use that might be made of data appearing therein.

MODIFICATION CONTROL			
Version	Date	Status	Author
1.0	06/10/2014	Draft	Manolis Tsiknakis Emmanouil G. Spanakis
1.1	08/12/2014	Draft	Emmanouil G. Spanakis Haridimos Kondilakis Kostas Marias
1.2	10/01/2015		Feng Dong Norbert Graf Zhikun Deng Xia Zhao Sarah Jensen Nikolaos Christodoulou Georgios Stamatakos Eleni Georgiadi Dimitra Dionysiou
1.4	15/02/2015		Feng Dong Emmanouil G. Spanakis Kostas Marias Tsiknakis Manolis
2.0	15/3/2015	Final	Manolis Tsiknakis, Emmanouil G. Spanakis

List of contributors

- Tsiknakis Manolis (TEI-CRETE)
- Emmanouil G. Spanakis (FORTH-ICS)
- Norbert Graf (USSAR)
- Feng Dong (BED)
- Sarah Jensen (LUH)
- Zhikun Deng (BED)
- Xia Zhao (BED)
- Haridimos Kondilakis (FORTH-ICS)
- Kostas Marias (FORTH-ICS)
- Nikolaos Christodoulou (ICCS)
- Eleni Georgiadi (ICCS)
- Dimitra Dionysiou (ICCS)
- Georgios Stamatakos (ICCS)

Table of Contents

1 Executive Summary

MyHealthAvatar proposes a solution for access, collection and sharing of long term and consistent personal health status data through an integrated environment, which will allow more sophisticated clinical data analysis, prediction, prevention and *in silico* treatment simulations tailored to the individual citizen. The technical architecture of the MHA system should be able to support: efficient information collection, long term management of integrated citizen-specific data, effective access mechanisms for data sharing, as well as innovative data analysis using integrated toolboxes. In more details, the MHA technical infrastructure should provide mechanisms for efficient, and long term data management in internal data repositories storing individual data for the avatars; links to external sources; model repositories, information extraction from the web and data collection using mobile apps; semantic data harmonization to support the data/model searching and reasoning.

MyHealthAvatar follows recommendations from relevant VPH activities on "Digital Patient". MyHealthAvatar architectural platform is designed as a multifunctional integrated facility. Its distinctive features include:

- Data and model repositories to provide rich resources of data and models
- ICT services to support data collection with minimal user input, including web information extraction, mobile apps, etc.
- ICT toolbox in support of clinical decision making by using multiscale models and visual analytics.
- Ontology and RDF repositories to support data integration, search and reasoning.
- An ICT architecture that allows for access to data from a range of different sources, and integration of the repositories, the toolbox and the ICT utilities.
- A local cloud solution to support the storage and computational requirements for the avatars without remote data transfer.
- Specifications of open MyHealthAvatar APIs for external, third-party developers.

In this deliverable we present an updated version of MyHealthAvatar's architectural blueprint and implementation activities. We have adopted the provisions of the IEEE 1471 standard that defines an architecture as "…*the fundamental organization of a system embodied in its components, their relationships to each other and to the environment and the principles guiding its design and evolution*".

The current version of the Deliverable also documents the steps taken on capturing stakeholder needs; the process of making a series of architectural design decisions that resulted in a solution meeting those needs, its assessment against the stakeholder needs, and the refining of the solution until it is adequate and captures the architectural design decisions in a complete "Architectural Description". These iterative activities formed the core of the architecture definition process, and are reported in detail.

2 Introduction

The complexity of the MHA technical infrastructure and the corresponding software system is challenging. Complexity, therefore, is a key concern that we would like our system architecture to address. This complexity presents itself in two primary ways:

- ***Intellectual intractability.*** The complexity is inherent in the system being built, and arises from the broad scope and sheer size of the system, the novelty, interdependencies, and the range of technologies employed. Through the MHA architecture we intend to make the system more understandable and intellectually manageable by a) providing abstractions that hide unnecessary detail, b) providing unifying and simplifying concepts, c) decomposing the system into its elementary parts so as to allow for reasoning about its structural properties.
- ***Management intractability.*** This type of complexity lies in the organization and processes employed in building the system, and arises from the size of the project (number of people involved in all aspects of building the system), dependencies in the project, geographically distributed teams, etc. Our system architecture intends to make the development of the system easier to manage by enhancing communication, providing better work partitioning with decreased and/or more manageable dependencies.

Given that we need to decompose the system to address complexity, we must make sure to address the new problems that emerge during this decomposition process. Namely:

- *How do we break this down into pieces?* A good decomposition must satisfy the principle of loose coupling between components, a key objective of the MHA system – given its distributed development; it is facilitated simplifying the problem by dividing it into reasonably independent pieces that can be tackled separately and by designing clean interfaces that enable "component integration"
- *Do we have all the necessary pieces?* The structure must support the functionality or services required of the system. Thus, *the dynamic behavior of the system* must be taken into account when designing the architecture.
- *Do the pieces fit together?* This is a matter of interface and relationships between the pieces. But good fit - that is fit that maintains system integrity - also has to do with whether the system, when composed of the pieces, has the right global and/or emergent properties.

3 Methodology

In achieving the previously stated objectives with respect to architecting the integrated MHA system, through decomposition on the one hand and integration on the other, appropriate methodological processes and tools should be established.

According to Bass et al "…the software architecture of a system is the structure or structures of the system, which comprise software components, the externally visible properties of those components, and the relationships among them"[3]. In parallel, the IEEE Std-1471-2000 defines an architecture as "… the fundamental organization of a system embodied in its components, their relationships to each other and to the environment and the principles guiding its design and evolution"[4].

In addressing these challenges, we have made a selection for the design process and proceeded to the identification of the major views adopting a strict engineering approach, in compliance with the terminology of the IEEE 1471 standard and current best practice in complex system development.

We have decided to follow an agile system development process, that values the efficient delivery and change in the system and software by focusing on the continuous communication with the stakeholders, the iterative design, and the frequent release cycle.

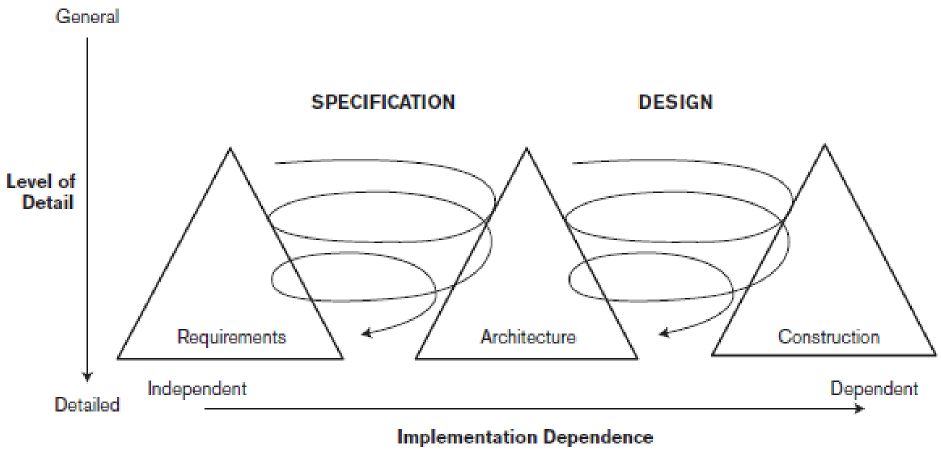


Figure 1: Architecture Definition Context – the three Peaks Model (based on Nuseibeh[5])

Such an incremental and iterative approach is the one proposed by the "Twin Peaks" model of Nuseibeh [Nuseibeh 2001]. Using the author's own words: *"… the spiral life-cycle model addresses many drawbacks of a waterfall model by providing an incremental development process, in which*

[3] Bass L., Clements P. and Kazman R., Software Architecture in Practice, Addison- Wesley, 1997.

[4] IEEE Computer Society. Recommended Practice for Architectural Description. IEEE Std-1471-2000. October 9, 2000. http://standards.ieee.org/reading/ieee/std_public/description/se/1471- 2000_desc.html

[5] Nuseibeh, Bashar. "Weaving Together Requirements and Architectures.", IEEE Computer, 34(3):115–117, March 2001.

developers repeatedly evaluate changing project risks to manage unstable requirements and funding". This approach is refined in the "*three peaks*" model shown in figure 1, which guides our specification, design and system construction efforts.

This interplay between requirements and architecture is justified by the following observations:

- Requirements analysis provides the context for architecture definition by defining the scope and the system's desired functionality and quality properties.
- Architecture definition often reveals inconsistent and missing requirements and also helps stakeholders understand the relative costs and complexities of meeting their concerns. This feeds back into requirements analysis to clarify and add requirements and to prioritize these when trade-offs are made between stakeholders' aspirations and what can be achieved given time and budget constraints.

Therefore this spiral type of architectural specifications has been adopted in MyHealthAvatar. This document presents the second formal iteration of requirements and how these are mapped into architectural decisions; subsequent versions of the architecture specification will elaborate more on the specifics of each architectural view, until our final specification.

According to this process, the following steps were made in order to capture needs, requirements and dependencies between system components as well as with external systems:

- *Capturing stakeholder needs*, that is, understanding what is important to stakeholders (possibly helping them reconcile conflicts such as functionality versus cost) and recording and agreeing on these needs (D2.3, D7.1, D9.1)
- Making a series of *architectural design decisions* that result in a solution that meets these needs, assessing it against the stakeholder needs, and refining this solution until it is adequate (D3.2)
- *Capturing the architectural design decisions* made, in evolving Architectural Description documents (D3.1) and the present updated deliverable.

These activities form the core of the architecture definition process were performed iteratively. In order to comprehend how the complex MHA system will be materialized, one needs to understand what each of its important parts does, how they inter-cooperate exchanging data and sharing services, and how they interact with external systems and applications. The MHA Architectural description is a complex, continuous process of progressive collection and update of items/products to document the architecture covering several views referring to all stakeholders concerns, needs and requirements.

In deliverable D3.2 we described our initial architecture description and methodology. Our methodological approach was, and still remains, centered on the notion of stakeholders, views, and viewpoints in conformance to the provisions of the ISO/IEEE 42010 standard.

In this document we proceed to define an updated version of the envisaged architecture following their approach. In any case the underlying principle of our methodology is to define "just enough architecture", which means to continue until the basic requirements are met and the identified risks are addressed[6].

During this process, we have focused on the following key questions and ultimately requirements of the MyHealthAvatar system, based on deeper and more elaborate understanding of requirements stemming out of the selected use cases and demonstration scenarios – reported in Chapter 4:

i. How does data from external sources is pulled or pushed into the MHA system and populates its repositories? This is topic is addressed in Chapter 5.

ii. Once data are pulled into the MHA environment, how they are semantically aligned and integrated. The related issues and technical approaches are addressed in Chapter 6.

iii. What are the required computational services to guarantee that whatever processing of personal data takes place within the MHA system's boundaries are in line with the legal and ethical framework established for the project and imposed by the European legislation? These are presented in detail in Chapter 7.

iv. What are the computational requirements for the deployment of the MHA technical platform? These issues, i.e. the design and development of a private cloud infrastructure to host the deployment version of the platform are presented in Chapter 8.

v. What are the provisions so that the designed MHA system is accessible and usable by external, third-party developers that wish to develop external, loosy-coupled value-added services? The current level of technical specifications with regards to this very important issue is presented in Chapter 9.

3.1 Purpose of this document

This deliverable aims to present the technical details in relation to the design, implementation and deployment of the MyHealthAvatar technical platform. It is a second version of such a documentation. It represents a naturally evolution of the previous, original attempt, which is in line with the provisions of the methodological framework adopted for use for iterative design the architecture based on evolving requirements and crystallized user scenarios of the project – as they gradually become available. The audience of this Deliverable is both internal and external. The internal use of the document is to act as a point of reference for all development teams regarding technical and scientific decisions taken with respect to numerous critical issues with respect to both functional and well as non-functional requirements of the system. The external use of the document relates mainly to documenting in adequate technical details the ways in which the system and its core services can be exploited by external applications and added-value services.

[6] George H. Fairbanks, Just Enough Software Architecture: A Risk-Driven Approach Marshall & Brainerd. 2010

4 MyHealthAvatar Architecture definition and design

As described in our previous deliverable we have decided to follow an approach that conforms to the ISO/IEEE 42010 *"Systems and software engineering - Architecture description"* standard. In particular we have chosen to base the architecture definition process on the set of viewpoints proposed by Rozanski and Woods[7]. The architectural description document is a live document in the sense that it evolves as the development of the actual system proceeds, as new requirements emerge, or previous decisions are reconsidered. It is therefore natural that some views are not fully described or have been totally eliminated because currently there's not input to drive them. Thus this document is the updated architecture definition description.

4.1 Stakeholders

A stakeholder is anyone who has an interest in or concerns about the system that we actually building. In MyHealthAvatar the most important stakeholders are:

- ✓ ***Patients/Citizens***. In MyHealthAvatar where personalized provision of health and patient empowerment services are to be delivered, patients represent a prominent type of stakeholders.
- ✓ ***Researchers/Medical personnel expert domain Users***. These can be further classified in bioinformaticians, clinicians, users of clinical trials management systems, etc.
- ✓ ***Software Engineers/Developers***. The people who actually build the system.
- ✓ **Administrators */Maintainers***. The people that evolve and sustain and guarantee the good operation and sustainability of the system.

In this document and at the current version thereof we mostly focus on the domain users and the patients, and secondly on the developers stakeholders. Focusing on the expert users/patients means that we elaborate on their concerns, which mostly have to do with the functionality, and some of its quality attributes such as security and usability. On the other hand the developers' concerns relate to the development process, its phases (e.g. design, code, test), and various "satellite" issues like the choice of the programming environment, the development tools, etc. The project has carried out analysis of detailed end-user requirements, reported in D3.1, and needs by collecting an initial set of Scenarios / Use Cases (D2.2). These cases were collected by consortium members through interaction with all MyHealthAvatar's system stakeholders, including citizens/patients, clinical doctors and clinical and IT researchers. From the initial set of scenarios the consortium defined MHA high end clinical demos that were selected for further implementation and evaluation. These demonstration scenarios are close related to the prioritized and final set of Use Cases / Scenarios reported in D7.1 and D9.1.

[7] Rozanski, Nick, and Woods, Eoin. Software Systems Architecture: Working with Stakeholders Using Viewpoints and Perspectives, 2nd Edition. Addison Wesley, 2011.

This user centered approach allowed us to us to select and describe in more details the Scenarios / Use Cases and related requirements / user need proposed for implementation and final demonstration for MyHealthAvatar. Each of these cases addresses a use scenario from a particular user perspective, either as a patient, or as a doctor, or as a clinical or IT researcher.

4.2 Architectural Views

The system context provides an overview of the system and the actors and other systems that it interacts with. A context diagram for the system, when considered a single, unified system, can be seen in Figure 4-1. The main feature of this type of diagram is that it shows clearly the channels of communication with all sub-systems and external systems. It clearly presents the key components that must be built during our work plan. The key components of the related working task include: 1) user requirements – objective 1; 2) Internal data repositories and an internal model repository – Objective 2.1-2.2; 3) ICT architecture that support data access to internal and external resources and data management – Objective 2.3; 4) Data collection utilities – Objective 2.4; 5) Semantics and RDF repository to support data search and reasoning – Objective 2.5; 6) Simulation and data analysis toolbox – Objective 2.6, 2.7 and 2.8; 6) Demo & evaluation – Objective 2.9; 7) Investigation of the legal and IPR aspects of the avatars – Objective 3; 8) Understanding of clinical acceptability – Objective 4; 9) Recommendations for the future work – Objective 5 ; 10) Dissemination and exploitation of the results to influence the future healthcare system – Objective 6

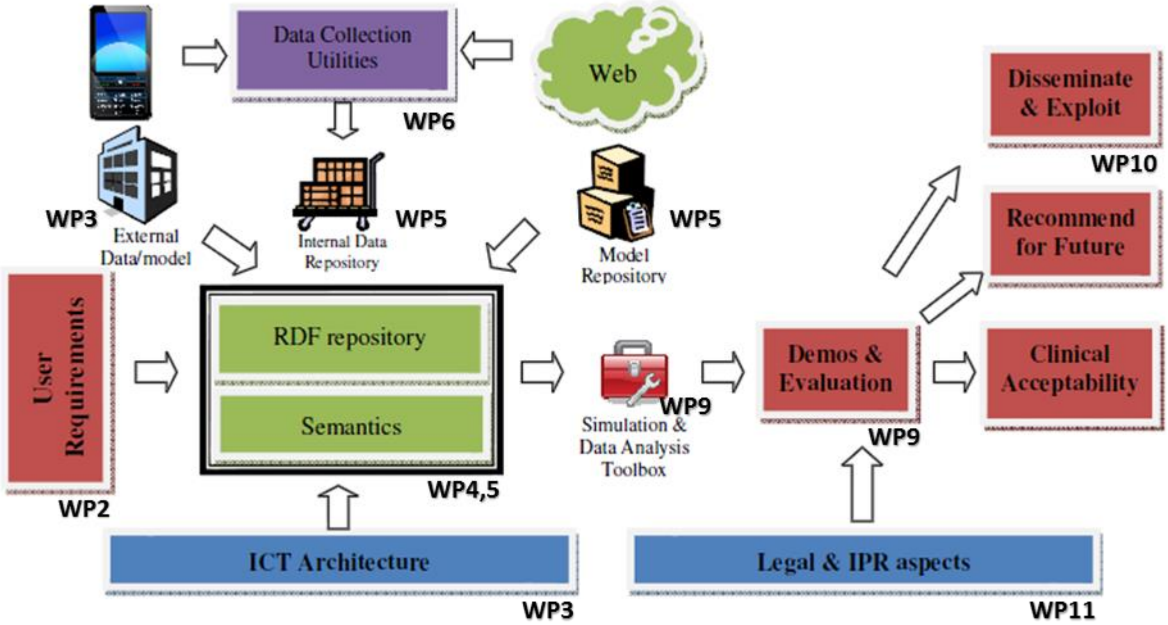


Figure 4-1: MyHealthAvatar platform as a unified system and its interactions with external entities

It is our goal that the requirement analysis of MyHealthAvatar will clearly delineate the borders of the system architecture. According to these requirements, the following are external entities are currently considered for the system under development:

- HIS, EHR and PHR systems
- Drug data
- Social Networks and other sources of online activity of the users
- Model repositories that contains simulation models
- PubMed Repository, Clinical Trials information, news articles, etc.
- Clinical processed data and data from external Warehouse (lab results, images etc.)
- eCRF with filed in data from ObTiMA. Health Avatar with clinical trial related data (i.e. laboratory results, pre-operative state, etc.)
- Third party application (external to MyHealthAvatar) data
    - Diabetes and Emergency Demo
    - Personalized CHF Related Risk Profiles and "Real-Time Monitoring" (CHF)
    - Osteoarthritis (OST)
    - Nephroblastoma (Wilms Tumour) Simulation Model and Clinical Trial (UC-NEPH): In-silico Profiling of Patients and Predictions

A logical view of the architecture based on the required functionality already defined in the project's description of work can be seen in D3.2. At this abstraction level we don't explicitly depict the components' functionality, the details of their interactions, and their dynamic behavior (e.g. when these interactions take place, etc.). In the following paragraphs we are going to describe some of the identified scenarios and the responsibilities of the components, their interactions, etc. will become clearer. MyHealthAvatar platform will be distributed along many computational nodes due to its complexity, functionality, and heterogeneity of components. Our deployment view diagram is shown below:

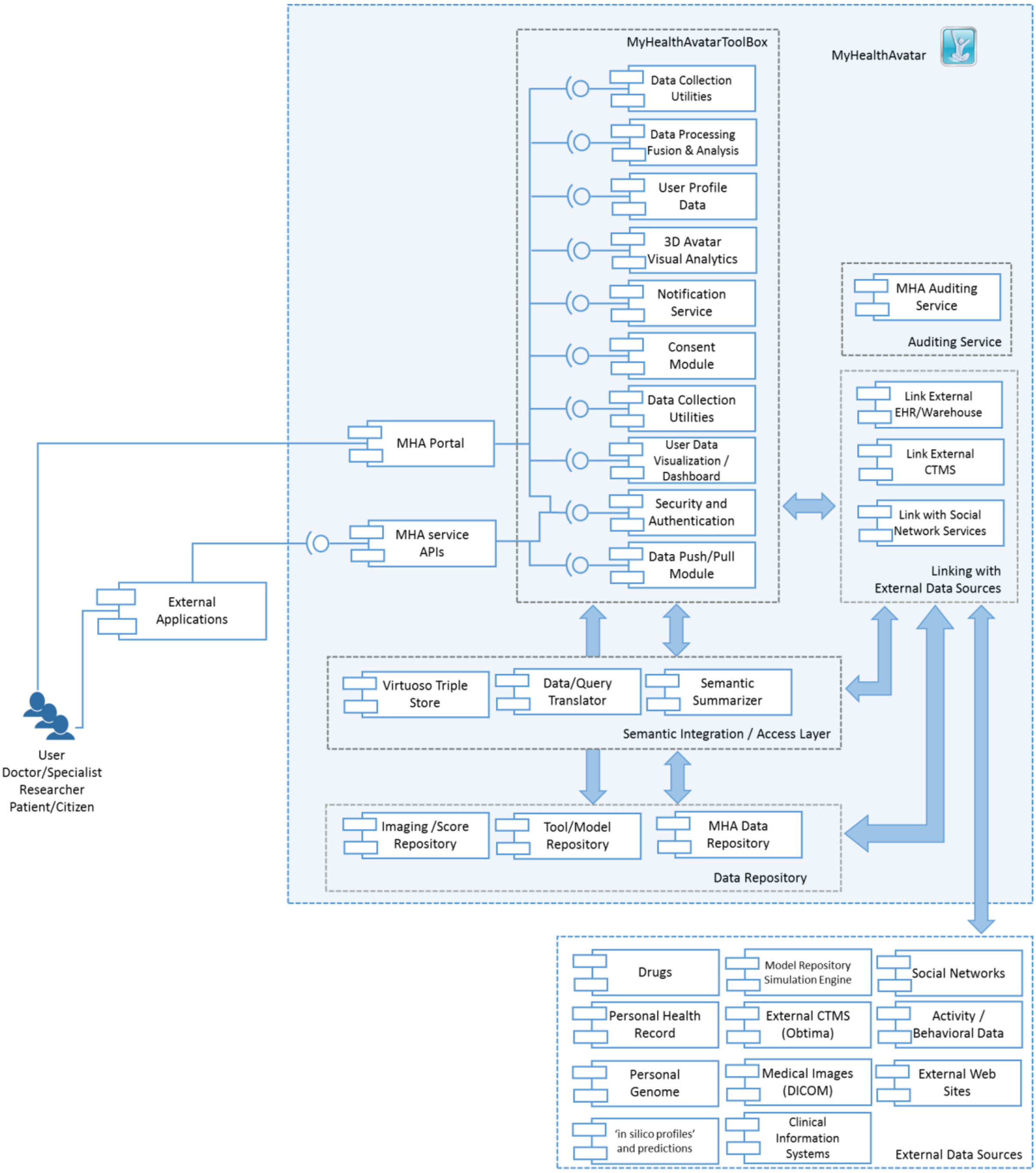


Figure 4-2: The main components of the system and their interactions

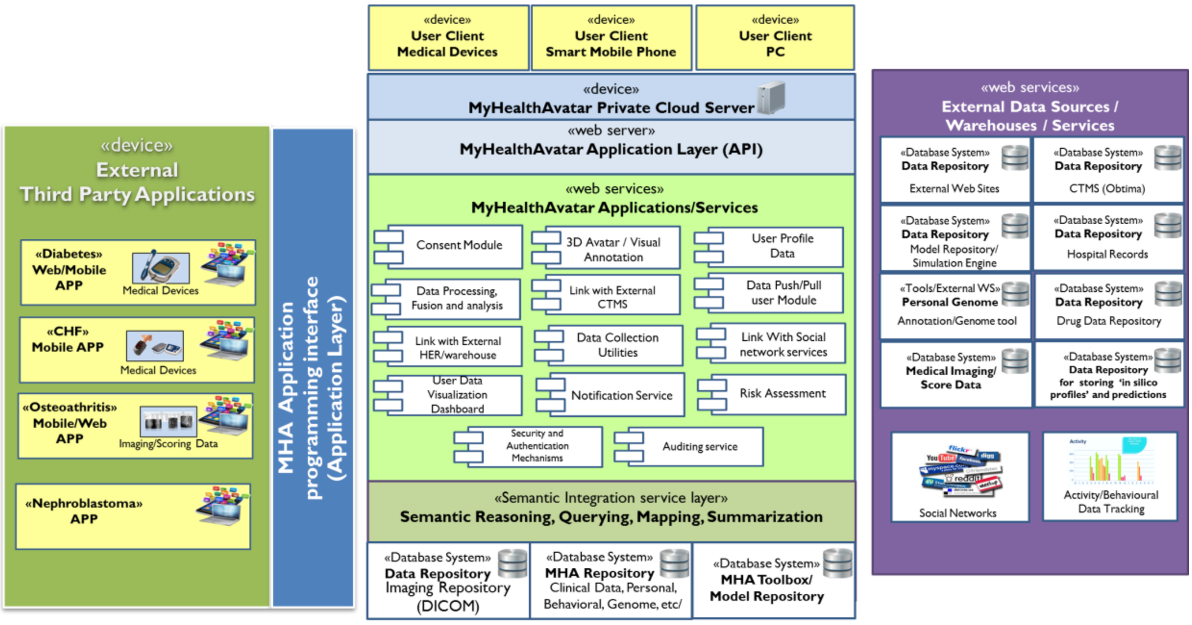

«device»
User Client
Medical Devices
«device»
User Client
Smart Mobile Phone
«device»
User Client
PC
«device»
MyHealthAvatar Private Cloud Server
«web server»
MyHealthAvatar Application Layer (API)
«web services»
MyHealthAvatar Applications/Services
Consent Module
3D Avatar / Visual Annotation
User Profile Data
Data Processing, Fusion and analysis
Link with External CTMS
Data Push/Pull user Module
Link with External HER/warehouse
Data Collection Utilities
Link With Social network services
User Data Visualization Dashboard
Notification Service
Risk Assessment
Security and Authentication Mechanisms
Auditing service
«Semantic Integration service layer»
Semantic Reasoning, Querying, Mapping, Summarization
«Database System»
Data Repository
Imaging Repository (DICOM)
«Database System»
MHA Repository
Clinical Data, Personal, Behavioral, Genome, etc/
«Database System»
MHA Toolbox/ Model Repository
«device»
External
Third Party Applications
«Diabetes» Web/Mobile APP
Medical Devices
«CHF» Mobile APP
Medical Devices
«Osteoathritis» Mobile/Web APP
Imaging/Scoring Data
«Nephroblastoma» APP
MHA Application programming interface (Application Layer)
«web services»
External Data Sources / Warehouses / Services
«Database System»
Data Repository
External Web Sites
«Database System»
Data Repository
CTMS (Obtima)
«Database System»
Data Repository
Model Repository/ Simulation Engine
«Database System»
Data Repository
Hospital Records
«Tools/External WS»
Personal Genome
Annotation/Genome tool
«Database System»
Data Repository
Drug Data Repository
«Database System»
Medical Imaging/ Score Data
«Database System»
Data Repository for storing 'in silico profiles' and predictions
Social Networks
Activity/Behavioural Data Tracking

Figure 4-3: MHA deployment diagram for the system and external third party applications for the demos

MyHealthAvatar Semantic Integration Layer: In Section 0 we present in detail all the information about the semantic integration layer, the common information model and the ontology the reasoning, mapping and query re-writing, and summarization

<u>MyHealthAvatar Database Repository</u>

The data repository is designed to use a public and a private cloud selecting a hybrid approach, see deliverable D3.5 for more details.

- ***MHA central repository (Casandra):*** The MHA central repository is used to store the data of the users of the MyHealthaVatar platform. This includes a range of user activities, health status related information, health profile and medical history. Cassandra is chosen for MHA data repository as described in *D6.2 Design for Data and RDF repositories*. A two-datacenter Cassandra cluster is deployed in the MHA hybrid cloud.
- ***DICOM repository:*** For the needs of the MyHealthAvatar project, DCM4CHEE[8] was chosen to be used as a medical imaging DICOM repository. *DCM4CHEE* is a free and open - source DICOM archive and image manager, forming the server side of a PACS system. It is actively developed and updated, with modules including HL7 and WADO, and is based on JEE, JMX and the JBOSS Application Server. Administration is through a web-based interface and is compatible with a wide range of databases (PostgreSQL, MySQL, SQL Server and Oracle). The DCM4CHEE server

[8] http://www.dcm4che.org/confluence/display/ee2/Home

has been already installed on a LINUX virtual machine (located at FORTH with IP address 139.91.210.41) and can be used directly by the partners for the needs of the project.

- ***Model repository:*** In order to further facilitate citizen and patient self-assessment and education through the use of general models (such as diabetes risk, or educational oncosimulator implementations), the WP5 Tool/Model repository is divided into two parts. Within the MHA platform the Generic Tool/Model Repository part will reside. It will store these models and communicate with the platforms other parts such as the MHA central repository to utilize the collected data. The Nephroblastoma use case will utilize the other part, named the Onco-simulation Tool/Model repository.

MyHealthAvatar Services/Application description

- ***Consent Module:*** Consent will not be sought by using a hardcopy written consent form, but by an online registration because the usual individual who wishes to join the MyHealthAvatar platform will be at home rather than in a clinical environment. The individual will not be able to enter the platform before he has granted consent. The consent form will explain the purpose of the platform, including the explanation of what data the user can upload to the platform and how the data will be used. Furthermore, it will be explained how the user's data will be protected and, following that, the rights of the user pursuant to Article 10 of the Data Protection Directive will be explained and described. To ensure that the user has understood all provided information a well-trained contact person will answer questions. Since consent has to be given explicitly according to Article 8 (2) lit. a of the Data Protection Directive, the consent form will not include a pre-ticked box. Instead, the user will have to take some positive action to signify consent.[9]
- ***3D Avatar / Visual Annotation:*** The 3D avatar is used as a tool for two purposes, namely for patient education, and for the visualization of the patients' health on a 3D anatomical model.
- ***User Profile Data:*** The user profile data include general user information as well as their health related information such as medical history, family history.
- ***Data Processing, Fusion and analysis (upload/store/retrieve):*** The data processing, fusion and analysis tools will validate and process the data within the MyHealthAvatar data repository. The fusion module will integrate the data from different sources. The validation module will look into the data reliability issue, and data analysis will offer functionalities to discover interesting data patterns and assess the quality of life of the users.
- ***Link with External CTMS:*** In relation to MyHealthAvatar ObTiMA can provide data from clinical trials to the platform. This is of utmost importance for the high-end scenario: 'Nephroblastoma (Wilms Tumour) Simulation Model and Clinical Trial (UC-NEPH)' where response to a given treatment is analysed. Through a gateway ObTiMA will be able to download clinical trial data in CDISC-ODM format and these data will be uploaded into the MyHealthAvatar platform for reuse. According to the UC-NEPH scenario it is possible to

[9] Opinion 15/2011, p. 35.

download clinical trial data also from other trials, if this is allowed by the legal framework and informed consent by the patients.

- ***Data Push/Pull user Module:*** With this module users will be able to upload their health profile data onto the platform as part of their profiles.
- ***Link with External HER/warehouse:*** This component enables MHA linking with a hospital information system to allow the exportation of the health related data of the patients into MHA platform. Although the abundance of standardized technologies accommodate the support of a large number of uses cases, due to the time constraints and the primary objectives of the project we focus on the use of epSOS Patient Summary for the retrieval of clinical data. The predominant issues relate with security and transformation of the data followed by the proper annotation in order to be compliant with the syntactic and semantic principles of the system. This component described the technical implementation in respect to the legal work presented in WP11.
- ***Data Collection Utilities:*** This includes the data collection functionality for self-monitoring of user data and behaviours, including step count, calories, locations and other health related information such as heart rate, blood pressure.
- ***Link with Social network services:*** Through this module users will be able to publish their data in social network, allowing for the promotion of healthcare related community activities to exchange relevant experiences.
- ***User Data Visualization Dashboard:*** This module is to visualize the captured user behaviour data, allowing for easy data browsing and self-behaviour monitoring.
- ***Notification Service:*** The avatar will allow for the planning of user activities for healthcare, such as medication, physical exercise. This functionality will send notifications to the users according to the plan.
- ***Risk Assessment:*** We use existing models from Framingham studies which will allow self-risk assessment by the patient themselves in major diseases such as diabetes and hypertension with the purpose of raising self-awareness to improve self-behaviours.
- ***Security and Authentication Mechanisms:*** In chapter 7 we present in detail the security framework of MyHeathAvatar, its principles, privacy and legal framework, trust and security mechanisms employed as well as the the technical details of the deployment framework for MHA cloud security.
- ***Auditing service:*** For cloud computing auditing is important in order to be able to mitigate risks introduced by the Cloud, evaluate the efficiency of controls related to the Cloud and the underline offered platform service and to be able to receive feedback for continuous improvement of all internal process, procedures and service tools for MHA platform. Our security framework contains a (centralized) audit service. Each authentication attempt (both successful and failed), resource access/change, available issue (e.g. server exceptions), etc. needs to be logged by the audit service. Next to the logging functionality, the audit service needs to present the logs to the administrators in such a way that they can be easily consulted and interpreted. MHA will use Could Auditing Data Federation (CADF) data format and interface specification as event model and audit log record specification. MHA audit service focuses on the following areas: i) Identity and Access Management: Verify that only approved

personnel are granted access to service based on their roles and that access is removed in a timely manner upon the personnel's termination of employment and/or change in their roles that does not require the said access; ii) Data Protection: Sufficiency of the data protection policies, procedures and practices at the Cloud Service providers as well as the user organization; iii) Technology Risks: Unique risks related to the use of virtual operating system with cotenants; iv) IT Support: Assess procedures related to incident management, problem management, change and access management in context of use of Cloud services and v) Regulatory: Compliance with regulatory requirements over the protection of information.

4.3 MyHealthAvatar demonstration Use cases

In this section we present the selected use cases for demonstration trying to identify the second phase of the various requirements refining the updated set of requirement that will be used to describe and design the architecture structure of MyHealthAvatar platform. A complete description of these use cases is reported in Deliverable D9.1. In MyHealthAvatar two general categories of scenarios are investigated: System use cases: these are the cases that describe the functionalities of the MyHealthAvatar system from the perspectives of both clinicians and citizens/patients and Clinical use cases: these are the cases that describe how to use the data from the MyHealthAvatar system in real clinical scenarios.

4.3.1 Diabetes and Emergency Demo

This use case is set to empower citizens by providing a supportive environment for the self-management of lifestyles for general health and wellbeing. Particular focus will be cast on risk analysis for diabetes, enabling more effective pre-diabetic care in terms of risk reduction through improving compliance with healthy lifestyle recommendation. The demonstration will allow the users to play a key role in monitoring and managing their own health. Allow multi-modal intervention of lifestyle in a shared decision manner between the doctor and citizens/patients. In the case of pre-diabetes, MyHealthAvatar will be able to demonstrate to the citizens/patients the relations between the outcomes of the self-management/treatment using prediction models. “Behaviour prescription” will be issued based on clinical guidelines and trusted sources (such as NICE), which is expected to include a set of targets in terms of daily activities, calorie intake and energy consumption, etc.

4.3.1.1 Objectives

Enrich the functionalities of the MyHealthAvatar platform in terms of enhancing user experiences in behaviour monitoring and facilitating their lifestyle management by incorporating:

- A verified risk assessment models for diabetes into the MyHealthAvatar platform.
- Personal behaviour intervention modules that allow for planning and remaindering services for daily physical exercises, diet control and medication where necessary.
- A mobile app that allows for easy access to the platform via mobile devices. The mobile app will be built that will allow users to access the MyHealthAvatar platform, especially their personal health data in the MyHealthAvatar data repository from mobile devices such as smart phone.

- An emergency identification service that will allow the attending doctor to identify the patient, using an appropriate way (e.g. his finger prints), in emergency cases where patient is not in a state to provide ID and/or consent. The summarised information made available to the doctor in such cases will help improve management regimen and avoid any complications. The service, called My Emergency Identifier, will be provided by MyHeathAvatar to grant limited access (snapshot view) to data of the patient in case the patient is passed out (unconscious). The attending doctor will use his own ID and the patient's finger print to generate a unique identifier. The MyHeathAvatar service will recognise this identifier and provide summarised information such as:

4.3.1.2 Targeted Users

This demonstration will target healthy citizens to facilitate their self-management of lifestyles for general health and well-being, with a particular focus on the risk assessment for diabetes and lifestyle management for reducing the risk.

4.3.1.3 Data involved

- Medical history, known allergies and sensitivity to drugs
- Insulin resistance for prediabetics
- Information of pregnancy in case of females e.g., if they are in first trimester and certain drugs are contraindicated,
- Impaired glucose tolerance tests during gestational diabetes in females,
- Medications being currently taken by the patient to prevent giving contraindicated medication,
- HBA1C results and fasting and random glucose levels,
- Life style such as food and beverage intake history for the last 1-2 day which could help in identification of cases such as intoxication,
- Exercise history for the last 1-2 days (number of steps etc.) which can be useful in establishing decreased glucose levels, ketoacidosis etc.,
- Also, the contact information of next-of-kin whose consent is legally required in case of relevant medical/surgical intervention.

4.3.1.4 Functional requirements/components

- ***Monitoring***: sensors and mobiles will be used to enable users to easily upload their own health data into the platform. This will include a wide range of data including their activities, movement, step accounts, diet and other health-related behaviours and events. We will allow the users to do this at any time, requiring them to take very simple actions such as press a single button.
- ***Personal Diary***: Storage and management of the health status of the individual and their behaviours. This will rely on techniques of self-monitoring, which will monitor a wide range of daily activities and behaviours of the citizen/patient, including their locations, movements, diet, quality of life, environment, mood, blood pressure, glucose, alcohol, smoking, and other symptoms, etc. Visual analytics will be used to display individual/aggregated data items to allow easy interpretation of the data from the patients. With the search bar of the system, the users can easily send queries about their activities, movements, diet, etc.

- ***Risk assessment and warning***: The system will allow for the progress review of the individual by comparing the personal diary with the behaviour prescription as men. The avatar system will send reminder messages at various priorities in one of the following occasions: medication reminder, due hospital visit (for screening etc.), sign of change of conditions, early sign of one of developing diabetes with constant scored as high risk.
- ***Incorporating personal behaviour intervention modules*** that allow for planning and remaindering for daily physical exercise, diet and medication where necessary.
- (Optional) ***Education intervention***: the life style intervention will be provided with relevant educational materials. The healthy life style for diabetes patients has been intensively researched. There are general recommendations available in terms of diet and life style[10], and other target recommendations for people at certain conditions, such as age and medication11. MyHealthAvatar will deliver these materials to the citizens in needs. For example, when the diary shows that individual did not reach the planned amount of exercise, the contents of the recommended weekly exercises will be delivered to the individual as a reminder.

4.3.1.5 Design & components

The implementation of the PDIAB-EME will be closely coupled with the implementation of the MyHealthAvatar platform. Many functionalities that will be utilised by the targeted end-users of PDIAB-EME will be implemented directly as inherent components of the platform, such as self-monitoring, emergency identifier and behaviour intervention. The implementation will involve technologies able to facilitate the self-management, self-monitoring of patients, to support patient empowerment and engagement, offering a supportive environment from the user's patients by means of offering advice, agreement, assistance assessment and arrangement; and by means of allowing health promotion.

This case is related to the following MHA architecture components

- MyHealthAvatar portal
- User management and consent
- Data repository
- Semantics repository
- Tool/Model repository
- Data collection utilities
- (Optional) link to hospital system

[10] http://www.hopkinsmedicine.org/gim/core_resources/Patient%20Handouts/ (November 2014)

[11] Exercise and Type 2 Diabetes, joint position statement of 'American College of Sports Medicines' and American Diabetes Association

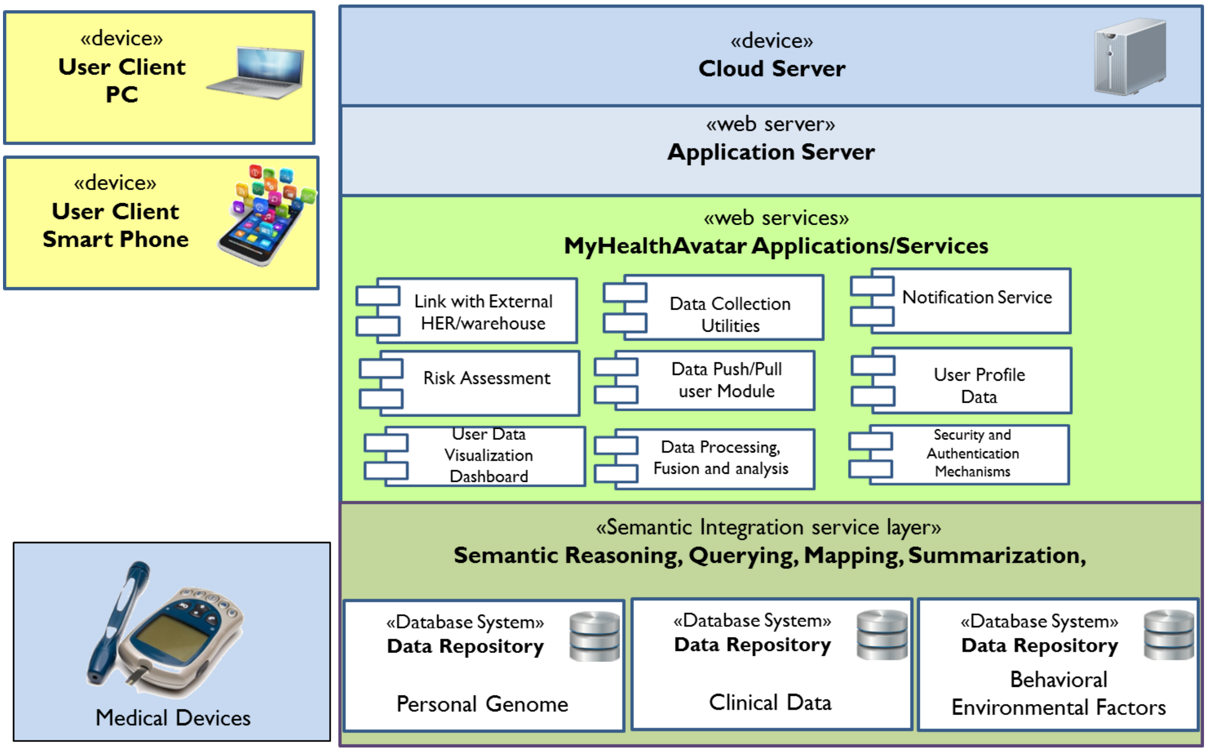


Figure 4-4: Deployment View of the Diabetes Use Case

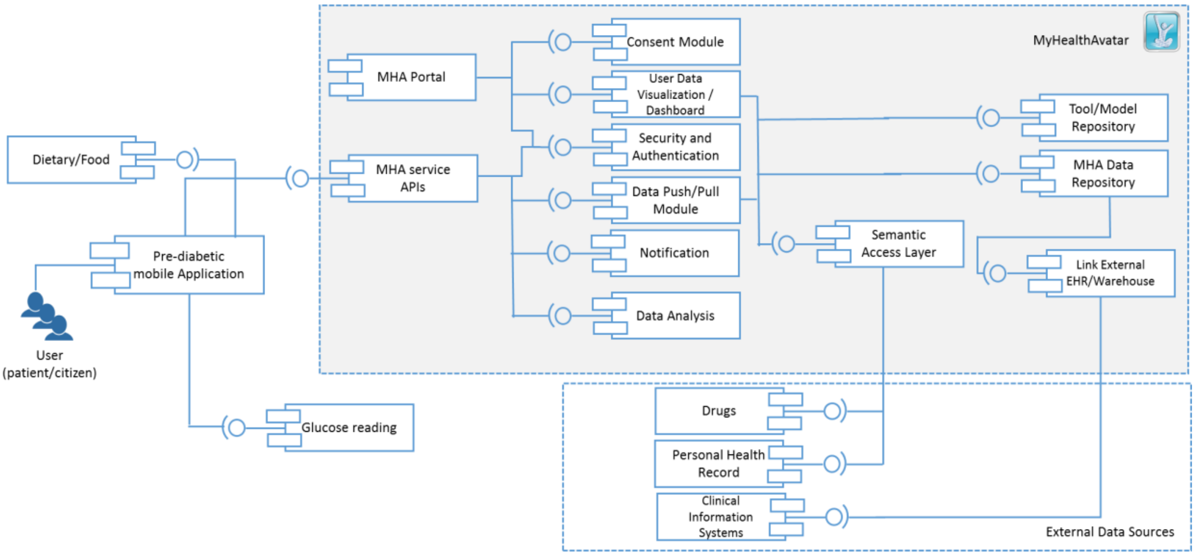


Figure 4-5: Functional View of the Diabetes Use Case

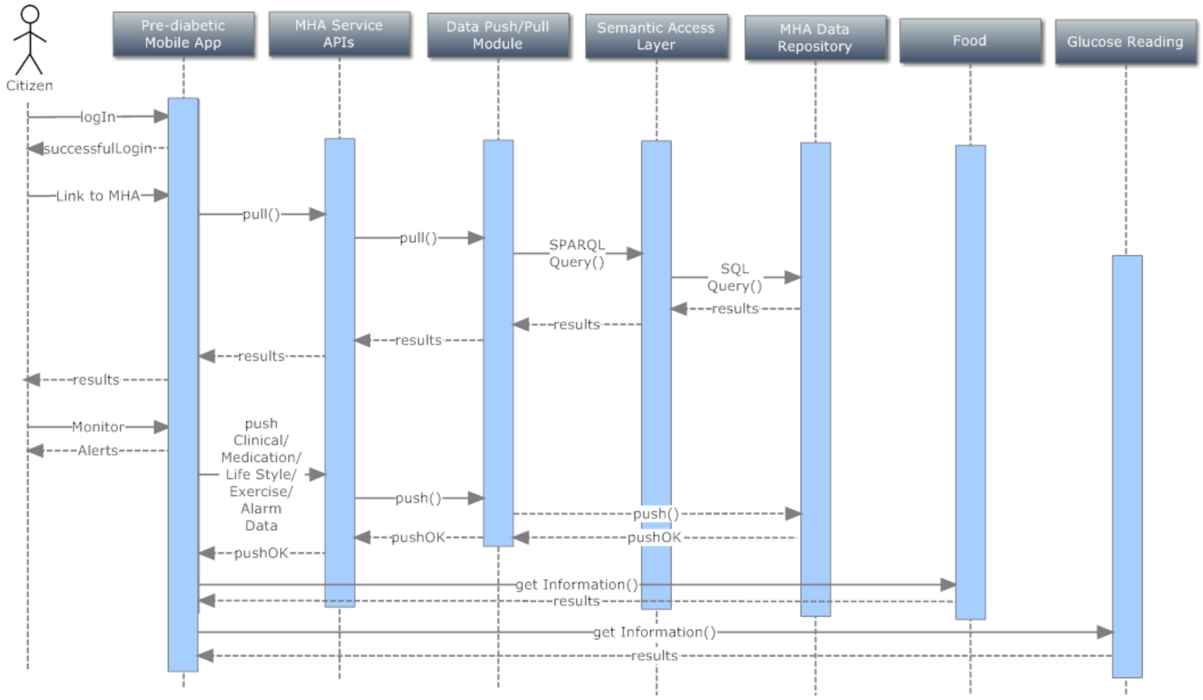


Figure 4-6: Sequence Diagram of the Diabetes Use Case

4.3.2 Personalized CHF Related Risk Profiles and "Real-Time Monitoring" (CHF)

Generally, cardiovascular disorders as chronic diseases require a continuous everyday record for patient’s status. Congestive heart failure (CHF) is a state in which the heart cannot provide sufficient cardiac output to satisfy the metabolic needs of the body. It is commonly termed congestive heart failure (CHF) since symptoms of increase venous pressure are often prominent. Its pathogenesis factors include: Age, Gender, Increased blood pressure, Smoking, Alcohol, Family and medical history, Genetic predisposition, Diabetes, Diet habits and Atherosclerosis. It’s a pathophysiologic state in which the heart, via an abnormality of cardiac function (detectable or not), fails to pump blood at a rate commensurate with the requirements of the metabolizing tissues or is able to do so only with an elevated diastolic filling pressure. Common causes include: coronary heart disease; hypertension; valvular heart disease and the general symptoms a patient/citizen can observe are: shortness of breath, leg swelling and exercise intolerance. The diagnosis is based on physical examinations and echocardiography. The management of the disease primarily involves intervention in order to improve the symptoms and the preventing disease progression mainly by altering the lifestyle (Diet, smoking, alcohol, moderate physical activity) and in many cases by pharmacological interventions.

The proposed scenario is built on the following pillars:

CHF Risk Assessment: In order to tailor the proposed system to the patient's profile and assist physicians in selecting people who are predisposed by coronary disease, hypertension, or valvular heart disease; we build a CHF related risk profile based on a risk appraisal function that is based on the diagnostic criteria [i.e. the Framingham Heart Study (486 heart failure cases during 38 years of follow-up)]. The predictors used are based on Age, Coronary heart disease and Valve disease status provided by the patient Electronic Health Record (EHR), as well as on HR, on blood pressure and on Body Mass Index (BMI) provided by the pulse oximeter, the blood pressure monitor and the weight scale, respectively. The calculated risk probability may be used to alter the default threshold values (higher risk probability adds more constraint on the physiological patterns). Furthermore, we present what else data regarding patients' health status could be embed into the platform towards the creation of a profile with necessary information for both patient and treating physicians. To this respect an approach of presenting data regarding demographic, physiology, diagnostic test results and disease management (i.e. prescribed drugs) is provided.

Real-time patient monitoring: In addition to the above the dedicated clinical personnel should be contacted immediately and possibly intervene in time before an acute state is reached, by changing medication, or any other interventions, in order to ensure patient safety. There is a need to support real-time remote monitoring of patients diagnosed with congestive heart failure and MHA, enhanced with semantic technologies, may host personalized, accurate and up-to-date clinical information. To this end we built a real-time patient/ doctor alarming will be built according to rule-based alarms enabling intelligent alerting of the dedicated physician in case of an emergency. The alarming process will be based on vital signs monitoring and specifically Heart Rate (HR), Pulse Oximetry, and Blood Pressure acquisition, adapted according each specific patient's medical history and age, and even risk predictor's outcome.

4.3.2.1 Objectives

The outcome of this use case in to create a demonstration service able to empower citizens, patients and doctors by providing a supportive environment for the self-management of patients/citizens with cardiovascular disease risks.

We aim at the following objectives:

- Incorporating verified risk assessment models for cardiovascular diseases into the MyHealthAvatar platform
- Enhance and expand the functionalities of MyHealthAvatar platform in real time health monitoring and management
- Create a CHF Real Time mobile monitoring app to allow for easy access to the platform alarming will be built according to rule-based alarms enabling intelligent alerting of the dedicated physician in case of an emergency.

4.3.2.2 Targeted end-users

This demonstration activity of MHA project targets healthy citizens/patients to facilitate their self-management CHF risk assessment for lifestyle management in order to reduce the foreseen risks and doctors that would be able to assess patient's health status. Synthetic (chimeric) patients will be used to demonstrate the CHF real time patient monitoring scenario.

4.3.2.3 Data involved

Citizens/Patient's data required

- Demographic
    - Gender
    - Age
    - Height
    - Weight
    - BMI/BSA
- Genetic
    - CHF related genome data
    - Pharmacogenomic data
- Physiology- Pathology
    - Blood pressure
    - Cardiac flow (BP and pulse)
    - Kidney function
    - Blood test results
    - Coagulation factors
    - Atherosclerosis level
- Differentiating tests
- Associated diseases
- Physical examination
- Imaging
- Electrocardiography
- Echocardiogram
- Protocols/References regarding disease diagnosis/treatment
- Available regimens for prescription and potential alternatives
- Medline references
- Patients hand out from regulatory organizations for disease management

4.3.2.4 Functional requirements/components

We define the "*CHF Real-time patient monitoring"* and the "*CHF Risk Assessment"* service provided by MyHealthAvatar platform in order to:

- **Assist individualized** self-monitoring of patient's own health-status
- Provide **risk analysis for personal risk monitoring** for developing a cardiovascular related episode in the future
- (Optional) **Provide comorbidities and drug interaction information** in both the treating physicians, but also the patient him/ herself regarding negative drug interactions
- **Expand** MyHealthAvatar platform functionalities by **creating an external monitoring** tool for Personalized CHF risk assessment enhancing user experiences. We use a number of medical devices such as Heart Rate, SpO_2, ECG, Blood Pressure, medical sensors together with a mobile application and MHA's schematics layer in order to enable users' ability to easily upload their own health data into the platform and monitor their health status.

- **Link the tool through MHA with external clinical information** systems to acquire specific EHR patient related data providing risk assessment and update data to MHA repository through the ontology-driven semantic integration layers capable of reasoning on the patient data.
- **Incorporate verified risk assessment models for CHF** through MyHealthAvatar platform. We will build a CHF related risk profile based on a number of risk appraisal functions based on specific diagnostic criteria [i.e. the Framingham Heart Study (486 heart failure cases during 38 years of follow-up)]. The predictors used are based on Age, Coronary heart disease and Valve disease status provided by the patient Electronic Health Record (EHR), as well as on HR, on blood pressure and on Body Mass Index (BMI) provided by the pulse oximeter, the blood pressure monitor and the weight scale, respectively. The calculated risk probability may be used to alter the default threshold values (higher risk probability adds more constraint on the physiological patterns).
- **Create a mobile app** that allow for easy access to the platform via smart phones, mobile devices and portable tablet computers. The CHF risk management mobile app allows *MHA users to access MyHealthAvatar platform*, especially their personal health data in the MyHealthAvatar data repository from mobile devices such as smart phone. The application is able to acquire data from medical sensors used by the user to gather biomedical data that can be used to assess if a CHF episode is eminent and to provide appropriate notification to the user. The application sends appropriate notification events to MHA platform with all CHF episodes detected.

4.3.2.5 Design & components

We will implement the functionalities give the following scenario flow of information:

a. Gather all the necessary patient data

b. Creating/connecting MyHealthAvatar profile for this patient

c. Real-time patient data updates to detect possible deviations from normal values

d. Alarm Doctor for possible intervention

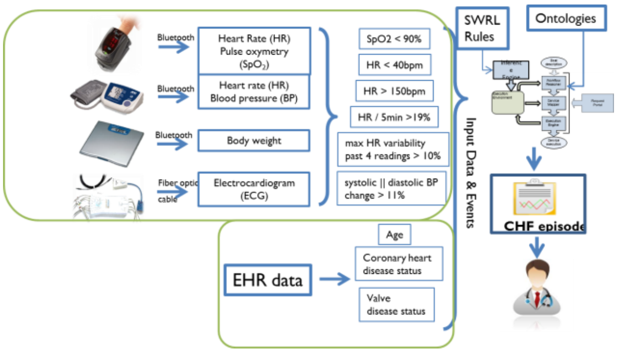
e. Alternative flows will be followed if patient data are not provided in full.

Correspondingly, research and technologies will be developed to facilitate the self-management, self-monitoring of patients, to support patient empowerment and engagement, offering a supportive environment from the user's patients by means of offering advice, agreement, assistance assessment and arrangement; and by means of allowing health promotion. More specifically, these will include:

- Real time CHF self-monitoring and alarm, which will facilitate the biomedical monitoring of the health indicators and biometrics collected by a number of medical devices and a wide range of daily activities and behaviours of the patients, including their location movements, diet, quality of life, environment, mood, blood pressure, heart rate, SpO_2, ECG, alcohol, smoking, and other, etc.
- Visual analytics techniques that allow for easy data browsing and self-behaviour review from the users, including:
  - Dashboard, which presents a summarisation of personal health status in graphs.
  - Timeline, which will display health events and issued allarms along the timeline at different scales.

- Risk assessment of the users based on CHF related risk profiles and a risk appraisal function that is based on the diagnostic criteria [i.e. the Framingham Heart Study (486 heart failure cases during 38 years of follow-up)]. *Where necessary, alerts will be issued to ask the users to check their health with doctors.*
- Comorbidities and Drug Interaction: there is a need for providing information in both the treating physicians, but also the patient him/ herself regarding negative drug interactions.

This case is related to the following MHA architecture components



- MyHealthAvatar portal
- User management and consent
- Data repository
- External link to Electronic Health Records, Clinical Information Hospital system repository
- Semantics layer/repository, data extraction and data query translation
- Tool/Model repository
- Data collection utilities

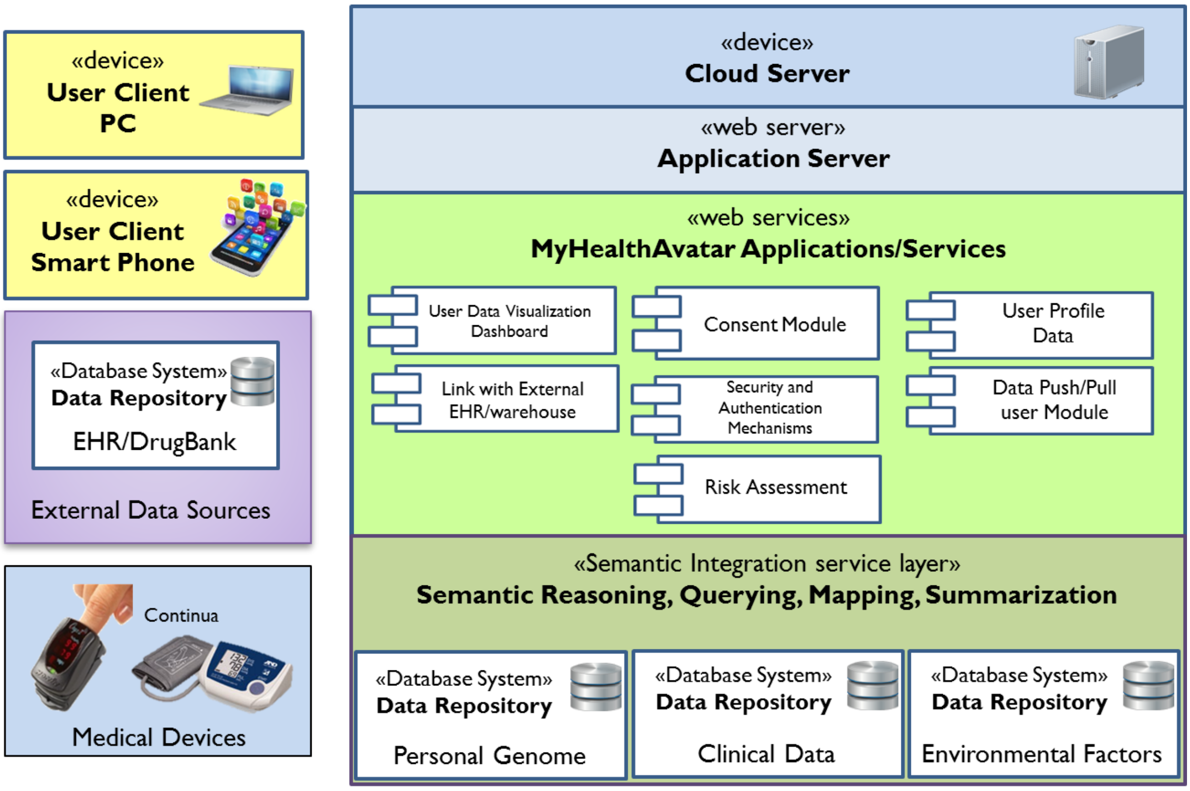


Figure 4-7: Deployment View of the CHF demonstration use case

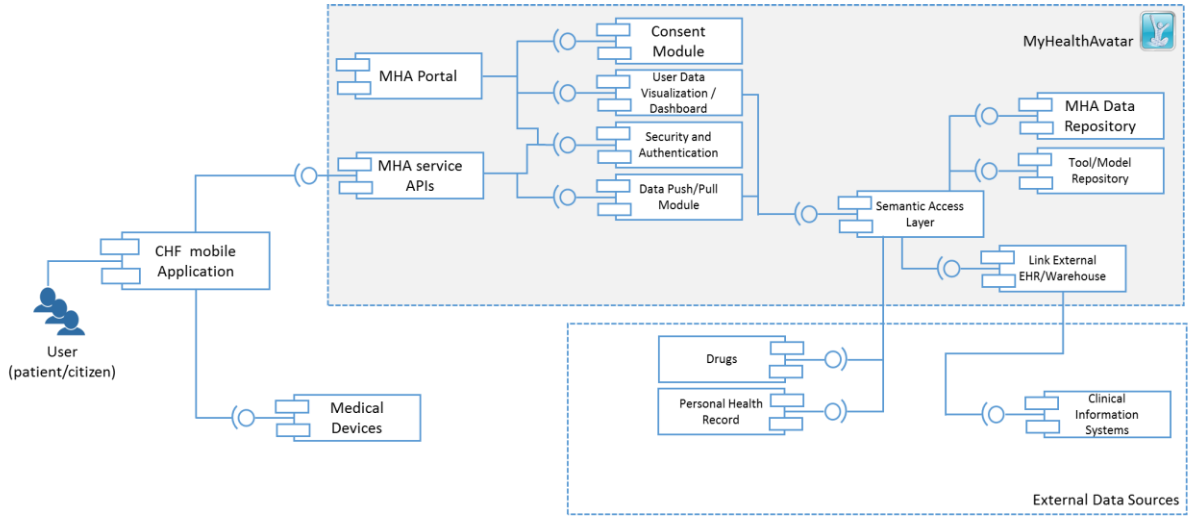


Figure 4-8: Functional View of the CHF Use Case

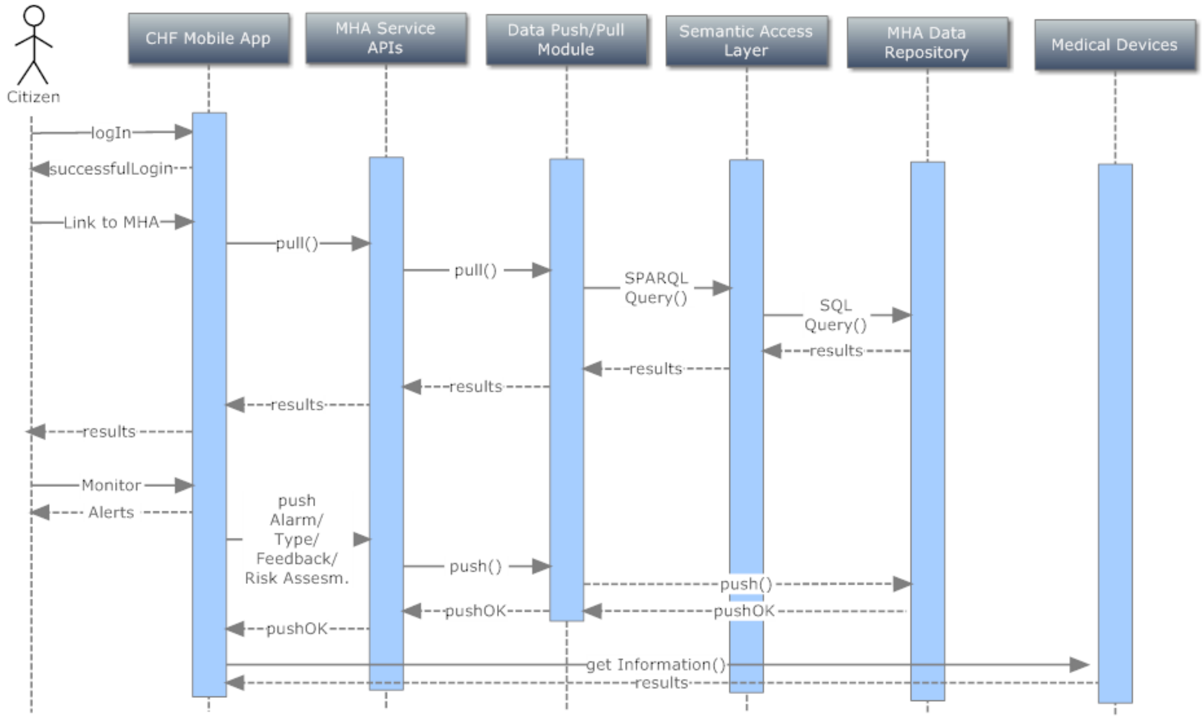


Figure 4-9: Sequence diagram of the CHF Use Case

4.3.3 Osteoarthritis (OST)

MHA offers a one-stop service for citizens for data collection and self-management such as monitor, record and education. Precisely, the system will support the storage of behaviours and daily activities of citizen. It will function as a supportive environment to empower normal citizens in looking after their own health, raising their self-awareness of any potential risk of developing diseases while encouraging their healthy lifestyles in terms of doing routine daily exercises, stopping smoking and controlling their diet. Therefore, naturally many existing functionalities in MHA can be directly used for the needs of osteoarthritis use case. In addition, we will incorporate genetic predisposition evaluation services for examining if an increased risk of developing osteoarthritis exists, which will be used by the citizens in order to understand their personal risk of developing osteoarthritis, and the impact of their behaviour and lifestyles towards the risk.

4.3.3.1 Objectives

The objective of this demo are to empower both patients/citizens and medical professionals by providing a supportive environment for the long-term management of osteoarthritis condition. Medical professionals (such as GPs) will be able to review together with patients a plethora of clinical and personal health information regarding the health status of patients/citizens through MHA platform. The related data (medical history, clinical examination, imaging data, evaluation metrics for measuring knee pain range of motion of the knee joint) will be properly visualized and presented using interactive multi-scale visualization techniques. This blend of medical imaging metrics and personal activity information, will give a better insight of the condition regarding OA diagnosis or progress and will allow the clinician to assess the situation in a more personalised fashion. In case where a GP is reviewing this information, it may also act as a baseline for better assessing if a referral to an expert is needed. In the suggested scenario, advanced personalized healthcare will also be enhanced by genomic predisposition evaluation for developing osteoarthritis. Although this might not be applicable at present, it is important to include it in the scenario in order to emphasise the vision on how MHA can really influence decision support in the future.

Patients/Citizens will be able to access a platform that will monitor their daily dietary and ambulatory activity and warn them, if they do not meet the recommendations that have been given to them (e.g. target activity, supplements etc.). Moreover, semi quantitative metrics, regarding knee pain and range of motion of the knee joint, will be collected periodically. The monitoring will rely on techniques of self-life logging for enhancing the patient engagement. Also, the platform will function as a supportive environment to the patients by means of offering advice and assistance. It is expected that a good knowledge of the condition will lead to enhanced patient behaviour. Thus, the demonstration will focus on how the users can play a key role in monitoring and managing their own health and become co-producers of their OA health management together with their GP.

4.3.3.2 Targeted end-users

This demonstration will target medical professionals, patients and healthy citizens with high risk of developing osteoarthritis in order to provide a complete, long-term management of the condition through lifestyle monitoring. Focus will also be on the genomic predisposition evaluation for developing osteoarthritis and lifestyle management for reducing the risk.

4.3.3.3 Data involved

The demonstration will include data for:

- Multi-level chimeric Patient information with osteoarthritis will be used in order to demonstrate the scenario in a case (e.g. progression of OA symptoms), and the course of actions that it will be taken for the health management. All this data in the platform will be used to highlight the necessary interplay of personal health information (e.g. activity, nutrition) to the medical imaging metrics. Self-care schemes will be adopted to help users improve their compliance to a healthy lifestyle.
- Healthy citizens with high risk of developing osteoarthritis in order to show the sequence of actions that will be taken in order to alert the citizen about being in the risk group and the course of actions for health management. Behaviour intervention schemes will be suggested to the citizen for compliance to healthy lifestyle.

4.3.3.4 Functional requirements/components

- A visualization toolbox providing proper interactive, multi-scale visualization techniques of the data related to osteoarthritis disease (e.g., medical history, clinical examination, imaging data, evaluation metrics). These techniques will offer a useful input to medical professionals in order to carry out personalized medicine and for better follow-up.
- **Visual analytics** will also be used to display aggregated lifestyle data aiming to easy interpretation by both citizens (patients and healthy) and medical professionals.
- **Data collection**: The users will be able to easily upload their own health data (e.g. activities, movements, step counts, diet etc.) into the platform using their mobiles. Part of this data are also related and will be used by the osteoarthritis use case (e.g. activities, movements, diet). Evaluation metrics, strongly related with the osteoarthritis condition, will also be collected periodically. These metrics include a number with regard to a certain pain-scale and a number with regard to the range of motion of knee joint (in degrees).
- **Personal Diary**: This diary presents and manage the patients/citizens' health status and behaviours, including diet, movement, environment, mood, smoking, symptoms etc. Visual analytics will be used in order to display individual or aggregated data for easy interpretation from the patients/citizens. The personal diary will be used by the self-care module described below.
- **Guided interventions** for patients/citizens through the MHA osteoarthritis service which will review the progress of the patient/citizen by comparing the data originated from the patient diary module with the guidelines given to patient/citizen by the medical professional. If the patient/citizen did not manage to reach these special, periodic targets, a reminder service we warn the patient. The general recommendation for the osteoarthritis condition will be delivered to the patients/citizens in needs, as it is expected that a good knowledge of the condition will lead to enhanced patient behaviours.
- MHA will incorporate **genomic predisposition** evaluation for estimating the risk of developing osteoarthritis for a patient/citizen. When a high risk of developing osteoarthritis is revealed for a healthy patient, he will be informed and guided for modifying and adopting a healthier lifestyle.

4.3.3.5 Design & components

The figure on the right is a diagram that shows the implementation of the Osteoarthritis case and its relationship with the MHA platform. It shows the interactions between MHA platform, external resources and the users regarding the use case that was described (patient and GP reviewing the blended information for a more personalised assessment of OA risk or progression). It should be stressed that the implementation of the scenario will be closely coupled with the implementation of the MHA platform. Many functionalities that will be utilised by the targeted end-users of the Osteoarthritis case will be implemented directly as inherent components of the platform.

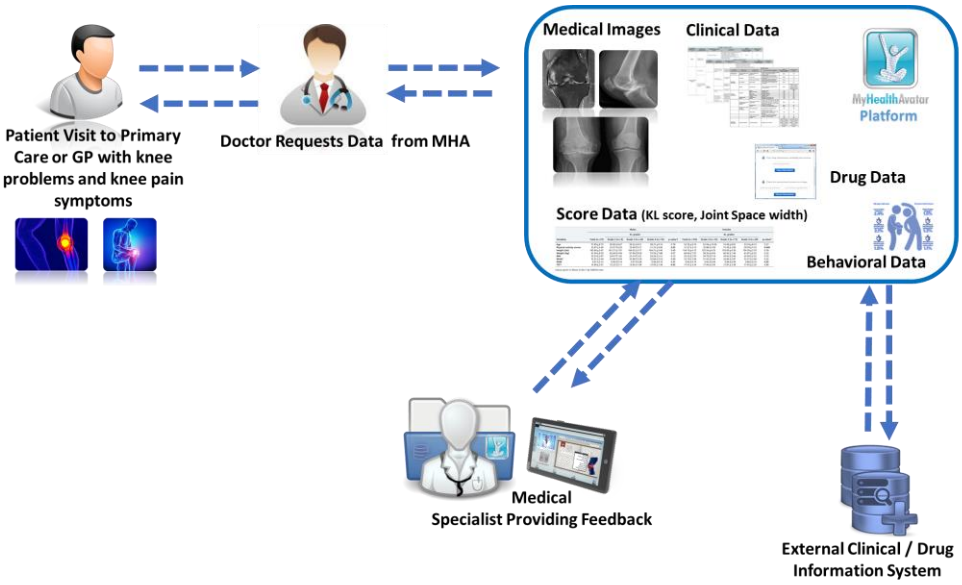

Patient Visit to Primary Care or GP with knee problems and knee pain symptoms
Doctor Requests Data from MHA
Medical Images
Clinical Data
MyHealthAvatar Platform
Drug Data
Score Data (KL score, Joint Space width)
Behavioral Data
Medical Specialist Providing Feedback
External Clinical / Drug Information System

Figure 4-10 Osteoarthritis case (left) the link with MHA Platform

In line with the scenario described, a number of R&D activities will be performed in order to facilitate the implementation of the functionalities described in detail in D9.1.

This case is related to the following MHA architecture components:

- MHA portal
- User management and consent
- Data/ Imaging repository
- Semantics search/repository
- Tool repository
- Data collection utilities

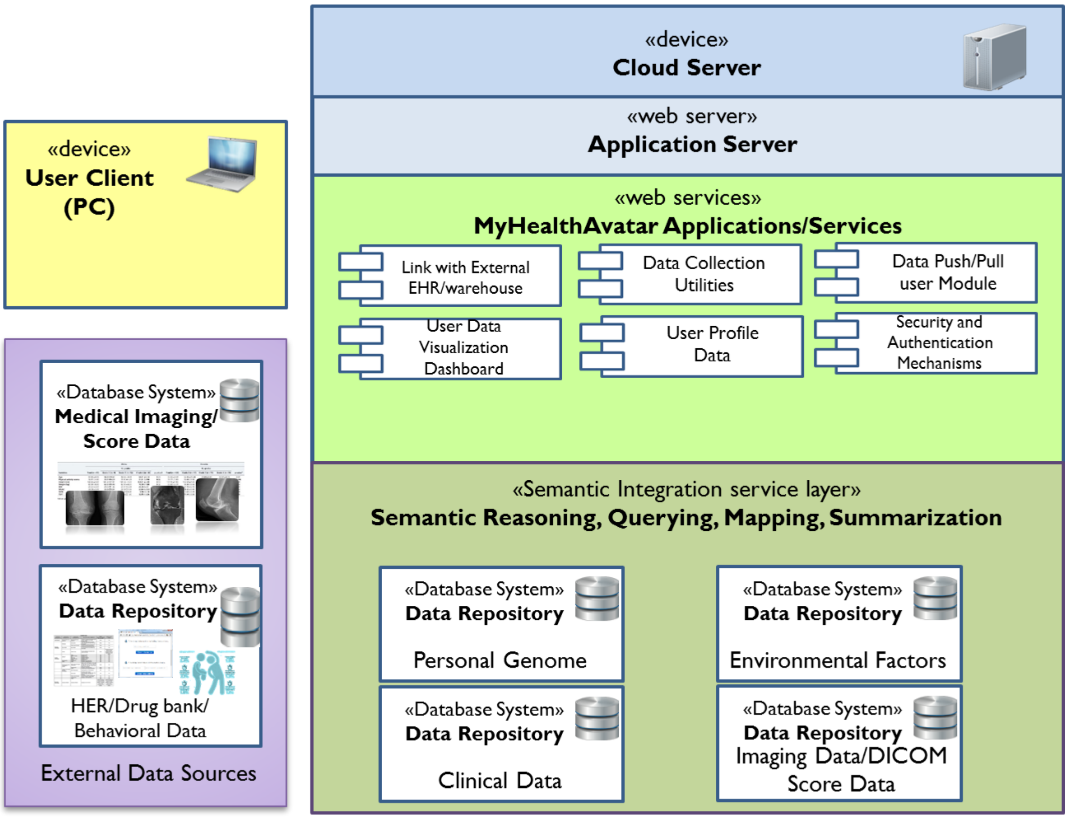

«device»
Cloud Server
«web server»
Application Server
«web services»
MyHealthAvatar Applications/Services
Link with External EHR/warehouse
Data Collection Utilities
Data Push/Pull user Module
User Data Visualization Dashboard
User Profile Data
Security and Authentication Mechanisms
«Semantic Integration service layer»
Semantic Reasoning, Querying, Mapping, Summarization
«Database System» Data Repository
Personal Genome
«Database System» Data Repository
Environmental Factors
«Database System» Data Repository
Clinical Data
«Database System» Data Repository
Imaging Data/DICOM Score Data
«device»
User Client (PC)
«Database System» Medical Imaging/ Score Data
«Database System» Data Repository
HER/Drug bank/ Behavioral Data
External Data Sources

Figure 4-11: Deployment View of the OA use case

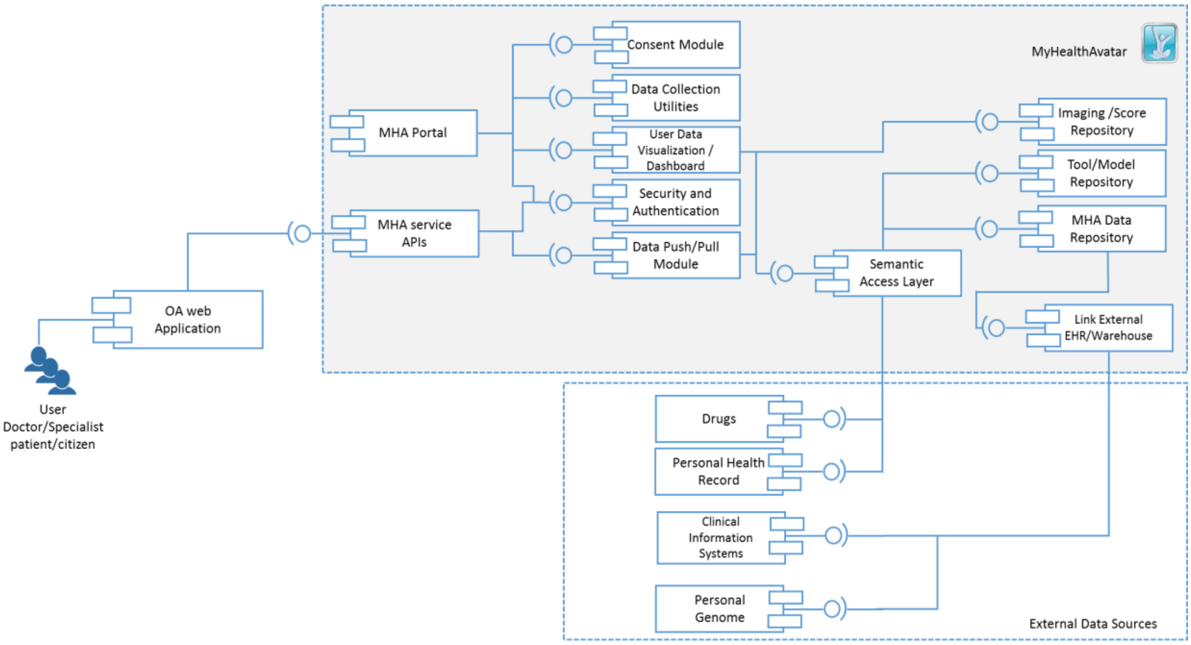


Figure 4-12: Functional View of the OA Use Case

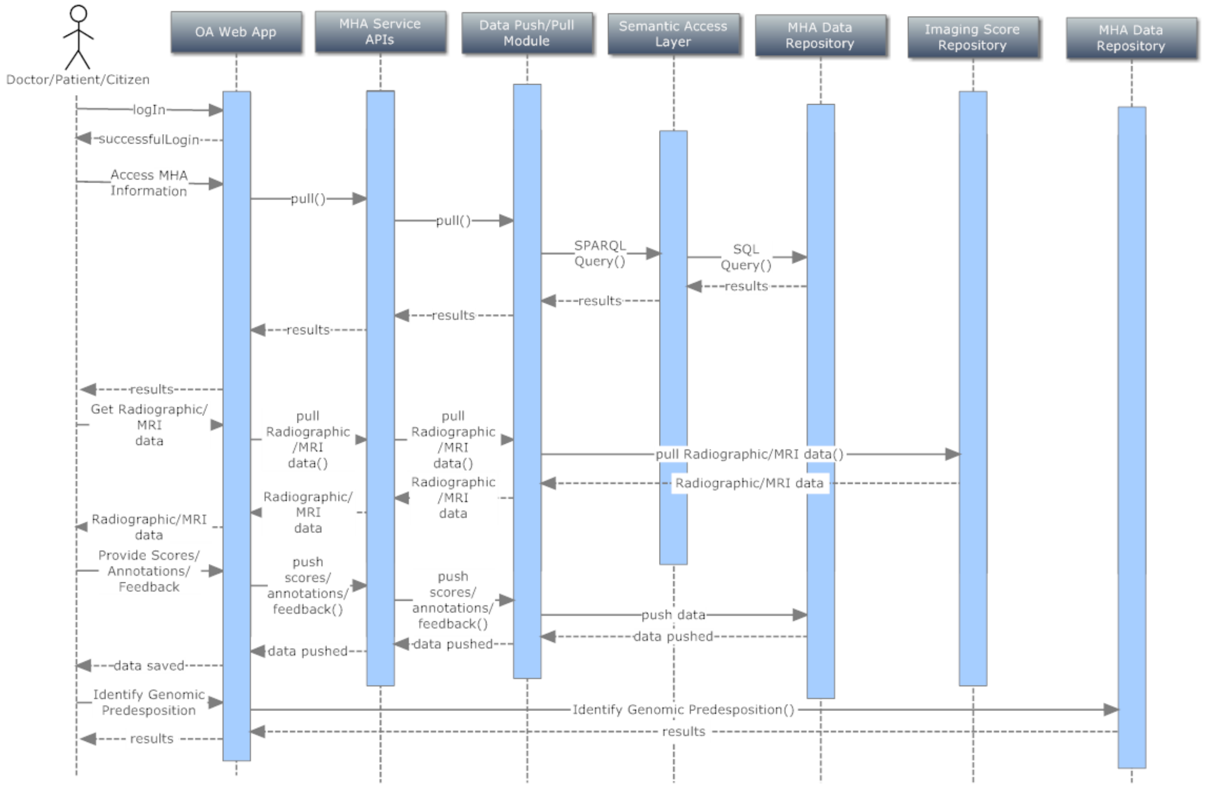


Figure 4-13: Sequence diagram of the OA Use Case

4.3.4 Nephroblastoma (Wilms Tumour) Simulation Model and Clinical Trial (UC-NEPH): In-silico Profiling of Patients and Predictions

The outcome of this high-end scenario is to provide a tool which produces the 'in-silico profiling' of nephroblastoma patients and performs 'in-silico' predictions of therapeutic schemes outcome.This can be used in a fourfold way:

1. To demonstrate to patients and/ or parents of patients how a given tumour will respond to preoperative chemotherapy. This will help in explaining diagnosis and treatment of nephroblastoma to patients and/or parents of patients. Such a demo will not use the actual data of the given patient.
2. To give physicians treating a patient with a nephroblastoma the ability to check how this specific nephroblastoma will respond to preoperative treatment with vincristine and actinomycin-D.
3. To provide clinical researchers and modelers a powerful tool to define an in silico patient profile and further exploit it in other modelling approaches and VPH projects. Moreover, it could serve as a statistical tool to categorize patients (by associating their clinical and *in silico* profiles) and define ranges of model parameter values to guide the process of model adaptation for new patient cases.

4. To demonstrate to citizen what 'in silico' models/tools can do today. This can serve as a learning environment for 'in silico' models and will help to disseminate the importance of 'in silico' models in medicine to the public, to medical stakeholders, industry and funding agencies. It is pointed out that the purpose of the in silico experimentation functionality is currently limited to the education of the public so that they can be prepared for the future translation of thoroughly clinically validated models to clinical practice.

4.3.4.1 Targeted end-users

The high-end scenario will target patients and parents of patients, paediatric oncologists, researchers and the public as given by the 4 different objectives.

4.3.4.2 Data involved

The demonstration will include data for:

The clinical partner (USAAR) will record data regarding radiology, histology, biological markers on blood and urine tests and genetic counselling. In addition biomaterial for molecular and genetic research will be available in order to find new biomarkers and targets for new compounds in the future. This will be done by storing and analysing biomaterial by a wide range of molecular and proteomic technologies from patients enrolled in the new nephroblastoma protocol and having given consent for this research. The aforementioned data will be pre-processed by the VPH modelling partner (ICCS) and transformed appropriately into Wilms Oncosimulator input. More specifically, clinical data, imaging data (MRI at the time of diagnosis and after preoperative chemotherapy) and molecular data (miRNA) will be exploited.

4.3.4.3 Functional requirements/components

The application of Nephroblastoma Oncosimulator into clinical practice will incorporate the following functionalities:

- **In-Silico Profiling of Patients**: Nephroblastoma diagnosis is based on a variety of multiscale data. These data constitute the multiscale clinical profile of the tumour. After the necessary preprocessing of the available data of nephroblastoma patients, the data are fed into the nephroblastoma simulation model. By integrating insights from the personalized multiscale clinical profile of the patient, numerical parameter studies and any information that can be gleaned from the experimental and theoretical biology literature, semiautomatic adaptation of the model parameters is conducted. The determined model parameter values serve as a patient record for in silico tumor characteristics and form the 'insilico profile' of the patient. Training the model with a patient's data gives a more accurate description of the specific kinetics of disease progression.

  The basic steps of the 'in silico profiling' demonstrator are:

  1. The citizen or the clinician logs into the ICCS onco-simulation system
  2. If a clinician is the one that requests the generation of the 'in-silico profile', then a list with all the available citizens who have shared their data with this specific clinician is presented. The clinician chooses a citizen. If a citizen is the one that have requested the generation of the "in silico profile", then this step is omitted.

3. The citizen or the clinician requests the generation of an ‘in silico profile’. A notification email is sent to bio-informatician (modeler).
4. The bio-informatician (modeler) logs into the platform.
5. The bio-informatician (modeler) uses the MHA platform in order to access and retrieve the necessary data from MHA data repository, hospital information systems(HIS), clinical trial management systems (CTMS) etc. If the citizen participates in a clinical trial, his/her clinical data are anonymized/pseudonymized.
6. The bio-informatitian generates the ‘in silico profile’ of the citizen (offline) and uploads it to platform. He may also alter an existing ‘in---silico profile’.
7. A notification email is sent to the citizen or the clinician that the requested ‘ in silico profile’ is ready.
8. The citizen can monitor the status of ‘in silico profiling’ by entering a status page.
9. The citizen or the clinician retrieves the ‘in silico profile’.

- **In-silico predictions:** This high end-scenario will demonstrate the response of preoperative chemotherapy in nephroblastoma. The demonstration will be graphically based. The user can select between 2 drugs treatment (actinomycin-D and vincristin in localized unilateral tumours over 4 weeks) and 3 drug treatment (addition of doxorubicin in metastatic tumors over 6 weeks). The tumour volume change will be dynamically shown over time (4 to 6 weeks, depending on the metastatic state). The user can interrupt the display at any timepoint between diagnosis and end of preoperative treatment. Additionally UC-NEPH scenario will give the tumour volume at diagnosis and at any time during the preoperative treatment. The difference of the tumour volume between diagnosis and end of preoperative treatment will be displayed. Depending on the availability of DWI-MRI, cell density will be shown by mapping ADC values on the tumour. This can be given as a parameter to the scenario. The clinician (paediatric oncologist) runs a number of experiments in silico (= on the computer) simulating the most likely response of the tumour to the most relevant candidate chemotherapeutic schemas. The outcomes of the simulations (predictions) help the oncologist decide the appropriate treatment plan.
  The basic steps of the demonstrator of the prediction simulations are:

1. The clinician logs in into the ICCS onco-simulation system.
2. A list with all the available citizens who have shared their data with this specific clinician is presented. The clinician chooses a citizen.
3. The clinician can create an “in silico profile”, if the citizen doesn’t have one. If the citizen already has an ‘in-silico’ profile then it can be reviewed and/or altered.
4. The clinician chooses the treatment scheme that will be simulated. Three options are available:
   a. Free Growth,
   b. Actinomycin-Vinctristine,
   c. Actinomycin-Vinctristine-Doxurubicin.

5. The clinician chooses the clinical question that the simulation will address. Three options are available:
   a. When to start Chemotherapy?
   b. How many days after chemotherapy to proceed to survey?
   c. Which chemotherapeutic scheme (administration points) is preferable? This step is omitted if in case 3 “Free Growth” is selected.
6. The clinician sets the simulation parameters. Depending on the treatment scheme and the clinical question that the simulation must address, different parameters are requested.
7. The clinician triggers the simulation.
8. A notification email is sent to the clinician when the simulation has been completed.
9. The clinician can monitor the status of the simulations by entering the status page.
10. The clinician reviews and retrieves the predictions.

4.3.4.4 Design & components

The core technology components of this use case are:

- MR: Onco-Simulation Tool/Model Repository.
- Simulation Engine
- DR: Data Repository for storing ‘in silico profiles’ and predictions.
- Email Notification Service

Architectural support

This use case is related to the following MHA architecture components:

- MHA portal
- User management and consent
- External link to Hospital Information System (HIS) and CTMS: Clinical Trial Management System.
- Data collection utilities

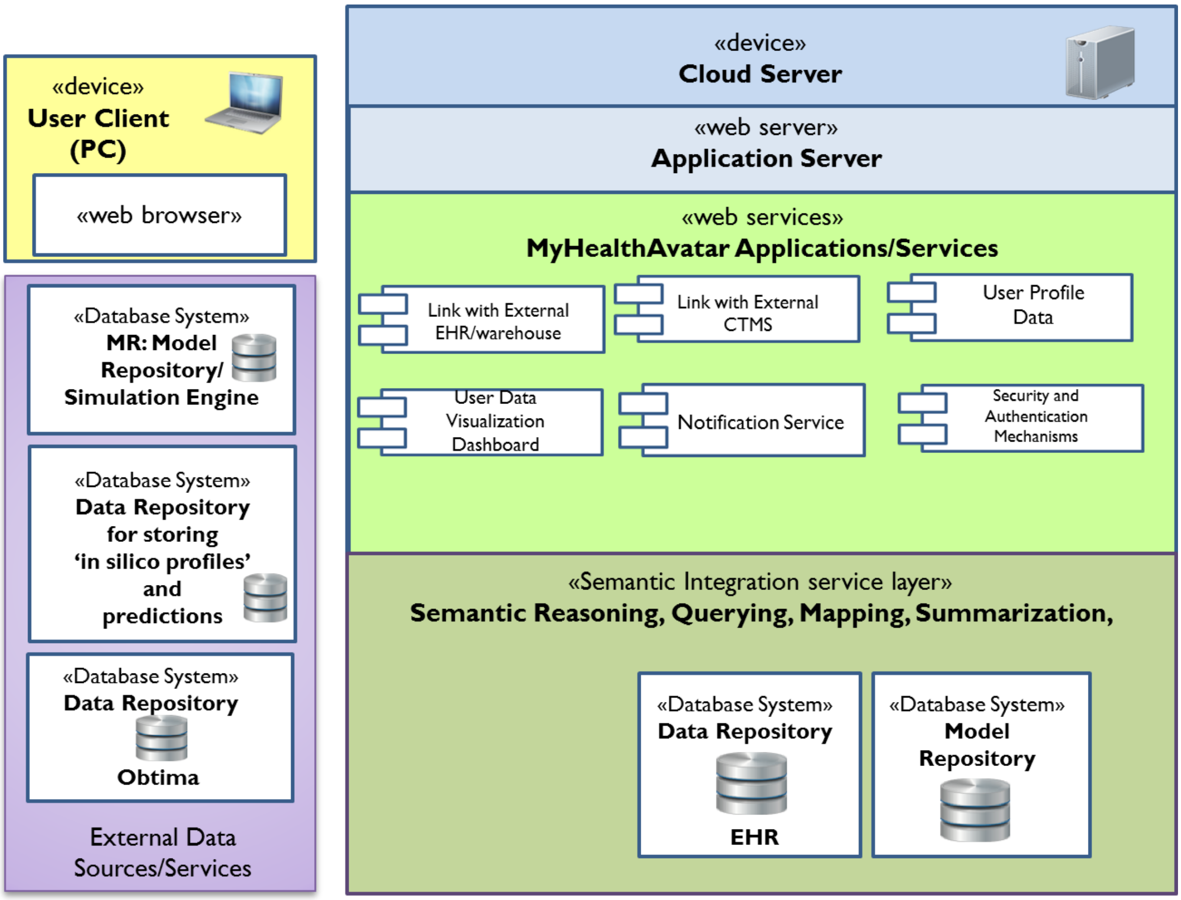


Figure 4-14: Deployment View of the Nephroblastoma use case

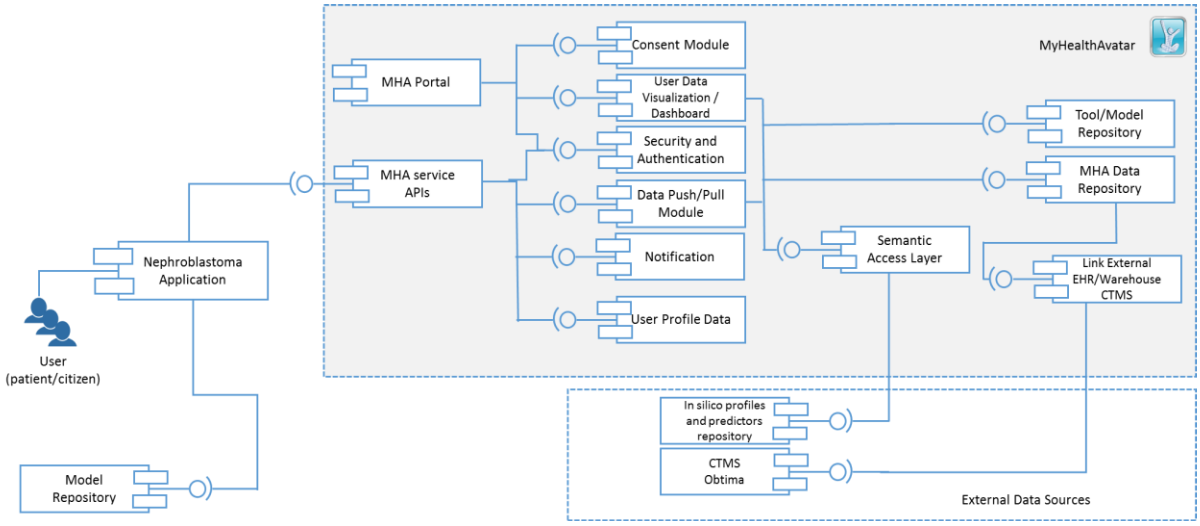


Figure 4-15: Functional View of the Nephroblastoma Use Case

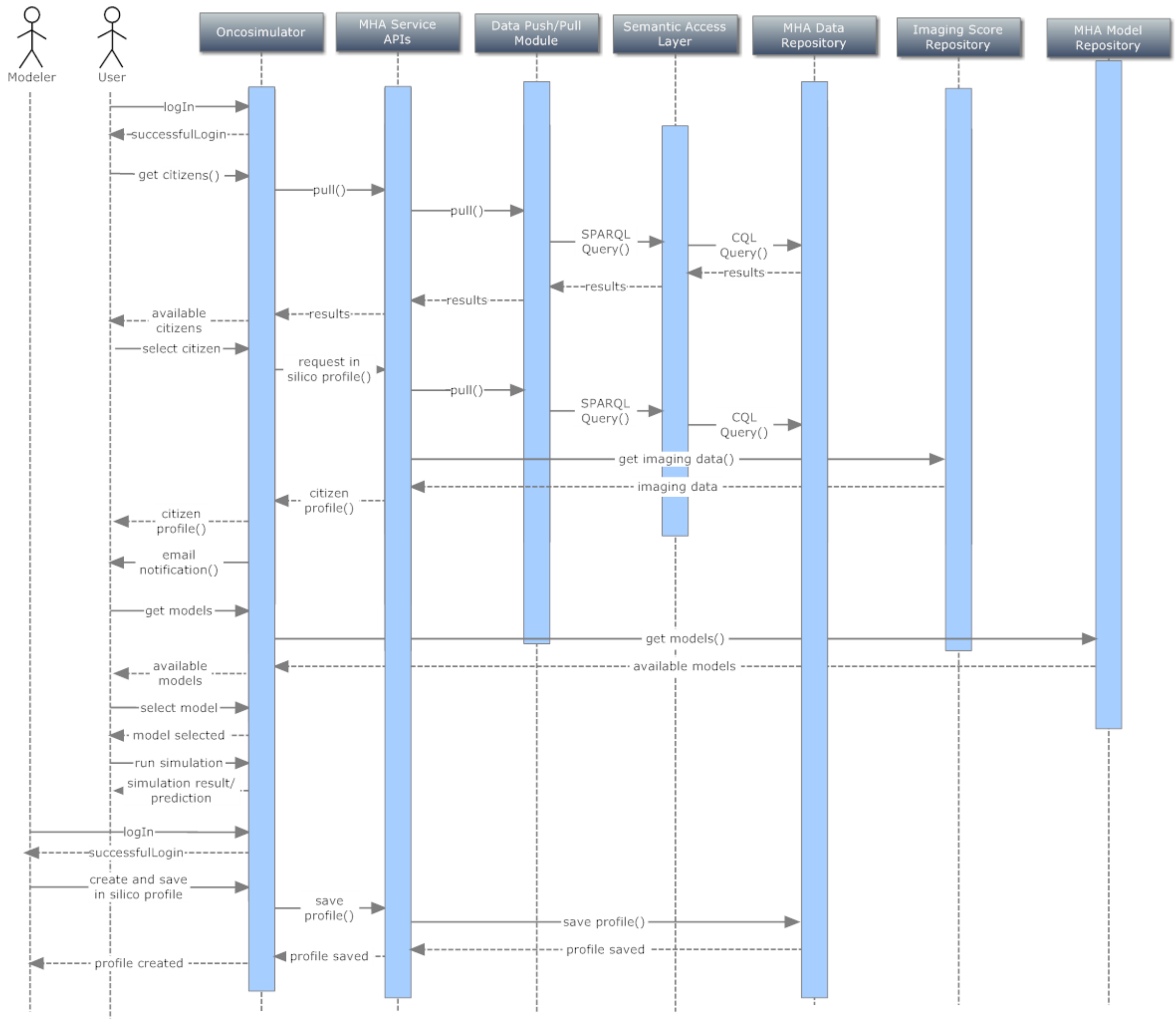


Figure 4-16: Sequence diagram of the Nephroblastoma Use Case

5 Link with external data sources

The platform of MHA aims to support the 4D digital representation of a given patient but of course parts of the patient's clinical and social history are already stored and managed by third party systems. For this reason proper mechanisms and infrastructure should be in place for retrieving relevant user information from these external data sources. Whenever it's possible such "linking" with the third party systems should be based on available standard interfaces since they allow the building of generic ports and interfaces and the reuse of existing code bases. Figure 5-1 shows some notable examples for the realization of these links to external resources: Clinical data can be retrieved from Hospital Information Systems (HIS) through the Clinical Document Architecture (CDA[12]) guidelines and set of specifications, clinical trial specific patient data can be acquired using the Operational Data Model (ODM) of the Clinical Data Interchange Standards Consortium (CDISC[13]), whereas cross-border healthcare provisioning is supported by the adoption of epSOS[14] Patient Summary interfaces. Additional well-known and widely supported standards and quasi-standards include Digital Imaging and Communications in Medicine (DICOM[15]) and the "transactions" defined by the Integrating the Healthcare Enterprise (IHE[16]) initiative.

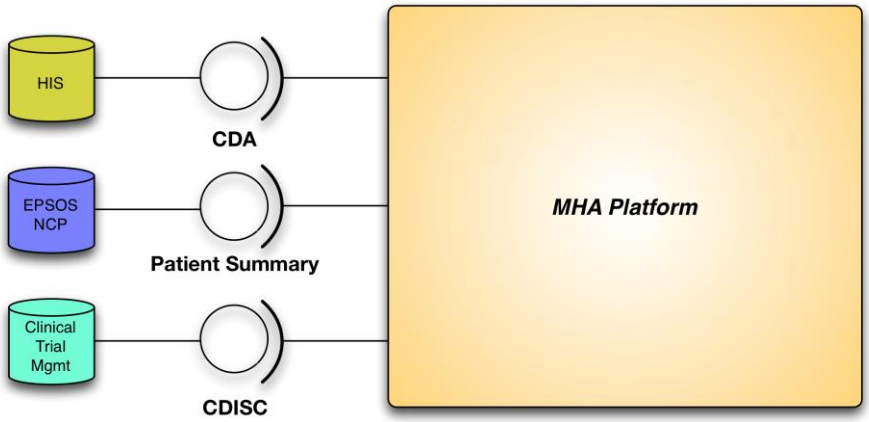


Figure 5-1: Linking MyHealthAvatar with external systems through well-defined interfaces

In the following picture we present the main external components that will be linked with MHA platform and will feed multilevel data to MHA repository but we do so in a generic way. A new architectural layer is introduced that hosts the adapters, gateways and other components which are

[12] http://www.hl7.org/Special/committees/structure/index.cfm

[13] http://www.cdisc.org/

[14] European Patients Smart Open Services (epSOS), http://www.epsos.eu/

[15] http://dicom.nema.org/

[16] http://www.ihe.net/

responsible for the linking with the external data sources. The components belonging to this layer interact with the main backbone of the MHA platform i.e. the Cassandra based data repository, the semantic infrastructure, and other repositories. In this figure we emphasize also in the semantic integration layer and the semantic transformation of these data in order to be uniformly accessible, through MHA common information data model, via MHA published APIs.

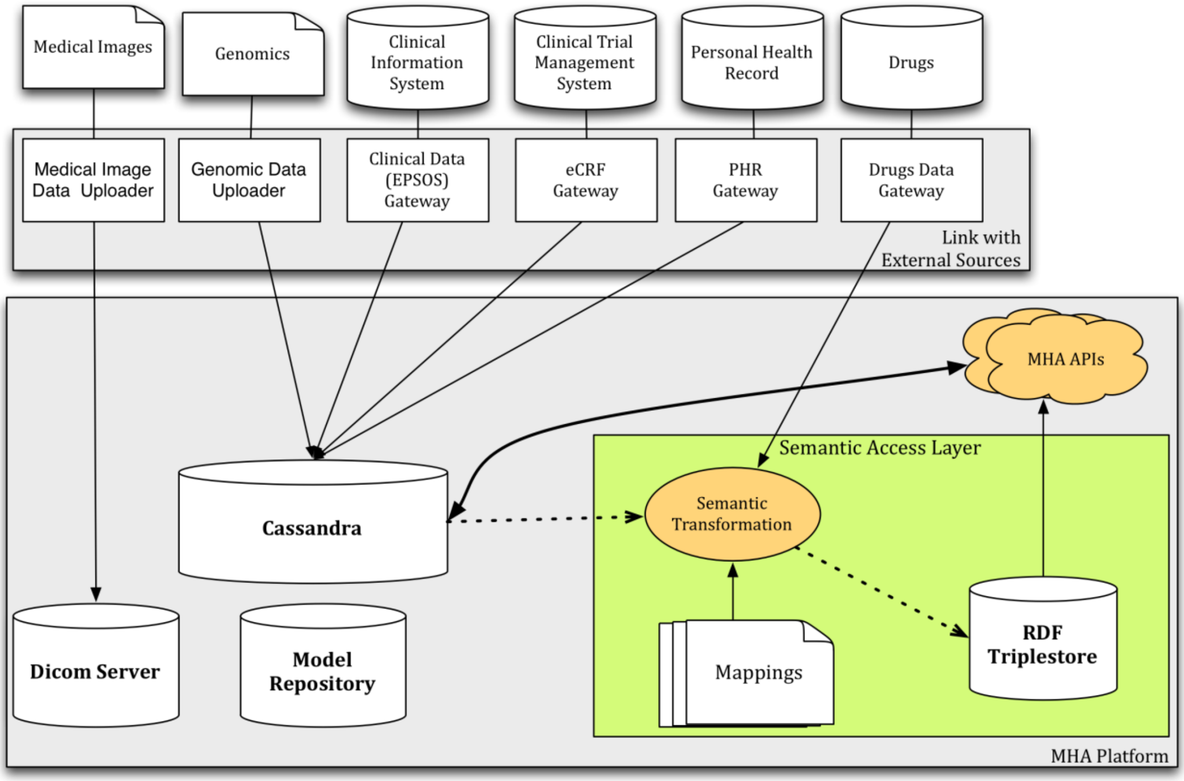


Figure 5-2: MHA external data source linkage

5.1 Link with hospital information Systems

In the following figure we show the flow of the clinical records patient data through the epSOS gateway. According to epSOS guidelines and proposed architecture the MHA epSOS Gateway retrieves "Patient Summaries" from a "National Contact Point (NCP)". The National Contact Point is an organization and IT infrastructure to support internal (in the same country) or cross-border (across countries) communication for the exchange of patient data. The patient summaries retrieved through the NCP contain essential information needed for the continuity of care such as the most important clinical patient data (e.g. allergies, current medical problems, medical implants, or major surgical procedures), a list of current medication including all prescribed medication that the patient is currently taking, etc.

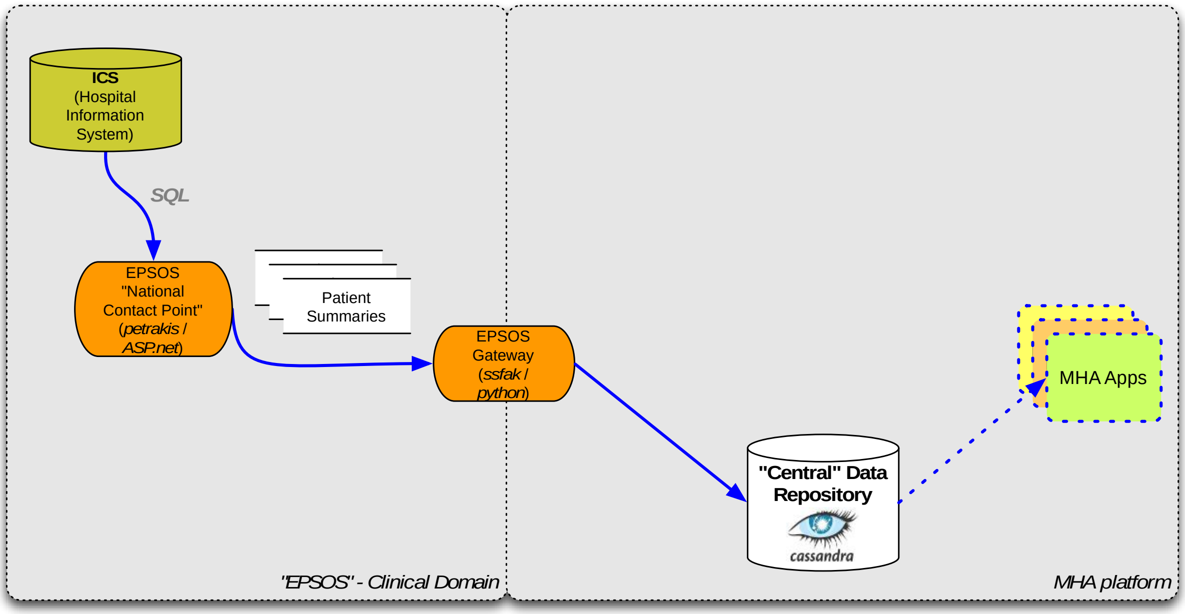


Figure 5-3: Retrieval of clinical records data using the epSOS Gateway

In the current deployment of this use case the implementation of the NCP retrieves the clinical data from a local installation of a production level Hospital Information System[17] using direct SQL queries to the system's database. Nevertheless, this is an internal implementation detail of the epSOS clinical domain. The important thing is that the MHA epSOS Gateway operating in the borders of the clinical domain and the MHA platform uses the epSOS Patient Summary documents to feed the central data repository of the platform. For more information on EpSOS Patient Summary Dataset see APPENDIX A. The Gateway filters and keep the relevant patient data, currently the medications and the vital signs measurements.

5.2 Link with clinical trial management Systems (Obtima)

ObTiMA is an ontology based clinical trial management application that is developed within different European projects (e.g.: ACGT[18], p-medicine[19]). The ontology-based creation of CRFs in the Trial Builder is one of ObTiMA's major functionalities. A user interface allows defining content, navigation, and layout of CRFs to capture all patient data during a trial, e.g., medical findings or diagnostic data. The resulting descriptions are based on concepts from the Health Data Ontology Trunk (HDOT)[20] for each CRF item along with metadata, e.g., data type and measurement unit, and are used to setup the

[17] Integrated Care Solutions (ICS), http://www.ics.forth.gr/ceha/index_main.php?l=e&c=471

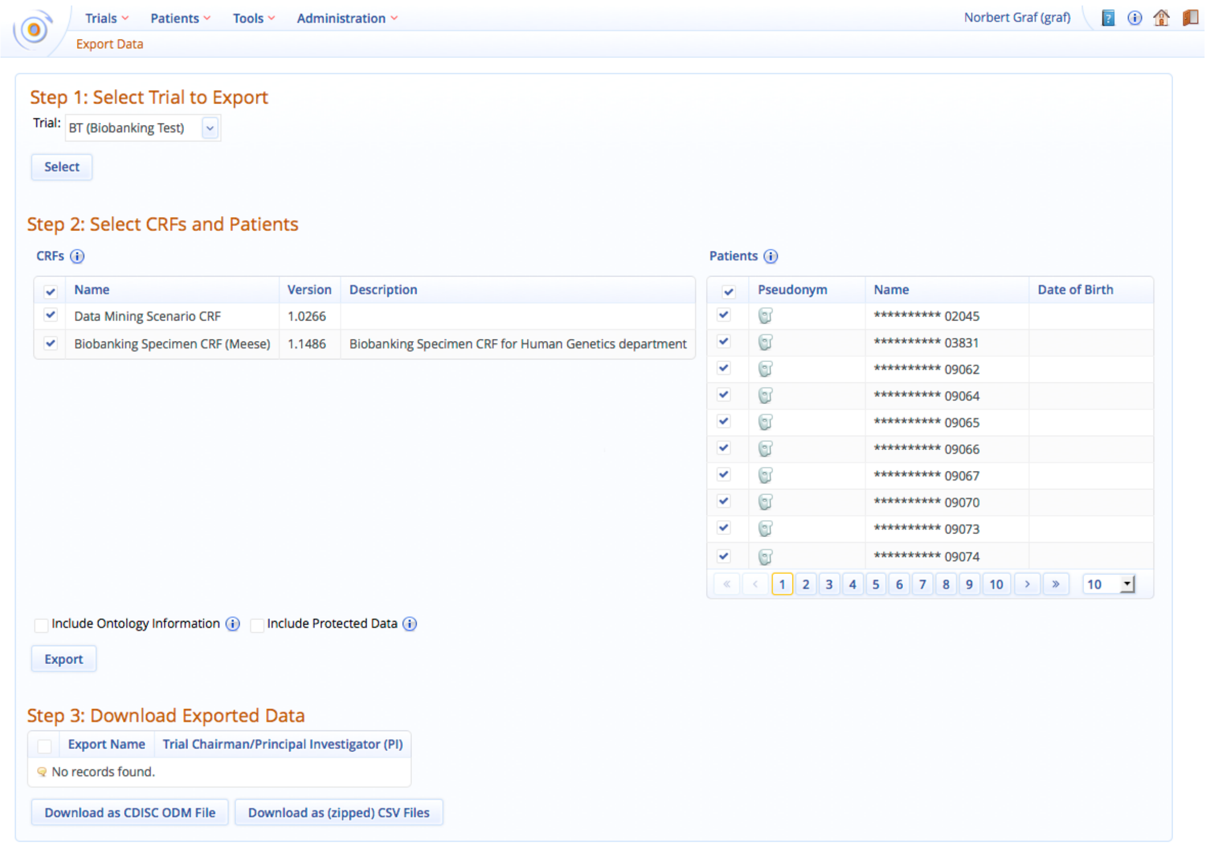
[18] http://acgt.ercim.eu/

[19] http://p-medicine.eu/

[20] https://code.google.com/p/hdot/

trial database. The user interface makes the underlying aspects of the ontological metadata transparent to users and tries to close the gap between clinical practice and the actual logical representation of ontological concepts. If an item has been created based on a concept, its attributes are determined automatically, e.g., label, data type or answer possibilities, but it can be adopted manually.

Since many trials collect similar or equal data, it is possible to store components of or complete CRFs in a repository as templates. When setting-up a clinical trial appropriate CRF templates can either be directly reused or can be quickly created by composing them from existing CRF components. This in turn fosters the CRF standardization since CRFs can then readily be compared on the level of single items (through ontological concepts) and also on component level or in their entirety.



Step 1: Select Trial to Export

Trial: BT (Biobanking Test)

Step 2: Select CRFs and Patients

CRFs

✓	Name	Version	Description
✓	Data Mining Scenario CRF	1.0266	
✓	Biobanking Specimen CRF (Meese)	1.1486	Biobanking Specimen CRF for Human Genetics department

Patients

✓	Pseudonym	Name	Date of Birth
✓		*********** 02045	
✓		*********** 03831	
✓		*********** 09062	
✓		*********** 09064	
✓		*********** 09065	
✓		*********** 09066	
✓		*********** 09067	
✓		*********** 09070	
✓		*********** 09073	
✓		*********** 09074	

« ‹ 1 2 3 4 5 6 7 8 9 10 › » 10

□ Include Ontology Information □ Include Protected Data

Step 3: Download Exported Data

□	Export Name	Trial Chairman/Principal Investigator (PI)
No records found.		

Figure 5-4: Download clinical trial data from ObTiMA

The second major functionality is the patient data management system supporting clinicians during a clinical trial. It is automatically setup based on the items defined in the Trial Builder in the design phase. ObTiMA itself is composed of different modules. In addition to the above described basic components, a DICOM server and DICOM viewer, a SAE and SUSAR reporting tool and a consultation tool are integrated. These tools are optional to handle images used in clinical trials or to simplify the SAE and SUSAR reporting according to GCP criteria. The consultation tool stores all consultations in a standardized way in the trial database. ObTiMA itself fulfills GCP criteria, including an Audit Trail. Data safety and security are guaranteed as pseudonymization of private data is implemented according to roles and rights assigned to users of ObTiMA. It is of utmost importance that ObTiMA will be certified for the use within clinical trials

In relation to MyHealthAvatar ObTiMA can provide data from clinical trials to the platform. This is of utmost importance for the high-end scenario: 'Nephroblastoma (Wilms Tumour) Simulation Model and Clinical Trial (UC-NEPH)' where response to a given treatment is analysed. ObTiMA is able to download clinical trial data in CDISC-ODM format (figure 5-4).

These data can then be uploaded into the MyHealthAvatar platform for reuse in the high-end scenario. According to the UC-NEPH scenario it is possible to download clinical trial data also from other trials, if this is allowed by the legal framework and informed consent by the patients.

5.3 Link with Social Networks

MHA infrastructure supports linking to social networks. MHA provides social web mechanisms and encourage the patients/citizens to adopt those in order to define their digital avatar. This requires integration with the social web accounts that the patients maintain already and the extraction of the social graph and other information. MHA is able to collect data from online patient diary using the utilities provided by T6.1. Volunteers will be organized to participate the research in this task. The task will continue towards the end of the project, leading to a considerable collection of information from the participants. The popularity of social media allows users to link their account profiles from social networks like Twitter, Facebook or PatientsLikeMe with MHA platform that become a communication hub for collecting people's personal stories and life experiences. MHA will then be able to contain a large volume of potential personal health information. The use of data mining techniques in the exploration of personal health information from these social networking services is the goal of this specific research task. The key problem we identified and focus on is on how to extract meaningful information from a large volume of data from popular social media services like Facebook, Twitter, etc. Details on the storage of these information have already been reported in D6.1.

We have explored and implemented the connection to the main social networks including Facebook, Twitter and Google +, we support connect the social network by their API (through OAuth 1 and 2). The detail of how it is connected and architecture graphic is described in previous deliverables. We have taken into consideration of Identity Federation with social networking, namely Login with Facebook/Twitter/Google, which would ease the user in terms of manage their credentials (this is work in progress). With users' explicit authorisation, MHA is able to read user's post, friend list, etc. and also able to post information back to the social network. E.g. Their daily activities from MHA. The post back to social network will increase the MHA exposure to general public, and potentially attracts new users to MHA. MHA also consider to allow users to invite friends from his social network, contact lists to MHA.

6 Semantic Integration Layer

The semantic integration layer consists of the following components shown in Figure 6-2:

- **Common Information Model:** The MyHealthAvatar Semantic Core Ontology is used as the virtual schema of all data stored within MyHealthAvatar, able to semantically describe the different types of data required and processed by the platform. It is consisted of 34 sub-ontologies shown in Figure 6-1 integrated through the Extended TMO ontology (eTMO). The integration is achieved by introducing terms from these subontologies to the eTMO ontology or via equivalences identified between each sub-ontology and the eTMO ontology.

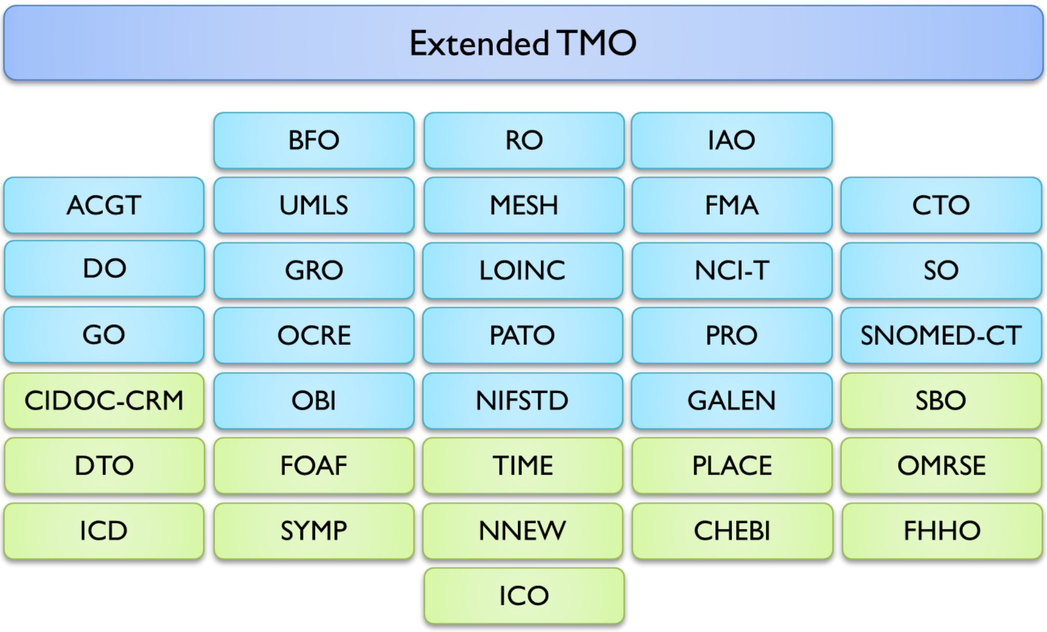


Figure 6-1. The modules of MyHealthAvatar Semantic Core Ontology (in green the extensions provided for MyHealthAvatar)

The explanation of the acronyms of the aforementioned ontologies and the domains they cover are shown in Table 1. The interested reader is forwarded to D4.1 for a description of each sub-ontology.

Table 1. Ontologies within MyHealthAvatar Semantic Core Ontology

Acronym	Ontology Title	Domain Described
ACGT	ACGT Master Ontology	Health Status & Clinical Information, Molecular Data, Medication Data
BFO	Basic Formal Ontology	Upper Layer Ontology
CHEBI	Chemical Entities of Biological Interest	Biological & Molecular Data
CIDOC-CRM	CIDOC Conceptual Reference Model	Upper layer ontology for documentation & Information Artifacts

CTO	Clinical Trial Ontology	Clinical Information
DO	Human Disease Ontology	Clinical Information
DTO	Disease Treatment Ontology	Clinical & Medical Information
FHHO	Family Health History Ontology	Personal Information & Lifestyle, Health Status & Clinical Information
FMA	Foundation Model of Anatomy	Health Status & Clinical Information
FOAF	Friend of a Friend Ontology	Social Information
GALEN	Galen Ontology	Medical Information
GO	Gene Ontology	Genomic Information
GRO	Gene Regulation Ontology	Genomic Information
IAO	Information Artifact Ontology	Upper Layer Ontology for Information Artifacts
ICD	International Classification of Diseases	Health Status & Clinical Information, Medication Data
ICO	Informed Consent Ontology	Health Status & Clinical Information
LOINC	Logical Observation Identifier Names and Codes	Health Status & Clinical Information
MESH	Medical Subject Headings	Health Status & Clinical Information, Medication Data
NCI-T	NCI therarurus	Health Status & Clinical Information
NIFSTD	Neuroscience Information Framework Standardized ontology	Molecular Data
NNEW	New Weather Ontology	Personal Information & Lifestyle
OBI	Ontology for Biomedical Investigation	BioMedical Infomation
OCRE	Ontology for Clinical Research	Clinical Information
OMRSE	Ontology of Medically Related Social Entities	Personal Information & Lifestyle
PATO	Phenotypic Quality Ontology	Clinical Information
PLACE	Place Ontology	Life-Style Information
PRO	Protein Ontology	Protein Information
RO	Relation Ontology	Upper layer Ontology
SBO	Systems Biology Ontology	Molecular Data, Systems Biology Models
SNOMED-CT	SNOMED clinical terms	Health Status & Clinical Information, Medication Data
SO	Sequence Ontology	Sequencing Information
SYMP	Symptom Ontology	Clinical & Health Information
TIME	Time Ontology	Lifestyle Information
UMLS	Unified Modeling Language System	Health Status & Clinical Information, Medication Data

- **Virtuoso Triple Store[21]:** Virtuoso is an innovative enterprise grade multi-model data server, developed by Openlink. It is used for reasons of efficiency in order to store the semantically enhanced data and to offer query functionalities over them.
- **Query/Data Translation Module:** *exelixis*[22,23] is a novel data integration engine that achieves query answering by accepting SPARQL queries. The queries are then rewritten according to the source schemata and forwarded to the sources to be answered. To achieve this rewriting the proper mappings should be established between the source schemata and the ontology. Instead of query rewriting there is also the possibility to materialize all generated information using the aforementioned mapping similarly to query rewriting. In both cases the results are provided as input to the RDF Triple Store where they are available for further queries.

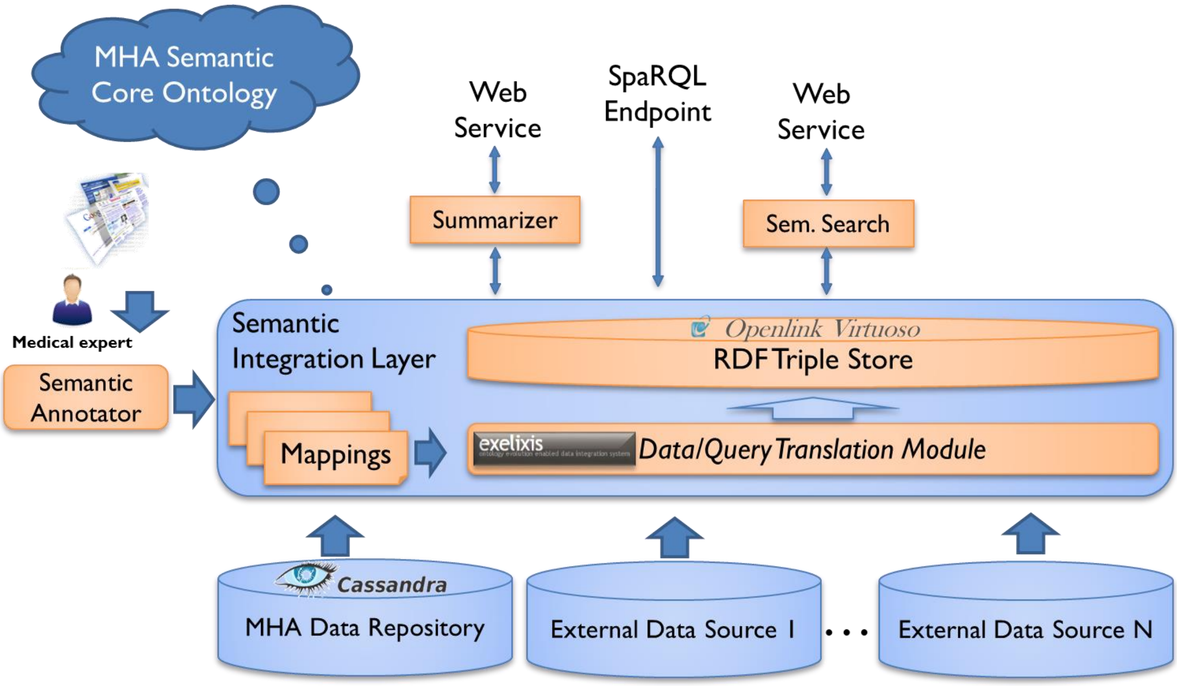


Figure 6-2: The Semantic Integration Layer of MyHealthAvatar platform

[21] http://virtuoso.openlinksw.com/

[22] Kondylakis, H., Dimitris, P.: Exelixis: Evolving Ontology-Based Data Integration System. SIGMOD, pp. 1283-1286 (2011)

[23] Kondylakis, H., Plexousakis, D.: Ontology Evolution without Tears, Journal of Web Semantics (2013), 19, pp. 42-58, Elsevier

- **Summarizer:** Ontology summarization aspires to produce an abridged version of an RDF/S document in order to highlight the most representative concepts of it. Within MyHealthAvatar, we developed RDF Digest[24], a novel platform that automatically produces summaries of RDF/S Knowledge Bases. A summary is a valid RDFS document that includes the most representative concepts of the schema adapted to the corresponding data instances. To construct this graph, our system exploits the semantics and the structure of the schema, as well as the distribution of the corresponding data/instances.
- **Semantic Annotator:** The Semantic Annotator focuses on the effective exploitation of the numerous textual sources which are available on the web. It uses the MHA Semantic Core Ontology to annotate high-quality documents with ontology terms. Those documents have been selected by the domain experts (doctors).
- **Semantic Search:** After annotating the web documents using ontology terms we are able to provide search functionalities to the database of documents using advanced information retrieval algorithms.

As stated also in D4.2[25], the architecture adopted is in a way based on the *command-query responsibility segregation* principle[26]. The core of that principle is that one used a different model to update information than the model one is using to read. Although the mainstream approach people use for interacting with an information system is to treat it as a CRUD data store, as the needs become more sophisticated state of the art approaches steadily move away from that model since we may want to look at the information in a different way to the record store, perhaps collapsing multiple records into one, or forming virtual records by combining information for different places. On the update side we may find validation rules that only allow certain combinations of data to be stored, or may even infer data to be stored that are different from that we provide.

Specifically in our case, the NoSQL databases adopted to be the main data repository imposes strict restrictions on the way the information is stored and queried. The main restrictions are the following:

- Joins are now allowed
- You cannot project the value of a column without selecting first the key of the column.
- Unlike the projection in a SQL SELECT, there is no guarantee that the results will contain all of the columns specified because Cassandra is schema-optional. An error does not occur if you request non-existent columns.

[24] Troullinou, G., Kondylakis, H., Daskalaki, E., Plexousakis, D.: RDF Digest: Efficient Summarization of RDF/S KBs, Extended Semantic Web Conference (ESWC) 2015

[25] The MyHealthAvatar Consortium, D4.2 Extension of the Semantic Core Ontology, February 2015

[26] http://en.wikipedia.org/wiki/Command%E2%80%93query_separation

And although the *exelixis* system is able to reconcile those restrictions by enabling the combination of information in different tables and systems the aforementioned restrictions make the whole process really time consuming and non-efficient when large amounts of data are available. This is why we use *exelixis* system *offline* to integrate and reconcile the data and the result is then stored in a native RDF triple store.

Using the aforementioned architecture there are two main benefits. Firstly complexity is handled more efficiently since the two layers have different characteristics. The other main benefit is in handling high performance applications by separating the load from reads and writes allowing to scale each independently. In addition, different optimization strategies can be applied to the different components.

7 Security Framework

7.1 Privacy and legal issues

Since sensitive health and lifestyle data are going to be processed in the MHA project, measures need to be taken in order to protect the stored data against unauthorised disclosure or access, accidental or unlawful destruction or accidental loss or alteration according to Article 17 (1) of the Data Protection Directive (Directive 95/46/EC[27]). Moreover the legal framework of MHA needs to ensure a lawful and fair data processing. The following section provides an overview of the most important legal aspects that need to be considered for the architecture design. The further details are explained in D11.1.

7.2 Legal framework for privacy and data protection

For determining the legal framework of MHA it is crucial to know if personal data is going to be processed because the Data Protection Directive as the major source for the legal requirements is applicable only if personal data are processed. If this is the case the processing of personal data is forbidden, except if there is a legal basis or the data subject has given informed consent (see 5.2.2). Moreover, the need for fair processing requires to de-identify the data whenever possible and to use anonymised or at least pseudonymised data. According to the principle of limited retention the data must be erased as soon as it is no longer needed for the purposes for which they were collected. The principle of data minimization states that personal data should be collected only if really needed and the principle of purpose limitation that also needs to be considered means that data must generally not be processed in a way incompatible with the purposes at collection.

The architecture design must ensure that personal data is protected against accidental or unlawful destruction or accidental loss, alteration, unauthorized disclosure or access by technical and organizational measures. As discussed further in D11.1, measures for achieving this are set out in section 9.2 the Recommendation No. R (97) 5 on the Protection of Medical Data[28], and shall ensure an appropriate level of security taking account of the technical state of the art and also of the sensitive nature of medical data and the evaluation of potential risks.[29]

[27] http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31995L0046:en:HTML

[28] http://www1.umn.edu/humanrts/instree/coerecr97-5.html

[29] Recommendation No. R (97) 5 on the Protection of Medical Data; at 9.2.

7.2.1 Personal Data (citizen)

Pursuant to Article 2 (a) of the Data Protection Directive, personal data is "*any information relating to an identified or identifiable natural person ('data subject'); an identifiable person is one who can be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity*". To determine whether a person is identifiable recital 26 states that account should be taken of all the means likely reasonably to be used either by the controller or by any other person to identify said person. So a person is not identifiable when *"all the means likely reasonably to be used"* do not exist and consequently the information cannot be considered as personal data.[30]

If citizens store data in their personal digital avatar on platform, e.g. data collected by apps and devices, the person behind the data is identifiable and thus the rules of the Data Protection Directive apply, amongst others that there must be no data processing without consent or another basis in law.

7.2.2 Medical Data (patient)

Since certain information is more important for the privacy of a person than other kinds of information, the Data Protection Directive provides special rules for sensitive data such as data concerning health. Here the strict rules of Article 8 must be taken into account in addition to the basic principle that the processing of personal data is forbidden, except if there is a legal basis or the data subject has given informed consent.

In this regard, Member States are permitted, subject to providing suitable safeguards, to allow processing of sensitive data **or reasons of substantial public interest** according to Article 8 (4) of Directive 95/46/EC. Public interest includes public health and scientific research as mentioned in recital (34) of the Directive. Although this exemption could be applicable in terms of MHA and especially with regard to the use cases, the solution of first choice is always to ask the data subject for consent.

Furthermore, Article 8 (3) allows the processing of sensitive data if it "*is required for the purposes of preventive medicine, medical diagnosis, the provision of care or treatment or the management of health-care services, and where those data are processed by a health professional subject under national laws or rules established by national competent bodies to the obligation of professional secrecy or by another person also subject to an equivalent obligation of secrecy".* Since it remains open how far this exemption is applicable for the use of digital avatars that go beyond the direct individual treatment of patients, the architecture design of MHA will not rely on Article 8 (3).

[30] Opinion 4/2007, p. 15.

7.2.3 Clinical Data (HIS)

In terms of clinical data and hospital information systems data linkage will be a big topic. For determining a legal framework, details which are not yet known need to be considered. In general, the rules that are outlined above for sensitive data such as data concerning health, also apply for clinical data that are processed by hospital information systems. As far as the data is not anonymized, the data subject should grant explicit consent. Moreover, a secure system is important to avoid unauthorized access.

7.2.4 Genomic Data [31]

The biggest challenge today is to develop computer infrastructure and systems able to gather genomic data sets from individuals for comprehensive analyses for biomedical research and the clinical practice, as MHA describes. In this context genomic data should also be seen as a sensitive category of data, whose processing is governed by Article 8 of the Data Protection Directive. As regards processing mechanics, cloud computing seems to be the optimum and most acceptable solution to integrate data from genomics, systems biology and biomedical data mining. Although cloud computing provides several benefits such as lower costs and greater capacity and efficiency, as described in D11.1, it also raises legal and ethical issues. This is especially true in terms of public clouds because they bear the risk that it is not or only with difficulties possible to track the data access and processing. In this sense a solution will need to take proper account of the following: data control; data security, confidentiality and transfer; and the use preferably of a private cloud which is located within the EU/EEA and processes the data only within this territory.

7.2.5 Activity/Social Data

Since activity/social data is personal data, and as the citizen is at the center of the relevant data generation, we will use a consent-driven approach when the processing of this data is needed.

7.3 Consent

As discussed above, a consent-driven approach is the best solution to apply not only with legal, but also with ethical requirements and to show respect for the data subject's self-determination. Moreover, this is in line with the core vision of MyHealthAvatar as a resource for citizen empowerment. Consent has to be granted voluntarily and requires that the data subject has understood how his data shall be used. To achieve this, the data subject should be given the possibility to ask a well-trained person questions. Moreover, in terms of sensitive health data consent needs to be given explicitly.

[31] http://www.nature.com/ejhg/journal/vaop/ncurrent/full/ejhg2014196a.html

7.4 Anonymity and Pseudonymity

As explained in 5.2.1, the need for a fair processing requires that data should be de-identified as far as possible. Pursuant to recital 26 the Data Protection Directive is not applicable to data rendered anonymous in such a way that the data subject is no longer identifiable. As elaborated in D11.1, the Article 29 Working Party under the Data Protection Directive has defined anonymous data as “*any information relating to a natural person where the person cannot be identified, whether by the data controller or by any other person, taking into account all of the means likely reasonably to be used either by the controller or by any other person to identify that individual. Anonymized data would therefore be anonymous data that previously referred to an identifiable person, but where that identification is no longer possible*.”[32]

Although the Data Protection Directive does not mention the term of “pseudonymous data” the Article 29 Working Party refers to pseudonymous data in its opinion on the concept of personal data and defines pseudonymisation as “*the process of disguising identities*“[33]and gives some examples how pseudonymisation can be done, e.g. by using correspondence lists for identities and their pseudonyms or by using two-way cryptography algorithms (whereas the use of one-way cryptography usually creates anonymised data). In its view, whilst still qualifying as personal data, the risk of re-identification – where pseudonymisation is coupled with other appropriate technical and organisational safeguards – are sufficiently low to justify a flexible approach to processing such data in compliance with the Data Protection Directive.

7.5 Privacy Option analysis in MHA

Consent is both legally and ethically the keystone for the processing of personal data. While exemptions to the requirement of consent exist, they are exemptions that should only be relied on where obtaining consent is not possible. Consequently, the MHA architecture must focus on developing and implementing a robust consent module. This module must reflect all the elements necessary for valid consent. The further details are described in D11.1.

7.6 Trust

In order to foster trust, within MyHealthAvatar we want to provide clear and objective information to users about the benefits and risks associated with the use of the system. The user should have as

[32] Opinion 4/2007, p. 21.

[33] Opinion 4/2007, p. 18.

complete an understanding as possible of the consequences and eventual risks when using the system. However the risk that information are too detailed and might overburden the user must be taken into account as well, as this may result in users developing unfounded expectations that cannot and will not be met by the system. Furthermore, this may result in users blaming the system for not living up to their expectations. Therefore it is important that the user has the possibility to ask questions and get information repeatedly to find a balance between the risk of a loss of information and the lack of understanding.[34]. Care must also be given to the aspect *when* information should be presented. If users are presented once with some long hardly comprehensible text before they are allowed to execute a certain ’pressing’ action, like acknowledging license rights before software installations, they are likely to ignore the text and pursue the action anyhow[35]. If such long texts cannot be avoided, they should be presented when the user is not engaged in routine tasks which the user expects to have finished in a few minutes. It could be helpful to *regularly remind the user of important topics by displaying brief, random pieces of information*, e.g., the user could be shown a hint when her device is busy and she cannot do anything else except watching the display and wait for the device to complete its task. Of course, such displayed information would not be complete but could help to ’educate’ the user over time and also remind her of places where to look for more details.

7.6.1 User Information

In order to foster trust, within MyHealthAvatar we want to provide clear and objective information to users about the benefits and risks associated with the use of the system. Of course, not every usage aspect and/or risk can and should be communicated to users as this would probably be too much information to comprehend and will likely lead users to ignore the information. The latter should naturally be avoided, as this may result in users developing unfounded expectations that cannot and will not be met by the system. Furthermore, this may result in users blaming the system for not living up to their expectations. Care must also be given to the aspect *when* information should be presented. If users are presented once with some long hardly comprehensible text before they are allowed to execute a certain ’pressing’ action, like acknowledging license rights before software installations, they are likely to ignore the text and pursue the action anyhow[36]. If such long texts cannot be avoided, they should be presented when the user is not engaged in routine tasks which the user expects to have finished in a few minutes. It could be helpful to *regularly remind the user of important topics by displaying brief, random pieces of information*, e.g., the user could be shown a hint when her device is busy and she cannot do anything else except watching the display and wait for the device to

[34] Nikolaus, Forgó, Marian Arning, Tina Kruegel, Imme Petersen. Ethical and Legal Requirements for Transnational Genetic Research. C.H.Beck, p. 28. 2010.

[35] Rachna Dhamija and Lisa Dusseault. The Seven Flaws of Identity Management: Usability and Security Challenges. *IEEE Security & Privacy*, 6(2):24–29, March/April 2008.

[36] Rachna Dhamija and Lisa Dusseault. The Seven Flaws of Identity Management: Usability and Security Challenges. *IEEE Security & Privacy*, 6(2):24–29, March/April 2008.

complete its task. Of course, such displayed information would not be complete but could help to ’educate’ the user over time and also remind her of places where to look for more details.

7.6.2 Trust Model

A trust model defines what can be expected from some party with respect to some item of interest. Here, ’expectation’ can also be read as ’assumption’. In other words, the trust model makes explicit the set of otherwise hidden assumptions we rely on when giving guarantees to others. Trust models can be compared with respect to their relative strengths. Apart from general assumptions, the trust model also encompasses assumptions on the well-behaviour of certain entities in some regard. Hence, a trust model also has to qualify in which regard someone or something is trusted, similar to stating against whom anonymity or pseudonymity is directed. MyHealthAvatar may have more than one trust model because the many use case scenarios, running in different environments with different requirements that will influence their respective trust model. It is also the case that in both use cases we can have different usage scenarios, which may give rise to different trust models, too.

In this section, we present some assumptions with respect to security which will have to be satisfied to allow for the level of security we aim for. Some of the assumptions are necessary because they cannot be technically enforced or controlled by users — at least not with reasonable effort. Other assumptions are necessary because without them some of the scenarios could not be realised — at least not with a reasonable expectation of security and privacy.

Legal processing of personal data. We assume that MyHealthAvatar platform processes personal data only if the data subject has given her consent to the processing of her personal data or if there is other legal ground covering the processing.

No misrepresentation. We assume that the deployment server’s operator does not deliberately misrepresent to data subjects the purpose or scope of some intended processing of personal data.

No permanent loss of secrets. We assume that reasonable steps have been taken to ensure that secret information, e.g., passwords or cryptographic keys, are regularly backed up such that the loss of the secrets, e.g., through theft or damaged media, is not permanent. Note that this does not imply that some data protected by the secret information should still be regarded as secure. The assumption merely ensures that important data does not become inaccessible for the user.

Wireless Key provisioning. In case devices need to be provisioned with secret keys over an ’air interface’, we assume that steps have been taken such that attackers cannot overhear the keys without considerable effort.

Trusted personal devices (affects the portability of MyHealthAvatar to personal portable mobile smart phones and tablest). We assume that the user’s device, e.g., the Android device in the inpatient scenario, acts as a trusted device, e.g., we assume that no malicious software is present on the device, e.g., key logger, viruses, Trojan horses, etc. We also assume that the device’s installed software is up to date, i.e., has no vulnerability that an attacker could exploit to gain access to the device or its data. Furthermore, we assume that if the user passes the device on to another person, this person is also trusted by the device's owner.

The purpose of the above assumptions above is to clearly state the limitations of the security architecture, i.e., what cannot be dealt with or what is beyond the scope of the security architecture for the purpose of this work. It is understood that these assumptions are not necessarily final or complete and only reflect the architecture as presented in this document. When the scenarios progress and/or more implementation work has been done, more assumptions may be necessary or some of the existing ones may need to be revised. An update if necessary will be provided at the final architecture definition deliverable

7.7 Security

The objective of a secure system is to protect sensitive information from unauthorized access, manipulation, misuse, etc. In MyHealthAvatar the information to protect are, patients citizen personal data stored in MHA repositories, medical measurements, private patient data, medical history, activity data, medical images, model repositories data etc. The protection goals which define the requirements of a secure system are defined by the following objectives[37, 38].

Authentication: In information systems authenticity typically means the genuineness and credibility of data or an entity. The act of confirming the authenticity is called authentication. In a system where certain data or entities are required to be authentic there have to be authentication mechanisms. Authentication is based on providing some proof in the form of a credential, such as a username and password in the case of user authentication that allows the verifier to assess the genuineness of the property claimed by the credential.

Authorisation: In secure information systems access to sensitive information has to be restricted to privileged entities. Therefore, permission rights have to be specified and assigned to the privileged entities. This act is called authorization. If an entity has been granted the permission to access a specific information, the entity is said to be authorized for this access.

Confidentiality: A system ensures confidentiality, if an entity cannot learn information which is not intended for it. The assurance of this objective requires control mechanisms to ensure that information is accessible only to those who are authorized to have access to the information. There

[37] Claudia Eckert. *IT-Sicherheit – Konzepte, Verfahren, Protokolle*. Oldenbourg Verlag, 5th edition, 2007. ISBN 978-3-486-58270-3.

[38] Alfred J. Menezes, Paul C. van Oorschot, and Scott A. Vanstone. *Handbook of applied cryptography*. CRC Press series on discrete mathematics and its applications. CRC Press, 1997. ISBN 0-8493-8523-7.

are numerous approaches to providing confidentiality, ranging from physical protection to cryptographic algorithms which render data unintelligible.

Integrity: Data integrity is preserved if it is not possible that an unauthorized entity can unnoticeably modify some information. For environments in which an unauthorized manipulation cannot be prevented a priori, mechanisms must be used to detect the manipulation a posteriori and thus, limit the damage incurred by further processing the manipulated data.

Accountability: Accountability is the ability to trace an action to a specific entity, such as a user, a process, or a device, and then hold them accountable or responsible for their actions. To ensure accountability, mechanisms are required to monitor and record relevant actions performed within the system.

Non-repudiation: A system that ensures non-repudiation prevents an entity from denying previous commitments or actions. Performed actions can be retroactively traced to a unique entity which can then be held accountable for its actions. When disputes arise, due to an entity denying to have committed certain actions, some sort of proof is necessary to resolve the situation. That proof must verifiably attest that these actions were actually performed by the person concerned. Non-repudiation is particularly useful for accountability.

7.8 Technical Enforcement of Security through MyHealthAvatar Platform

In practical systems these security objectives can be enforced by various mechanisms. In the following the most common technical means are introduced.

7.8.1 Authentication Procedures

For entities, such as users or machines, authentication technologies are subdivided into categories which are based on the knowledge of some specific information, such as a password or a PIN, personal possession, such as a key card or a software token like a digital certificate, or biometrical characteristics, such as fingerprints. To improve security and to utilize the advantages of different techniques, authentication technologies are often combined. Authenticity of data can be ensured by applying Message Authentication Codes (MAC)[39] typically based on block ciphers or cryptographic

[39] ISO/IEC 9797-1:1999. Information technology – Security techniques – Message Authentication Codes (MACs) – Part 1: Mechanisms using a block cipher, 1999.

hash functions (HMAC)[40]. Another way of ensuring data authenticity is to use digital signatures [66]. Both approaches produce an authenticator which attests the authenticity of some data with respect to its originator. The difference between these techniques is that MACs and HMACs are symmetric, which means that all participants use the same secret to produce the authenticator, whereas, digital signatures are asymmetric, which means that each participant has its own secret to produce the authenticator. Both, entity and data authentication are relevant for MyHealthAvatar. Knowledge and possession-based mechanisms are more established than biometrics and thus, will be favoured for entity authentication. For data authentication several mechanisms will apply depending on the kind of data to be protected. [41]

7.8.2 Authorization & Access Control

Authorization covers the specification and the assignment of permissions to access specific resources or allowance to perform certain operations. Access control, on the other hand, is the enforcement of the permission assignment, i.e., the act of checking and granting or denying a requested action. In access control, objects are resources to protect, such as data but also processes, memory, etc. Subjects are entities, such as users or processes that want to access an object. Permissions are specific access rights on an object, such as reading or modifying a file. An Access Control Strategy determines how the permissions are granted. The concrete implementation of an access control strategy or a combination of strategies is called an Access Control Model. The following list introduces the three major strategies and some representative models:

- In Discretionary Access Control (DAC) the authorisation originates from the owner or creator of an object and is passed on to other subjects. DAC is not suitable to enforce any system wide security policies because it is not possible to control the way an object owner manages its permissions. DAC is used, e.g., in UNIX-based systems.
- In Mandatory Access Control (MAC) permissions are managed centrally and ordinary users of the system cannot change the permissions. The disadvantage of MAC is that it can be difficult and expensive to manage large amounts of objects and users.
- In Role Based Access Control (RBAC) permissions are granted by the system not with focus on the subject itself but rather on the task or purpose a subject has. This means the subjects are categorised into roles according to their tasks and each role is then assigned to a corresponding set of permissions. RBAC is suitable for controlling systems with a large number and different types of users, like MyHealthAvatar, because organisational structures of institutions can often

[40] ISO/IEC 9797-2:2002. Information technology – Security techniques – Message Authentication Codes (MACs) – Part 2: Mechanisms using a dedicated hash-function, 2002.

[41] Chen CL1, Yang TT, Chiang ML, Shih TF., A privacy authentication scheme based on cloud for medical environment, J Med Syst. 2014 Nov;38(11):143. doi: 10.1007/s10916-014-0143-9. Epub 2014 Oct 15.

be mapped to roles in a straightforward manner. For example, users of a hospital's information system could be assigned to the roles *administrator*, *physician*, *nurse*, patient etc.

- OAuth 2.0 delegate access to third parties application, which reduces password sharing between users and third parties. It enable user to not only delegate access, but also revoke access. OAuth 2.0 provides specific authorization flows for various different situations, and supports web application, mobile application, desktop applications and other devices.

7.8.3 Confidentiality

Confidentiality can be achieved by using encryption mechanisms in order to transform information in such a way that an unauthorized person is not able to successfully interpret the data without knowing a relevant secret, for example the decryption key.

Encryption schemes can be categorised into two major classes:

- Symmetric cryptography refers to encryption schemes in which the same secret key is used for encryption and decryption. In the case of encrypted communication, the problem of a secure key exchange between sender and receiver arises. Symmetric cryptography mainly relates to the study of block ciphers, such as the Advanced Encryption Standard (AES)[42], and stream ciphers, such as RC4[43].
- Asymmetric cryptography, also called public key cryptography, on the other hand does not require a secure initial exchange of one or more secret keys. For asymmetric algorithms, a key pair — comprised of a private and a public key— is usually needed, where the private key is kept secret and the public key may be published. A message is encrypted by using the receiver's public key and the receiver uses her private key to decrypt the message. The crux of public key cryptography though is the distribution and management of public keys. Prominent representatives of public key cryptograhpy are the ElGamal encryption scheme[44] and the RSA encryption scheme[45].

Symmetric and asymmetric cryptography can also be combined to a hybrid cryptosystem, which take advantage of the fact that symmetric cryptography is more efficient than asymmetric encryption schemes. In a hybrid cryptosystem, the data is typically encrypted using a symmetric encryption scheme and the symmetric data encryption key is encrypted using an asymmetric scheme. This allows

[42] U.S. National Institute of Standards and Technology NIST. Advanced Encryption Standard (AES). FIPS PUB 197, November 2001

[43] Alfred J. Menezes, Paul C. van Oorschot, and Scott A. Vanstone. *Handbook of applied cryptography*. CRC Press series on discrete mathematics and its applications. CRC Press, 1997. ISBN 0-8493-8523-7

[44] Taher El Gamal. A public key cryptosystem and a signature scheme based on discrete logarithms. In *Proceedings of CRYPTO 84 on Advances in cryptology*, pages 10–18, New York, NY, USA, 1985. Springer-Verlag New York, Inc

[45] Ron L. Rivest, Adi Shamir, and Leonard M. Adleman. A method for obtaining digital signatures and public-key cryptosystems. *Communications of the ACM*, 21(2), February 1978

the receiver to first (asymmetrically) decrypt the data encryption key, using her private key, and subsequently decrypt the (symmetrically) encrypted data. The security of encryption schemes does not only depend on the secrecy of the secret keys but also on the lengths of the used keys. Using short keys enables adversaries to do an exhaustive search over the whole key space. With increasing technological advancements and growing computational power the cryptographic keys have to increase in length in order to resist these attacks. To ensure confidentiality in the long run, sufficient key lengths have to be chosen. An increase in key lengths on the other hand also increases the computational costs of encrypting and decrypting messages.

7.8.4 Integrity

Typically, data manipulation can be detected by using cryptographic hash functions. On the one hand, using hash functions enables the receiver of a message to detect if a message has been manipulated, given that she knows the hash value of the original data. On the other hand, an adversary is not able to bypass the hash function and thus, to manipulate a message in such a way that it cannot be detected by the receiver when she compares it to the original data's hash value. Hash function-based mechanisms to detect unauthorised data manipulation are often called Manipulation Detection Code (MDC), Message Digest, Digital Fingerprint, cryptographic checksum, or Message Integrity Check (MIC). These mechanisms cannot prevent unauthorised data manipulation but merely make it retroactively detectable. Examples of established cryptographic hash functions are RIPEMD160[46], SHA-1, and SHA-256 [47]. Integrity can also be ensured by applying Message Authentication Codes (MAC) or digital signatures.

7.8.5 Accountability

To ensure accountability it is necessary to monitor and record relevant actions executed within a system, e.g., when did they happen and who executed them. A lower level of accountability can typically be provided by auditing techniques, such as audit logs, which record the time and date, the performed action and who performed that action, etc. In MyHealthAvatar, accountability is a property that is especially relevant for audits as their purpose is to check whether past actions have been conducted in accordance with a given policy, e.g., laws or other regulations. Therefore, audit logs and other mechanisms will be applied in full detail.

[46] Hans Dobbertin, Antoon Bosselaers, and Bart Preneel. RIPEMD-160: A Strengthened Version of RIPEMD, 1996

[47] U.S. National Institute of Standards and Technology NIST. Secure Hash Signature Standard (SHS). FIPS PUB 180-2, August 2002.

7.8.6 Non-repudiation

Non-repudiation is a property typically achieved with the help of cryptographic methods, which prevents an individual or entity from denying having performed a particular action. Several mechanisms for non-repudiation are specified[48, 49, 50]. These can be based on symmetric techniques, such as Message Authentication Codes (MAC), as well as asymmetric cryptographic techniques, such as digital signatures. These techniques include non-repudiation certificates, tokens, and protocols and rely on additional Trusted Third Parties (TTPs), such as secure time-stamping services and Certification Authorities (CAs) of Public Key Infrastructures (PKIs).

7.8.7 Web security access

7.8.7.1 Authentication and Authorization for MHA

The trust principles of security, availability and processing integrity of MHA system is guided by Service Organization Control (SOC 2), which allows MHA system processing to be authorized and complete. ISO/IEC 27002:2013 code of practices for comprehensive information security control and risk management is considered during the design and implement of MHA security system design.

MHA's security of password infrastructure could be further augmented by the Universal Second Factor (U2F) protocol, which adds a strong second factor to user login. The commonly used U2F application is Google Authenticator of Android platform. The "meta protocol" OAuth 2 provides a very useful foundation for other protocols (e.g. OpenID Connect, NAPS and UMA). OAuth 2 forms the security foundation of for MHA API, which delegate access to third party applications. Detail of the API security can be found from deliverable 3.6. Various cutting edge security protocols and practises are evaluated and listed in following sections with their pros and cons.

7.8.7.2 SOAP over HTTPS

Simple Object Access protocol, is a protocol specification for exchanging structured information in the implementation of web services in computer networks. It uses XML Information Set for its message format, and relies on other application layer protocols, most notably Hypertext Transfer Protocol

[48] ISO/IEC 13888-1:2009. Information technology – Security techniques – Non-repudiation – Part 1: General, 2009

[49] ISO/IEC 13888-2:2010. Information technology – Security techniques – Non-repudiation – Part 2: Mechanisms using symmetric techniques, 2010

[50] ISO/IEC 13888-3:2009. Information technology – Security techniques – Non-repudiation – Part 3: Mechanisms using asymmetric techniques, 2009

(HTTP) or Simple Mail Transfer Protocol (SMTP), for message negotiation and transmission. HTTPS simply means that the HTTP protocol is tunnelled through the TLS protocol[51]. TLS can be run on top of any reliable transport protocol, e.g., TCP.

Authentication in TLS requires the relying party (RP), i.e., the verifier, to verify the validity and trustworthiness of some end entity's (EE) certificate and this certificate's issuer. A certificate's issues is also called a *certificate authority* (CA). Verifying the validity of a certificate includes, at least, checking some information presented by the sender that should prove that she really is the holder of the certificate, that the certificate is not expired, and that the certificate was issued by a CA that the RP deems trustworthy. Such a trusted CA, more specifically its certificate, is usually called a *trust anchor*. It is also possible that the issuer of the EE's certificate is not some CA directly trusted by the RP but by at least one of its trust anchors. This leads to some form of 'indirect trust' because the RP trusts a certain CA to only issue certificates to trustworthy entities and, similarly, this CA trusts another CA to do the same, and so forth. In practice, this leads to a so called *certificate path* which runs from the EE's certificate over some intermediate CAs' certificates to a trust anchor of the RP. Note that a certificate path trusted by the RP *must* be terminated by a trust anchor. This also warrants the term "trust anchor" because no more checking is done beyond the point where a trust anchor is encountered on a certificate path — beyond this point only trust remains. However, if no trusted path can be built, the RP would reject the EE's certificate. Building a certificate path is what is done by default by major browsers when they encounter a URL with HTTPS in the URL's protocol part. In this case, the EE's certificate is the certificate of the site being addressed by the URL, the RP is the user/her browser, and the trust anchors are all certificates which are included in the browser's certificate store. If the site requires mutual authentication the roles are reversed, i.e., the site becomes the RP, the user becomes the EE, and the trust anchors are all certificates configured in the certificate store of the RP's Web server.

Validating a certificate usually involves a revocation check, too. Any certificate, not only EE certificates but also CA certificates, can be revoked, e.g., because some guarantee given by the certificate does no longer hold. For instance, if every employee of a hospital receives a certificate when starting to work for the hospital, the certificate would normally be revoked when the employee quits, as the guarantee that the holder of the certificate is employed by the hospital no longer holds. However, a digital certificate cannot be made unusable like a credit card that is being cut when it becomes invalid. Therefore, the certificate's serial number is blacklisted by its issuing CA when the certificate is revoked. Such a blacklist is called certificate revocation list (CRL) and typically made available by the issuing CA. Now, if an RP wants to validate an EE's certificate, it will perform the certificate path building and check that the serial number of every certificate on the path has not been put on its issuing CA's CRL.

[51] Tim Dierks and Eric Rescorla. The Transport Layer Security (TLS) Protocol — Version 1.2. RFC 5246, August 2008

Another method for checking the revocation status of a certificate are online status checks using the Online Certificate Status Protocol (OCSP). Roughly speaking, the difference between OCSP and CRL checks is that CRL checking can be done offline once the current CRL has been retrieved while OCSP requires the RP to be online to inquire the OCSP service about the status of the submitted serial number of the certificate in doubt.

Channel Security, Given that the restrictions and requirements from the previous paragraphs are satisfied, a SOAP message sent over HTTPS is transmitted confidentially and the message's authenticity can also be checked by the recipient. However, TLS is a connection-oriented protocol, which means that the security guarantees are really attached to the communication channel and not to the SOAP message. That is, the SOAP message's authenticity can be *derived* from the authenticity of the connection but cannot be inferred from the SOAP message alone. As a consequence, the security guarantees given by the communication channel are lost as soon as the message exits the channel, i.e., when it is received by the Web service. Furthermore, the security properties derived from the channel —confidentiality and authenticity— cannot be verified by a third party after the message was received. Thus, if accountability with non-repudiation is desired, TLS is not an option to achieve this. In addition, if the content of the received message was meant only for the eyes of a specific person, e.g., a patient's attending doctor, then TLS will not help either because the message will be available in clear for the receiving end, i.e., some Web server. Thus, *end-to-end security* —from the sending person to the receiving person— is not possible with TLS, either. For MyHealthAvatar, all of the above is required if HTTPS/TLS is used to secure the communication between the platform and a third party service.

7.8.7.3 SOAP with WSS

Another option to provide authenticity, integrity, and confidentiality for messages exchanged between a client and MyHealthAvatar platform is the use of Web Service Security (WSS)[52]. The use of WSS would provide additional security guarantees which cannot be given by HTTPS, e.g., end-to-end security. Unlike HTTPS, WSS is message/document oriented, i.e., the security properties are attached to the transmitted document itself and not to the communication channel. WSS incorporates XML Signature[53] and XML Encryption[54] which provide authenticity/integrity and confidentiality, respectively. In other words, WSS provides security functionality comparable to HTTPS, though at a different level. Another difference to HTTPS is that WSS necessarily changes the SOAP message by

[52] Anthony Nadalin, Chris Kaler, Ronald Monzillo, and Phillip Hallam-Baker (Editors). Web Services Security: SOAP Message Security 1.1 (WS-Security 2004). OASIS Standard Specification, 1 February 2006, February 2006

[53] Mark Bartel, John Boyer, Barb Fox, Brian LaMacchia, and Ed Simon. XML Signature Syntax and Processing (Second Edition). W3C Recommendation 10 June 2008, June 2008

[54] Takeshi Imamura, Blair Dillaway, and Ed Simon. XML Encryption Syntax and Processing. W3C Recommendation 10 December 2002, December 2002

adding a security header which contains the security information, e.g., digital signatures and certificates. If confidentiality is desired, it will be necessary to actually change parts of the SOAP message and replace them by their cipher texts.

WSS, like HTTPS, provides a number of state-of-the art cryptographic algorithms for encryption, message authentication, and key exchange. However, the cryptographic algorithms included in WSS have not been found weak, yet. The use of WSS does not necessarily ensure authenticity/integrity *and* confidentiality because it is allowed to have the former without the latter and vice versa. Hence, one could end up in a similar situation as with TLS cipher suites where no encryption algorithm is given.

WSS' approach to security is quite flexible as it permits to include or exclude specific elements from the SOAP message in a signature or in the cipher text. Even more, it is possible to encrypt different parts of the same message for different receivers, which allows for *separation of duty*. *For instance, a patient's insurance number and address could be encrypted for the administration staff while the patient's health data would be encrypted for the medical personnel only.* Conceptually, ***WSS provides security at the application level*** while ***HTTPS provides it at the transport level***. This could mean more complexity for an application because apart from dealing with application data it would have to deal with security as well. With HTTPS, applications delegate the security to a lower level layer, i.e., TLS, just as they delegate reliable transport to a lower level, e.g., TCP. However, a similar approach can also be taken for WSS and, in fact, this is the idea for MyHealthAvatar security layer in order to enable transparent application call to a Web service.

7.8.8 Connection with third party data source for activity data retrieval and sensor

7.8.8.1 Mobile Devices

Apart from stationary systems, such as personal computers and servers, MyHealthAvatar aims for an integration of mobile devices, external mobile applications and web access applications. Mobile devices, such as smart phones or wireless sensors, have the major benefit of providing the users with flexibility and mobility. Mobile devices are small in size, can easily be carried with, and come with their own power supply. With today's technical progress mobile devices, such as mobile phones, have become a constant companion technically almost equal with personal computers. Thus, it seems natural for MHA to make use of existing devices and also let the patients benefit from utilizing the flexibility of mobile devices.

In contrast to stationary systems, mobile devices are exposed to several constraints having an impact on security. First of all, mobile devices have less computational power which can be an issue when considering expensive computations of cryptographic operations. Another critical constraint is the limited power of the device's battery. Therefore, it has to be taken into account that a mobile device runs out of power in critical situations. Even though the use of mobile devices provides local independence it has to be considered that some areas, especially rural ones, provide a limited bandwidth or even lack network coverage and hence, render the device useless for some applications.

Apart from technical limitations mobile devices are more strongly exposed to the threat of physical access than stationary systems, e.g. at home. Also, mobile devices are easily lost and far more subject to theft than stationary systems. Thus, it should not be overlooked that a mobile device might fall in the hands of an adversary far more easily than other platform components. Physical access, in turn, opens the door for additional attacks, such as getting direct access to the device's file system, memory card, or system memory.

Today, many mobile consumer devices, such as mobile phones or set-top boxes, enable the user to utilize online offers and services by connecting to the Internet. However, the Internet, can also be a loop hole for further threats to the device. Modern mobile operating systems are equal to their stationary counterparts in complexity and performance. Thus, mobile devices are also subject to remote attacks, such as session hijacking or the installation and execution of malicious code.

Device Classification. Mobile devices, such as laptops, which are with regard to technical capabilities not distinguishable from stationary systems can be neglected when considering particular security consideration which would typically arise for mobile devices with technical limitations. For these devices the same security considerations as for stationary systems apply.

For mobile devices with strong technical constraints, such as wireless sensors (e.g., glucose meter, SpO_2 meter, HR meter, BP meter, etc.), special assumptions concerning security have to be made. These devices are typically very limited in computational power or networking capabilities so that some security mechanisms cannot be implemented in practice. Due to the technical constraints, however, the threat of adversary installing and executing malicious code on these devices is unlikely.

Device Security. When considering security it has to be taken into account that some mobile platforms are innately equipped with security measures that protect from attacks which require physical access. For example, some mobile operating systems provide secure storage by encrypting the device's file system. Thus, some threats can be neglected depending on the underlying platform, i.e., the mobile device's operating system. Considering the spectrum of mobile platforms, the provision of complete device security is out of the scope of MyHealthAvatar. Therefore, the security of a mobile device's operating environment is rather assumed than enforced with in the terms of MyHealthAvatar project.

7.8.8.2 Protocols and Standards

In this section, we present some details on protocols and related security aspects that are envisaged to be used within the scope of MyHealthAvatar mobile devices communication and medical device communication. For more details the reader is referred to the respective standard's publication.

7.8.8.2.1 Continua

The Continua Health Alliance is a non-profit, open industry alliance of several healthcare and technology companies. Continua's mission is to establish a system of interoperable personal solutions that offer a more independent and improved management of health and wellness. One of Continua's main objectives is develop design guidelines that will enable vendors to attend a product certification program in order to build interoperable sensors, home networks, eHealth platforms, and health and wellness services. The Continua Health Alliance’s Design Guidelines[55] are based on existing standards and specifications that Continua selected for ensuring end-to-end interoperability of devices. It also contains additional guidelines for interoperability that reduce or extend options in the underlying standards or specifications. By adhering to these guidelines, products of different manufacturers can exchange health information seamlessly. For MyHelathAvatar, the major benefit of complying with the Continua Design Guidelines is the end-to-end interoperability based on established industry standards. Using Continua certified devices allows for seamless communication between heterogeneous devices over heterogeneous networks. Interoperability also helps to stimulate technological improvements and enables users to use devices of their choice. Healthcare providers are able to easily switch and integrate other device manufacturers, reduce costs and, finally, optimise the patient’s treatment. The Continua Design Guidelines aim at the interoperability of devices with a focus on the transport and the data level. The goal is to empower users to control how their health information is shared and used in a personal electronic health eco-system.

For MyHealthAvatar, additional concepts regarding privacy, trust, and security play an important role. As mentioned, for adoption of personal eHealth systems, trust, security and privacy are very important. The same holds for compliance to legislation like EU Directive 95/46 and HIPAA[56]. Continua acknowledges the importance of these issues amongst others through its E2E Security Task Force. Initial security and privacy issues have been addressed in Continua version 1 guidelines for the PAN and HRN interfaces. Continua version 1.5 guidelines added security features for the WAN and LAN interfaces with e.g. TLS for secure communication and SAML 2.0 tokens for authentication of AHD users. With personal eHealth systems emerging today people are able to participate in their own care supported by an open distributed system with health services. This poses new end-to-end security and privacy challenges. In Koster et. al.[57] new end-to-end security requirements and presented with a design for consent management in the context of the Continua Health Alliance architecture.

[55] Continua Design Guidelines - Version 2015, http://www.continuaalliance.org/products/design-guidelines

[56] http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31995L0046:en:HTML

[57] Koster P, Asim M, Petkovic M., End-to-end security for personal telehealth, Stud Health Technol Inform. 2011;169:621-5.

7.8.8.2.2 ZigBee

ZigBee is a very popular wireless communication standard for low-cost, low-power devices secure communication especially among medical devices[58]. ZigBee makes use of the security mechanisms defined by the IEEE 802.15.4 standard[59]. It also complements by providing security management functions for different key types. ZigBee supports keys for the following functions: a) key establishment (also known as key agreement); b) key transport (for master keys, network keys, link keys); c) frame protection (regarding integrity, authenticity, and confidentiality) and d) device management (configuration). In general, ZigBee distinguishes two key types, *network keys* and *link keys*. Both key types are 128 bit long and make use of the Advanced Encryption Standard (AES)[60] operated in the CCM mode (Counter with CBC-MAC[61]) which provides *authenticated encryption*. The former key type is used for broadcast communication within a device network and the latter one is used for unicast communication between two (paired) devices. Link keys are acquired by a device or through pre-installation. A network key is shared among all devices from the same network and used to secure communication (broadcast or point-to-point) between these devices. Network keys come in two flavours, *standard* and *high-security*. The difference between the two is only how the network key is being distributed and, possibly, how frame counters for messages are initialised.

7.8.8.2.3 Bluetooth

Bluetooth is a well-known open standard for short range radio frequency communication, primarily used to establish wireless personal area networks (WPAN). Bluetooth is based on a so called master-slave architecture allowing point-to-point communication from the master to the slave. In Bluetooth, the level of security highly depends on the version of the standard implemented by the device because early versions use procedures which are not considered secure[62], e.g., downgrading of encryption key sizes. Bluetooth supports keys for the following purposes: device authentication and message

[58] Frehill P, Chambers D, Rotariu C., Using Zigbee to integrate medical devices, Conf Proc IEEE Eng Med Biol Soc. 2007;2007:6718-21.

[59] IEEE Computer Society. IEEE Standard for Information technology – Telecommunications and information exchange between systems – Local and metropolitan area networks – Specific requirements, Part 15.4: Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications for Low-Rate Wireless Personal Area Networks (WPANs). IEEE Std 802.15.4–2006, September 2006

[60] U.S. National Institute of Standards and Technology NIST. Advanced Encryption Standard (AES). FIPS PUB 197, November 2001

[61] Doug Whiting, Russ Housley, and Niels Ferguson. Counter with CBC-MAC (CCM). RFC 3610, September 2003

[62] Karen Scarfone and John Padgette. Guide to Bluetooth Security. NIST Special Publication 800121, September 2008. Recommendations of the National Institute of Standards and Technology

encryption. The authentication and encryption procedures employ symmetric algorithms called E_1 and E_0, respectively, which are specific to Bluetooth and have also been found to be weak[63]. Note that Bluetooth does not offer frame or message authenticity. It only employs CRC checksums to guard against transmission errors in noisy channels. However, correct checksums can be computed by anyone and thus, messages can be injected without the recipient being able to detect this. Also note that encryption is not a replacement for proper message authentication[64].

7.9 Cloud Security

Security is a primary concern when we choose the cloud service. A private cloud allows IT to control the perimeter and ensure the highly secured data not to be transferred out of the premise; but it's also responsible for staying on top of a rapidly shifting security landscape and making all required fixes, updates, and upgrades. While by using public clouds, data is protected by both managed security on a software and physical level, since large-scale data centres like those used by public cloud providers have state-of-the-art security.

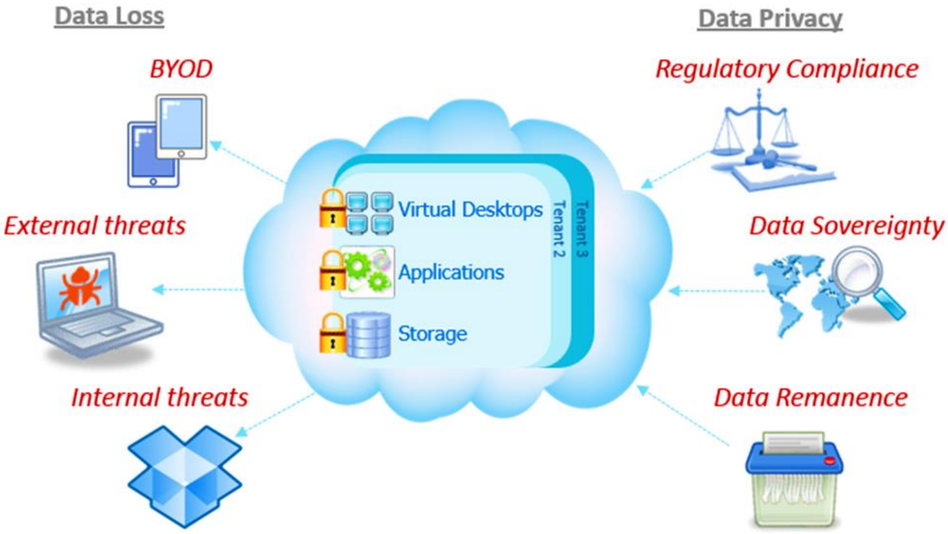


Figure 7-1: Securing data cloud infrastructure

Service availability and disaster recovery is another key consideration. Public cloud typically operates in multiple data centres at multiple geo locations, which provides high level service availability especially in the scenarios of disaster recovery. Private cloud typically resides in one premise and

[63] Yi Lu, Willi Meier, and Serge Vaudenay. The Conditional Correlation Attack: A Practical Attack on Bluetooth Encryption. In Crypto 2005, number 3621 in LNCS. Springer Verlag, 2005.

[64] Michael McIntosh, Martin Gudgin, K. Scott Morrison, and Abbie Barbir. Basic Security Profile Version 1.0. Web Services Interoperability Institution (WS-I), March 2007

requires expensive proposition in order to ensure the constant service availability. There are important differences between each model in terms of merged features, complexity and security. Cloud service providers can provide the basic security architecture; consumers are responsible for implementing and managing the provided security features. Cloud architectures are built upon underlying technology. A cloud built over the Internet inherits all of the internet's inherent security risks. Even if an enormous amount of security is established in the cloud, data must still be transmitted via the underlying internet technology. Therefore, the security concerns threatening the internet also threaten the cloud. However, the risks to cloud computing are especially great. The vulnerability consideration and asset value of the resources and asset value of the resources and their nature of them settling together. Cloud systems still use normal internet protocols and security standards but require greater levels of security. Although secure protocols and encryption cater to current needs to a certain extent, they are not context oriented. A strong set of policies and protocols is necessary to secure data transmission within the cloud. Concerns regarding the intrusion of external non-users into cloud databases should also be considered. Standards should be established to construct a secure, private and isolated cloud environment in the internet that is capable of avoiding attacks. Below we try to investigate on security issues for SaaS, PaaS and IaaS.

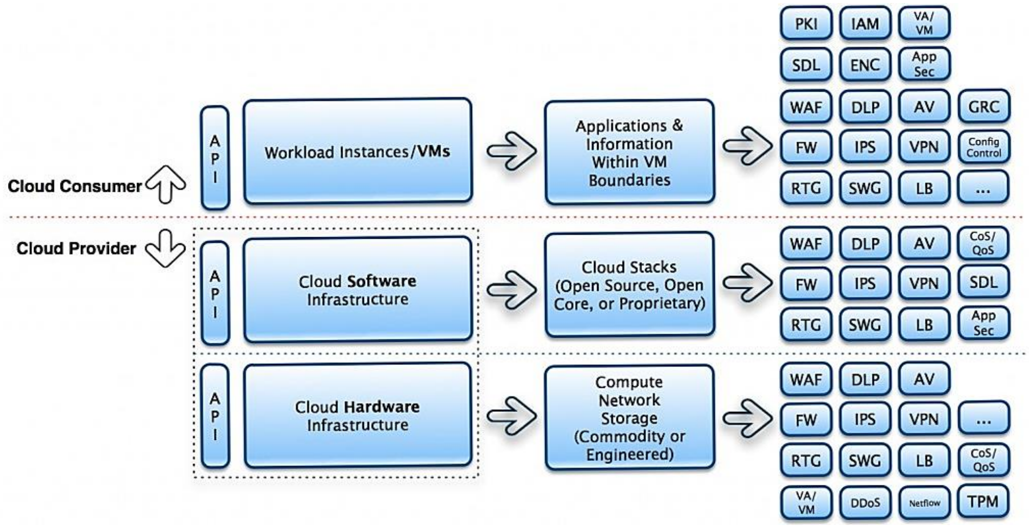


Figure 7-2: Cloud Security elements (examples)[65]

[65] http://www.ncbi.nlm.nih.gov/pubmed/25732083

7.9.1 Security Issues for cloud delivery models (SaaS, PaaS, IaaS)

In **SaaS**, the client's security measures are dependent on the provider. The provider should ensure that each user's data are hidden from all other users. Security measures must be in place and the client must be confident that the application will be ready for use when needed. In SaaS, the cloud client will often replace old software applications with newer ones letting the focus of security to be upon protecting or developing the security functionality of applications/service and attaining successful data relocation[66]. There is great concern that problems involving data availability or data breaches could lead to financial and legal liabilities[67]. Security components should be considered essential parts of the SaaS application development and data deployment processes, including security, network security, locality, integrity, segregation, access, authentication and authorization, confidentiality, web application security, breaches, virtualization vulnerability, availability, backup, identity management and sign-on processes (see Figure 2).

In **PaaS**, developers build applications on a computing platform controlled by the provider. In addition, any security issues beneath the application level, such as network and host intrusion prevention, are under the control of the provider, who must offer strong guarantees that the data cannot be accessed by other applications[3]. As a result, PaaS offers more flexibility than SaaS at the expense of customer-ready features. This trade-off extends to security features and capabilities, in that built-in capabilities are less complete, but, simultaneously, there is more flexibility to incorporate additional security.

In **IaaS**, the developer has the best control over security, as long there is no security hole in the Virtualization Manager (VM). While in theory virtual machines might be able to address these issues as they arise, there are many security problems in practice. An additional factor is the reliability of the data stored in the provider's hardware. Due to the growing virtualization of the information society, enabling owners to maintain control over their data regardless of its physical location will become a topic of extreme interest. To obtain maximum trust and security on a cloud resource, several techniques need to be practiced, as describe in Descher *et al*., 2009 [68]. The security obligations of both

[66] Subashini, S. and V. Kavitha, 2011. A survey on security issues in service delivery models of cloud computing. J. Netw. Comput. Appli., 34: 1-11. DOI: 10.1016/j.jnca.2010.07.006

[67] Anding, M., 2010. SaaS: A Love-Hate Relationship for Enterprise Software Vendors. In: Software-as-aService: Anbieterstrategien, Kundenbedürfnisse und Wertschöpfungsstrukturen, Benlian, A., T. Hess and P. Buxmann (Eds.), Springer DE, Wiesbaden, ISBN10: 383498731X, pp: 43-56.

[68] Descher, M., P. Masser, T. Feilhauer, A.M. Tjoa and D. Huemer, 2009. Retaining data control to the client in infrastructure clouds. Proceedings of the International Conference on, Availability, Reliability and Security, Mar. 16 19, IEEE Xplore Press, Fukuoka, pp: 9-16. DOI: 10.1109/ARES.2009.78

the provider and the consumer vary greatly between cloud service models. Amazon's Elastic Compute Cloud (EC2) infrastructure presents an example in which the vendor's responsibility for security extends only to the hypervisor. This means that they can only address security controls such as virtualization security, physical security and environmental security. The consumer is responsible for the security controls corresponding to the system, including the applications, OS and data. IaaS gives rise to security issues whose severity depends on the cloud deployment model through which the services are delivered. The physical security of the infrastructure is extremely important; disaster management plans are necessary to prevent damage, either natural or intentional, to the infrastructure. Infrastructure includes not only the hardware in which data are computed and stored but also the paths by which it is obtained or transmitted. In a standard cloud environment, data will be transmitted from source to destination through numerous third-party infrastructure devices[69].

Today given that the development of cloud computing services is speeding up the rate in which the organizations outsource their computational services or sell their idle computational resources. One of the most important aspect when organizations or private companies try to exploit the possibility of deploying their services in to the cloud is security: while some cloud computing security issues are inherited from the solutions adopted to create such services, many new security questions that are particular to these solutions also arise, including those related to how the services are organized and which kind of service/data can be placed in the cloud.

7.9.2 Cloud security Taxonomy

Aiming to give a better understanding and classify main security concerns and solutions in cloud computing, Gonzales et. al[70] discuss security concerns and solutions for cloud computing and propose a taxonomy of security in cloud computing, giving an overview of the current status of security in this emerging technology. Their analysis on security concerns in the context of cloud computing solutions shows that each issue brings different impacts on distinct assets. Aiming to create a security model both for studying security aspects in this context and for supporting decision making, in this section we consider the risks and providing a cloud security taxonomy. The architecture dimension is subdivided into network security, interfaces and virtualization issues, comprising both user and administrative interfaces to access the cloud. It also comprises security during transferences of data and virtual machines, as well as other virtualization related issues, such as isolation and cross-VM

[69] Ristenpart, T., E. Tromer, H. Shacham and S. Savage, 2009. Hey, you, get off of my cloud: Exploring information leakage in third-party compute clouds. Proceedings of the 16th ACM Conference on Computer and Communications Security, Nov. 09- 13, Chicago, IL, USA, pp: 199-212. DOI: 10.1145/1653662.1653687

[70] Gonzalez, Nelson Mimura Icon ; Miers, Charles Christian Icon ; Redígolo, Fernando ; Simplicio Junior, Marcos Antonio Icon ; Carvalho, Tereza Cristina Melo de Brito Icon ; Näslund, Mats ; Pourzandi, Makan, A quantitative analysis of current security concerns and solutions for cloud computing, Journal of Cloud Computing: Advances, Systems and Applications, Heidelberg, v.1, 2012

attacks. The architecture group allows a clearer division of responsibilities between providers and customers, and also an analysis of their security roles depending on the type of service offered (Software, Platform or Infrastructure).

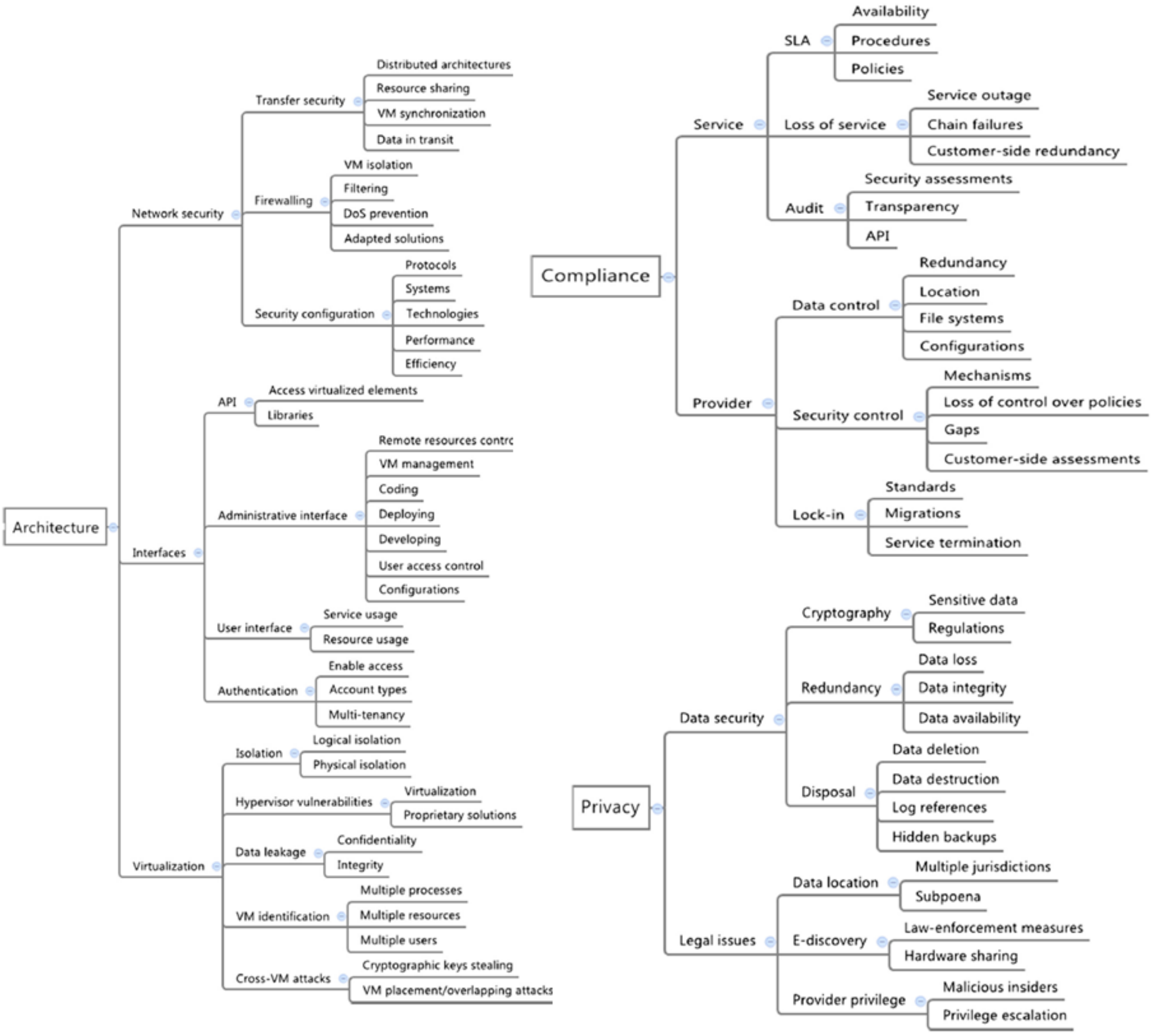


Figure 7-3: Security taxonomy (Architecture, Privacy, and Compliance), (Gonzales et. Al.)

This suggests that the security mechanisms used must be clearly stated before the service is contracted, defining which role is responsible for providing firewalling capabilities, access control features and technology-specific requirements (such as those related to virtualization). The compliance dimension introduces responsibilities toward services and providers. The former includes SLA concerns, loss of service based on outages and chain failures, and auditing capabilities as well as transparency and security assessments. The latter refers to loss of control over data and security policies and configurations, and also lock-in issues resulting froml ack of standards, migrations and service terminations. The privacy dimension includes data security itself (from sensitive data, regulations and data loss to disposal and redundancy) and legal issues (related to multiple jurisdictions derived from different locations where data and services are hosted). The concerns in this dimension cover the complete information lifecycle (i.e., generation, use, transfer, transformation, storage, archiving, and destruction) inside the provider perimeter and in its immediate boundaries (or interfaces) to the users.

7.9.3 Security in MHA Cloud infrastructure

7.9.3.1 Private Cloud (OpenStack)

An overview of security measure for MyHealthAvatar is summarized below (Details on the deployment of OpenStack can be found below):

Physical security

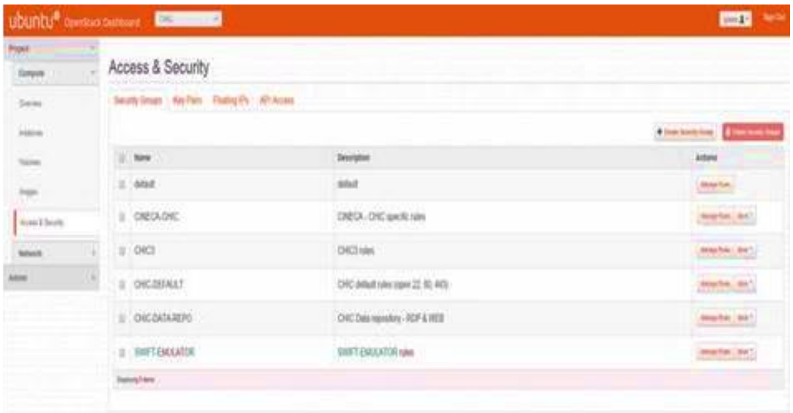
- Machine room in a restricted access area
- Smartcard & PIN access control
- Heavy duty air conditioning system
- UPS (uninterruptible power supply)
- Intrusion detection alarm
- Temperature fluctuation alarm
- Fire alarm
- Automatic CO2 fire extinguisher
- Automatic power generator in case of power shortage
- Administrative stuff available on nearby office

Network security

- Firewall
- IP based access control
- Certificate-based login (optional)

Data security

- RAID controllers (hardware)
- Automatic replication of data (software -Openstack)
- Tape backups



Access & Security

Name	Description
default	default
CINECA-CHIC	CINECA - CHIC specific rules
CHIC3	CHIC3 rules
CHIC-DEFAULT	CHIC default rules (open 22, 80, 443)
CHIC-DATAREPO	CHIC Data repository - RDF & WEB
SWIFT-EMULATOR	SWIFT-EMULATOR rules

Figure 7-4: MHA cloud security control panel (OpenStack)

User provisioning and identity is the process of registering a new user with a given system and user de-provisioning is the process of removing a user from the system. OpenStack object storage “swift” offers significant automation of user data management tasks by using authentication/authorization systems referred to as “tempAuth” and “swAuth”. The difference between “tempAuth” and “swAuth” lies in the backend storage of user data. TempAuth uses a configuration file in which user data are saved as plain text. On the other hand, swAuth is meant to be a “scalable authentication and authorization system that uses swift itself in a backing store. The characteristics of user management are based on OpenStack object storage “swift”. The following characteristics are present:

- Users are not given administrative power over any other users
- Provider Admins have admin agreements with all accounts but cannot add other provider admins
- Super admins are powerful users who are able to perform all user management procedures, including adding provider admins

TempAuth and swAuth often use a username and password for the authentication process. When authentication is successfully performed, the user receives a token that will identify him to the system for a period of time. The provided token has a configurable expiration time, the default value of which is set to 4-6 h. All cloud security documents must, allow authentication by accepting confirmations in SAML format; however, this feature is not yet available in OpenStack . Because all OpenStack projects use a password and username system to authenticate users, password strength requirements should receive greater scrutiny.

Authentication tokens play similar roles as identifiers for web applications. An API, such as an OpenStack service, is used to authenticate a user. Successful authentication generates a token that is used to authorize service requests. The password and username are given as input to the API interface. When authentication succeeds, the resulting feedback includes an authentication token and service catalogue. Note that tokens remain valid for 12 h. Issued tokens become invalid in two situations: If the token is expired or if the token has been cancelled. It is important that the authentication be executed over a secure channel, such as Transport Layer Security (TLS); otherwise, an attacker could obtain a user token by executing a man-in-the-middle-attack and remove the user who received the token from the authentication system. Lastly most cloud providers do not encrypt data before saving it to a cluster. In fact, OpenStack does not provide any data encryption at all; thus, users would need to encrypt their data before uploading it and manage their encryption keys themselves. Given the difficult of track security issues in cloud computing environments the following table provides a summary of security issues, which are divided into five categories and listed with their implications.

Security issues	Implications of Security
Trust	This is interrelated to the designated deployment modal because the control of the data and applications is directly supervised by the strict control of the owner
Availability	The capacity of a system to operate upon the demands of a certified entity. This notion implies that the system should be able to function even in the presence of authorities that disobey the regulations. Furthermore, the system must also maintain the capacity to operate even in the existence of a security breach
Integrity	Resources can only be reformed by approved individuals and through official procedures.
Software	The diverse resources include data, software and hardware
● Data (Authentication, Authorization and Access control AAA) Confidentiality	Data in cloud computing are more vulnerable because of the increase in the number
● Software	of individuals, devices and applications that use cloud computing which will in turn
● Data	increase the number of access points. Consequently, authorized individuals and systems are the only entities that are allowed to access the protected data
Privacy	An individual's need to govern the entree to his/her personal information

7.9.3.2 Public Cloud

Linode London cloud infrastructure is built based on Telecity group's Powergate data center, which is designed, built and operated by top engineers to the highest industry standards. Powergate data center features high-density capabilities, which support the power, cooling and management requirements of Linode's cloud servers. Following will refer to the physical Powergate data center as Linode cloud.

Power

- Flexible and scalable power densities and configuration of up to 20kW
- Fully redundant and resilient power supplies at a minimum of N+1
- Power distribution through dedicated A&B supplies
- Separate UPS system which covers 15 minutes of grid-power failure
- Backup diesel generation at N+1 for event of grid-power failure, which is always warm up and work in seconds

Cooling

- Robust heating, ventilation and air conditioning (HVAC) system
- Minimum N+1 redundancy on water-cooling system
- Close Control Unites (CCUs) provide conditioned air at minimum N+2 redundancy
- Fully concurrently maintainable, dual-path secondary water distribution

Fire detection and suppression

- High sensitivity smoke detection system (VESDA)
- Fully addressable two-stage fire detection and monitor system
- Water mist fire suppression system
- Network monitored fire-protection system

Security and access

- Protected by security personnel and multi-layered physical security
- Integrated digital video camera surveillance throughout the exterior and interior of data center
- Independent client-card and biometric-identification access system
- 24/7/265 manned security with unified security-breach alarm

8 MyHealthAvatar platform deployment

8.1 Cloud deployment infrastructure

Cloud computing is a model for enabling convenient, on demand network access to a shared pool of configurable computing resources that can be rapidly provisioned and released with minimal management effort or service provider interaction. MyHealthAvatar, collecting, accessing, managing and possibly sharing healthcare related data are not only important to individuals who can manage their own health, but also important for clinicians and other healthcare workers for patient monitoring and providing suitable in-time care. The consideration of cloud technology is vital to ensure the long term scalability and performance of such systems. MyHealthAvatar cloud infrastructure is a vital component in the data management and service management architecture.

Cloud computing utilizes three delivery models in which different types of services are delivered to the end user. The three delivery models are Software as a Service (SaaS), Platform as a Service (PaaS) and Infrastructure as a Service (IaaS), which provide infrastructure resources, application platforms and software as services to the consumer[71] shown in Figure. 1. These service models also place different levels of security requirements upon the cloud environment. IaaS is the foundation of all cloud services, PaaS builds upon IaaS and SaaS, in turn, builds on PaaS. As capabilities are inherited by successive models, so too are information security issues and risks.

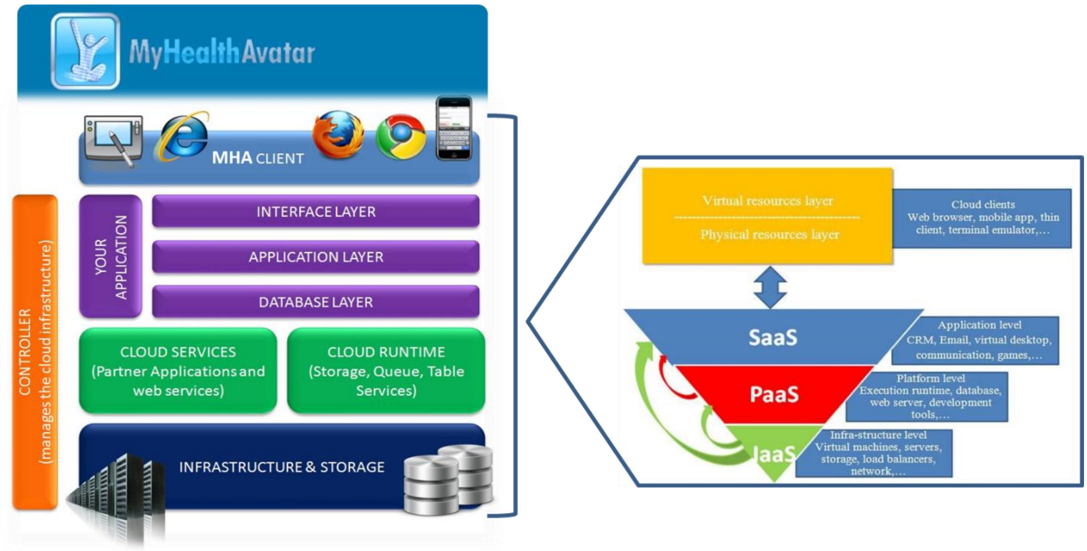

MyHealthAvatar
MHA CLIENT
CONTROLLER (manages the cloud infrastructure)
YOUR APPLICATION
INTERFACE LAYER
APPLICATION LAYER
DATABASE LAYER
CLOUD SERVICES (Partner Applications and web services)
CLOUD RUNTIME (Storage, Queue, Table Services)
INFRASTRUCTURE & STORAGE
Virtual resources layer
Physical resources layer
Cloud clients
Web browser, mobile app, thin client, terminal emulator,…
SaaS
Application level
CRM, Email, virtual desktop, communication, games,…
PaaS
Platform level
Execution runtime, database, web server, development tools,…
IaaS
Infra-structure level
Virtual machines, servers, storage, load balancers, network,…

Figure 8-1: Cloud computing architecture considered for MHA

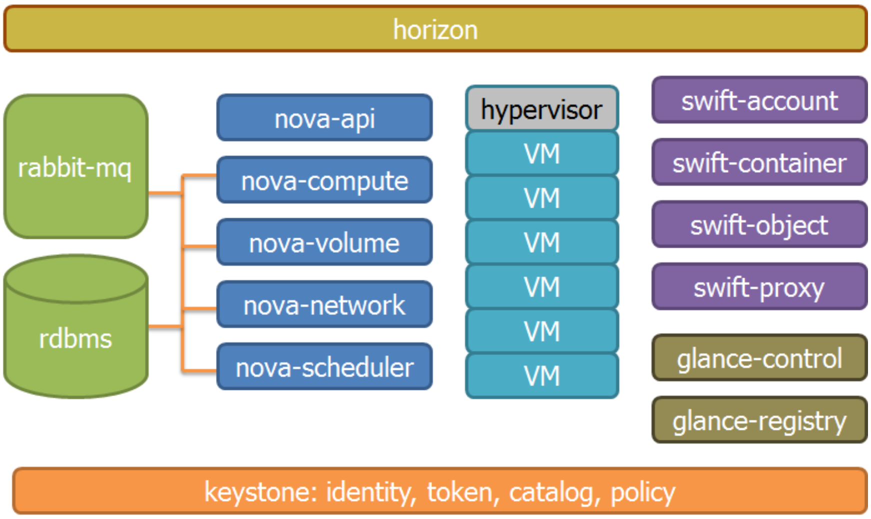
[71] Shey, H., R. Wang, J.P. Garbini and E. Daley, 2009. The State of Enterprise Software: 2009. Forrester Research, Inc.

The main approach is to work on a locally-deployed cloud infrastructure which can utilize local computing power as well as maintain the ability to outsource the infrastructure to commercial cloud computing facilities (e.g. Amazon EC2). This could be potentially used by the organizations that host the data as the architecture to perform data storage and analysis services to provide information for remote users without transferring the data out. This way we can provide stable and production ready resources to support project needs regarding medical data preservation. The swift services are highly autonomous so the whole architecture is flexible enough to allow different deployment scenarios. The four main services are: Proxy Services, Object Services, Container Services and Account Services. Proxy Service plays role of the contact point (API) with users and 3rd party services, while other three kind of services manage files, containers and accounts so are used to manage physical data and logical structure. The major difference between locally deployed cloud (also known as private cloud) and public cloud is the control and access of the resource. In private cloud, resources are controlled and accessed by the premise only. On the other hand, in public cloud, resources are controlled by the cloud provider but are accessible for public users. Therefore, a hybrid cloud infrastructure is adopted in MyHealthAvatar with both public and private cloud facilities available.

8.1.1 Private Cloud deployment for MHA (OpenStack)

OpenStack is a cloud operating system which basically, provides IaaS. It is actually designed to manage data centers and has no proprietary hardware or software requirements.

OpenStack structure, is presented at the illustration on the right:

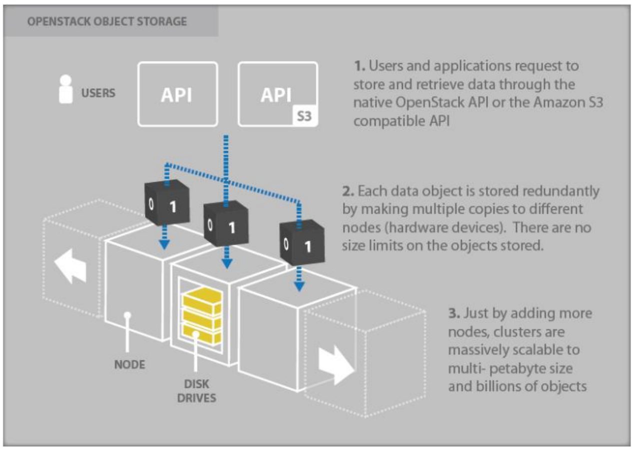


OpenStack is built upon nine different services:

Service	Codename	Description
Dashboard	Horizon	A web based interface for OpenStack services
Compute	Nova	Provides virtual servers on demand
Identity	Keystone	A framework for authentication and authorization

Network	Neutron	It provides "*network as a service*" between interface devices (e.g., vNICs) managed by other OpenStack services (e.g., nova). The service works by allowing users to create their own networks and then attach interfaces to them. Quantum has a pluggable architecture to support many popular networking vendors and technologies
Image Service	Glance	Catalog and repository service for virtual disk images
Block Storage	Cinder	Traditional block storage services to VMs (iSCSI, FC etc.)
Object Storage	Swift	Object store allows you to store or retrieve files. It provides a fully distributed, API-accessible storage platform that can be integrated directly into applications or used for backup, archiving and data retention.
Monitoring	Ceilometer	Monitors and meters the OpenStack cloud for billing, benchmarking, scalability, and statistics purposes.
Orchestration	Heat	Orchestrates multiple composite cloud applications by using the AWS CloudFormation template format, through both an OpenStack-native REST API and a CloudFormation-compatible Query API.

Storage options that OpenStack provides are of great significance. The object storage is exposed via the relevant API, so that it can be integrated into applications or even used as backup and data archiving. The OpenStack object storage, has several capabilities:



1. Users and applications request to store and retrieve data through the native OpenStack API or the Amazon S3 compatible API

2. Each data object is stored redundantly by making multiple copies to different nodes (hardware devices). There are no size limits on the objects stored.

3. Just by adding more nodes, clusters are massively scalable to multi- petabyte size and billions of objects

- Object storage is provided using clusters of servers capable of storing large amount of data of data. The storage has characteristics of redundancy and scalability.
- It's a distributed storage system for static data such as virtual machine images, photo storage, email storage, backups and archives.
- Objects and files are written to multiple disk drives spread throughout servers in the data center
- Storage clusters scale horizontally simply by adding new servers.
- Should a server or hard drive fail, OpenStack replicates its content from other active nodes to new locations in the cluster.

On the other hand, Block storage uses traditional storage options over iSCSI or FC protocols. In brief:

- The block storage system manages the creation, attaching and detaching of the block devices to servers.
- It has unified storage support for numerous storage platforms including Ceph, NetApp, Nexenta, SolidFire, and Zadara.

- Block storage is appropriate for performance sensitive scenarios such as database storage, expandable file systems, or providing a server with access to raw block level storage.
- Snapshot management provides powerful functionality for backing up data stored on block storage volumes. Snapshots can be restored or used to create a new block storage volume.

OpenStack networking of VMs can be deployed with either as flat networking or VLAN networking. To summarize:

- OpenStack Networking manages IP addresses, allowing for dedicated static IPs or DHCP. Floating IPs allow traffic to be dynamically rerouted to any of your compute resources, which allows you to redirect traffic during maintenance or in the case of failure.
- Users can create their own networks, control traffic and connect servers and devices to one or more networks.
- The pluggable backend architecture lets users take advantage of commodity gear or advanced networking services from supported vendors.
- Administrators can take advantage of software-defined networking (SDN) technology like OpenFlow to allow for high levels of multi-tenancy and massive scale.
- OpenStack Networking has an extension framework allowing additional network services, such as intrusion detection systems (IDS), load balancing, firewalls and virtual private networks (VPN) to be deployed and managedOpenStack is designed to manage and automate pools of compute resources. It takes advantage of the popular KVM or Xen hypervisors and it also supports ARM and alternative hardware architectures. Finaly OpenStack provides API for both EC2 and S3. For more details

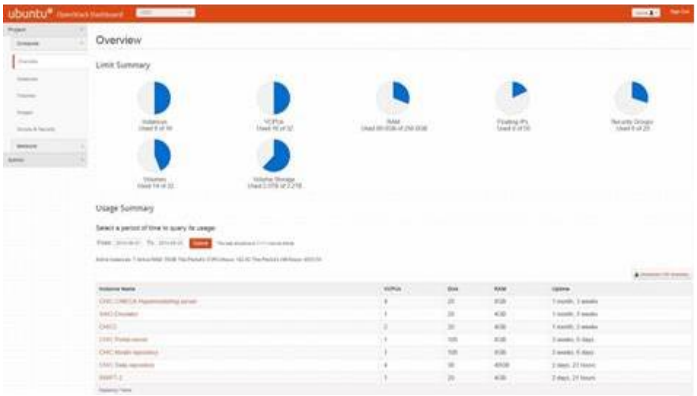


Figure 8-2: MHA private cloud control Panel

Deployment

MHA platform will be deployed in the premises of FORTH, in the form of a private computational and storage cloud.

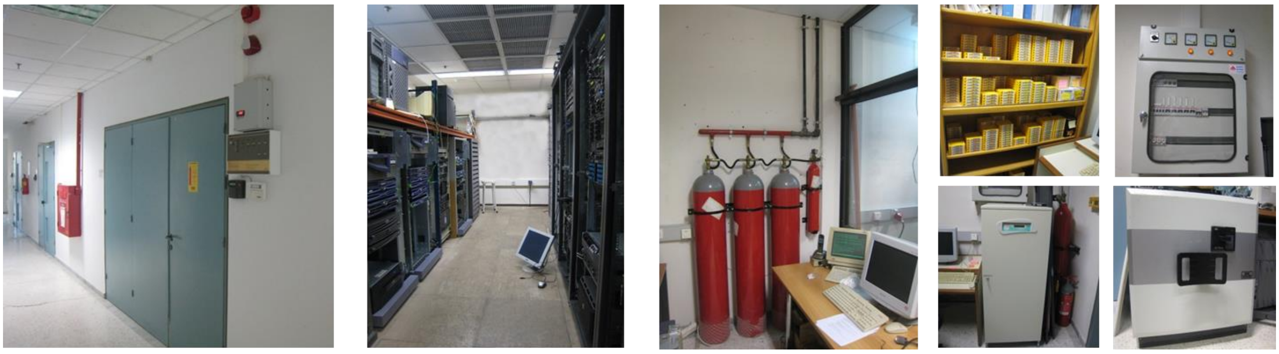


Figure 8-3: Physical installation

In terms of hardware resources, the cloud infrastructure allows for maximum elasticity and flexibility by effectively adapting to the load of any given time. The current minimal specifications include:

- 300 GB of RAM
- 9TB of storage
- 16 cores Intel® Xeon® Processor E5-2690 and 4 cores Intel® Xeon® Processor E7520 (Dell PowerEdge R720 and SC 1425 Servers series)

In terms of the software the OpenStack15 open source cloud computing software has been installed on the machines using the Linux Ubuntu 12.04 operating system16. **Fort MHA we can have one or more VMs for 4GB RAM and 100GB for storage (or any other different setting for each VM), according to MHA deployment demands. We are able to serve MHA need as the project progress.**

8.1.2 Public Cloud deployment for MHA

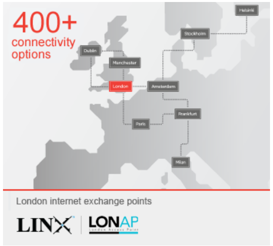
Linode is a cloud hosting which provides high performance SSD Linux servers for MHA public cloud evaluation and deployment. MHA mainly uses Linode's UK datacentre based in London, which is located at Telecity Group's state-of-art Powergate data centre. Many well-known providers use the same facility, which includes Deutsche Telekom, Interoute, Level3, PCCW-BTN, Telia and Tiscali.

The main features of Linode (Powergate) data centre are:

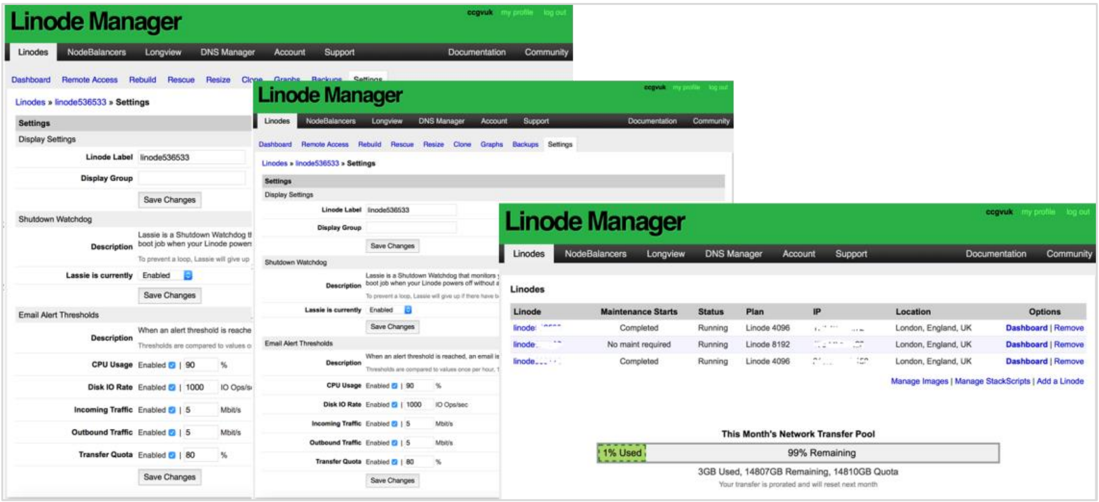
- 5000m^2 data centre
- 25MVA incoming supply
- 10MWV of resilient power
- 4kW of power density per rack
- High Levels of connectivity, including the London Internet Exchange (LINX)

Connectivity

- Direct connectivity to a wide range of networks
- Global peering and connectivity
- Managed connectivity with multiple carriers
- Pan-European connectivity across the data center network



Management interfaces



Linode Manager

Linodes » linode536533 » Settings

Settings

Display Settings

Linode Label: linode536533

Display Group:

Save Changes

Shutdown Watchdog

Description: Lassie is a Shutdown Watchdog that monitors y boot job when your Linode powers off without a

To prevent a loop, Lassie will give up if there have b

Lassie is currently: Enabled

Save Changes

Email Alert Thresholds

Description: When an alert threshold is reached, an email is

Thresholds are compared to values once per hour. T

CPU Usage	Enabled ☑	90	%
Disk IO Rate	Enabled ☑	1000	IO Ops/sec
Incoming Traffic	Enabled ☑	5	Mbit/s
Outbound Traffic	Enabled ☑	5	Mbit/s
Transfer Quota	Enabled ☑	80	%

Save Changes

Linodes

Linode	Maintenance Starts	Status	Plan	IP	Location	Options
linode[redacted]	Completed	Running	Linode 4096	[redacted]	London, England, UK	Dashboard | Remove
linode[redacted]	No maint required	Running	Linode 8192	[redacted]	London, England, UK	Dashboard | Remove
linode[redacted]	Completed	Running	Linode 4096	[redacted]	London, England, UK	Dashboard | Remove

Manage Images | Manage StackScripts | Add a Linode

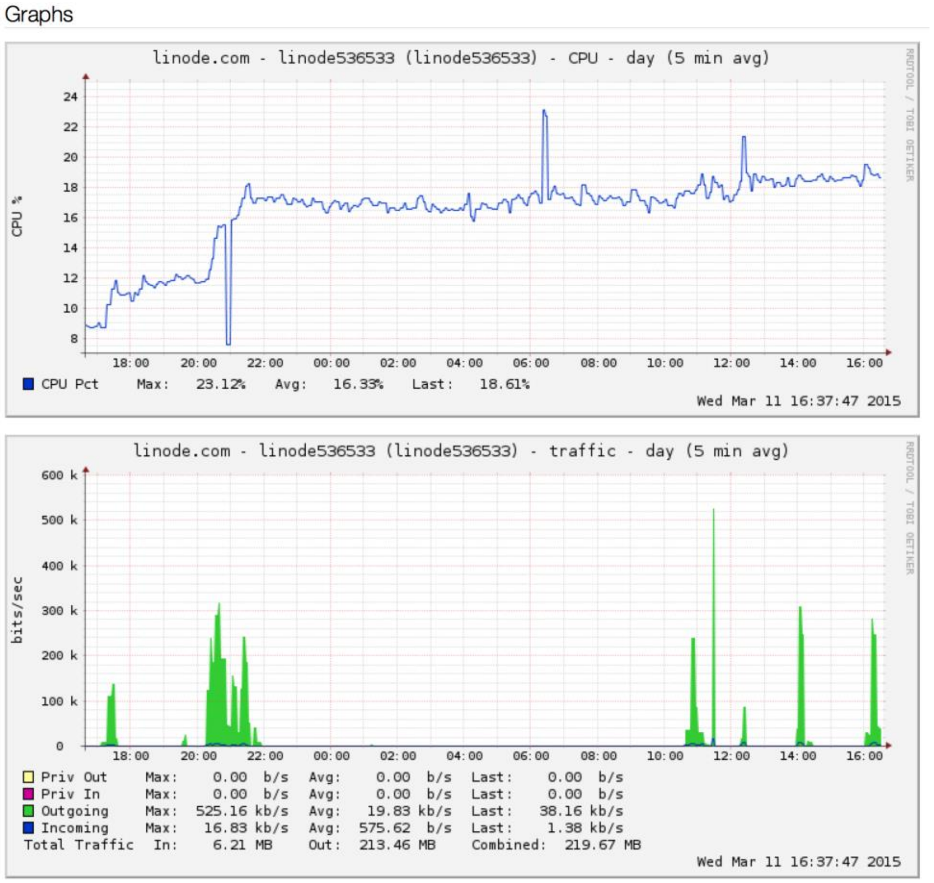
This Month's Network Transfer Pool

1% Used 99% Remaining

3GB Used, 14807GB Remaining, 14810GB Quota

Your transfer is prorated and will reset next month

Monitoring interface



Graphs

```
■ CPU Pct     Max:   23.12%   Avg:   16.33%   Last:   18.61%
                                              Wed Mar 11 16:37:47 2015
```

```
□ Priv Out     Max:    0.00  b/s   Avg:    0.00  b/s   Last:    0.00  b/s
■ Priv In      Max:    0.00  b/s   Avg:    0.00  b/s   Last:    0.00  b/s
■ Outgoing     Max:  525.16 kb/s   Avg:   19.83 kb/s   Last:   38.16 kb/s
■ Incoming     Max:   16.83 kb/s   Avg:  575.62  b/s   Last:    1.38 kb/s
Total Traffic  In:    6.21 MB      Out:  213.46 MB     Combined:  219.67 MB
                                                 Wed Mar 11 16:37:47 2015
```

8.2 MHA Model Repository

Following the directives of *D5.1: Model and clinical data repositories design*, the MHA Generic Tool/Model repository part is being implemented utilizing the Model-View-Controller architectural pattern. A combination of Python’s Django web framework and MySQL is being used for the front and back end, respectively. Taking into consideration its functionality within the MHA platform, the Generic Tool/Model repository (MHA Model Repository) will contain a subset of the tables that were originally defined in D5.1. The basic E-R schema concerning the tool description and classification, files that are related to tools and Basic Authentication and Authorization as well as sessions and logging tables will be included.

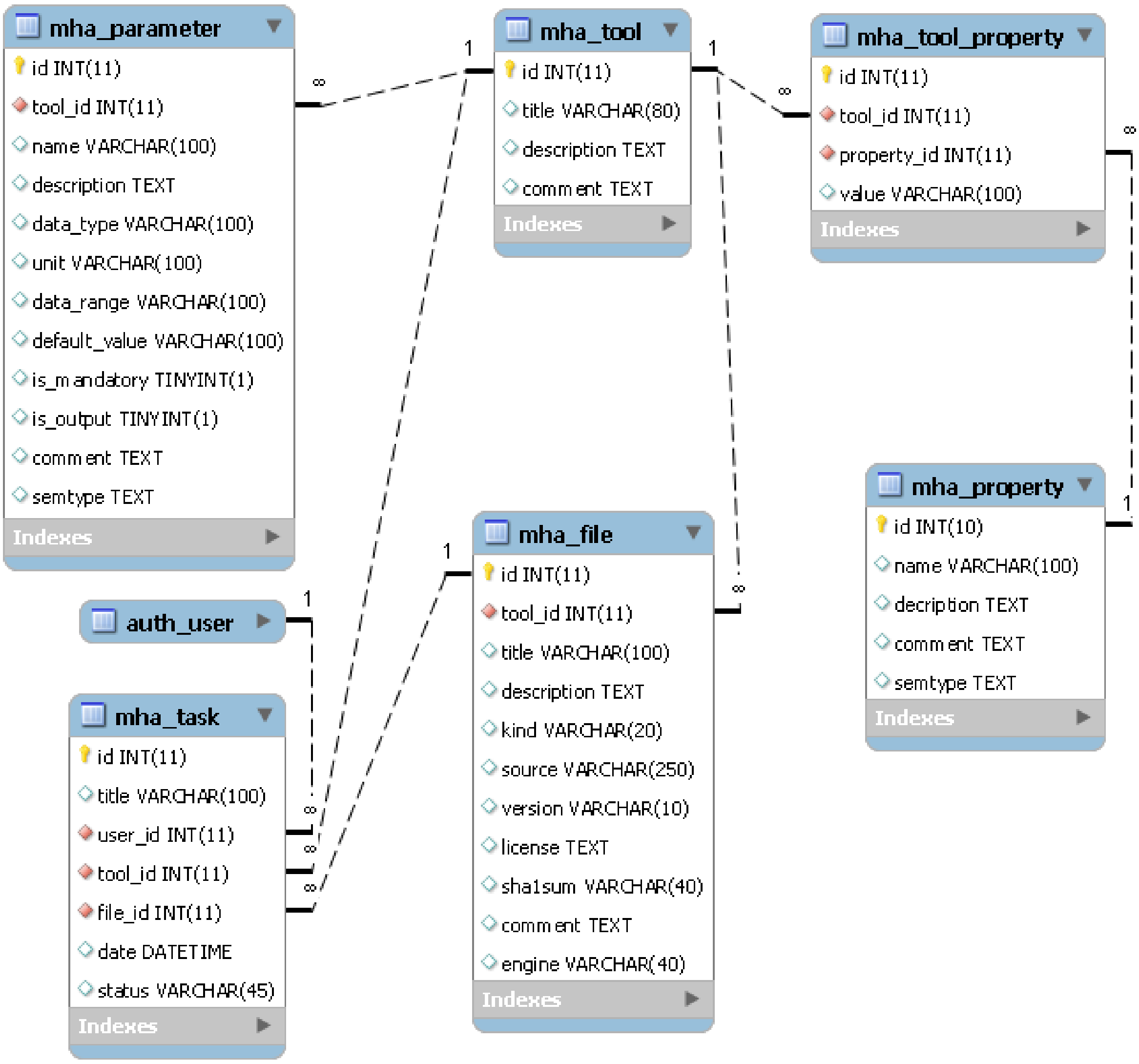


Figure 8-4: E-R diagram of tables of the tool repository

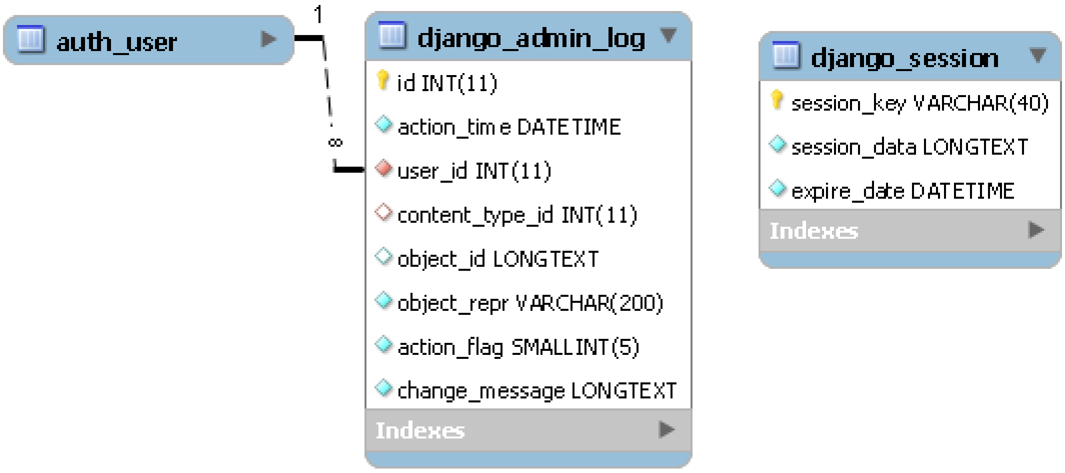


Figure 8-5: E-R diagram of tables of the tool repository related to logging and sessions

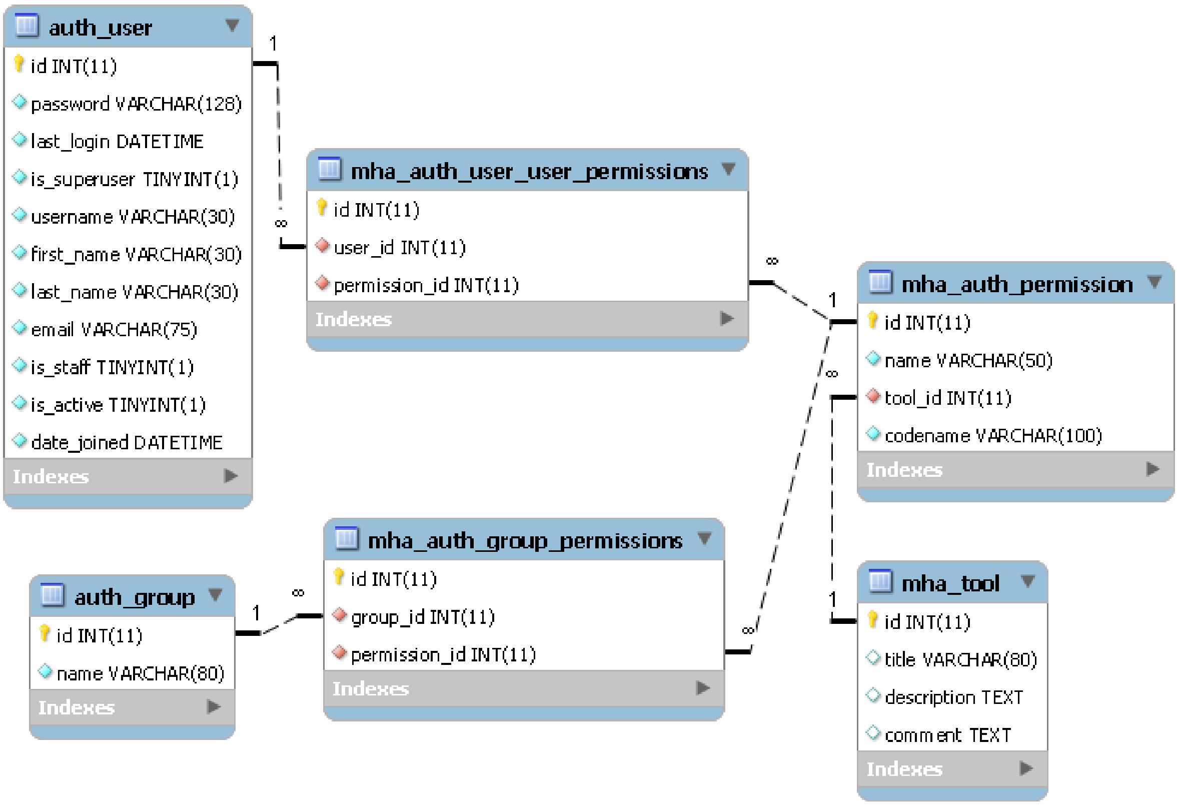


Figure 8-6: E-R diagram of tables associated with authentication and authorization (tools specific permissions)

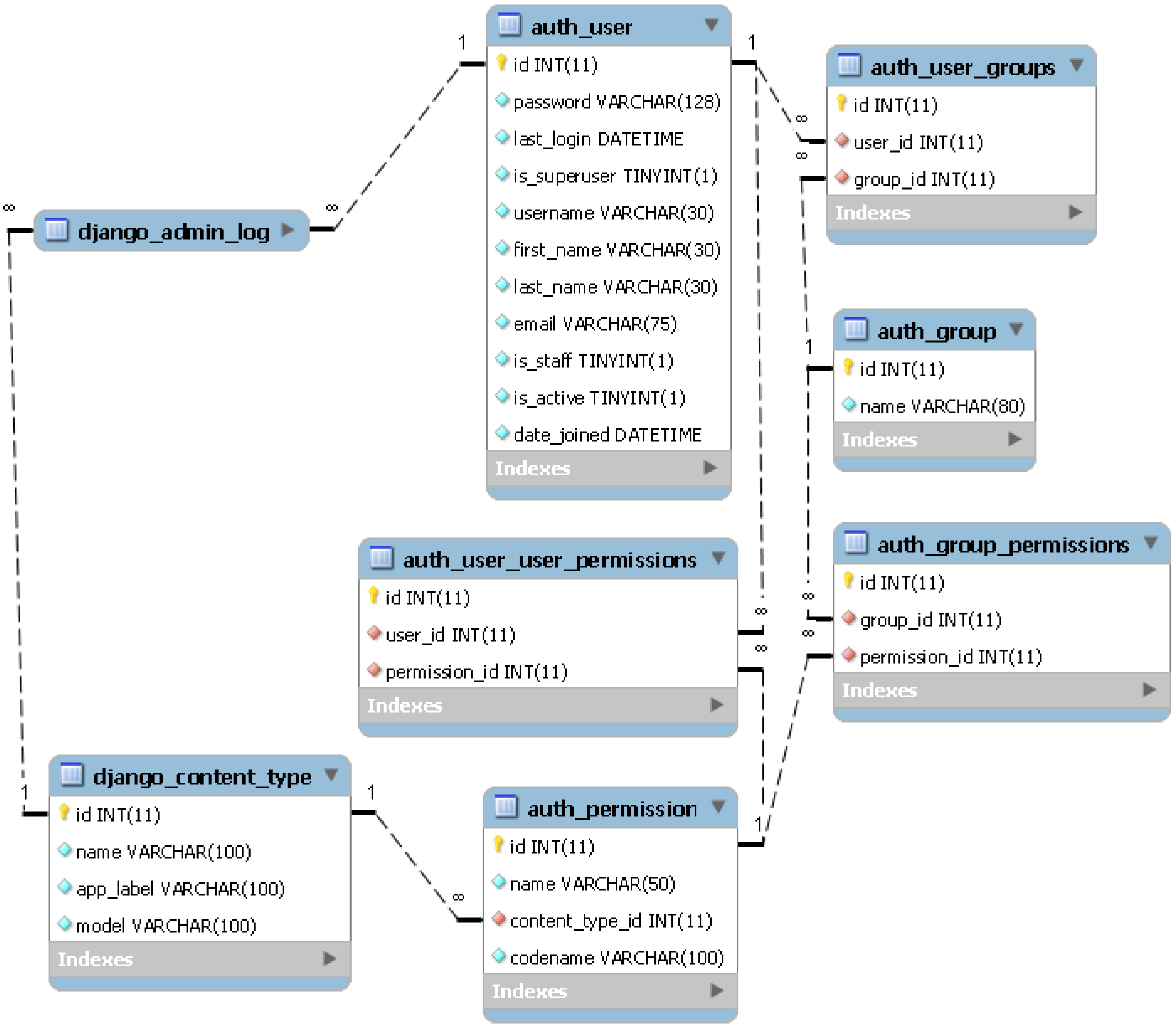


Figure 8-7: E-R diagram of tables associated with authentication and authorization (interface permissions)

8.3 Development view

The Development view of the system defines any constraints on the software development process that are required by the architecture. This includes the system's module organization, common processing that all modules must implement, any required standardization of design, coding and testing and the organization of the system's code base. Our opinion is that in this stage of the project, where the architecture is still under definition, the definition in detail of the development processes, technologies and constraints is premature. After the definition of the platform's functionalities and context properties, and taking into account the interactions between the various components, it is appropriate first to define the interfaces of these interactions and based on these interfaces to conclude in specific development decisions.

Nevertheless, in the context of Task T3.1 (monitoring of standards) and the deliverable D3.1, some development decisions have been taken or at least seem to be the most appropriate, based on the evaluations of the various standards and the best practice techniques. This documentation can be found in the Deliverable 3.1 in full extent, and we copy here only a selected set of guidelines, technologies and architectural styles that seem to be the best choices amongst the various alternatives. However, we stress out the fact that we do not exclude the usage of any technology, we only encourage the usage of specific technologies for easier integration and interoperability reasons as preferable, whenever the ability to select the development technology is given.

8.4 Open standards and technologies

Due to the distributed nature of the platform, which is composed by many different tools and services, the most logical choice for the development of the platform is by using open standards and open technologies. This way, the platform can be able to easily adopt externally provided solutions and interoperate with other projects, organizations, data providers and end users. Examples of such proposed open technologies are:

- The usage of HTML5 for the development of web interfaces, instead of proprietary techniques and tools.
- The usage of XML for data exchange instead of proprietary or non-standardized data formats.
- The usage of HTTP as the transfer protocol and the adoption of the principles of the REST architectural style.
- The usage of LAMP/LAPP (Linux, Apache, MySQL/PostgreSQL, Perl/PHP/Python) solutions instead of proprietary server solutions.

8.5 Operational View

The Operational view defines how the system will be installed into its production environment and how it will be configured, managed, monitored, controlled and maintained. This will be defined later in the project and will be reported final deliverable.

8.6 System Qualities

System qualities capture in essence the non-functional requirements of the entire platform. Usually these qualities are orthogonal to functionality and are observable properties of the system. They are usually “systemic” in the sense that there’s not a single place that has the responsibility for each of them. Instead these non-functional requirements or quality properties of the system emerge from the architecture and the design.

8.7 Topology

There are no specific requirements for the geographical location for the platform to be installed. Currently two place of installation exist. One is located in FORTH’s private cloud and another on at a public cloud. Five VMs are established in this hybrid cloud environment: three from FORTH's OpenStack private cloud provision (8GB RAM, 100GB hard disk each), two from commercial cloud provider Linode (https://www.linode.com) (one is 4GB RAM, 96GB SSD; the other is 8GB RAM, 192GB SSD).

8.8 Performance and scalability

The requirements regarding performance and scalability have not yet specified. The specific requirement will be available together with the availability of the data in the data repository and with the definition of the queries that the users (clinicians and citizens) will do. The computational requirement will also depend on where the simulation will take place. It is envisaged that the MyHealthAvatar platform will have access to simulation models in various model repositories by clinicians. A possible scenario is to have access to the computing service provided (or linked to) by the model repository in order to run these models, in which case no additional computing resources are required for the simulation. In subsequent versions of the architecture we expect to be clarified in detail. Nevertheless, the system should provide a good quality of service and responsiveness

8.9 Usability

Usability plays an essential role in the whole development process of the project p-medicine. The main objective of the usability methodology in the beginning of a project is to describe the task with the whole context of use of the end users. To assure that the software used in MyHealthAvatar will meet the high demands of the end users and that the platform fulfils the requirements for usability of the main target groups, the software has to be evaluated by the users throughout the development period. Taking user needs into account early in the project development can reduce implementation costs and avoid loss of time.

9 MyHealthAvatar API

MyHealthAvatar APIs can be accessed by implicit grant and authorization code grant. Client id and secret may be required for access.

MHA API Endpoints

The complete description of the MHA API can be found in *D3.6 Report on the Review of Open Source APIs for MHA*. This section lists the available endpoints for each API.

Activities

1) [GET] http://myhealthavatar.org/mha/api/v2/user/activities
2) [GET] http://myhealthavatar.org/mha/api/v2/user/activities?from={yyyy-mm-dd}
3) [GET] http://myhealthavatar.org/mha/api/v2/user/activities?to={yyyy-mm-dd}
4) [GET] http://myhealthavatar.org/mha/api/v2/user/activities?from={yyyy-mm-dd}&to={yyyy-mm-dd}
5) [GET] http://myhealthavatar.org/mha/api/v2/user/activities?date={yyyy-mm-dd}
6) [POST] http://myhealthavatar.org/mha/api/v2/user/activities?source=[fitbit, moves, withings]

Activity Daily Summary

1) [GET] http://myhealthavatar.org/mha/api/v2/user/activities/summary
2) [GET] http://myhealthavatar.org/mha/api/v2/user/activities/summary?from={yyyy-mm-dd}
3) [GET] http://myhealthavatar.org/mha/api/v2/user/activities/summary?to={yyyy-mm-dd}
4) [GET] http://myhealthavatar.org/mha/api/v2/user/activities/summary?from={yyyy-mm-dd}&to={yyyy-mm-dd}
5) [GET] http://myhealthavatar.org/mha/api/v2/user/activities/summary?date={yyyy-mm-dd}

Activity Segments

1) [GET] http://myhealthavatar.org/mha/api/v2/user/activities/segments
2) [GET] http://myhealthavatar.org/mha/api/v2/user/activities/segments?from={yyyy-mm-dd}
3) [GET] http://myhealthavatar.org/mha/api/v2/user/activities/segments?to={yyyy-mm-dd}
4) [GET] http://myhealthavatar.org/mha/api/v2/user/activities/segments?from={yyyy-mm-dd}&to={yyyy-mm-dd}
5) [GET] http://myhealthavatar.org/mha/api/v2/user/activities/segments?date={yyyy-mm-dd}

6) [GET] http://myhealthavatar.org/mha/api/v2/user/activities/segments?start_time={yyyy-mm-dd hh:mm:ss}

7) [GET] http://myhealthavatar.org/mha/api/v2/user/activities/segments?start_time={yyyy-mm-dd hh:mm:ss}&end_time={yyyy-mm-dd hh:mm:ss}

Diary

1) [POST] http://myhealthavatar.org/mha/api/v2/user/diary
2) [GET] http://myhealthavatar.org/mha/api/v2/user/diary
3) [GET] http://myhealthavatar.org/mha/api/v2/user/diary?time=[yyyy-MM-dd'T'HH:mmssZ]
4) [GET] http://myhealthavatar.org/mha/api/v2/user/diary?from=[yyyy-MM-dd'T'HH:mmssZ]
5) [GET] http://myhealthavatar.org/mha/api/v2/user/diary?from=[yyyy-MM-dd'T'HH:mmssZ]&to=[yyyy-MM-dd'T'HH:mmssZ]
6) [POST] http://myhealthavatar.org/mha/api/v2/user/diary/delete?id=[local_id]

Measurements

1) [GET] http://myhealthavatar.org/mha/api/v2/user/measurements
2) [GET] http://myhealthavatar.org/mha/api/v2/user/measurements?from={yyyy-mm-dd}
3) [GET] http://myhealthavatar.org/mha/api/v2/user/measurements?to={yyyy-mm-dd}
4) [GET] http://myhealthavatar.org/mha/api/v2/user/measurements?from={yyyy-mm-dd}&to={yyyy-mm-dd}
5) [GET] http://myhealthavatar.org/mha/api/v2/user/measurements?date={yyyy-mm-dd}
6) [POST] http://myhealthavatar.org/mha/api/v2/user/measurements?source=[*]

General Health

1) [POST] http://myhealthavatar.org/mha/api/v2/user/general
2) [GET] http://myhealthavatar.org/mha/api/v2/user/general
3) [GET] http://myhealthavatar.org/mha/api/v2/user/general?from={yyyy-mm-dd}
4) [GET] http://myhealthavatar.org/mha/api/v2/user/general?to={yyyy-mm-dd}
5) [GET] http://myhealthavatar.org/mha/api/v2/user/general?from={yyyy-mm-dd}&to={yyyy-mm-dd}
6) [GET] http://myhealthavatar.org/mha/api/v2/user/general?date={yyyy-mm-dd}

User Profile

1) [POST] http://myhealthavatar.org/mha/api/v2/user/full_profile
2) [GET] http://myhealthavatar.org/mha/api/v2/user/full_profile
3) [GET] http://myhealthavatar.org/mha/api/v2/user/personal_information

4) [GET] http://myhealthavatar.org/mha/api/v2/user/insurance

5) [GET] http://myhealthavatar.org/mha/api/v2/user/allergy

6) [GET] http://myhealthavatar.org/mha/api/v2/user/diagnosis

7) [GET] http://myhealthavatar.org/mha/api/v2/user/medication

8) [GET] http://myhealthavatar.org/mha/api/v2/user/vital_sign

General Insurance Contact

1) [GET] http://myhealthavatar.org/mha/api/v2/insurance_contact

2) [GET] http://myhealthavatar.org/mha/api/v2/insurance_contact?id=[uuid]

Vital Sign Codes

1) [GET] http://myhealthavatar.org/mha/api/v2/vital_sign_codes

Sharing

1) [POST] http://myhealthavatar.org/mha/api/v2/user/followers/request?user_name=**

2) [POST] http://myhealthavatar.org/mha/api/v2/user/followees/request?user_name=**

3) [POST] http://myhealthavatar.org/mha/api/v2/user/followers/accept?user_name=**

4) [POST] http://myhealthavatar.org/mha/api/v2/user/followees/accept?user_name=**

5) [POST] http://myhealthavatar.org/mha/api/v2/user/followers/update?user_name=**

6) [POST] http://myhealthavatar.org/mha/api/v2/user/followees/update?user_name=**

7) [POST] http://myhealthavatar.org/mha/api/v2/user/followers/delete?user_name=**

8) [POST] http://myhealthavatar.org/mha/api/v2/user/followees/delete?user_name=**

9) [GET] http://myhealthavatar.org/mha/api/v2/user/followers

10) [GET] http://myhealthavatar.org/mha/api/v2/user/followees

11) [GET] http://myhealthavatar.org/mha/api/v2/user/followers/request

12) [GET] http://myhealthavatar.org/mha/api/v2/user/followees/request

10 Conclusion

MyHealthAvatar proposes a solution for access, collection and sharing of long term and consistent personal health status data through an integrated environment, which will allow more sophisticated clinical data analysis, prediction, prevention and *in silico* treatment simulations tailored to the individual citizen. In this second deliverable we presented an updated version of MyHealthAvatar's architectural blueprint and implementation activities. We have adopted the provisions of the IEEE 1471 standard that defines an architecture as "…*the fundamental organization of a system embodied in its components, their relationships to each other and to the environment and the principles guiding its design and evolution*". We described all technical architecture details for the implementation of MHA platform should be able to support: *efficient information collection*, *long term management of integrated citizen-specific data*, *effective access mechanisms for data sharing*, as well as *innovative data analysis using integrated toolboxes*. We followed am agile software development process that values the efficient delivery and change in the software by focusing on the continuous communication with the stakeholders, the iterative design, and the frequent release cycle.

In more details, the MHA technical infrastructure provides mechanisms for efficient, and long term data management in internal data repositories storing individual data for the avatars; links to external sources; model repositories, information extraction from the web and data collection using mobile apps; semantic data harmonization to support the data/model searching and reasoning. MyHealthAvatar follows recommendations from relevant VPH activities on "Digital Patient". MyHealthAvatar architectural platform is designed as a multifunctional integrated facility including: Data and model repositories to provide rich resources of data and models; ICT services to support data collection with minimal user input, including web information extraction, mobile apps, etc.; ICT toolbox in support of clinical decision making by using multiscale models and visual analytics; Ontology and RDF repositories to support data integration, search and reasoning; An ICT architecture that allows for access to data from a range of different sources, and integration of the repositories, the toolbox and the ICT utilities; A local cloud solution to support the storage and computational requirements for the avatars without remote data transfer; and Specifications of open MyHealthAvatar APIs for external, third-party developers.

The current deliverable documents all steps taken on capturing stakeholder needs; the process of making a series of architectural design decisions that resulted in a solution meeting those needs, its assessment against the stakeholder needs, and the refining of the solution until it is adequate and captures the architectural design decisions in a complete "Architectural Description". These iterative activities formed the core of the architecture definition process, and are reported in detail. We provide an updated view of MHA platform component diagram, functional diagram as well as the common information model used. We emphasize on all security aspects following guidelines from related workpackages on trust, privacy, legal and ethical requirements and present the actual deployment with in the cloud of the security framework. We present how the linking with external information sources is achieved through MHA and highlight the architectural design and implementation steps towards developing the selected high end demonstration scenarios. Continuing our work we will be able to report the final architectural description results of MHA platform with an evaluation report within the deliverable D3.7.

11 Appendix 1 – EPSOS patient Summary Dataset [72]

PATIENT ADMINISTRATIVE DATA [73]				
Variable (nesting level 1)	**Variables (nesting level 2)**	**Variables (nesting level 3)**	**DEFINITION AND COMMENTS**	**BASIC (Basic)/ EXTENDED (Ext) DATASET**
Identification [1]	National healthcare patient ID	National healthcare patient ID	Country ID, unique to the patient in that country. Example: ID for United Kingdom patient	Basic
Personal information	Full name	Given name	The first name of the patient (example: John). This field can contain more than one element.	Basic
		Family name/surname	This field can contain more than one element. Example: Español Smith Note: some countries require surnames to be the birth name [to avoid potential problems with married women's surnames).	Basic
	Date of birth	Date of birth	This field may contain only the year if the day and month are not available, e.g. 01/01/2009	Basic
	Gender	Gender code	This field must contain a recognized valid value.	Basic
Contact information	Address[74]	Street	Example: Oxford Street	Ext
		House number	Example: 221	Ext
		City	Example: London	Ext
		Post code	Example: W1W 8LG	Ext
		State or province	Example: London	Ext
		Country	Example: UK	Ext
	Telephone no.	Telephone no.	Example: +45 20 7025 6161	Ext
	e-mail	e-mail	Example: jens@hotmail.com	Ext
	Preferred HP/HPO to contact[75]	Name of the HP/HPO	Name of the HP/ HPO that has been treating the patient. If this is an HP, the structure of the name will be the same as described in 'Full name' (given name, family name/surname).	Basic
		Telephone no.	Example: +45 20 7025 6161	Basic
		e-mail	e-mail of the HP/legal organization	Basic
	Contact person/ legal guardian (if available)	Role of that person	Legal guardian or contact person	Ext
		Given name	The first name of the contact person/guardian (example: Peter). This field can contain more than one element.	Ext

[72] http://www.epsos.eu/uploads/tx_epsosfileshare/D3.2.2_Final_Definition_Functional_Service_Req_Patient_Summary.pdf

[73] Dataset that enables the univocal identification of the patient

[74] May vary by country

[75] A health professional in country A may need a contact (health professional/healthcare provider) who knows the patient.

		Family name/surname	This field can contain more than one element. Example: Español Smith	Ext
		Telephone no.	Example: +45 20 7025 6161	Ext
		e-mail	e-mail of the contact person/legal guardian	Ext
Insurance information	Insurance number	Insurance number	Example: QQ 12 34 56 A	Ext

PATIENT CLINICAL DATA				
Variable (nesting level 1)	**Variables (nesting level 2)**	**Variables (nesting level 3)**	**DEFINITION AND COMMENTS**	**BASIC (Basic)/ EXTENDED (Ext) DATASET**
Alerts	Allergy	Allergy description	Description of the clinical manifestation of the allergic reaction. Example: anaphylactic shock, angioedema (the clinical manifestation also gives information about the severity of the observed reaction)	Basic
		Allergy description ID code	Normalized identifier	Basic
		Onset date	Date of the observation of the reaction	Ext
		Agent	Describes the agent (drug, food, chemical agent, etc.) that is responsible for the adverse reaction	Basic
		Agent ID code	Normalized identifier	Basic
	Medical alert information (other alerts not included in allergies)	Healthcare alert description	Medical alert information: any other clinical information that is essential to know so that the life or health of the patient does not come under threat. Example 1: Intolerance to aspirin due to gastrointestinal bleeding. Example 2: intolerance to captopril because of cough (the patient is not allergic but cannot tolerate it because of persistent cough).	Basic
		Healthcare alert ID code	Normalized identifier	Basic
Medical history	Vaccinations	Vaccinations	Contains each disease against which the patient has been immunized	Ext
		Brand name		Ext
		Vaccination ID code	Normalized identifier	Ext
		Vaccination date	Date when the immunization was given	Ext
	List of resolved, closed or inactive problems	Problem description	Problems or diagnoses not included in the definition of "current problems or diagnosis". Example: hepatic cyst (the patient has been treated with an hepatic cystectomy that solved the problem, which is therefore a closed problem)	Ext
		Problem ID code	Normalized identifier	Ext
		Onset time	Date of onset of problem	Ext
		End date	Problem resolution date	Ext

		Resolution circumstances	Describes the reason for which the status of the problem changed from current to inactive (e.g. surgical procedure, medical treatment, etc.). This field includes "free text" if the resolution circumstances are not already included in other fields such as surgical procedure, medical device, etc., e.g. hepatic cystectomy (this will be the resolution circumstances for the problem "hepatic cyst" and will be included in surgical procedures).	Ext
	Surgical procedures prior to the past six months	Procedure description	Describes the type of procedure	Basic
		Procedure ID (code)	Normalized identifier	Basic
		Procedure date	Date when procedure was performed	Basic

PATIENT CLINICAL DATA				
Variable (nesting level 1)	**Variables (nesting level 2)**	**Variables (nesting level 3)**	**DEFINITION AND COMMENTS**	**BASIC (Basic)/ EXTENDED (Ext) DATASET**
Medical problems	List of current problems/diagnoses	Problem/diagnosis description	Problems/diagnoses that fit these conditions: conditions that may have a chronic or relapsing course (e.g. exacerbations of asthma, irritable bowel syndrome), conditions for which the patient receives repeat medications (e.g. diabetes mellitus, hypertension) and conditions that are persistent and serious contraindications for classes of medication (e.g. dyspepsia, migraine and asthma)	Basic
		Problem ID (code)	Normalized identifier	Basic
		Onset time	Date of onset of problem	Basic
	Medical devices and implants	Device and implant description	Describes the patient's implanted and external medical devices and equipment upon which their health status depends. Includes devices such as cardiac pacemakers, implantable fibrillators, prostheses, ferromagnetic bone implants, etc. of which the HP needs to be aware.	Basic
		Device ID code	Normalized identifier	Basic
		Implant date	Date when procedure was performed	Basic
	Major surgical procedures in the past six months	Procedure description	Describes the type of procedure	Basic
		Procedure ID (code)	Normalized identifier	Basic
		Procedure date	Date when procedure was performed	Basic
	Treatment recommendations	Description of recommendations	Therapeutic recommendations that do not include drugs (diet, physical exercise constraints, etc.)	Basic
		Recommendation ID (code)	Normalized identifier	Basic

	Autonomy/invalidity	Description	Need for the patient to be continuously assessed by third parties; invalidity status may influence decisions about how to administer treatments	Basic
		Invalidity ID code	Normalized invalidity identifier (if any, otherwise free text)	Basic
Medication summary	List of current medicines (All prescribed medicines whose period of time indicated for the treatment has not yet expired whether it has been dispensed or not)	Active ingredient Exemption: brand name	Substance that alone or in combination with one or more other ingredients produces the intended activity of a medicinal product. Example: "paracetamol" Brand name if a biological medicinal product or when justified by the health professional (ref. Commission Directive 2012/52/EU)	Basic
		Active ingredient ID code	Code that identifies the active ingredient	Basic
		Strength	Content of the active ingredient expressed quantifiably per dosage unit, per unit of volume or per unit of weight, according to the pharmaceutical dose form. Example: 500 mg per tablet	Basic
		Pharmaceutical dose form	Form in which a pharmaceutical product is presented in the medicinal product packaging (e.g. tablet, syrup)	Basic
		Number of units per intake	Number of units per intake that the patient is taking. Example: 1 tablet	Basic
		Frequency of intakes	Frequency of intakes per hour/day/week/month. Example: every 24 hours	Basic
		Duration of treatment	Example: 14 days	Basic
		Date of onset of treatment	Date when patient needs to start taking the medicine prescribed	Basic

Variable (nesting level 1)	Variables (nesting level 2)	Variables (nesting level 3)	DEFINITION AND COMMENTS	BASIC (Basic)/ EXTENDED (Ext) DATASET
Social history	Social history observations	Social history observations related to smoking, alcohol and diet	Health-related "lifestyle factors" or "lifestyle observations" Example: cigarette smoker, alcohol consumption	Ext
		Reference date range	Example: from 1974 to 2004	Ext
Pregnancy history	Expected date of delivery	Expected date of delivery	Date on which the woman is due to give birth. Year, month and day are required (e.g. 01/01/2014).	Ext
Physical findings	Vital signs observations	Blood pressure	One blood pressure value, which includes systolic blood pressure and diastolic blood pressure	Ext
		Date when blood pressure was measured	Date when blood pressure was measured	Ext
Diagnostic tests	Blood group	Result of blood group	Result of blood group test performed on the patient	Ext

		Date	Date on which the blood group test was performed. This field may contain only the year if the day and month are not available (e.g. 01/01/2009).	Ext

PATIENT ADMINISTRATIVE DATA				
Variable (nesting level 1)	**Variables (nesting level 2)**	**Variables (nesting level 3)**	**DEFINITION AND COMMENTS**	**BASIC (Basic)/ EXTENDED (Ext) DATASET**
Country	Country	Country	Name of country A	Basic
Patient Summary	Date created	Date created	Date on which PS was generated	Basic
	Date of last update	Date of last update	Date on which PS was updated (date of most recent version)	Basic
Nature of the PS	Nature of the PS	Nature of the PS	Defines the context in which it was generated. Distinguishes between three methodological approaches for generating the PS: direct human intervention by an HP, automatically generated approach and mixed approach	Basic
Author organization	Author organization	Author organization	At least one author organization (HCP) shall be listed. If there is no HCP, at least one HP shall be listed.	Basic

12 Appendix 2 – Abbreviations and acronyms

ADaM Analysis Data Model CDISC standard supporting efficient generation, replication, and review of analysis results

AES Advanced Encryption Standard a specification for the encryption of electronic data based on a design principle known as a Substitution permutation network

AGPL Affero General Public License refers to two free software licenses. Affero General Public License, Version 1 and GNU Affero General Public License, version 3.

API application programming interface is a particular set of rules (’code’) and specifications that software programs can follow to communicate with each other

ASCII American Standard Code for Information Interchange a character-encoding scheme based on the ordering of the English alphabet

BMP Bitmap a raster graphics image file format used to store bitmap digital images

BSD Berkeley Software Distribution is a Unix operating system derivative developed and distributed by the Computer Systems Research Group (CSRG) of the University of California, Berkeley

CAS Central Authentication Service a single sign-on web protocol

CA Certification Authority an entity that issues digital certificates

CBC Cipher-Block Chaining a cryptographic mode of operation in which each block of plaintext is XORed with the previous ciphertext block before being encrypted

CCM Counter with CBC-MAC Mode a mode of operation for cryptographic block ciphers

CCZero Creative Commons licenses are several copyright licenses that allow the distribution of copyrighted works

CDASH Clinical Data Acquisition Standards Harmonization CDISC standard describing the basic recommended (minimal) data collection fields for 18 domains, including common header fields, and demographic, adverse events, and other safety domains that are common to all therapeutic areas and phases of clinical research

CDA Clinical Document Architecture is an XML-based markup standard intended to specify the encoding, structure and semantics of clinical documents for exchange

CDE Clinical Document Architecture an XML-based markup standard defined by HL7 intended to specify the encoding, structure and semantics of clinical documents for exchange

CDISC Clinical Data Interchange Standards Consortium - a global, open, multidisciplinary, non-profit organization that has established standards to support the acquisition, exchange, submission and archive of clinical research data and metadata

CDMI Cloud Data Management Interface - defines a functional interface that applications can use to create, retrieve, update and delete data elements from the Cloud

CDS Clinical Decision Support decision support software designed to assist physicians and other health professionals with decision making tasks, as determining diagnosis of patient data

CMWG Cloud Management Work Group focused on standardizing interactions between cloud environments by developing specifications that deliver architectural semantics and implementation details to achieve interoperable cloud management between service providers and their consumers and developers

CRISP-DM Cross Industry Standard Process for Data Mining a data mining process model that describes commonly used approaches that expert data miners use to tackle problems

CRL Certificate Revocation List a list of certificates that have been revoked, and therefore should not be relied upon

CSS Cascading Style Sheets is a style sheet language used to describe the presentation semantics (the look and formatting) of a document written in a markup language

CSV Comma-Separated Values a set of file formats used to store tabular data in which numbers and text are stored in plain-text form that can be easily written and read in a text editor

CWM Common Warehouse Metamodel a specification for modeling metadata for relational, non-relational, multi-dimensional, and most other objects found in a data warehousing environment

CeCILL CEA CNRS INRIA Logiciel Libre is a free software license adapted to both international and French legal matters, in the spirit of and retaining compatibility with the GNU General Public License

CellML Cell Markup Language is an XML based markup language for describing mathematical models

DICOM Digital Imaging and Communications in Medicine - a standard for handling, storing, printing, and transmitting information in medical imaging

DTMF Distributed Management Task Force brings the IT industry together to collaborate on the development, validation and promotion of systems management standards

EHR Electronic health record is an evolving concept defined as a systematic collection of electronic health information about individual patients or populations

EULA End-user licensing agreements An EULA is a legal contract between the manufacturer and/or the author and the end user of an application

EUPL European Union Public License the first European Free/Open Source Software (F/OSS) license

FMA F Foundational Model of Anatomy it is concerned with the representation of classes or types and relationships necessary for the symbolic representation of the phenotypic structure of the human body in a form that is understandable to humans and is also navigable, parseable and interpretable by machine-based systems

FieldMLField Markup Language is an XML based markup language for describing field models

GAS Gridge Authorization Service provides functionality that would be able to fulfill most authorization requirements of grid computing environments

GCM Galois/Counter Mode a mode of operation for symmetric key cryptographic block ciphers that has been widely adopted because of its efficiency and performance

GEM Guideline Elements Model an XML-based guideline document model that can store and organize the heterogeneous information contained in practice guidelines

GNU Gnu’s Not Unix is a Unix-like computer operating system developed by the GNU project, ultimately aiming to be a ”complete Unix-compatible software system” composed wholly of free software.

GO Gene Ontology is a major bioinformatics initiative with the aim of standardizing the representation of gene and gene product attributes across species and databases

GPL General Public License is the most widely used free software license, originally written by Richard Stallman for the GNU Project

GridFTP GridFTP is an extension of the standard File Transfer Protocol (FTP) for use with Grid computing

HL7 Health Level Seven is an all-volunteer, non-profit organization involved in development of international healthcare informatics interoperability standards

HMAC Hash-based Message Authentication Code a mechanism for message authentication using cryptographic hash functions

HTML Hypertext Markup Language is the predominant markup language for web pages. HTML elements are the basic building-blocks of webpages.

HTTPS Hypertext Transfer Protocol Secure is a combination of the Hypertext Transfer Protocol (HTTP) with SSL/TLS protocol to provide encrypted communication and secure identification of a network web server

IBM International Business Machines

ID-FF Liberty Identity Federation Framework an approach for implementing a single sign-on with federated identities based on commonly deployed technologies

ID-WSF Liberty Identity Web Services Framework a framework for identity-based web services in a federated network identity environment

IEC International Electrotechnical Commission is the world’s leading organization that prepares and publishes International Standards for all electrical, electronic and related technologies

IEEE Institute of Electrical and Electronics Engineers is a non-profit professional association headquartered in the United States that is dedicated to advancing technological innovation and excellence

IETF Internet Engineering Task Force a large open international community of network designers, operators, vendors, and researchers concerned with the evolution of the Internet architecture and the smooth operation of the Internet

IHE Integrating the Healthcare Enterprise - an initiative by healthcare professionals and industry to improve the way computer systems in healthcare share information

IPSec Internet Protocol Security a protocol suite for securing Internet Protocol (IP) communications by authenticating and encrypting each IP packet of a communication session

ISBN International Standard Book Number is a unique numeric commercial book identifier based upon the 9-digit Standard Book Numbering (SBN) code created by Gordon Foster

ISO International Organization for Standardization is an international standard-setting body composed of representatives from various national standards organizations

InSilicoML InSilico Markup Language is a markup language that can explicitly describe the multi-level hierarchical structures of the physiological functions in mathematical models

JDMP Java Data Mining Package an open source Java library for data analysis and machine learning

JPEG Joint Photographic Experts Group is a commonly used method of lossy compression for digital photography

JSDL Job Submission Description Language is an extensible XML specification from the Global Grid Forum for the description of simple tasks to non-interactive computer execution systems

KNIME Konstanz Information Miner a user-friendly and comprehensive open source data integration, process, analysis and exploration platform

LGPL Lesser General Public License is a free software license published by the Free Software Foundation

LOINC Logical Observation Identifiers Names and Codes is a database and universal standard for identifying medical laboratory observations

MAGE-ML Microarray and Gene Expression - Markup Language markup language format for the representation of gene expression data from microarrays to facilitate the exchange of information between different data systems

MAGE-OM Microarray and Gene Expression - Object Model data exchange model for the representation of gene expression data from microarrays to facilitate the exchange of information between different data systems

MAGE-TAB Microarray and Gene Expression - Tabular tabular format for the representation of gene expression data from microarrays to facilitate the exchange of information between different data systems

MIAME Minimum Information About a Microarray Experiment needed to enable the interpretation of the results of the experiment unambiguously and potentially to reproduce the experiment

MIASE Minimal Information About a Simulation Experiment common set of information a modeller needs to provide in order to enable the execution and reproduction of a numerical simulation experiment, derived from a given set of quantitative models

MIASE Minimum Information About a Simulation Experiment is an effort to list the common set of information a modeller needs to provide in order to enable the execution and reproduction of a numerical simulation experiment, derived from a given set of quantitative models.

MIBBI Minimum Information for Biological and Biomedical Investigations maintains a web-based, freely accessible resource for ”Minimum Information” checklist projects, providing straightforward access to extant checklists (and to complementary data formats, controlled vocabularies, tools and databases), thereby enhancing both transparency and accessibility

MIT MIT License is a free software license originating at the Massachusetts Institute of Technology

ML Markup Language is a modern system for annotating a text in a way that is syntactically distinguishable from that text

MOF MetaObject Facility the foundation of OMG’s industry-standard environment where models can be exported from one application, imported into another, transported across a network, stored in a repository and then retrieved, rendered into different formats

MPL Mozilla Public License is a free and open source software license

MS Microsoft is an American public multinational corporation headquartered in Redmond, Washington

MTOM Message Transmission Optimization Mechanism is the W3C Message Transmission Optimization Mechanism, a method of efficiently sending binary data to and from Web services

MedLEE Medical Language Extraction and Encoding system System to extract, structure, and encode clinical information in textual patient reports so that the data can be used by subsequent automated processes

NeuroML Neuro Markup Language is an XML (Extensible Markup Language) based model description language that aims to provide a common data format for defining and exchanging models in computational neuroscience

OASIS Organization for the Advancement of Structured Information Standards a not-for-profit consortium that drives the development, convergence and adoption of open standards for the global information society

OBO Open Biomedical Ontologies is an effort to create controlled vocabularies for shared use across different biological and medical domains

OGSA-BES Open Grid Services Architecture - Basic Execution Services defines Web Services interfaces for creating, monitoring, and controlling computational entities such as UNIX or Windows processes, Web Services, or parallel programswhat we call activities within a defined environment

OGSA-DAI Open Grid Service Architecture-Data Access and Integration allows data resources (e.g. relational or XML databases, files or web services) to be federated and accessed via web services on the web or within grids or clouds. Via these web services, data can be queried, updated, transformed and combined in various ways.

OSI Open Source Initiative is an organization dedicated to promoting open source software

OS Operating System is a set of programs that manages computer hardware resources, and provides common services for application software

OWL-S Ontology Web Language for web Services an ontology of services to discover, invoke, compose, and monitor Web resources offering particular services and having particular properties

OWL Web Ontology Language is a family of knowledge representation languages for authoring ontologies.

OpenID Open Identity provider of web-based SSO services

PAOS Reverse HTTP Binding for SOAP a binding that enables HTTP clients to expose services using the SOAP protocol, where a SOAP request is bound to a HTTP response and vice versa

PATO PATO an ontology of phenotypic qualities, intended for use in a number of applications, primarily defining composite phenotypes and phenotype annotation.

PHP PHP: Hypertext Preprocessor is a general-purpose server-side scripting language originally designed for web development to produce dynamic web pages

PKIX Public-Key Infrastructure Working Group was established in the fall of 1995 with the goal of developing Internet standards to support X.509-based Public Key Infrastructures

PMML Predictive Model Markup Language an XML-based language which provides a way for applications to define statistical and data mining models and to share models between PMML compliant applications

PNG Portable Network Graphics is a bitmapped image format that employs lossless data compression

POST POST is one of many request methods supported by the HTTP protocol used by the World Wide Web

RAD Rapid application development is a software development methodology that uses minimal planning in favor of rapid prototyping

RDF Resource Description Framework is a family of World Wide Web Consortium (W3C) specifications originally designed as a metadata data model

REST Representational state transfer is a style of software architecture for distributed hypermedia systems such as the World Wide Web

RFC Request for Comments is a memorandum published by the Internet Engineering Task Force (IETF) describing methods, behaviors, research, or innovations applicable to the working of the Internet and Internet-connected systems

RICORDO RICORDO is focused on the study and design of a multiscale ontological framework in support of the Virtual Physiological Human community to improve the interoperability amongst its Data and Modelling resources

RIM Reference Information Model is the cornerstone of the HL7 Version 3 development process and an essential part of the HL7 V3 development methodology

SAML Security Assertion Markup Language a standard, XML-based framework for creating and exchanging security information between online partners

SAS Business analytics software and service developer, and independent vendor in the business intelligence market

SAWSDL Semantic Annotations for WSDL defines mechanisms using which semantic annotations can be added to WSDL components

SBML System Biology Markup Language is a representation format, based on XML, for communicating and storing computational models of biological processes

SDTM Study Data Tabulation Model CDISC defining a standard structure for human clinical trial (study) data tabulations that are to be submitted as part of a product application to a regulatory authority

SED-ML Simulation Experiment Description Markup Language an XML-based format for encoding simulation experiments, following the requirements defined in the MIASE guidelines

SHA Secure Hash Algorithm a number of cryptographic hash functions published by the National Institute of Standards and Technology as a U.S. Federal Information Processing Standard

SLO Single Log-Out termination of a SSO action

SNIA Storage Networking Industry Association not-for-profit trade organization for companies and individuals in various sectors of the storage industry

SNOMED-CT Systematized Nomenclature of Medicine - Clinical Term is a systematically organised computer processable collection of medical terminology covering most areas of clinical information such as diseases, findings, procedures, microorganisms, pharmaceuticals etc

SOAP Simple Object Access Protocol is a protocol specification for exchanging structured information in the implementation of Web Services in computer networks

SOAP Simple Object Access Protocol a lightweight XML-based protocol for exchange of structured information in a decentralized, distributed environment

SOA Service-Oriented Architecture s a set of principles and methodologies for designing and developing software in the form of interoperable services

SPARQL SPARQL Protocol and RDF Query Language - query language for RDF

SQL Structured Query Language a standard language for accessing and manipulating databases

SSH Secure Shell is a network protocol for secure data communication, remote shell services or command execution and other secure network services between two networked computers that it connects via a secure channel over an insecure network

SSL Secure Sockets Layer a cryptographic protocol that provides communication security over the Internet, predecessor of TLS

SSO Single Sign-On a mechanism whereby a single action of user authentication and authorization can permit a user to access all computers and systems where he has access permission, without the need to enter multiple passwords

TCP/IP Transmission Control Protocol/Internet Protocol the first two networking protocols defined in the Internet Protocol Suite standard

TDD Test-driven development is a software development process that relies on the repetition of a very short development cycle.

TLS Transport Layer Security a cryptographic protocol that provides communication security over the Internet, successor of SSL

UML Unified Modelling Language a specification defining a graphical language for visualizing, specifying, constructing, and documenting the artifacts of distributed object systems

VDM Vienna Development Method is one of the longest-established Formal Methods for the development of computer-based systems

VPH-NoE Virtual Physiological Human - Network of Excellence is a project which aims to help support and progress European research in biomedical modelling and simulation of the human body

VPH Virtual Physiological Human is a methodological and technological framework that, once established, will enable collaborative investigation of the human body as a single complex system

WAV Waveform Audio File Format is a Microsoft and IBM audio file format standard for storing an audio bitstream on PCs

WS-* Web Services-* common prefix for the family of Web Services specifications

WSDL Web Services Description Language a way to describe the abstract functionalities of a service and concretely how and where to invoke it

WSMO Web Service Modelling Ontology ontology for describing Semantic Web Services

XFree86 A freely redistributable open-source implementation of the X Window System

XHTML Extensible HyperText Markup Language is a family of XML markup languages that mirror or extend versions of the widely-used Hypertext Markup Language (HTML), the language in which web pages are written

XML Extensible Markup Language - a format for encoding documents in machine-readable form, similar in syntax to HTML

XTS XEX-based Tweaked Codebook a mode of operation for cryptographic block ciphers

caBIG Cancer Biomedical Informatics Grid a virtual network of interconnected data, individuals, and organizations that work together to redefine how cancer research is conducted