Co-Funded by the Horizon 2020 programme of the European Union

Grant Agreement: 875358

FAITH
intelligent patient support

# Deliverable D3.3

# FAITH Data Privacy & Protection

| | |
|---|---|
| Work package: | WP3 – Hospital Infrastructures, Visualisation & Distributed Ledger Technology (DLT) |
| Prepared By/Enquiries To: | Fernando Ferreira – UNINOVA |
| Reviewers: | Maria Beltran – UPM<br>Tom Flynn – TFC |
| Status: | Draft |
| Date: | 14/06/2021 |
| Version: | 1.0 |
| Classification: | Public |

Authorised by:

Philip O'Brien
WIT

Authorised date: 21/06/21

Disclaimer:

This document reflects only authors' views. Every effort is made to ensure that all statements and information contained herein are accurate. However, the Partners accept no liability for any error or omission in the same. EC is not liable for any use that may be done of the information contained therein.



.

**FAITH Project Profile**

**Contract No H2020-ICT- 875358**

| | |
|---|---|
| **Acronym** | **FAITH** |
| **Title** | **a Federated Artificial Intelligence solution for moniToring mental Health status after cancer treatment** |
| **URL** | **https://h2020-faith.eu/** |
| **Twitter** | **https://twitter.com/H2020_Faith** |
| **LinkedIn** | **linkedin.com/company/faith-project** |
| **Facebook** | **https://fb.me/H2020.FAITH** |
| **Start Date** | **01/01/2020** |
| **Duration** | **36 months** |

**FAITH Partners**

List of participants

| Participant No | Participant organisation name | Short Name | Country |
|---|---|---|---|
| 1 (Coordinator) | WATERFORD INSTITUTE OF TECHNOLOGY. | WIT | Ireland |
| 2 | UPMC Whitfield, Euro Care Healthcare Limited. | UPMC | Ireland |
| 3 | Universidad Politécnica de Madrid. | UPM | Spain |
| 4 | Servicio Madrileño de Salud. | SERMAS | Spain |
| 5 | UNINOVA, Instituto de Desenvolvimento de Novas Tecnologias. | UNINOVA | Portugal |
| 6 | Fundação D. Anna de Sommer Champalimaud e Dr. Carlos Montez Champalimaud. | CF | Portugal |
| 7 | Deep Blue. | DBL | Italy |
| 8 | Suite5 Data Intelligence Solutions Limited. | SUITE5 | Cyprus |
| 9 | TFC Research and Innovation Limited. | TFC | Ireland |

*SC1-DTH-01-2019: Big data and Artificial Intelligence for monitoring health status and quality of life after the cancer treatment*

*H2020-SC1-DTH-2019*

FAITH is co-funded by the European Commission - Agreement Number 875358 (H2020 Programme)

**Document Control**

This deliverable is the responsibility of the Work Package Leader. It is subject to internal review and formal authorisation procedures in line with ISO 9001 international quality management system procedures.

| Version | Date | Author(s) | Change Details |
|---|---|---|---|
| 0.1 | 01/02/2021 | Fernando Ferreira (UNI) | Table of Contents defined. |
| 0.2 | 28/02/2021 | Fernando Ferreira (UNI) | First draft version. |
| 0.3 | 15/03/2021 | João Gião(UNI) | Contribution to DLT and services. |
| 0.4 | 30/03/2021 | Stefanos Venios (S5) | Contribution to App security. |
| 0.5 | 05/04/2021 | Ricardo Matias, Pedro Ferreira (FC) | Ethical and security aspects required by Fundação Champalimaud. |
| 0.6 | 07/04/2021 | Maria Eugénia Beltran, Diego Carvajal (UPM) | Ethical and security aspects required by Hospital General Universitario Gregorio Marañón. |
| 0.7 | 10/04/2021 | Philip O'Brien, Gary McManus (WIT) | Ethical and security aspects required by UPMC. |
| 0.8 | 30/04/2021 | Fernando Ferreira (UNI) | Revised version. |
| 0.9 | 31/05/2021 | Tom Flynn (TFC) | QA'd version. |
| 0.9.1 | 20/06/21 | Fernando Ferreira (UNI) | Final revision based on QA |
| 1.0 | 21/06/2021 | Philip O'Brien (WIT) | Final release for submission to European Commission portal. |

## Executive Summary

***Objectives:***

The main objective of this deliverable is to describe the main aspects of data privacy and data protection that for the FAITH research project. Data collection, analysis and proper handling of data sets are fundamental aspects in the fast-growing data driven research and innovation. In the health domain, data gathering, and its consistent analysis becomes critical mainly as it considers people and the management of sensitive data regarding each person and respective health condition. This task identifies the vulnerabilities and the preventive security measures to ensure data protection and the privacy of users within a FAITH environment. Privacy Protection implementation in the FAITH framework is central to the whole architecture taking in account the entire pervasive chain of data exchange and new knowledge generated. Privacy protection is built around a Distributed Ledger Technology (DLT) principle, ensuring that all data transactions entering and accessing within the framework will have full protection and auditability. We will utilize, whenever available, existing information on access and usage policies set by the data owners (hospitals) and will explore feasibility and potential benefits and challenges of existing semantics and sensitivity levels (e.g., GDPR-related data) of the datasets. UNINOVA bring their experiences to lead this task with heavy contributions from the trust & security experts in WIT but driven by the needs of the hospital partners in FAITH. Data protection is a concern for the FAITH project. It is a project that involves an elevated number of individuals from society, who are not scientists or researchers, since the defining characteristic is being individuals which already endured cancer treatments. In this sense, the FAITH project has several methodologies and approaches to the data handling that will reinforce the protection of the information during the scope of the project with reinforced recommendations, from lessons learned and for after the project has concluded.

***Results:***

This deliverable presents the results from the activities undertaken in Task 3.3 which comprise the Data Privacy, Trust & Protection framework for the FAITH Architecture. This framework resulted from an analysis of vulnerabilities and security checkpoints. This document reports those security aspects and proposes the most suited measures to tackle them, thus, enforcing security and data protection for patient data in FAITH.

# TABLE OF CONTENTS


1 ABBREVIATIONS AND ACRONYMS ..... 10

2 INTRODUCTION ..... 11

3 GDPR and Data Acquisition in eHealth ..... 12

3.1 Data protection ..... 13

3.2 GDPR ..... 13

3.3 Acquiring Data from an Hospital ..... 14

4 Data Privacy and Data Protection the Sate of Art ..... 15

4.1 General Principles for data handling ..... 15

4.2 Essential aspects for Data protection ..... 15

4.3 Data Anonymization ..... 16

4.3.1 Anonymization ..... 17

4.3.2 Data Masking ..... 17

4.3.3 Pseudo-Anonymization ..... 18

4.3.4 Differential privacy ..... 19

4.3.5 Personal Identifiable Data ..... 19

4.4 Communication Protocols ..... 19

4.5 DLT in eHealth ..... 21

4.6 Encryption ..... 22

4.7 Edge computing and Device Encrypted Cache ..... 25

5 Requirements for Data Acquisition in FAITH ..... 26

5.1 Data Protection in FAITH ..... 26

5.2 Cloud Services ..... 28

5.3 Server and Infrastructure ..... 29

5.4 Communications in FAITH ..... 29

5.5 Mobile Application ..... 30

5.6 Data Visualization Service ..... 31

5.7 Data privacy and Data protection compliance to Hospitals ..... 32

5.7.1 Hospital General Universitario Gregorio Marañón ..... 32

5.7.2 Centro Clínico Champalimaud - Fundação Champalimaud ..... 32

5.7.3 UPMC ..... 34
**5.8 DLT, Data privacy and Data protection ..... 34**
**5.9 Architectural design for privacy ..... 35**
5.9.1 Authentication Management ..... 36
Registration and Login ..... 36
5.9.2 User account information repository ..... 37
5.9.3 Data Sharing ..... 37
5.9.4 Data Limitation management ..... 38
5.9.5 Data Ownership Handling ..... 38
5.9.6 Id management ..... 38
5.9.7 Personal Identifiable data management ..... 39
5.9.8 Data Notification Centre ..... 40
5.9.9 FAITH data repository ..... 40
5.9.10 Data Integrity Validation ..... 40
5.9.11 Data Access Logs ..... 40
5.9.12 DLT Demo ..... 42
**6 DISCUSSION AND CONCLUSIONS ..... 44**
**7 APPENDIX ..... 46**
**A.1. Legal Framework for HGUGM ..... 46**
**8 Bibliography ..... 48**

**TABLE OF FIGURES**

Figure 1 - Production FHIR System [1] ..... 27
Figure 2 – ID Management in FAITH Trials ..... 39
Figure 3 - Security framework with DLT log registration ..... 41
Figure 4 - DLT Dashboard for data auditing ..... 43
Figure 5 - Interface for getting registration logs from the blockchain ..... 43

# 1 ABBREVIATIONS AND ACRONYMS

| Abbreviation | Description |
|---|---|
| Android: | A mobile operating system based on a modified version of the Linux kernel and other open-source software, designed primarily for touchscreen mobile devices such as smartphones and tablets. |
| Blockchain: | A system of recording information in a way that makes it difficult or impossible to change, hack, or cheat the system. |
| DLT: | Distributed Ledger Technology. |
| DM: | Data Model. |
| EMA: | European Medications Agency. |
| FDA: | Food and Drugs Administration. |
| FHIR: | Fast Healthcare Interoperability Resources. |
| GDPR: | General Data Protection Regulation. |
| iOS: | A mobile operating system created and developed by Apple Inc. |
| ISO9001-2015: | International Quality Management Systems. |
| QoC: | Quality of Care. |
| WP: | Work Package. |
| WPL: | Work Package Leader. |

# 2 INTRODUCTION

The Quality of Care (QoC) provided to citizens by professionals and institutions depends on the quality and availability of information. Early commencement of treatment and medication, and the decisions on how to proceed, depend a lot on patients' data in the different modalities available. It is also important to note that large pools of data help to inform health and wellbeing parameters for the large size community. To make that possible, it is necessary both to have the best hospital practices but also to get consent and collaboration of patients. In pursuing such a goal, it is necessary to use practices, which adhere to legal constraints and are transparent in handling data while transmitting those identified practices and protocols to professionals and patients. The presented document, (D3.3), aims to provide a framework envisaging the seamless application of the clinical procedures, following legal guidance, and making the process known, secure and trustworthy. It aims to contribute to clinical practice as well as clinical research, thereby contributing to Big Data analysis, ensuring trust and best clinical data handling in the process. The process of data collection and handling of sensitive data requires special procedures and strategies to ensure data protection and data security. This is especially imperative when data in connected with individuals; moreover, when such data refers to the individual's specific characteristics or is an individuals' identifiable data.

Deliverable, D3.3, reports on the activities ongoing under the scope of Task 3.3 in the project. This task is aimed at the implementation of the privacy protection and trust mechanism required by the hospital infrastructures for the FAITH environment. The Privacy Protection implementation in the FAITH framework is central to the whole architecture in order to take into account the entire pervasive chain of data exchange and new knowledge generated. This privacy protection is being built around a Distributed Ledger Technology (DLT) principle, ensuring that all data transactions entering and accessing within the framework will have full protection and auditability. As there are diverse types of concerns from the data management to be secure and protect the privacy of patients, the following chapters present the different angles and strategies to tackle potential risks. Furthermore, in chapter 3, an overview is presented based on the implications of GDPR in health data and its relevance for hospital data. In FAITH, data is provided to hospitals, in this case only in that direction so that no interference is made to the hospital data. Data is naturally collected and collated in according to GDPR rules. Chapter 4 presents state of art for data privacy and data protection regarding all other implications beyond GDPR. Chapter 5 is focused on the concrete aspects for FAITH data acquisition and data protection. In Chapter 6, draws conclusions in what respects to the work performed in FAITH regarding data protection and data security.

# 3 GDPR and Data Acquisition in eHealth

Data acquisition and its management are at the centre of both healthcare institutions and users, the citizens. Data, properly acquired and managed, provides the ground for the healthcare institutions to develop the best assessment of a citizen's health, deploy the proper assets and to develop the best treatment for each person's specific needs. On the other side, a patient needs to trust that health data referring to him is properly managed, ensuring privacy and data security. In this process, there is legislation to be followed and regulations to be observed in the interaction between the citizen and the hospital. Current practice shows that each hospital will generally has its own data protection policy and an ethics board, with local variations by country adjusted to the nature of the healthcare service. This chapter aims at providing guidance on how the interaction with the hospital can be engaged and what is needed to support such liaison, supported by a legal and ethical framework. In observing such process flow, the limitations can be tackled, the obstacles overcome towards the best interest of the citizen, the hospital, and the community. Hospitals are very careful about their data and patients are increasingly concerned that their data might be misused, and there is a risk that patients will exercise their right to 'opt out' of their data being used beyond their own care, which in turn will jeopardize the potential of science to apply 'big data' analysis. Therefore, to alleviate patient's concerns regarding data use is a core consideration. An important part of the process is to define the proper questions to ask and to identify among the specificities of the system, which will be the best answer to the specificities of each case to be addressed. This deliverable establishes a pathway and provides a step-by-step process of some of the issues encountered, and suggestions on how to deal with them while ensuring compliance; with reference to the current data protection policy and, in particular, the recently enacted General Data Protection Regulation (GDPR). A graphical schematization of the aspects to be addressed, in Figure 3, shows the flow of data requests and receiving information to be endorsed by each healthcare node including the DLT usage for log registry. Some important issues explored in more depth relate to data sharing and anonymization. A brief case study provides an example of data transfer from a hospital, including impact of GDPR on intelligent systems, such as artificial intelligence (AI). The main goal presented in this document is for the establishment of the framework for the management of the patient information and all that it includes. This is the case for the data collected, clinical decisions, medication, clinical workflows and the results of tests, examinations and other events related with the citizen's healthcare. It aims at providing guidance to research practice, hospital management and also for the development of clinical software. The document is divided as follows. Chapter 2 makes an overview of the data protection current practices as it become with the entrance of GDPR. Chapter 3 establishes a general guide for the process

of acquiring and managing data, while the remaining chapters make the overall analysis of the impact of such framework drawing the relevant conclusions.

## 3.1 Data protection

The heterogeneity of systems, devices and places, presents a multidimensional challenge that must be carefully addressed, especially when addressing personal data related aspects. The path through the different aspects that should be considered, regarding data privacy, are addressed in this and the following chapters. The first aspect to observe is the compliance with the GDPR. Diverse aspects are contemplated by the GDPR regulation and surely, following all these rules it is also necessary to address the need to record keep access to data and to provide the right to be forgotten. Those seem almost conflicting aspects of the same reality but in fact are not. It is necessary to have a clear registry of who had access to data and for that, the DLT can be a tool to have such records in the block chain. This is known to be inviolable and incorruptible mean to store securely relevant information. That is the case of the access logs. On the other hand, if a patient wants to opt out his/her data then there should be provided tools to remove the user and delete all associated personal data. Since data collected during a trial is provided or collected from the patients, the same procedures for hospital data treatment should be followed; in this regard, complying with the project's associated hospitals Ethical Boards in what concerns to the affiliated patients in the trial.

## 3.2 GDPR

To safeguard patient privacy and their personal data, there exist privacy standards in different regions of the world, such as General Data Protection Regulation (GDPR) [1] in Europe, Health Insurance Portability and Accountability Act (HIPAA) [2] in the United States or the highest-level standard framework on personal information in China, the Cybersecurity Law of the People's Republic of China [3]. This document only refers and is guided by GDPR, because its demonstration take place in European countries. It is also important to define what "personal data" is. According to Article 4(1) of the GDPR, personal data is now legally defined as follows: "'Personal data' means any information relating to an identified or identifiable natural person 'data subject'". For this reason, a stream of data that does not contain any information that relates to an identified individual, such as a name or an address, or it is not possible to associate to an individual, is considered as non-personal data. The problem is, however, that the line between personal and non-personal data is a moving target and the data that today is seen as non-personal data may become personal data in a near future (e.g. through analytical and technological advancements) [4].

The European Parliament designed GDPR to harmonize and define data privacy law across Europe (and the companies doing business in Europe) with the purpose to protect and empower EU citizens' personal

data and reshape the way organizations handle data. In particular, GDPR sets mandatory requirements in Electronic Health Records and Personnel Health Records systems as well the health data exchange. In [5], the authors synthetized the GDPR legislation and synthetized the following key factors that influence health systems, as the following:

- Data protection by design and by default
- Data portability
- Right to be forgotten—notification requirement
- Unambiguous consent
- Privacy notices
- Right to Access and Records of processing activities
- Explicit and formally represented policies

The GDPR aims to improve the regulation of data protection laws in Europe, and in doing so cope with the new challenges of data protection in the digital health era.

The framework in FAITH is focused on ensuring that users own and control their personal data. As such, the system recognizes the users as the owners of the data and the services as guests with delegated permissions

Instead, for the FAITH framework at any given time, the user may alter the set of permissions and revoke access to previously collected data.

## 3.3 Acquiring Data from an Hospital

The process of acquiring data from a hospital is quite complex, mostly due to the patient doctor privilege but then to the Ethical Boards that supervise and restrict any outbound data. That is the main reason for the strategic decision for the FAITH operations model to avoid data acquisition from the hospitals as a less viable option and also because, in fact, it is not needed as long as the doctors have access to the patient's data as regularly happens in consultation. The doctor receives data and analytics from FAITH, which can be compared and enrich with existing data from the hospital, about that patient, without integrating such input to FAITH and thus, avoiding any changes to data privacy and safety conditions existing in the hospital.

# 4 Data Privacy and Data Protection the Sate of Art

## 4.1 General Principles for data handling

A challenge created by GDPR stems from its treatment of “anonymous” versus “pseudonymous” data. GDPR states, in Recital 26, that “the principles of data protection should apply to any information concerning an identified or identifiable natural person” and pseudonymised data “should be considered to be information on an identifiable natural person”. This means that, not only personal data and pseudonymised data are under the GDPR legislation, but also scientific research activities where a key is needed to re-identify the user’s data. On these cases, GDPR takes the position that all pseudonymized data are considered personal data, regardless of whether they are, or ever will be, in the hands of the person who holds the key needed for re-identification, thus, it is required direct authorization from the data-subject [6].

Since health data are considered as personal data, the software infrastructure must be capable to allow the right to be forgotten or right to erasure rule. According to this rule, the user has the right to request for the complete erasure of the user’s personal data.

Since GDPR proposes both anonymization and pseudonymization techniques as possible data protection measures; both are described in the next sections.

## 4.2 Essential aspects for Data protection

Sometimes there are flaws in daily practice that may endanger the security of the whole system. Those can be caused by over-confidence, ease the procedures due to routine and lack of awareness about potential threats. Those are mistakes in which any person can fall and promoting the right amount of awareness could easily become the first level of protection for computational devices, their associated services and therefore reinforcing a data protection framework. Among those problems some are a common place as can be seen in the next list.

**Problems with passwords** – As it is easier to remember birthdays or other relevant dates and the name of loved ones, it is sometimes a matter of some research about personal facts to discover access passwords. Those are a vulnerability along with the tendency to use simple passwords (e.g., 12345, password, 0000, etc.) or the usage of default passwords of equipment and services.

**Outdated software** – Security is a sort of cat and mouse chase; hackers discover new weaknesses and software manufacturers issue updates tackling those weaknesses. Operating systems and other critical software need to be updated so that security is at the level of current threats.

**Antivirus and firewall** – Sometimes systems lack adequate protections, either because they are not central to the computational systems or because remote usage transfers the problem to personal computers. Adding to this sometimes protections are left disabled as they would restrain from installing some software packages or specific software seen as a threat by the computational defences. It is therefore important that systems are not vulnerable either at the hospital or at home to avoid illegitimate access or putting in risk the whole infrastructure.

**Insecure protocols** – Some older versions of communication protocols are still in use (e.g. older TLS, and SSL) this way putting the system at threat. Moreover, hospital systems have sometimes outdated pieces of hardware and associated software whose replacement would be expensive and, anyway they perform as needed. Those are aspects that add insecurity to the computational systems and need to be properly addressed once identified.

**Amorphous networks** – the usage of the same VLAN (Virtual Local Area Networks) for different devices, such as medical equipment, printers and office computers add a layer of vulnerability as networks mix with each other and different pieces of equipment are connected to the same network. This exposes the whole system to threats since many of those devices don't have the level of security that a workstation or personal computer should have.

**Exposure to external threats** – Healthcare institutions (HI) are an elected target for ransomware attacks. Hackers know that the trust is essential to HI since patient confidence plays an important role in the patient doctor relationship and the patient to the hospital relationship. In that sense, sensitive data is stored in the hospital databases and may include confidences from patients to their assisting doctor. Ransomware attacks constitutes a multiple threat as hackers may encrypt data that becomes inaccessible for the hospital and may threaten to expose personal data unless the hospital pays a given amount of money. It is therefore necessary that servers and databases become protected by encryption and adequate control access, including distributed servers and adequate firewalls and VPNs (Virtual Private Networks)

## 4.3 Data Anonymization

Privacy is something that people praise as fundamental right even if it is hard to find a definition as such. The central notion is that a person's intimacy must be protected, and that person must have control over that information, who has access to what portions of that data in what time frame. This is the case of clinical information where a patient may provide access to clinicians of the necessary personal data items

during the scope of a clinical intervention. In fact, definitions of privacy tend to be too broad or too fine since it may provide general considerations, missing some specificities or if it goes into detail data will become useless by its limitations with potential risk for that person (e.g., misdiagnosis, false positive and false negative). Considering privacy as a protection of a person's personal sphere, it is difficult to define such sphere as it depends on context. However, it is possible to elicit some possible definitions of privacy that may be complementary one of the others. Privacy can be defined as the protection of a person's personal information. Privacy is also to keep safe what information regards to the boundaries of a person's sphere.

The usage of data needs sometimes to ensure protection of the sources. That occurs in many areas of science and business but is especially relevant when regarding to people. In that case, when personal data such data is then anonymized.

Several techniques can be applied to ensure data anonymity protecting patient's data. Those techniques include Anonymization, Masking, Pseudo-Anonymization and Differential Privacy.

### 4.3.1 Anonymization

Data collection, storage and processing has the objective of ensuring that people's identity is preserved. That relies on a compromise between providing most of the available data about a citizen and risking identifiable traces that lead to that person and actively making the person untraceable but significantly damaging the value of data. That compromise relies on the option for the best anonymization technique that provides enough data for a given purpose while ensuring that data cannot provide identification of any citizen having data in that set. In that sense there, probably the simplest approach is Data masking as is the one that requires less effort and actions to be taken.

Data is not officially confidential until it is anonymized in such a way that it is impossible or extremely difficult to recognize the subject. This can be done by removing the associated personal data, in such a way the data subject it not or no longer identifiable.

### 4.3.2 Data Masking

This technique consists in hiding some of the fields in a database. Data Masking is a simple way of removing the fields that are not relevant for the current purpose leaving only those that will be processed, analysed, and stored. Data masking introduces an uncertainty that aims at reducing the probability of a person being identified. In that sense the technique poses a hard intervention in terms of making disappear most of the fields that would lead to the identification of the person. This may become a

complicated choice once leaving some fields may conduct to an identification and removing too much will make such date less useful, perhaps without relevance to studies or assessments that would, otherwise, be possible. Nevertheless, is a technique that may be useful in many situations to identify the persons involved. It is also important to notice that from GDPR there is the recommendation to minimize data to what is necessary and, for that, masking fields that are not necessary or not so relevant will enforce the protection over people's identity. Masking is useful for anonymization of pseudo anonymization as presented in the next sub-sections.

### 4.3.3 Pseudo-Anonymization

Pseudo-anonymization consists in making a replacement of the subject by a pseudo identity. The objective is to hide the identification of a person but, there is a possibility that the identity can be traced back to the person, if needed. For that purpose, a pseudonym can be used so that the name is replaced by another name, non-related to that subject, in a way that only certain persons can have access to that bridge between the original person and the assigned name or code. This technique may be useful even inside institutions where information will necessarily flow, electronically or with any other support, in cases that people have legitimate access to such data (e.g. for transport, storage, management) but should not have access to the identity of the persons reported in those documents. Data masking is a possible intervention over data so that fields that identify people's identity are removed thus blocking the possibility of identifying or tracing back to the individuals.

According to Article 4(5) of the GDPR, the term "pseudonymisation" is defined as: "processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person". In order to make a pseudonymisation, the data subject must be first assigned to a pseudonym, e.g. user digital identification. This process can be done by the data subject himself, the controller or an independent trusted third party. Then, an assignment rule must be created, e.g. through a reference table. Thereafter, the additional identification that leads to subject's identification must be kept separately from the data and be separated from the pseudonym. The data subject must be in- formed about the pseudonymisation process and it must be clarified who generates the pseudonym, who owns the assignment rule and under what circumstances an identification may take place. This is because pseudonymised data continues to be personal data [7].

### 4.3.4 Differential privacy

The compromise between privacy and ensuring that there will be enough information to be analysed and processed to generate relevant markers has blurred borders where a clear definition is hard to establish without loss. The effort to pursue plain anonymity may be achieved at the expense of discarding relevant data and maybe jeopardizing the chances of finding relevant insights on a patient status or even of reaching new knowledge. Differential privacy consists in analysing data of such a person without knowing who that person is. It is possible then to know that such a person exists with a certain set of characteristics without knowing who that person really is. This is achieved by some techniques, more or less complex, but in general it could include to insert data about individuals that do not exist but enlarge the pool without compromising the data. Meaning that we can add a greater degree of uncertainty about who a person really is without compromising the value of data in use.

### 4.3.5 Personal Identifiable Data

The other type of personal data that must have special attention are Personal identifiable data. These can include dates and timestamps that are associated with user's birthdate or hospital appointments, or even locations that can be related with user's residence or work. In [8], the authors analysed some health wearable technologies and concluded that some of these devices disclose personal information, such as email address, phone number or social media account information, when using them. As such, it is necessary to first define which data is necessary to acquire and inquire from the user and verify if any of these data can be used to identify its owner. Possible solutions include deletion of these values or reduction of the level of detail.

## 4.4 Communication Protocols

A data communication protocol deals with the rules that allows two or more within a system, to transmit data via two or more points (also called nodes). The protocol defines the rules syntax semantics and synchronization of communication and possible error recovery methods. Protocols may be implemented by hardware software or a combination of both [1].

Transmission Control Protocol is the most popular standard for exchanging data over the Internet Protocol (IP), and it's often referred to as TCP/IP.

[1] "Wireless communication protocol", issued 2004-12-01
[2] https://www.manageengine.com/network-monitoring/network-protocols.html

Next in Table 2 those layers are presented according to their positioning

**Table 1 - Communication Protocol Layers**

| | |
|---|---|
| Layer 7: Application layer network protocols | • Provides standard services such as virtual terminal, file, and job transfer and operations. |
| Layer 6: Presentation layer network protocols | • Masks the differences in data formats between dissimilar systems.<br>• Encodes and decodes data, encrypts and decrypts data, and compresses and decompresses data. |
| Layer 5: Session layer network protocols | • Manages user sessions and dialogues.<br>• Establishes and terminates sessions between users. |
| Layer 4: Transport layer network protocols | • Manages end-to-end message delivery in networks.<br>• Renders reliable and sequential packet delivery through error recovery and flow control mechanisms. |
| Layer 3: Network layer protocols | • Routes packets according to unique network device addresses.<br>• Renders flow and congestion control to prevent network resource depletion. |
| Layer 2: Data link layer network protocols | • Frames packets.<br>• Detects and corrects packet transmit errors. |
| Layer 1: Physical layer network protocols | • Interfaces between network medium and devices.<br>• Defines optical, electrical, and mechanical characteristics. |

Remote access in a secure form is, in many cases, ensured by the Hyper Text Transfer Protocol Secure (HTTPS). HTTPS is a standard protocol to secure the communication among two computers one using the browser and other fetching data from web server. HTTP is used for transferring data between the client browser (request) and the web server (response) in the hypertext format same in case of HTTPS except that the transferring of data is done in an encrypted format. So, it can be said that https thwart hackers from interpretation or modification of data throughout the transfer of packets.

On the internet there are several common use protocols, some are used without notice others are so frequent that become widespread in daily activities that pass unnoticed. Those protocols can be divided into layers according to their relevance for their aim; Application, Transport, Internet and Link layer. For all those layers, there are in numerous options that we will take the most suited and needed for the FAITH project.

## 4.5 DLT in eHealth

Conventional healthcare systems are centralized since all health-related data is controlled and stored by a central entity. Most of these systems have sharing problems, data types are not compatible. Therefore, given their lack of interoperability, sharing the patient's health data can be difficult with the traditional healthcare systems.

Since Distributed Ledger Technology (DLT) emerged, being the most visible application the crypto coins, in particular Bitcoin, many applications have been proposed to this new technology. The healthcare domain is one of those application domains.

Recently, Kosba et al. [9] introduced in 2016 the idea to use automatic scripts from DLT to help data sharing. Distributed ledger technology (DLT) systems conceptually emerged in 1982, before Bitcoin and blockchain technology. In 1982, The Byzantine Generals Problem, theorised by Lamport et. al. in [10], described how 'computer systems must handle (. . . ) conflicting information' in an adversarial environment [11]. The Byzantine Generals Problem is a classical problem that demonstrates the consistency problems faced a distributed system. This is derived from a lack of a general consensus as to what the state of the system is at any given time and may be subjected to an adversarial attack.

Distributed Ledger Technology (DLT) is generating significant interest of applicability by a wide range of enterprises, interested in security process, approve or validate monetary transactions and other type of data exchange. This technology is similar to a spreadsheet where a record of transactions and other account information is accessible and transcribed. This information is owned by every node of a Peer to Peer (P2P) network, where their users have available a consensus mechanism to guarantee the integrity of the stored information [12]. To understand the objective of a distributed ledger, it is necessary to clarify what a ledger is. Ledgers are and have been historically used to record payments or contracts for the transaction of goods or properties. These actions were stored in a trusted place and have moved from being recorded on clay tablets to papyrus, vellum, and paper. Thus, ledgers only allow new data to be appended, making impossible to delete or update appended data. For FAITH project, a ledger is used to mention a group of electronic records, held by a significant proportion of the network participants, which were previously approved as true and are unlikely to be erased or amended during its life cycle. Accordingly to the author in [11], what differentiates DLT systems and traditional distributed databases are features defined on their design, allowing a DLT to maintain data integrity in an environment where

their parties don't fully trust each other to accept and maintain consensus about a set of shared information.

As mentioned before, an important benefit that a DLT brings to healthcare systems is to remove the need to rely on a trusted third party, e.g. a central authority, achieving decentralization. In addition to the fact that data is replicated and shared amongst all nodes in the network, it can provide transparency and ensure a trustful environment between the system's members [13].

Another relevant feature a DLTs can bring for the healthcare domain is data integrity. This fact happens because each participant node has a copy of the ledger and changing past information of one ledger will differentiate this node content with the other nodes. This immutable nature is suitable for high number of scenarios, in which accurate and honest records of information are necessary [12].

## 4.6 Encryption

In cryptography, encryption is the process of encoding information with the objective of protecting the information by making its contents non-understandable for those without the proper key. This process consists in the conversion of original representation of the information, known as plaintext, into an alternative form known as ciphertext. Encryption is thus the process of transforming data from a readable format into an encrypted format that can only be read by authorized users who can convert the encoded data back to original data and access the original data. The process of converting an original message into an encoded format by the sender is known as encryption and the process of converting the encoded message back to its original format is known as decryption.

Encryption becomes essential to protect data in order to restrict its access to those with permissions but also, in a context of increasing cybercrime over institutional data, to protect data from cyberattacks. The fact is that data exposure mines the credibility of institutions as people feels less prone to provide personal data to an institution that suffered a cyberattack or that does not provide clear information about the existence of a data protection framework. One of the strategies used for shared data across a network is the usage of a public and a private key. In simple terms, the private key provides access to the encrypted contents while the public key ensures the validity of the private key. Encryption is a need for different purposes as next listed:

- Authentication – The access to websites can be used in a fraudulent way if secure connections are not established sensitive data may be exposed. Protocols such as HTTPS provide insurance about the type of connection to a website and the type of website a user is accessing. In such cases the user understands that, at least in a first approach, security measures are implemented by that website.
- Security – Electronic Health Records (EHR) need to be protected even in restricted circulation (e.g., in the hospital). It is important that data is not readable so that anyone inside the institutions can access information and eventually, without encryption EHR could leak exposing patients and mining confidence in the institution.
- Confidentiality – Encryption ensures that even in the eventuality of an attack, of any order, intruders would not have access to data. The risk in such case stays in the possibility of ransomware in the sense that data could be hijacked, at least data would not be exposed to hackers. In this case a policy of regular backups would be enough to override any criminal intentions.
- Legal Framework – In certain countries the law impose confidentiality or even to comply with national ethical regulators or even local ethical boards it is necessary to ensure data encryption to protect every user's privacy and data security.

In what regards to Encryption algorithms there are several techniques some with higher level of adoptions but in general all have the same purpose of protecting the contents from unauthorized access.

- Symmetric Encryption Techniques – This method consists in having the same key for encryption and decryption for both sender and receiver, it implies that both have the key prior to usage. This method is also called private key cryptography.
- Asymmetric Encryption Techniques – Also known as public-key cryptography is a method using two keys, the public and the private. The public is publicly available but only with the private key data can be decoded.
- Hashing – Is a process of converting an input of any length to a fixed size string of text by means of a mathematical function. Hashing can also be used to verify the integrity of the data. Hash functions can be classified into unkeyed hash functions and keyed hash functions. Unkeyed hash functions use the message as a single input whereas keyed hash functions take two distinct inputs, the message and a fixed length secret key.

Another cryptographic primitive that can improve a system's security is a *message authentication code*, or MAC. Its purpose is to check the integrity of the message as well as its authenticity, without the use of any additional security mechanism. To accomplish that, MAC function takes two arguments, a fixed-size key (symmetric) and a message, and returns a fixed-size MAC value, called MAC code or tag. To authenticate the message, the sender sends both the message and the tag. The receiver computes the tag of the received message with the symmetric key and verifies if the result of the previous function matches with the received tag. If they match, it turns out that the original message was not modified(Ferguson, Schneier and Kohno, 2015).

Similar to handwriting signatures, a Digital Signature must provide identical properties, such as singularity (i.e., no one can forge other's signature), non-repudiation (threats that are concerned with the users who deny after performing an activity with the data), possibility to prove the ownership of the signed content and ensure the data or contract did not changed after it has been signed. A digital signature can be used to verify if the message was altered during the transfer over the network, therefore, can be effectively used to verify the integrity of the message(Subramanya and Yi, 2006).

A comparison between some primitives that are used to enhance security and were described along the previous sections is made in Table 2 and a detailed comparison can be found in (Pranitha, 2019).

**Table 2: Comparison between security techniques**

| **Characteristic** | **Encryption** | **Hash** | **MAC** | **Digital Signature** |
|---|---|---|---|---|
| Key type | keyed | Unkeyed | Symmetric-key | Public-key |
| Verification | - | Anyone | Key-pair holder | Anyone |
| **Security objective** | | | | |
| Integrity | No | Yes | Yes | Yes |
| Authentication | No | No | Yes | Yes |
| Non-repudiation | No | No | No | Yes |
| Confidentiality | Yes | No | No | No |
| Examples | DES, AES | MD5, SHA | HMAC, CBC-MAC | RSA, DSA |

The presented security primitives provide different approaches that can be used to increase security. As such, in FAITH project, it is crucial to ensure data protection from any kind of unauthorized access, but it is also important to guarantee that the data that is being used, either for building the prediction models or being used by data scientists, is correct or was not modified by an attacker.

## 4.7 Edge computing and Device Encrypted Cache

The aim of protecting data has multiple architectural options that can secure, protect, and avoid data theft or any exposure to misuse or unknown risks. By assessing those risks, becomes obvious that encrypted data is less accessible to others and data that is not sent to the network cannot be intercepted and stolen in the path. In trying to keep privacy for every user, strategies can be used to ensure security of personal data and anonymity of sent data. Thus, processing data behind the edge of the local network (e.g., device connected to smartphone) before sending it and keeping encrypted records seems to be an important strategy to adopt. As such, it is important to notice that data collected by devices may characterize the subject that carries or wears such devices at different levels, from metadata to physiological records. In that sense, processing data and send only the resulting biomarkers encrypting before dispatch to the network are strong protective measures. Encrypted cache is a mechanism for storing sensitive data on the user side, implemented using HTML5 web storage technology, which allows data to be saved locally and retrieved on subsequent used for the application's services.

The data is encrypted and stored by a combination of a user-provided key and a randomly generated token that is retrieved from the server, thus increasing the security level. Encrypted cache is like a secure deposit box, it remains open until an action closes the cache protecting the whole storage.

# 5 Requirements for Data Acquisition in FAITH

In the healthcare complex ecosystem, it is important to identify the flows for sensitive data and where vulnerabilities may exist so that measures could be taken to prevent the associated risks either intentional or casual. Within the FAITH ecosystem, a serious analysis was made to the infrastructure that supports the projects developments and the possible flows of data so that all flanks are covered in what regards to ensure data protection during and after the FAITH project. In that sense, this chapter goes deeper into the concrete aspects of the FAITH project infrastructure in what concerns to the data management processes, creating awareness to the problems and issuing recommendations and strategies aiming to prevent from harm and tackle potential data vulnerabilities. Those aspects, covering background knowledge instantiated to the FAITH project framework, are covered in the next sections.

## 5.1 Data Protection in FAITH

Data protection is a two-sided quest, for one side it is important to empower themselves against attacks and any compromise to data integrity on the other side it is necessary to build trust among users, both clinical and, especially, patients who provide their personal data expecting advice and treatment in strict confidentiality and security.

In the healthcare technical domain, there is a common concern among healthcare personnel about data privacy and the accuracy of ICT systems, including wearable devices, to monitor the health status of patients. This is one of the reasons why the adoption of this type of system in different medical areas is being slowed down, especially among the most conservative. Data privacy is an issue that also concerns citizens (end users). In this sense, FAITH guarantees compliance with all European and national regulations on privacy and data security, including the EU General Data Protection Regulation 2016/679, which guarantees full privacy and protection of sensitive data.

From the definition of the FAITH protocol, the necessary variables have been identified and with them the different types of data present in the Data Model (DM). These variables are demographic, clinical or compliance variables and can be collected by clinical staff or from the mobile application. To achieve interoperability, the data exchange is carried out using the Fast Healthcare Interoperability Resources (FHIR) standard to describe the variables of the FAITH DM. Although FHIR is not a security protocol, it defines exchange protocols and content models that must be used with various security protocols defined elsewhere.

Implementing a production FHIR system requires the use of a security subsystem to manage users, user authentication, and user authorization.

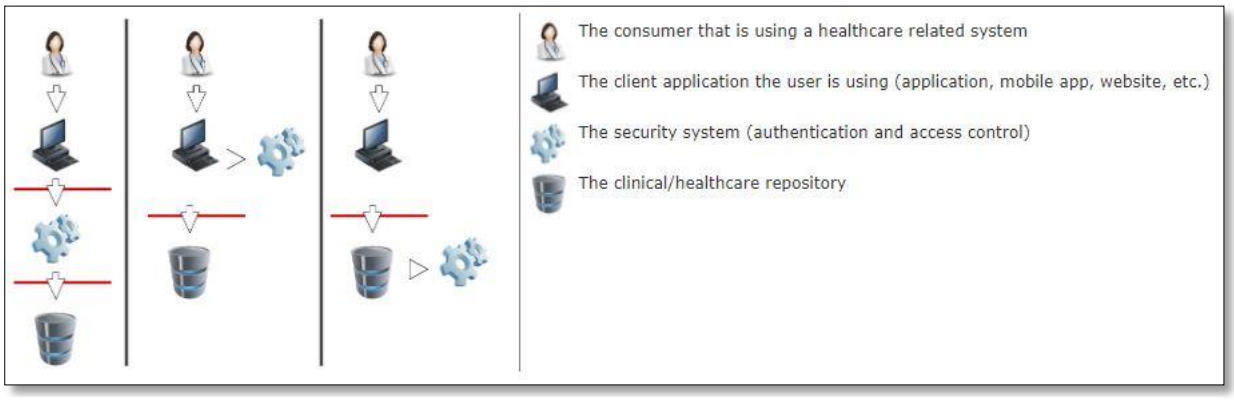


**Figure 1 - Production FHIR System [1]**

Figure 1 shows three different scenarios of how applications or network components can be assembled. The red lines represent FHIR interfaces. From the perspective of the FHIR API, the client (FHIR service consumer) can interact with a security system that manifests itself as an FHIR server and relies on a downstream FHIR interface to provide the actual storage, either the client or the server interacts with the security system independently. The FHIR specification assumes that a security system exists and that it can be implemented in front of or behind the FHIR API (HL7.org, s.f.). The security system should include the following subsystems:

- Authentication: Identifies and authenticates the user.
- Access Control decision engine: Decides whether FHIR operations are allowed.
- Audit Log: records actions to allow for subsequent review and detection of intrusion or inappropriate usage.

In this way, it is intended to ensure the confidentiality of the data of all FAITH users and at the same time preserve the integrity of the information they contain.

Another important issue is the set of good practices that should be followed by systems' administrators as well as by users as mentioned in 4.2. In this direction, Workpackage 4, shall have a best practices manual with the operating parameters of the platform, including recommendations for the administrators and users in terms of passwords and also for the network administration.

## 5.2 Cloud Services

Clouds become a useful asset for scalable storage and computational resources management. A cloud is an important asset by allowing escalation of the infrastructures and for providing the needed resources, as needed, without implying the early investment in hardware and services locally. Clouds can be developed locally, so that a pervasive access is provided for the needs of a framework and can also be requested to a provider with many options in the market. In fact, to have support for a ready to use solution and to quickly escalate the present setup a cloud from a provider in the market can be the most adequate solution. It is however necessary to pay attention to the risks, at different levels, including security threats.

The choice of using a cloud service implies the need of awareness for the threats and possible problems that may arise from those services, a summary of those risks becomes relevant to include while preparing a framework for security where the cloud may be an option. Therefore, it is important to identify risks associated with the usage of cloud. Since data is handed to an external company it is possible that the following can occur:

**Theft or loss of intellectual property**

While depositing information in an external cloud, it is possible that data can be stolen, if some casualty leads to a security breach, data may be stolen. The other problem is that intellectual property can be at risk since the only insurance is the belief in the company's trust policy.

**Malware attacks**

Companies are sometimes subject to malware attacks. Since the information is handed to another company the owned loses all possibility of protecting data. The responsibility is from the company that owns the cloud, if some problem happens those who trusted data to the cloud may be in trouble.

**Contract breaches with clients and/or business partners**

If there is some contract breach it is not clear what could happen to such data and even worse if there are third parties involved in the cloud company's operations.

**Shared vulnerabilities**

The usage of a cloud services makes the co-responsibility for data safety and integrity but also shares de vulnerabilities of both systems. Since the systems depends on both the Cloud server and the local infrastructure, vulnerabilities are shared between both resulting in a major exposure for the overall architecture.

**Attacks to deny service to legitimate users**

Clouds as any other infrastructure can be attacked thus denying users to access their databases.

**Insecure APIs**

APIs are another vulnerability since even with a secure cloud, the interaction with it may be exposed to risk. This is an additional fragility that is part of the cloud system services.

**Concluding on Clouds**

All the risks pointed out to the cloud do not diminish the value of the cloud as an infrastructure widely adopted that solves problems of scalability and anywhere access. The cloud solves the issues with scalability and makes possible for a system to have a reliable infrastructure in terms of robustness and services provided. It is however important to establish some barriers for cloud usage; on one side, only anonymized data should be stored and managed in the cloud. Another important aspect is that mainly if the cloud is provided by a contracted service, it is important to ensure the measures of security and trust of the adopted services (e.g., servers behind firewall, encryption and secure connections).

## 5.3 Server and Infrastructure

Security aspects are at the foundations of the platform architecture. This is assured on stacked levels: software architecture data centre architecture and network architecture.

In what regards to the data access level, after the necessary capture, analysis and processing, data is encrypted before inserted at the database. Thus, in the case of a security breach data would be unreadable and completely useless to safeguard privacy and data security.

To ensure the security of the system, the architecture is designed so that the inner layer comprises the FAITH servers inaccessible from the outside WWW (internet). Both servers “FAITH Server” and “Auth FAITH Server” exist only in the inner network. Data is located in the “Data FAITH Server” and all sensitive data is encrypted in the database which is not directly accessible from the outside and even in the remote possible breach data is encrypted and database not accessible to unauthorized users. On the other side, the “Main FAITH Server” allows access to data over “REST over HTTPS” after request authorization to “Auth Server” depending on the type of permission to each specific user.

## 5.4 Communications in FAITH

In FAITH the technological architecture prevents the risks posed to communications by promoting the anonymization of data before sending packages of information that is, nevertheless encrypted to enforce security of communications. Since that data will be sent, there are several options in what regards to

protocols, being used, once those play different roles. Those protocols can be divided into layers according to their relevance for their aim, being those the Application, Transport, Internet and Link layer. For all those layers there are in numerous options to ensure that the most suited and needed are adopted for the FAITH project as described next.

Communication between the FAITH mobile App and the cloud-based backend system are designed so that they performed based on state-of-the-art Hypertext Transfer Protocol Secure (HTTPS). HTTPS is a natural evolution of the HTTP protocol and is used for secure communication over computer networks, being ideal for the execution of REST API services. The security layer of HTTPS is the Transport Layer Security (TLS) (formerly known as Secure Sockets Layer (SSL), and the overall concept is based on using long-term public and private keys which in turn generate short-lived session keys, with the latter being used to secure the data flows between a client (in our case the FAITH App) and a server (the FAITH Cloud Platform). The preferred TLS version to be used is TLS 1.3 using the AES-256 cipher. Thus, the whole FAITH system is behind a firewall and for the clients the only access is to the “Main FAITH Server” over the port 443 HTTPS and port 80 for HTTP all other ports must be closed.

The only connection allowed to “Auth and Data FAITH Servers” is by the main server’s expectedly in the FAITH network infrastructure. Thus, the whole FAITH system is behind a firewall and for the clients the only access is to the “Main FAITH Server” over the port 443 HTTPS and port 80 for HTTP all other ports are closed. Data from the patients is pre-processed analysed and encrypted behind the Edge and sent to the “FATH Server” via HTTPS. Furthermore, the FAITH Trial Managers and Researchers will access the “FAITH Server” via HTTPS meaning that the connection are encrypted all the time ensuring the security of such connection.

Additionally, the fragmentation of the system allows to protect the data of the data server from possible corruption in case of DDOS (Denial of service attack) since the Main Server can only make requests to the Data Server after authenticating the user.

## 5.5 Mobile Application

With regards to the mobile application, there are several aspects to ensure data protection and privacy for the user. The first is that information is pseudo-anonymized before leaving the smartphone, the

second is that there will be no intrusion in the devices from FAITH, all data is collected, anonymized, and processed and then sent to FAITH. That way the risk is reduced to the normal risk of a smartphone.

The mobile application (FAITH App) which will be used by the different individuals will make sure that the data which are collected and stored adheres to the security and privacy requirements set by the project. In terms of the App this is done in four different layers, as described below:

- Data Access: In terms of data access, it is made explicit that access to the data is only possible to the user of the App. The user will be able to log into the App using a combination of a username and of a secure password which will be only known to him and will be stored encrypted in an online identity provider, in order to allow the user to reset this password in case it is forgotten. No other Applications will have access to the data of the users, without the user allowing them to do so, which is in conformance with the application sandboxing methods employed in Android and iOS.
- Data Acquisition: The data that to be collected by the FAITH App relay on inputs coming from the user, as well as on inputs coming from pre-defined data sources that are used to automatically retrieved metrics which are collected by third party devices (in the case of sleep tracking) or by the different sensors of the mobile phone. As such, is retrieved directly from different sources within the phone's operating environment, or through REST APIs that will be performed over secure communication channels, using specific user tokens, as for example the calls necessary to retrieve the activity data from the Google Fit service.
- Data Storage: The data that is stored in the mobile phone will be encrypted and as protected from any data leakage.
- Data Exposure: Data will be exposed only towards the FAITH cloud-based engine, which will be necessary for the initial training of the algorithms, prior to constructing and putting in operation the federated learning algorithms. From that moment on, no captured data is exposed, apart from the federating learning model attributes, and the only data that is exposed towards the cloud-based engine would be an identifier of the user in case an alert regarding his health status is triggered. In the latter case, the actual data of the user remains on the phone and can be offloaded to a device owned by a medical institution only upon user's command.

## 5.6 Data Visualization Service

The interaction of FAITH with the doctors will be supported by a Data Visualization Interface within task T3.4. This service will be held via HTTPS a usual practice for services to be deployed by banks or other entities where security is a critical demand. The servers will be the same as for the whole infrastructure

so that no additional measures are needed. The service will be pseudo-anonymised since graphics and displayed data will be identified by a code that only the hospital will have the correspondence to real people. It is however important to notice that aimed function of the platform is to alert for identified potential patients at risk of mental health decline. The objective, beyond the need to provide alerts is to

## 5.7 Data privacy and Data protection compliance to Hospitals

The trials will be executed with local populations from each hospital. This fact implies that for each Trial a set of regulations need to be enforced since it must respect that hospital ethical board's request. It is therefore necessary a consultation to each hospital about applicable law and ethical board's requests in terms of data acquisition and Management. However, since the trials will be executed from the faith platform the resulting framework will be a merge of all requests since technology will be applied to most of the technological settings, unless some request is made for a particular condition. In the next sub-sections it is reported the requests from each hospital, but the overlapping is not needed since the requests will be the sum of all requests and not specific to each trial.

### 5.7.1 Hospital General Universitario Gregorio Marañón

HGUGM follows as minimum the Madrid Resolution on worldwide privacy standards, the International Conference of Data Protection and Privacy Commissioners adopted privacy by design "as a holistic concept that may be applied to operations throughout an organisation, end-to-end, including its information technology, business practices, processes, physical design and networked infrastructure" (the so-called "Privacy by Design Resolution[2]" as well as the main principles from the European legal data protection context (Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, including: lawfulness, consent, necessity and data minimisation, transparency and openness, rights of the individual, information security, accountability, data protection by design and by default.

### 5.7.2 Centro Clínico Champalimaud - Fundação Champalimaud

The Champalimaud Foundation data privacy and data protection guidelines are built in accordance with current national and European legislation on the protection of individuals, in particular the General Data

[2] 32nd International Conference of Data Protection and Privacy Commissioners. Privacy by de-sign resolution, October 2010. 27-29 October 2010, Jerusalem, Israel

Protection Regulation (GDPR). These guidelines, in summary, define that personal data should be processed lawfully, fairly and in a transparent way; the purposes for which data are collected should be specific, explicit and legitimate and should not be further processed in ways that are incompatible with the initial purposes; data should be processed in an adequate and relevant manner while also being limited to what is necessary; data should be accurate and kept up to date; data should be kept in a format that only allows the identification of the data subjects for the necessary amount of time during its processing; data should be processed in a secure manner, and all efforts must be taken to ensure protection against unauthorised or unlawful processing, including the use of appropriate technological and organisational means.

The data collected during FAITH will be exclusively used for the objectives of the project, or for additional scientific research conducted with the purpose of further validating instruments developed during FAITH. All information generated during FAITH will be kept in paper and digital formats. In order to maintain data privacy subjects will be pseudonymised, meaning that they will not be identified using their real name, but with a code that is exclusive to them. If results of FAITH are published or if data are shared with other investigators (in the format described above, e.g. to further validate instruments developed during FAITH), subjects personal data will be kept confidential. The principal investigator responsible for FAITH at FC will have access to a list that associates the identifying code and the subjects' personal information, and this list will be kept physically and virtually separated from the remaining data collected during FAITH. The regulatory agencies and members of the ethical committee (e.g., Data Protection Officer) will be granted access, upon request, to the associative list and any other subjects' personal data collected during FAITH. If data are shared with other investigators, the associative list will not be provided.

An informed consent form will be provided to each and every subject participating in FAITH, and will be requested to sign it if in agreement with all the information there present. By signing the informed consent, subjects will agree to participate in FAITH and allow for their data to be collected and processed. By signing the informed consent, subjects will also provide access to clinical information collected prior to FAITH (during cancer treatment) in case this information proves fundamental to the development of FAITH.

Subjects participating in FAITH are entitled to request access from the principal investigator to all data concerning them. Subjects are also entitled to withdraw from the study at any time, to ask for rectification or for erasure of data. Any of these decisions will have no consequences on present or future treatments or clinical follow-up at the Champalimaud Foundation.

### 5.7.3 UPMC

The management of personal data is a central aspect for the data management on UPMC units. The preservation of the patients' data is a matter of concern for managers and systems dealing with the flows of data, in different modalities, along with the standards for data integrity and subsequent data analysis as a source of information for decision support systems in the clinical environment. UPMC follows GDPR as a major regulation for patient handling and additionally applying national and local regulations in each unit. The strict preservation of patients' data is an insurance of conformity with the legal framework but also a matter of trust for users, in all clinical interventions, so that confidence is ensured with the clinical staff and the computational systems. UPMC applies the best practices and recommendations from the World Health Organization (WHO), European Medications Agency (EMA), U.S. Food and Drug Administration (FDA) and other local regulatory bodies. It is therefore of most importance for the Ethical Board to evaluate well documented proposals for clinical trial management, ensuring full conformity with legislation and regulations.

## 5.8 DLT, Data privacy and Data protection

IoT systems, in particular for healthcare, use devices in order to get data for applications and use a middleware layer component where data can be stored. As mentioned before, a DLT can be used to store data, and it can use two different types of storage, on-chain and off-chain storage. The first type of storage is used when the data are directly stored on the ledger. It has the advantage of their data being immutable, sharable and accessible through all nodes, since it is replicated across multiple nodes and it is not possible to change their value. In [14], the authors use a DLT as a general-purpose database distributed system to store the IoT sensorial data that is used by multiple parties. This method to store data allows the system to always have the information accessible by third-party components and using the latest data, by automatically synchronize all nodes. However, this approach may directly conflict with GDPR privacy rights presented above, such as the "right to be forgotten", requiring the deletion of user's stored health records from DLT. This right clashes with the immutability objective of this technology. Apart from that, the data must not be visible to the other users without the authorization of information owners, which is a violation of privacy. Scalability of DLTs in the healthcare domain is another challenge because there is a high volume of data involved. For health applications that depend on latency, DLT can incur considerable processing delays, especially if the data load is significant.

The second type of storage is an off-chain storage. This type of storage is independent from the DLT and it can be an SQL or NoSQL database. When is used an external component to store information, such as on a cloud infrastructure, the user should encrypt that information before transferred it [15]. In such a case, health data can be stored off-chain, and the user exercises the “right to be forgotten”, the system has the ability to delete the information that is stored off-chain. This explains why most of the state-of-the-art on medical data sharing, as mentioned in [16], use off-chain storages to store health data, while data query strings and hash values are stored on-chain for authenticity and integrity verification. Even if the pointer to the off-chain storage cannot be erased, due to the immutability of the ledger, the health data that was stored can be deleted anytime. On the other hand, only storing a hash of data on-chain and keeping the contents off-chain will improve confidentiality and may allow the storage of big files quicker, but partly undermines the distinctive benefit of a DLT in providing distributed trust. This may create a single point of failure or reducing system availability and reliability. Another type of off-chain storage are hash tables. A hash table is a data structure that is used to store key-value pairs by resorting to a hash function. The key value is hashed and the hash value is used as the index of the table, where the data will be inserted or searched for. Hash tables can provide a constant time for search and insert operations. The indexes of the hash table act as pointers to the address of the stored data. Moreover, by storing the hash value on the blockchain, the authors could ensure the integrity of the stored data.

Both on-chain and off-chain types of storage bring advantages and disadvantages for the system, and for this reason they should be chosen carefully. In [17], the authors made a comparison between off-chains and on-chains storages.

## 5.9 Architectural design for privacy

Taking into consideration IoT systems, in particular for the Healthcare domain, use multiple devices and produce and consume a large number of different services, they are prone to be more susceptible to attacks than the rest of the Internet [9]. For this reason, a central characteristic of every IoT system is security, which must be considered when designing every layer of the system architecture. Accordingly to [10], the main principles an IoT system must address for gaining security are the following:

- Authenticity: It guarantees the origin of a service request, a piece of data or a message, the identity of a service provider or the creator of a piece of information. Assigning a unique identifier for the devices and users is the basis for the authentication step and the consequent authorization phase. The authenticity for a piece of data is done by recognizing the creator’s signature.

- Confidentiality: Available information is not available or disclosed to unauthorized individuals, entities, or processes. The data access of the IoT system must be controlled mainly by means of cryptographic mechanisms and users access lists [11]. As an example, this security feature allows two users to communicate with each other and be sure that nobody apart from these users can read the data or information of the messages.
- Integrity: An IoT system is based on exchanging data and information between different types of devices, applications or end-users. That is why it is important to ensure the accuracy of the data. During the data life cycle, i.e. since the device collects data until this data is erased, the system must maintain the consistency, accuracy and trustworthiness of their data.
- Availability: Users and system's components should have all the data available whenever is needed. Not only that, but the devices and existing services must also be reachable and available when needed.

The FAITH platform is aimed to securely handle and store identifiable, personal health and health-related data. For this reason, it must follow a privacy and secure by design approach. As mentioned in deliverable D3.1 (Hospital Cloud-Network Infrastructure, Visualisation & Distributed Ledger Technology (DLT)), this platform is destined to be used by different users, such as clinical, patients or data scientists. For this reason, the platform receives the data through mobile apps, hospitals, and results from data analysis, including explainable Artificial Intelligence.

The expected behaviour from the system in terms of data privacy is described during this section, considering sensitive data is handled across all platform's components and for this reason, it must fulfil the data protection standards, as described in section 4.2.

### 5.9.1 Authentication Management

**Registration and Login**

A secure registration and authentication must be made available for all end-users of FAITH platform, by using a user-friendly credentials, such as the mobile phone number or the user's email. The patients will use this functionality which provides a security token that will be used by the FAITH App to interact with FAITH platform in a secure way. This connection is made through an HTTPS channel.

At the registration phase, each user provides personal details, such as email or mobile phone. For this reason, these personal data should be stored and securely protected (e.g., through encryption) in a

separated repository. The user must also receive an informed consent form and agree with the presented conditions.

### 5.9.2 User account information repository

Data repository responsible to contain the information of each registered user. It will contain personal data, such as the user's names, emails, and address. This sensitive information is encrypted and only decrypted by the data owner or users with pre-agreed data access. This repository does not contain data that is used by any component of the FAITH platform (e.g., user's weight) to ensure that it is compliant with GDPR data minimization principle.

### 5.9.3 Data Sharing

In general, traditional data sharing transactions contain four main concepts, namely:

- Data owner - has the right to specify which user(s) will have access to the data.
- Data requester - any user that has the intention to get and use the data shared by a data owner.
- Intermediary - responsible to prove the identity of data requester and the credibility of data owner. The intermediary also supervises the establishment of the transaction.
- Transaction data - data that will be shared.

In order to have a patient-centric approach, the developed framework ensures the patient's own and control their personal data. As such, on FAITH platform, the patients are considered as the owners of the data and at any given time, they can change the set of permissions and revoke the access of the provided data. This can be achieved by protecting all relevant data by using techniques such as encryption or pseudonymisation. At this point we describe two access control mechanisms that are widely used in the literature, namely Attribute Based Access Control (ABAC) and Role Based Access Control (RBAC).

In order to have a patient-centric approach, the developed framework ensures the patient's own and control their personal data. As such, on FAITH platform, the patients are considered as the owners of the data and at any given time, they can change the set of permissions and revoke the access of the provided data. This can be achieved by protecting all relevant data by using techniques such as encryption or pseudonymisation. As an example, Attribute Based Access Control (ABAC) and Role Based Access Control (RBAC) are two access control mechanisms that are widely used in the literature, which allows to make a fine-grained choice regarding access to user's data. They also allow to define which other services/applications/trusted end-users the user explicitly grants access to their data, in addition to default rights the user has over his/her own data (e.g., the data owner can retrieve, update or delete its

data). The data sharing permissions shouldn't be static but dynamic because the data owner can allow or abolish the authorizations he/she gave in the past.

### 5.9.4 Data Limitation management

The user's data should be stored for a defined period, based on the necessity of using its data and no more than that. The same does not apply to reasoning or data extracted from that and other patients, once ensured anonymization process, for data that can become a legacy to scientific knowledge. As such, a component that manages the expiration date for the user's data should be used on FAITH platform. As an example, user's personal data should be erased when user no longer is enrolled on the trial or if the data is no longer needed, and for this reason, the data should be scheduled to be deleted upon FAITH conclusion.

### 5.9.5 Data Ownership Handling

Users own their data. For this reason, they must be able to update or delete their data whenever they need. Moreover, each user should be able to get their data that is stored on FAITH repository. It should also be available to the end-user the possibility to inactivate its account and give the user the option to later choose to activate or delete the account. This would result on specific measures on its account, such as to freeze its information and subsequently pause the data collection.

### 5.9.6 Id management

FAITH platform uses the patient's Personal Health Records (i.e., sleep monitoring device, nutrition, outlook, and activity) gathered during the trial to enable the creation of the prediction models. For this reason, they will contain a pseudonym (as mentioned in sub-section 4.3.3) such as an ID that will be used to associate with a unique patient. This method ensures that the stored data cannot be attributed to any patient, without being associated with additional information. For this reason, the association between the user's identification and its pseudonym must be stored separately from the patient's data and the hospital, where the patient is followed, is the only identity that holds the connection between patient's ID and its pseudonym. Next in Figure 2 such approach is presented where the names are pseudonyms for one user at each pool which will be replicated for all users in a pool with assigned pseudonyms.

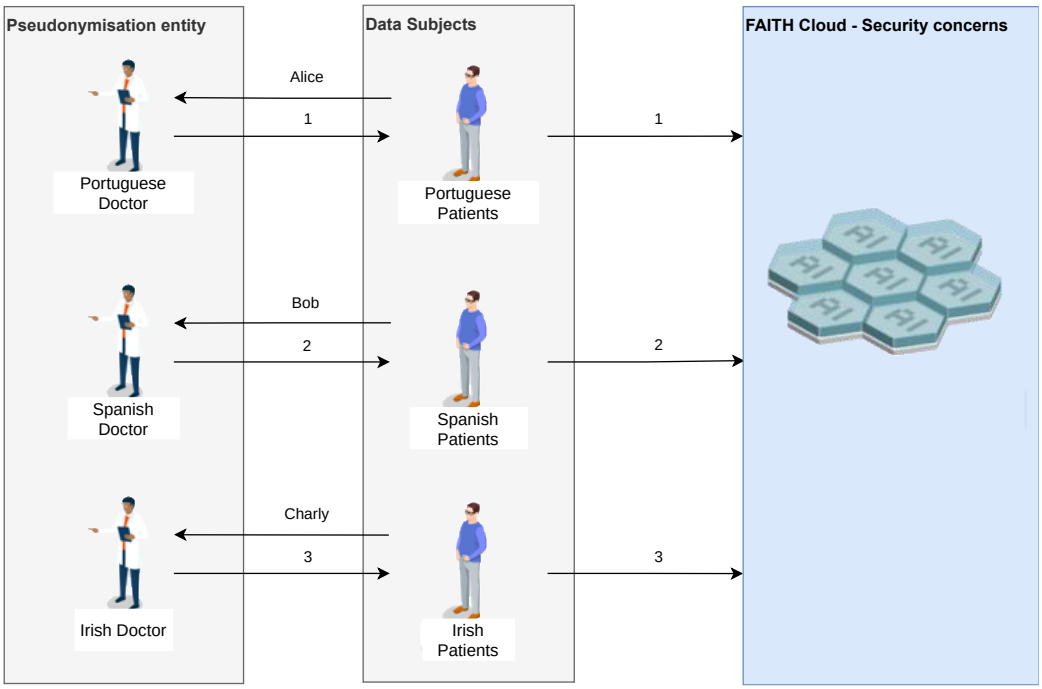


**Figure 2 – ID Management in FAITH Trials**

Data pseudo-anonymization, as depicted in Figure 2, will ensure compliance with the Ethical Boards and to the premises of FAITH to comply with GDPR and ethical standards for Trial execution, by physically separating user's identity form patient's pseudonym.

### 5.9.7 Personal Identifiable data management

As mentioned in section 4.3, some data can be used to identify the patient, and for this reason, additional measures should be taken, such as masking data and promote differential privacy (explained in sub-section 4.3.4). For this reason, both mobile app and FAITH platform need to have this kind of component. This FAITH modules implied in personal data management will be responsible to remove the personal identifiable data from the data streams. The thread of personal data management must cope with the need to identify relevant biomarkers of mental health decline with the need to preserve patient's anonymity. In this regard, during the execution of the trial the patient's identity will be preserved and data to be analysed. This will be accomplished since the patient's hospital is the only entity that can establish the link between data and a patient, for the benefit of the patient and the clinical process. It is therefore the Medical Doctor who follows a patient who needs to have access to the link between data and the patient. In the aftermath of the FAITH project, data protection will be reinforced in the sense that, without the need to develop additional markers, data protection no longer needs to be used beyond the identified markers with the algorithms proposed in the scope of the project's execution.

### 5.9.8 Data Notification Centre

Users must know if personal data breaches occur within FAITH environment, especially if their data was compromised. The data notification centre will be responsible to communicate to the end users when their personal data was compromised or as a confirmation notification when an action involving the user's data was performed (e.g. user requested to delete or update their personal data). The users should also be informed when a request to delete or update their data is made, in order to have a confirmation of those requests were fulfilled.

### 5.9.9 FAITH data repository

The data collected and produced during the trial is stored in the centralized FAITH platform will be stored after being subjected to high security procedures. First, a pseudonymous mechanism is made in the user data, which will replace the user's identification with the corresponding pseudonym. Secondly, the data stream is analysed in order to detect all personal identified data that can be used to identify the patient. Afterwards, its access is limited to specific role(s). With the aim to improve security of the patient data, two additional measures are considered by resorting to DLT functionalities. The first one is to guarantee the data that is used for the FAITH platform has integrity and the second one is an access log mechanism that will register which users viewed his/her pseudonymised data. To be in accordance with GDPR legislation, the proposed security architecture allows the FAITH platform to delete from its repositories all patient data when it is asked by the patient. Since the DLT will not be used to store any personal data or personal identifiable data, this immutable component continues to follow all GDPR legislations made until now.

### 5.9.10 Data Integrity Validation

Assimilating the integrity verification mechanism increases the sense of control over the data for the end users of the system. This privacy mechanism will create trust while using the registered data while resorting to DLT functionalities as depicted in section 4.5. Since the information stored on the distributed ledger is open and immutable for each node, a possible approach, for FAITH project, is to use an off-chain storage and only use the DLT to store a hash of the patient data.

### 5.9.11 Data Access Logs

A logging system for monitoring data access will be used in order to know which users accessed the data from the FAITH repository. By registering when someone tries to access data, it is possible to create logs of which part of the data was accessed, who accessed the data and when this action took place. The implementation of these logs with a DLT allows for almost unchangeable information that is available for each individual to consult.

Next in Figure 3 is depicted the proposed architecture of the security framework within FAITH project.

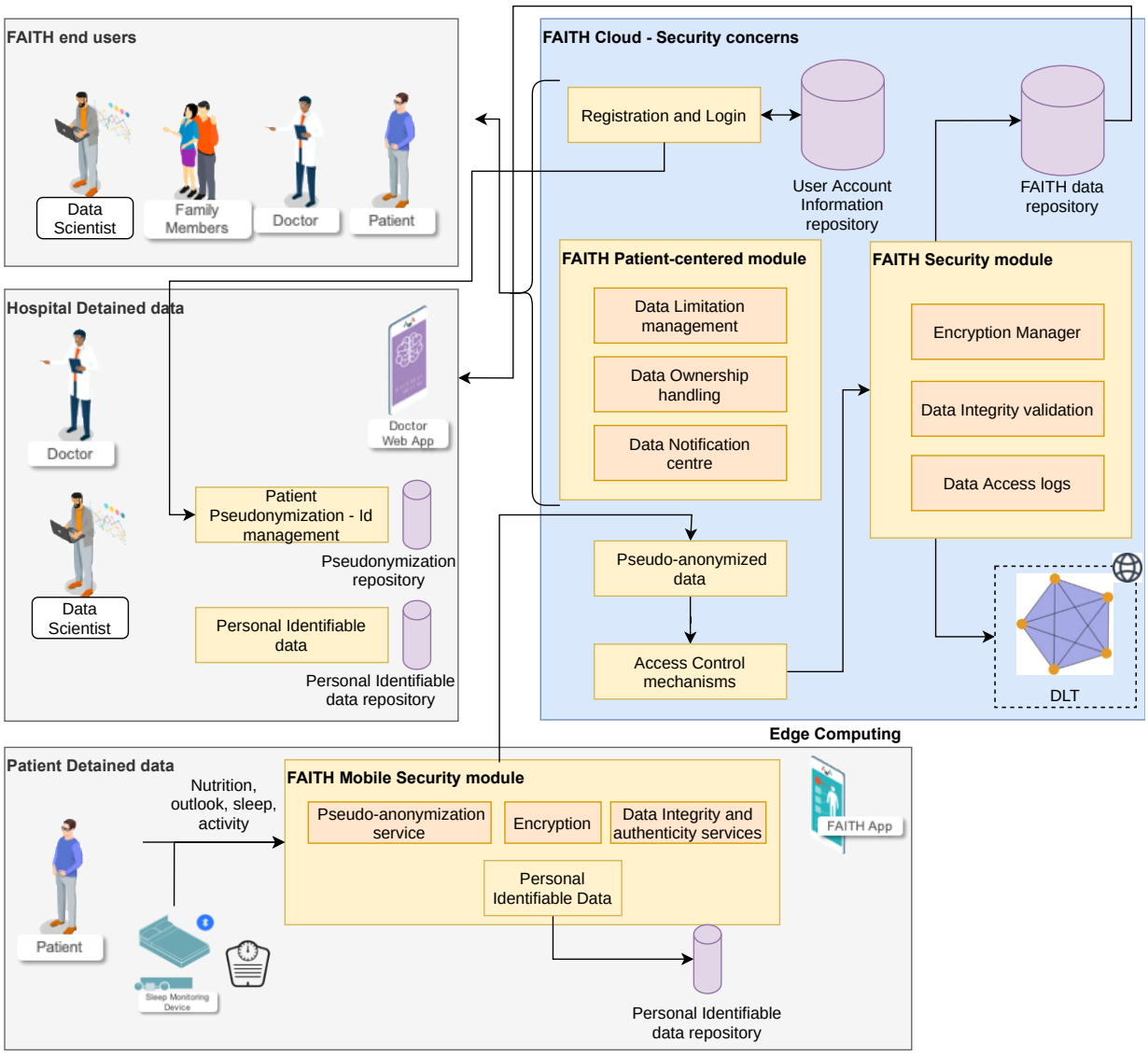


**Figure 3 - Security framework with DLT log registration**

The system will be backed by a reliable cloud infrastructure, such as Amazon Web Service (AWS) that will provide the necessary services with potential for scalation and adaptation to remote access needs.

The before presented Framework was designed to be used by several profiles such are: the Patients that are the focus of all, the Project and in particular, the Trials. In some cases, the patients are supported by the Family Members. From the hospital side there will be involved the Doctors and Data scientist profiles. Tentatively, the system's administrator from the side of the healthcare institution will be a Data Scientist

since those are qualified personnel in computer domains and thus more prone to solve any network problem at that level.

Users are the source of data, both when using devices and filling questionnaires. Data thus originated and detained by the patient is still behind the Edge of the local device. At the person's neighbourhood, data is personally identifiable and will be pre-processed at that person's device level. At that point, Data becomes pseudo-anonymized since their ID is replaced by an alternative identification that only that person's doctor and institution have the key to match the pseudo-ID with the real ID.

The FAITH Patient-centred module handles the limitations on data usage, at the same time it handles data ownership and the data Notification centre, so that evaluates the thresholds reached by data analytics, triggering notifications as needed. The FAITH Security Module will manage encryption procedures, validates Data Integrity, and manages the Data Access logs. Those logs stored in the DLT, in this case at Ethereum blockchain, are a fundamental piece towards the support for audits on data access on a protected way, the security provided by the DLT.

From the side of the hospital, there exists repositories of Personal Identifiable data that makes part of what the infrastructure knows about the patient. That information should be controlled by the Doctor, but that is under the traditional management infrastructure of the Hospital. In that sense, since the hospital has the conventional information about the patient and the information from the project, pseudo-identified, along with the keys to identify such information, becomes possible to reason, compare, and evaluate all data. This is the ultimate clinical objective for the Trial evaluation; to have data and knowledge about the patient to verify and evaluate a Patient's mental status and validate data generated by the FAITH framework.

The process is managed based on the information from the User Account Repository. This repository will serve as a base for the Registration and Login Process, ensuring that all data has integrity, through the support of the DLT thus closing the circle of obtaining data, managing it, supplying to the target audience (e.g. doctors and data scientists) regarding all safety and data protection mechanisms, duly registering all accesses in the DLT where it becomes securely stored and auditable anytime.

#### 5.9.12 DLT Demo

Under the scope of Task 3.3, a demonstrator was made to enable testing and software development for the Distributed Ledger Technology approach to support register and audit in the blockchain. The objective is to establish a service that will operate in the Ethereum Blockchain hereby developed as a sandbox for demonstration purposes.

The main window, presented in Figure 4 depicts the service with the hash representing the DLT account for the storage of the smart contract and the blocks used to registrate the transactions.

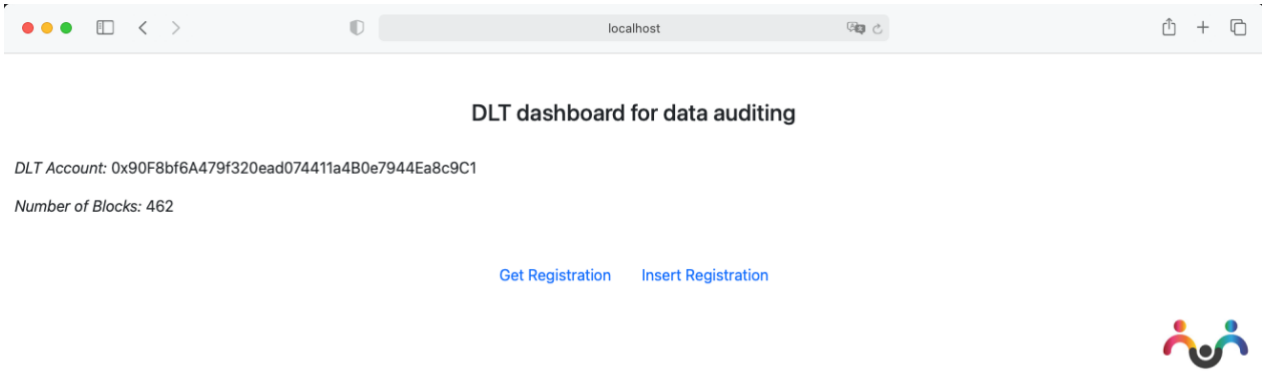


**Figure 4 - DLT Dashboard for data auditing**

Then it is possible to take actions to get registration of logs for the areas in FAITH patients monitoring such as Activity, Outlook, Sleep and Appetite/Nutrition, as presented next in Figure 5

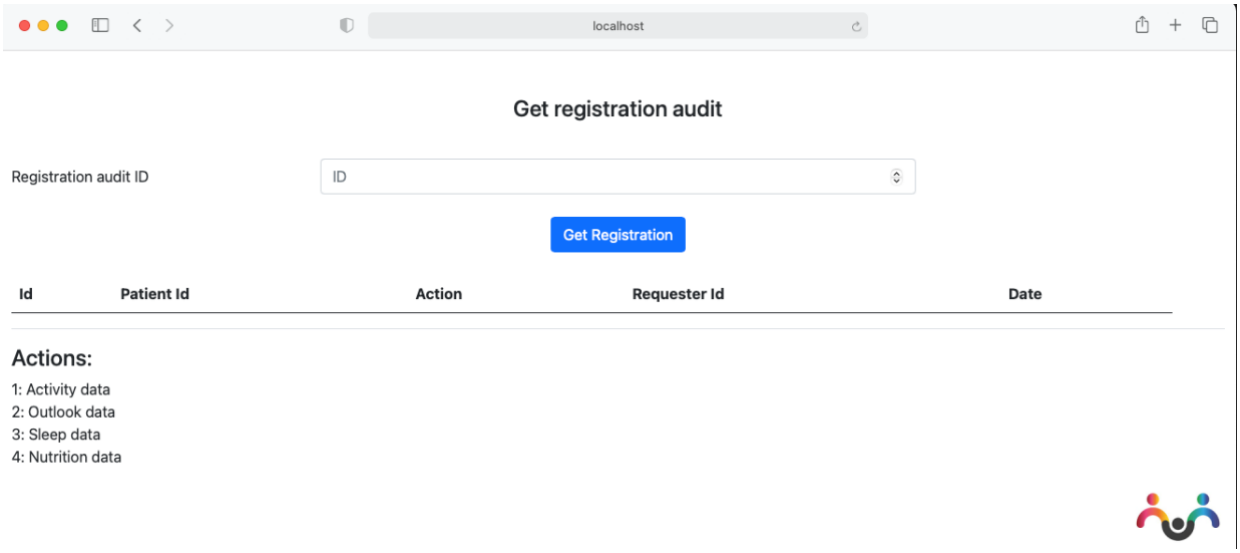


**Figure 5 - Interface for getting registration logs from the blockchain**

# 6 DISCUSSION AND CONCLUSIONS

This Deliverable presents the framework for data security and privacy at FAITH research project. The main goal is to provide a background about the main aspects to be addressed to ensure data privacy and security. The document initially presents the state-of-art in this regard with a generalist view of all aspects but aspects related to specifically to FAITH project circumstances are further explored. This deliverable establishes the foundation for the software development and deployment regarding all those aspects that are critical for the operation of a FAITH framework and in compliance with legislation and Ethical Board's requirements.

The rational on data privacy and GDPR compliance, described in chapter 3, is the base of all constructs towards the Framework for Data Security and privacy. This development can be seen as the technological implementation of what was analysed and widely discussed, in this regard, and reported in the FAITH Protocol Document. Following such guidance, the summary presented in chapter 3 highlights the relevance for the technical developments and it gives the background for the design and implementation of the FAITH privacy and security framework. In pursuing such goals, FAITH will observe a strict policy of anonymization, where the hospital is the keeper of patients' identity and FAITH platform with the data, only matched by the respective hospital personnel.

In pursuing the aim of strictly observing law, local regulations such as hospital requests as well as the European GDPR regulations, chapter 4 presents the state of art in technological deployments for the assurance of security and privacy on a framework. This is an accumulation of all information considered relevant for the implementation of a secure and trustable FAITH platform.

The analysis of the requirements and technical needs for FAITH is addressed in chapter 5, where the previously presented state-of-art is judiciously selected and adapted to the needs for the development of the FAITH platform.

In this sense, it is noted that, in what regards to Data protection as presented ahead in section 5.1. FAITH will comply with all the regulations, mainly GDPR and FHIR, as also presented in section 5.1. The Cloud becomes an important asset as support for services, storage and also due to its scalability. However, it is important to take in account the risks declared in section 5.2. The infrastructure, supported by the cloud, has the security and data protection as a main aspect and it will have serious concerns. Those are presented in section 5.3.

Communication protocols as a segment of the developments are analysed in section 4.4 where the potential options are presented giving place to an instantiated solution that meets the needs of FAITH as presented in section 5.4. The mobile application and the Visualization service were also objects of study in this deliverable with a presentation in sections 5.5 and 5.6 noting, however, that the development of

this applications will be adjusted according to the requests for data protection from the Ethical Boards and also adjusted to the needs of the users being those the doctors and data scientists and, in the case of the application, the final users the patients.

As the concerns mount for the security aspects to respect all the regulations and law it is also important to ensure the compliance with hospitals' Ethical Boards and for that, section 5.7 presents views from the three hospitals in the project, which will run the trials, so that it becomes possible to understand and converge to the Ethical Boards requests.

In section 5.8 and quite extensively in section 5.9 the design for security, privacy and data protection are presented, including the usage of the DLT. Those become the guidance in those matters based on the studies performed earlier in sections from 4.2 to 4.6. All those concerns were analysed in terms the needs for the FAITH platform and thus become instantiated in chapter 5.

In the conclusion of the technical part, in sub-section 5.9.11 it is presented in Figure 3, the overall view of how the architecture will be arranged with the particularity, there represented, of using the DLT for logs' registration. This figure gives a presentation of how the architecture is being designed and how the security aspects are implemented along with the specification of the different components.

**Conclusions**

The development of computational networks regarding all the aspects of anonymization, data privacy, secure connections and infrastructure security are complex and multidimensional. There are many aspects to be considered and a set of solutions for each potential vulnerability. Such solutions are not unique, and the best option depends on the system, its goals and, in some concerns, on the entities involved and their specific demands. This deliverable covered most of the relevant aspects proposing the most suitable options and thus, addressed the important aspects for FAITH technical deployment.

Finally, to conclude, it is important to reinsure that the observations from the developers' point of view and the interaction with ethical boards will shape the development process and the requirement's evolution. The tests and software evaluation will complete this process and provide information for the next edition of the present document.

# 7 APPENDIX

## A.1. Legal Framework for HGUGM

- Lawfulness: including GDPR principles "lawfulness, fairness and transparency" (Art. 5), and "lawfulness of processing" (Art. 6)
- Consent: GDPR definition of the "data subject's consent" (Art.4) and "conditions for consent" (Art. 7)
- Purpose of binding: GDPR principle "purpose limitation" (Art. 5 & Art. 21)
- Necessity and data minimisation: including amongst others GDPR principles "data minimisation", "storage minimisation" and "data protection by design and by default" (Art. 5, Art. 23)
- Transparency and Openness: including amongst others GDPR (Art. 5, Art. 10, Art. 11, Art. 12, Art. 13, Art 14, Art 15, including "general principles for data subject rights", "concise, transparent, clear and easily accessible policies", "standardised information policies", "information to the data subject", "right to access and to obtain data for the data subject", and defining the conditions for exercising data subject rights).
- Rights of the individual: including amongst others and GDPR (Art. 5 "effectiveness", Art. 7 including right to withdraw consent at any time, Art. 10 "general principles for data subject rights", Art. 13 "notification requirement in the event of rectification and erasure", Art. 17 "right to erasure", Art. 19 "right to object", Art. 12 including defining the conditions for exercising data subject rights).
- Information Security: including amongst others GDPR (Art. 5 principle "accuracy", principle "integrity"; Art. 30 "Security of processing", Art. 50 "Professional secrecy").
- Accountability: accountability is not directly stated, but aspects of the accountability principle are considered by including amongst others: DPD (Security of processing) or by mentioning the possibility of appointing a "personal data protection official" who should be responsible for ensuring the application of data protection law; in the GDPR: Arts. 5, 22, 33, 35
- Data protection by design and default: including amongst others GDPR (Art. 23 "Data protection by design and by default")
- Accountability: accountability is not directly stated, but aspects of the accountability principle are considered by including amongst others: DPD (Security of processing) or by mentioning the possibility of appointing a "personal data protection official" who should be responsible for

ensuring the application of data protection law; in the Data Protection Working Party (Art. 29) and GDPR: Arts. 5, 22, 33, 35

- Data protection by design and default: including amongst others DPD (Art. 17 security of processing) and GDPR (Art. 23 “Data protection by design and by default”)

# 8 Bibliography


[1] European Parliament and Council of European Union, "Regulation (EU) 2016/679 of the European Parliament and of the Council," 2016. https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32016R0679&from=EN (accessed Jan. 09, 2021).

[2] Centers for Disease Control and Prevention, "Health Insurance Portability and Accountability Act of 1996 (HIPAA)," 2018. https://www.cdc.gov/phlp/publications/topic/hipaa.html (accessed Apr. 12, 2021).

[3] R. Creemers, P. Triolo, and G. Webster, "Translation: Cybersecurity Law of the People's Republic of China (Effective June 1, 2017)," 2018. https://www.newamerica.org/cybersecurity-initiative/digichina/blog/translation-cybersecurity-law-peoples-republic-china/ (accessed Apr. 12, 2021).

[4] V. Janeček, "Ownership of personal data in the Internet of Things," *Comput. Law Secur. Rev.*, vol. 34, no. 5, pp. 1039–1052, Oct. 2018, doi: 10.1016/j.clsr.2018.04.007.

[5] A. Mense and B. Blobel, "HL7 Standards and Components to Support Implementation of the European General Data Protection Regulation (GDPR)," *Eur. J. Biomed. Informatics*, vol. 13, no. 1, pp. 27–33, 2017, doi: 10.24105/ejbi.2017.13.1.5.

[6] D. Peloquin, M. DiMaio, B. Bierer, and M. Barnes, "Disruptive and avoidable: GDPR challenges to secondary research uses of data," *Eur. J. Hum. Genet.*, vol. 28, no. 6, pp. 697–705, 2020, doi: 10.1038/s41431-020-0596-x.

[7] F. Menges *et al.*, "Towards GDPR-compliant data processing in modern SIEM systems," *Comput. Secur.*, vol. 103, p. 102165, Apr. 2021, doi: 10.1016/j.cose.2020.102165.

[8] E. Y. Arquitectura *et al.*, *Networked RFID Systems and Lightweight Cryptography*, vol. 53, no. 9. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008.

[9] A. Kosba, A. Miller, E. Shi, Z. Wen, and C. Papamanthou, "Hawk: The Blockchain Model of Cryptography and Privacy-Preserving Smart Contracts," in *2016 IEEE Symposium on Security and Privacy (SP)*, May 2016, pp. 839–858, doi: 10.1109/SP.2016.55.

[10] L. Lamport, R. Shostak, and P. Marshall, "The Byzantine Generals Problem," *ACM Trans. Program. Lang. Syst.*, 1982.

[11] M. Rauchs *et al.*, "Distributed Ledger Technology Systems: A Conceptual Framework," 2018. doi: 10.2139/ssrn.3230013.

[12] P. Ferraro, C. King, and R. Shorten, "Distributed Ledger Technology for Smart Cities, the Sharing Economy, and Social Compliance," *IEEE Access*, vol. 6, pp. 62728–62746, 2018, doi: 10.1109/ACCESS.2018.2876766.

[13] X. Yang, T. Li, X. Pei, L. Wen, and C. Wang, "Medical Data Sharing Scheme Based on Attribute Cryptosystem and Blockchain Technology," *IEEE Access*, vol. 8, pp. 45468–45476, 2020, doi: 10.1109/ACCESS.2020.2976894.

[14] M. Pincheira, M. Vecchio, R. Giaffreda, and S. S. Kanhere, "Cost-effective IoT devices as trustworthy data sources for a blockchain-based water management system in precision agriculture," *Comput. Electron. Agric.*, vol. 180, no. November 2020, p. 105889, Jan. 2021, doi: 10.1016/j.compag.2020.105889.

[15] I. Sukhodolskiy and S. Zapechnikov, "A blockchain-based access control system for cloud storage," in *2018 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (EIConRus)*, Jan. 2018, pp. 1575–1578, doi: 10.1109/EIConRus.2018.8317400.

[16] H. Jin, Y. Luo, P. Li, and J. Mathew, "A Review of Secure and Privacy-Preserving Medical Data Sharing," *IEEE Access*, vol. 7, pp. 61656–61669, 2019, doi: 10.1109/ACCESS.2019.2916503.

[17] T. Hepp, M. Sharinghousen, P. Ehret, A. Schoenhals, and B. Gipp, "On-chain vs. off-chain storage for supply- and blockchain integration," *it - Inf. Technol.*, vol. 60, no. 5–6, pp. 283–291, Dec. 2018, doi: 10.1515/itit-2018-0019.

[18] C. Pielli, D. Zucchetto, A. Andrea Zanella, L. Vangelista, and M. Zorzi, "Platforms and Protocols for the Internet of Things," *EAI Endorsed Trans. Internet Things*, vol. 1, no. 1, p. 150599, 2015, doi: 10.4108/eai.26-10-2015.150599.

[19] Q. Gou, L. Yan, Y. Liu, and Y. Li, "Construction and Strategies in IoT Security System," in *2013 IEEE International Conference on Green Computing and Communications and IEEE Internet of Things and IEEE Cyber, Physical and Social Computing*, Aug. 2013, pp. 1129–1132, doi: 10.1109/GreenCom-iThings-CPSCom.2013.195.

[20] R. Mahmoud, T. Yousuf, F. Aloul, and I. Zualkernan, "Internet of things (IoT) security: Current status, challenges and prospective measures," in *2015 10th International Conference for Internet Technology and Secured Transactions (ICITST)*, Dec. 2015, no. October 2016, pp. 336–341, doi: 10.1109/ICITST.2015.7412116.