D2.1. Technical recommendations for the FAIR4Health platform and agents implementation
Improving Health Research in EU through FAIR Data

**Document Information**

<table>
<thead>
<tr>
<th>GRANT AGREEMENT NUMBER 824666</th>
<th>ACRONYM: FAIR4Health</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full title</td>
<td>Improving Health Research in EU through FAIR Data</td>
</tr>
<tr>
<td>Horizon 2020 Call</td>
<td>SwafS-04-2018: Encouraging the re-use of research data generated by publically funded research projects</td>
</tr>
<tr>
<td>Type of action</td>
<td>Research and Innovation Action</td>
</tr>
<tr>
<td>Start Date</td>
<td>1st December 2018</td>
</tr>
<tr>
<td>Duration</td>
<td>36 months</td>
</tr>
<tr>
<td>Website</td>
<td><a href="http://www.fair4health.eu">www.fair4health.eu</a></td>
</tr>
<tr>
<td>Project Officer</td>
<td>Raluca Iagher</td>
</tr>
<tr>
<td>Project Coordinator</td>
<td>Carlos Luis Parra Calderón, Andalusian Health Service (P1–SAS)</td>
</tr>
<tr>
<td>Deliverable</td>
<td>D2.1. Technical recommendations for the FAIR4Health platform and agents implementation</td>
</tr>
<tr>
<td>Work Package</td>
<td>WP2: Comprehensive Analysis for FAIR Data policy implementation in Health research</td>
</tr>
<tr>
<td>Date of delivery</td>
<td>Contractual M06 Actual May 2019</td>
</tr>
<tr>
<td>Nature</td>
<td>Report Dissemination Level Public</td>
</tr>
<tr>
<td>Lead beneficiary</td>
<td>P6–HL7</td>
</tr>
<tr>
<td>Responsible Author</td>
<td>Catherine Chronaki</td>
</tr>
<tr>
<td>e-mail</td>
<td><a href="mailto:chronaki@HL7europe.org">chronaki@HL7europe.org</a> Phone number +30 6944594990</td>
</tr>
<tr>
<td>Reviewers</td>
<td>Patrick Weber (EFMI), Carlos Luis Parra-Calderón (SAS)</td>
</tr>
<tr>
<td>Keywords</td>
<td>FAIR, Technical Barriers, Security, Interoperability</td>
</tr>
</tbody>
</table>

**Document history**

<table>
<thead>
<tr>
<th>Version</th>
<th>Issue Date</th>
<th>Stage</th>
<th>Changes</th>
<th>Contributor(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>24/04/2019</td>
<td>Draft</td>
<td>First full draft</td>
<td>C. Chronaki (HL7), V. Foufi (UNIGE), M. Schweitzer (UMIT), L. Huber (UMIT), A. Sinaci (SRDC), A. Hörbst (UMIT), F. Núñez (SAS)</td>
</tr>
<tr>
<td>0.2</td>
<td>28/05/2019</td>
<td>Draft</td>
<td>Reviewers’ comments</td>
<td>P. Weber (EFMI), C. Parra (SAS)</td>
</tr>
<tr>
<td>1.0</td>
<td>30/05/2019</td>
<td>Final</td>
<td>Comments processed</td>
<td>C. Chronaki (HL7)</td>
</tr>
</tbody>
</table>

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 824666.
Table of Contents

1. Executive Summary ............................................................................................................. 3
2. Introduction ......................................................................................................................... 4
    2.1. Background: scope and alignment with other deliverables in WP2 ......................... 4
    2.2. Objective and target audience of this document ....................................................... 4
3. Methodology ....................................................................................................................... 5
    3.1. Technical barriers and overcoming strategies ............................................................. 5
    3.2. Technical aspects related to cybersecurity ................................................................... 6
        3.2.1. Scenario 1: Local access to raw data ................................................................. 7
        3.2.2. Scenario 2: Local access to FAIR data ............................................................... 7
        3.2.3. Scenario 3: PPDDM access to FAIR data ........................................................... 7
        3.2.4. Scenario 4: P2P access to FAIR data ................................................................. 8
        3.2.5. Scenario 5: Access to eHealth services .............................................................. 9
4. Technical recommendations for FAIR policy implementation in health research .......... 9
    4.1. Technical barriers identified for FAIR data policy implementation in EU Health research and proposed overcoming strategies ................................................................. 9
        4.1.1. Trustworthiness barriers ...................................................................................... 10
        4.1.2. Health data management life cycle barriers ......................................................... 11
        4.1.3. Interoperability barriers .................................................................................... 12
        4.1.4. Other technical barriers .................................................................................... 14
    4.2. Technical aspects derived from the security requirements ......................................... 14
        4.2.1. General Concerns ................................................................................................ 14
        4.2.2. Identity Management Module .......................................................................... 14
        4.2.3. Authentication & Authorization Module (Auth Module) ..................................... 15
        4.2.4. Audit Module .................................................................................................... 16
        4.2.5. PPDDM Security Module .................................................................................. 17
        4.2.6. End-to-end Security Module ............................................................................. 17
    4.3. Utilization of the Cybersecurity modules within the scenarios .................................... 17
5. Summary: Overcoming Strategies for Technical Barriers ............................................. 21
6. References ......................................................................................................................... 21

List of figures

Figure 1. Scenario 1 - Local access to raw data. ................................................................. 7
Figure 2. Scenario 2 - local access to FAIR data ............................................................... 7
Figure 3. Scenario 3 - PPDDM access to FAIR Data .......................................................... 8
Figure 4. Scenario 4 - P2P access to FAIR data. ......................................................... 8
Figure 5. Scenario 5 - Access to eHealth services. ......................................................... 9
Figure 6. Security Modules attached to Scenario 1. ......................................................... 18
Figure 7. Security Modules attached to Scenario 2. ......................................................... 19
Figure 8. Security Modules attached to Scenario 3. ......................................................... 19
Figure 9: Security Modules attached to Scenario 4. ......................................................... 20
Figure 10. Security Modules attached to Scenario 5. ......................................................... 20

List of acronyms

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATNA</td>
<td>Audit Trail and Node Authentication</td>
</tr>
<tr>
<td>CRUD</td>
<td>Create, Read, Update and Delete</td>
</tr>
<tr>
<td>DM</td>
<td>Data manager</td>
</tr>
<tr>
<td>DS/SD</td>
<td>Data Scientists/Software Developers</td>
</tr>
<tr>
<td>eIDAS</td>
<td>electronic IDentification, Authentication and trust Services</td>
</tr>
<tr>
<td>EHR</td>
<td>Electronic Health Record</td>
</tr>
<tr>
<td>EU</td>
<td>European Union</td>
</tr>
<tr>
<td>FAIR</td>
<td>Findable, Accessible, Interoperable and Reusable</td>
</tr>
<tr>
<td>FHIR</td>
<td>Fast Healthcare Interoperability Resources</td>
</tr>
<tr>
<td>HL7</td>
<td>Health Level 7</td>
</tr>
<tr>
<td>HP</td>
<td>Healthcare professionals</td>
</tr>
<tr>
<td>HR</td>
<td>Health researcher</td>
</tr>
<tr>
<td>IG</td>
<td>Implementation Guide</td>
</tr>
<tr>
<td>IHE</td>
<td>Integrating the Healthcare Enterprise</td>
</tr>
<tr>
<td>OID</td>
<td>Object Identifier</td>
</tr>
<tr>
<td>OSF</td>
<td>Open Science Framework</td>
</tr>
<tr>
<td>P2P</td>
<td>Peer-to-peer</td>
</tr>
<tr>
<td>PHI</td>
<td>Protected Health Information</td>
</tr>
<tr>
<td>PPDDM</td>
<td>Privacy-Preserving Distributed Data Mining</td>
</tr>
<tr>
<td>RPO</td>
<td>Research Performing Organisation</td>
</tr>
<tr>
<td>WP</td>
<td>Work Package</td>
</tr>
</tbody>
</table>
1. Executive Summary

This report “Technical recommendations for the FAIR4Health platform and agents implementation” presents in high level the technical issues that arise when implementing a FAIR data policy in health research. It provides general directions for overcoming these issues emphasizing the importance of developing a trust building strategy for health data reuse that addresses privacy and data security concerns, while fostering interoperability and open science principles.

On the technical aspects related to cybersecurity, five different scenarios included in the FAIR4Health use cases that elaborated alternatives related to identification and access to data, as well as user role were explored:

(a) local access to raw data  
(b) local access to FAIR data  
(c) PPDDM access to FAIR data  
(d) P2P access to FAIR data  
(e) access to eHealth services

The first technical barrier identified was trustworthiness linked to the dimensions of certification, security by design, and transparency.

The second technical barrier identified was health data management life cycle barriers and considers the dimensions of metadata acquisition, curation, and archiving.

The third technical barrier considered is interoperability, technical, semantic, and organizational. The issue of selection and indexing of metadata is particularly important.

The report proposes a set of key recommendations to address these technical barriers oriented towards an effective implementation and use of the FAIR4Health platform:

❖ To encourage public institutions to deploy FAIR demonstrators with health data  
❖ To reach a harmonization of FAIR metrics  
❖ To implement sound data provenance methods  
❖ To boost the availability of trusted data repositories  
❖ To develop a FAIR code of conduct  
❖ To encourage the development of a sustainability plan beyond the research project  
❖ To engage software providers to make data FAIR  
❖ To raise awareness and provide training on the use and management of metadata  
❖ To emphasize the use of community-based standards and ontologies  
❖ To set-up a best practice guide for the use of health information exchange standards  
❖ To address alignment and harmonization of meta data in FHIR IGs.  
❖ To democratize tools for mapping data with related standards and ontologies.
These key recommendations of the report are **for RPOs to consider** as they develop their FAIR data strategy. Therefore, the report complements and address deliverables D2.2 “Functional design of the FAIR4Health platform and agents - FAIRification workflow”, and D2.3 “Guidelines for implementing FAIR open data policy in health research”. This deliverable (D2.1) will be used to guide work in WP3 and WP4, responsible for the design and implementation of the FAIR4Health platform, respectively.

2. Introduction

2.1. Background: scope and alignment with other deliverables in WP2

In the frame of the FAIR4Health project, the main objective of WP2 “Comprehensive analysis for FAIR data policy implementation in health research” was to elucidate which are the current barriers, facilitators and potential overcoming mechanisms for the implementation of a FAIR data policy in EU health research institutions. For this purpose, information from a wide variety of domains (technical, ethical, security, legal, cultural, behavioural and economic) has been gathered in order to inform a guideline directed towards providing the optimal strategy for Research Performing Organisations (RPOs) to implement a FAIR data policy.

A first draft of this guideline has been released within the deliverable D2.3 "Guidelines for implementing FAIR data policy in health research", that will be further discussed among the FAIR data community seeking its endorsement and a wide consensus on this topic.

Furthermore, two other key outputs of this WP have been:

- The FAIRification workflow (D2.2) expressed as the set of processes and functionalities that should be implemented in order to adhere to a FAIR data policy in health research.
- The set of technical recommendations (D2.1) that would facilitate the implementation of these processes and functionalities considered in the FAIRification workflow.

2.2. Objective and target audience of this document

The main objective of this document is to gather all the findings related to general technical issues that could hinder the implementation of a FAIR data policy in EU health research. In this sense, several technical barriers have been considered and potential overcoming strategies have been analysed and described. In addition, given that reusing health data raises special concerns in terms of privacy and data security, and as long as this project aims at developing upon a trust building strategy, this document also stresses on the findings about technical recommendations from a cybersecurity point of view.

This report is intended for an audience interested in improving their degree of knowledge about the technical barriers and potential overcoming strategies that a RPO might face when trying to implement a FAIR data policy in health research. As stated in the background
section of this document, this report is complementary to deliverables D2.2 “Functional design of the FAIR4Health platform and agents - FAIRification workflow”, and D2.3 “Guidelines for implementing FAIR open data policy in health research”.

3. Methodology

This report deals with two main aspects regarding the implementation of a FAIR data policy in EU health research: on one hand, a general analysis of technical barriers that could hinder the implementation of such policy and proposed overcoming strategies, and, on the other hand, an specific analysis of the technical aspects related to cybersecurity strategies that should be put in place in order to guarantee data subjects’ privacy when reusing health data for research purposes.

3.1. Technical barriers and overcoming strategies

For the general analysis of technical barriers, the following key references in the field have been reviewed: “Turning FAIR into reality” report [1] and “Interoperability and FAIRness through a novel combination of Web technologies” paper [2]. As a result of this revision, several general barriers have been brought up and summarised in three main topics as follows:

1. Trustworthiness
   a. Certification. How can we measure the FAIRness of health datasets?
   b. Security by design. In a context of sharing and reusing health data, which are the most suitable security and cybersecurity measures?
   c. Transparency. Is it possible to audit every operation performed over every single piece of (meta)data? how should it be performed?

2. Health data management life cycle
   a. (Meta)data acquisition from multiple sources (images, text, bio-signals, etc). How should this be handled?
   b. (Meta)data curation. We need to curate, validate and add relevant metadata to our raw data. How to provide these services and keep updated data provenance?
   c. (Meta)data archiving. PIDs assignment, indexation, local vs non-local repositories. How to proceed to be FAIR and legally compliant?

3. Interoperability
   a. Technical. Different HW and architectures exchanging information at the same time. How should be done?
   b. Semantic. Use of common standards in the community. Is HL7 FHIR suitable for this task?
   c. Organisational. How could organisational interoperability be promoted from a technical point of view?

In order to elucidate potential overcoming strategies related to these barriers, a focus group of experts in the field was convened to discuss each one of them by teleconference.
The group of experts included the following members from the FAIR4Health consortium and from the External Scientific Advisory Board:

- **Alfonso Valencia**: Director of the Spanish National Bioinformatics Institute (INB-ISCIII).
- **Anil Sinaci**: Senior researcher and project manager at SRDC.
- **Catherine Chronaki**: General Secretary of HL7 Foundation.
- **Ronald Cornet**: Associate professor at the department of Medical Informatics in the Amsterdam Public Health research institute, Academic Medical Center - University of Amsterdam.
- **Mario Rodríguez**: Software analyst at Atos Research and Innovation Healthcare group.
- **Mark Musen**: Professor of Medicine at Stanford University, Director of the Stanford Center for Biomedical Informatics Research.
- **Marcos Martínez**: Senior research software engineer, MED/BMIR at the Stanford Center for Biomedical Informatics Research.

After the discussion, the members of the focus group reached consensus on a list of potential overcoming strategies. These strategies were then translated into technical and functional recommendations.

### 3.2. Technical aspects related to cybersecurity

Regarding the specific analysis of the technical aspects related to cybersecurity strategies that should be put in place in order to guarantee the privacy of data subjects when reusing health data for research purposes, members of the FAIR4Health project identified **five different scenarios** where security-related factors on data type, data access, and user role applied:

- **Type of data**:
  - Raw data, including personal identifiers.
  - FAIR data, making distinctions between:
    - Pseudonymised data: personal identifiers are replaced with artificial identifiers, or pseudonyms.
    - De-identified data: personal identifiers are prevented from being connected with information.

- **Type of access**:
  - Through PPDDM algorithms
  - Through P2P
  - Local access

- **User roles**:
  - Data managers (DM)
  - Health researchers (HR)
  - Data Scientists/Software Developers (DS)
  - Healthcare professionals (HP)
The development of the scenarios was based on the FAIR4Health open community and its foreseen information communication exchanges, resulting the following:

3.2.1. Scenario 1: Local access to raw data

In this scenario, users (data managers, DM) access locally to raw data stored in their local facilities to process it and make it FAIR by using the FAIR4Health local agent. Users will perform processes of pseudonymisation/de-identification, curation, validation, mapping to standard vocabularies, authoring and preservation of data in order to enable its sharing, reusability and actionability. In this scenario, only metadata is sent outside the users’ facilities to enable its discoverability.

3.2.2. Scenario 2: Local access to FAIR data

In this scenario, users (health researchers, HR) access locally FAIR data stored in their local facilities to perform health research related activities, such as updating their datasets or performing statistical analyses on them. In this scenario, only metadata is sent outside the users’ facilities to enable its discoverability.

3.2.3. Scenario 3: PPDDM access to FAIR data

In this scenario, users (data scientists, DS) trigger the execution of PPDDM methods to access locally FAIR data stored in the data owner’s facilities to perform algorithmic operations on these datasets. In this scenario, the outcomes of these algorithmic operations
will be sent to the FAIR4Health platform, which will be located outside the data owner’s facilities.

![Diagram](image)

**Figure 3.** Scenario 3 - PPDDM access to FAIR Data.

### 3.2.4. Scenario 4: P2P access to FAIR data

In this scenario, once the users (health researchers, HR) have received the appropriate permission from the data owner, they access to remote FAIR datasets through the FAIR4Health platform. In this case, FAIRified datasets (pseudonymised or de-identified) will be sent to the user, which will be located outside the data owner’s facilities.

![Diagram](image)

**Figure 4.** Scenario 4- P2P access to FAIR data.
3.2.5. Scenario 5: Access to eHealth services

In this scenario, users (healthcare professionals, HP) access eHealth services based on FAIR data through the FAIR4Health platform during their daily work. The following is an example of this scenario:

During a routine consultation, a medical doctor wants to make use of the “Prediction of 30-days readmission risk” eHealth service. After checking that his patient fulfills the inclusion and exclusion criteria, he completes the online form provided by the FAIR4Health platform with the required clinical observations and launches the execution of the eHealth service, receiving after a few seconds the 30-days readmission risk prediction according to the entered variables.

![Figure 5. Scenario 5 - Access to eHealth services.]

A list of technical recommendations was then drafted for each one of the scenarios, adding a complementary list of general technical recommendations related to cybersecurity and interoperability.

4. Technical recommendations for FAIR policy implementation in health research

4.1. Technical barriers identified for FAIR data policy implementation in EU Health research and proposed overcoming strategies

The FAIR principles constitute a powerful context to make data and services more useful and reusable. Although they are not directly linked to a standard or a specific format, the guiding principles may involve some challenges to be faced. The objectives of FAIR4Health require analysing and identifying the potential technical barriers for the generation of FAIR datasets.

The ambition of the project to bring the FAIR guiding principles to the Health Research community, frequently involves some technical barriers that are important to detect and try to overcome. In general, the barriers could be identified by a wide range of stakeholders involved in the context: from IT development, security, analytics, operations to legal departments, audit and regulatory compliance, data producers, health institutions, researchers, independent third parties, end customers and end users.

To address the study of the technical barriers that influence the implementation of FAIR data policies in Health research, an initial classification has been proposed that has served
as a skeleton to unify the treated topics and to structure the strategies associated with them.

4.1.1. Trustworthiness barriers

Some descriptions of the term Big Data indicate that it can be defined by referring to a series of characteristics that begin with the letter 'V' [3]: volume, variety, velocity. This can be extended with value, validity and veracity [4], which lead us to consider the accuracy of the data not just as the quality of the data, also the inconsistencies, abnormalities, and how trustworthy the data source, data processing of it is.

The quality of data involves essentially data validity, consistency and completeness. FAIR4Health needs to address this by determining a clear data flow and defining what is required from the different systems, what is expected and thus setting the basis for them to have clear understanding of data and its provenance.

The focus group that has studied the technical barriers in FAIR4Health has structured them in three topics:

- **Certification**: How can we measure the FAIRness of health datasets?
- **Security by design**: In a context of sharing and reusing health data, which are the most suitable security and cybersecurity measures?
- **Transparency**: Is it feasible to audit every operation performed over every single piece of (meta)data? how should it be performed?

The sharing of medical data is a complex issue. Currently, there are enough technological tools that facilitate the task of convincing citizens regarding the sharing of their medical data and can even demonstrate technically how to eliminate any traces that link medical data with personal information for research purposes. This barrier can be overcome by implementing existing technology that allows these results to be obtained. Once this technology is implemented, trust can be built through practical demonstrations.

Making the current data FAIR implies, in some cases, access to tools that can help in the process. If this supposes an extra expense of resources or is difficult to find, it will be a problem for the data producers. They must perceive the **advantages of offering their data without appreciable extra effort in infrastructure and processes**, regardless of the immediate benefits. Furthermore, data providers should be supported by software providers to make the FAIRification process easier.

Regarding the FAIRness of health databases, there are currently several proposals for the development of **FAIR metrics**. FAIR4Health as a FAIR initiative counts with them to harmonize these metrics. It is relevant to associate the FAIR metrics with the related trustworthiness and ensure that FAIR data are also data with accuracy and quality.
Security and transparency can also be associated with trustworthiness. In case of accountability, **auditing every operation and metadata could not be feasible**, and furthermore some national norms can affect.

In the field of research, trustworthiness is generated mainly by a thorough knowledge about how the data was created, under what conditions, and through a track about its operations. Again, the existence of **metadata to describe the data and its operations** is a key factor.

Due to the recurring importance observed in the metadata as well as in the data by the focus group, the objective of gaining trust in the data must therefore be extended to the metadata. After data acquisition, data curation and metadata assignments, the establishment of a seal that **guarantees the long-term stewardship** of the data is identified both as a barrier to implement it, and as a valid solution in itself, which has already been recently reflected in existing FAIRification recommendations [5].

**Proposed strategies for building trustworthiness in the use of FAIR data for health research:**

- To encourage public institutions to deploy FAIR demonstrators with health data
- To reach a harmonization of FAIR metrics
- To implement sound data provenance methods
- To boost the availability of trusted data repositories
- To develop a FAIR code of conduct
- To encourage the development of a sustainability plan beyond the research project
- To engage software providers to make data FAIR

These overcoming strategies will be used to advance the trustworthiness of the FAIR4Health platform in a number of ways: (a) in creating synergies with other EU funded and international projects (b) in supporting public institutions to create FAIR demonstrators that are measurable, trusted, and sustainable abiding to a shared code of conduct (c) in assisting software vendors to provide FAIR support as part of their products.

4.1.2. Health data management life cycle barriers

The life cycle of health data comprises several stages. Involving the FAIR principles in existing processes requires knowing and adapting them to align the cycle with the FAIR data.

We can consider that **the cycle of health data** consists of the discovery of the data, the capture and its normalization. Subsequently, the health data are aggregated and reported. Based on the information that is known, the data is interpreted, and actions based on the data can be taken.
Health data are considered in their dual nature for FAIR data: data and metadata. Therefore, the focus group has taken as reference the following topics:

- **(Meta)data acquisition** from multiple sources (images, text, bio-signals, etc). How should this be handled?
- **(Meta)data curation.** We need to curate, validate and add relevant metadata to our raw data. How to provide these services and keep updated data provenance?
- **(Meta)data archiving.** PIDs assignment, indexation, local vs non-local repositories. How to proceed to be FAIR and legally compliant?

The treatment of data is not a trivial matter. The first barrier that may arise is the need to clarify the use of metadata associated with the current life cycle of health data. This clarification is necessary both in the annotation of the data and in the analysis of datasets after the data has been generated based on metadata. The validation of the data and its quality, together with the analysis, suppose an obstacle linked to the lack of previous experience.

A new work position specialized in data curation is evidenced as necessary in this context. The research industry must therefore adapt to the use of metadata and, consequently, know the associated semantic and associated data web technologies. This **role of data curation** would correspond to the task of creating the correct metadata and registering the correct identifiers of the datasets. New tasks associated with the management of metadata should be separated from professionals with a medical profile.

The use of metadata brings with it the importance of using **standards, vocabularies and ontologies** to ensure that the data is comprehensible and, in the future, searchable. Therefore, using metadata must have the ambition to achieve high quality metadata.

---

**Proposed overcoming strategies for health data management life cycle barrier:**

- To raise awareness and provide training on the use and management of metadata
- To emphasize the use of community-based standards and ontologies

---

**4.1.3. Interoperability barriers**

The Healthcare Information and Management Systems Society defines interoperability as [6]: “The ability of two or more systems or components to exchange information and to use information that has been exchanged”.

Health organizations are complex and for different reasons the evolution of their information systems has not been uniform. Given the levels of specialization of health services, it is common to find different information systems even for similar functional needs. It is necessary to establish electronic data exchange processes, both in the same health organization and in projects with a broader focus, such as FAIR4Health. In these cases, there is a need to interoperate, exchange data and use the information exchanged. To exchange information fluidly, it is essential to adopt standards and common mechanisms on which all the systems involved coincide.
The focus group has faced interoperability in the following levels:

- **Technical interoperability**: Different HW and architectures exchanging information at the same time. How should be done?
- **Semantic interoperability**: Use of common standards in the community. Is HL7 FHIR suitable for this task? How can we address the needs for harmonizing or aligning metadata standards?
- **Organisational interoperability**: How could organisational interoperability be promoted from a technical point of view?

Different architectures, paradigms and standards are coexisting today, and it is unapproachable to be oriented towards a solution that fits all. Along with interoperability directly related to clinical standards, other common standards and mechanisms must be adopted to achieve good FAIRification of data, including transport and exchange protocols.

The barriers caused by interoperability can be addressed through interfaces built on existing environments, so that they allow the search and obtain data in a uniform manner for researchers. This requires the definition of a high level minimum content that will be exchanged. Subsequently, specific standards and terminologies will model their effective representation.

In terms of interoperability, all the parties are meant to reach an agreement. HL7 FHIR is identified as a good framework to work effectively with good metadata and best practices.

The tools to facilitate the adoption of the standards, terminologies and ontologies through which interoperability pursues are crucial. There is a barrier, which was predictably detected in the previous topic, which refers to the need for tools to achieve the objectives towards FAIR, in this case, that allow transforming, linking and annotating the data in the appropriate ontologies and following the standards. Metadata are particularly important for the sustainable take up of FAIR health data and addressing that problem with an HL7 FHIR accelerator program or an implementation guide is a topic worth considering.

**Proposed strategies to advance interoperability:**

- To set-up a best practice guide for the use of health information exchange standards
- To address alignment and harmonization of metadata in HL7 FHIR IGs.
- To democratize tools for mapping (meta) data with related standards and ontologies

The proposed strategies to advance interoperability will be taken up by WP3 and WP4 in designing and implementing the FAIR4Health platform.
4.1.4. Other technical barriers

Through the topics in which the work of the focus group has been structured, it is considered that the technical barriers have been sufficiently framed.

A brainstorm prior to the work in the focus group had thrown potential technical barriers in the inconsistency of data between institutions, possible problems incorporating the FAIR requirements in the data management lifecycle (from acquisition to storage, structuring, curation and categorization), adoption by healthcare professionals, the need for workforce with digital skills and the possible influence of regulatory and finally the legal and ethical concerns.

All these matters were treated by the focus group or derivatives to their corresponding study group when their content does not correspond essentially to technical barriers.

4.2. Technical aspects derived from the security requirements

The following sections are an excerpt from the “Report on security requirements for FAIR implementation and data reusability in health research” openly available at OSF\(^1\). The reader is referred to such report for a further and comprehensive understanding of the cybersecurity requirements developed by FAIR4Health.

4.2.1. General Concerns

This report recommends utilize/implementing open source protocols and standards while designing the cybersecurity related components of FAIR4Health. Security related leaks should be monitored continuously, this can be achieved through automated security tests in production environment. A dedicated cybersecurity interface can be implemented for the end-users so that they can always see a summary of the ongoing security related activities and the implemented precautions. This interface can also be designed as a security wizard so that the respective administrative users can manage the security related activities by turning modules/features on and off based on specific scenarios and requirements.

4.2.2. Identity Management Module

This module should be responsible for the unique identification of all kind of clients within the FAIR4Health system. Clients can be the users (Data Manager, Health Researcher, Data Scientist, Healthcare Professional) of the several components of FAIR4Health while the components themselves can be the clients interacting within the FAIR4Health system.

- All clients should be managed (CRUD operations).

\(^1\) https://osf.io/268f3/
Non-managed clients cannot make any transactions within FAIR4Health. Authentication & Authorization Module should ensure this by interacting with this Identity Management Module.

Object Identifiers (OIDs)\(^2\) are an identifier mechanism standardized by the International Telecommunications Union (ITU) and ISO/IEC for naming any object, concept, or "thing" with a globally unambiguous persistent name. OIDs can be used as a basis for the identification of the clients and components within this module.

Patient identities which exist in the raw data should be handled within a different mechanism than the client identification. If our scenarios require patient identification across different data sources, then we need a patient identification module. The following standards can be used to design the patient identification module.

- **IHE PIX**: The Patient Identifier Cross Referencing (Pix) Integration Profile\(^3\) supports the cross-referencing of patient identifiers from multiple Patient Identifier Domains. This profile enables a secure and trusted identification of patients although they have different identifiers within different data sources.

- **HL7 FHIR MPI**: A Master Patient Index (MPI)\(^4\) is a service used to manage patient identification in a context where multiple patient databases exist. Healthcare applications and middleware use the MPI to match patients between the databases, and to store new patient details as they are encountered. MPIs are highly specialized applications, often tailored extensively to the institution’s particular mix of patients. MPIs can also be run on a regional and national basis.

Beyond healthcare, eIDAS provides the framework for identification in a two-level structure that involves a panEuropean and a national framework. Open National Contact Points (NCPs) for eHealth are key to the implementation of cross-border health services under the Connected Europe Facility. The HealtheID project created a proof of concept eHealth connector for eIDAS which led to recommendations to the eHealth Network the body of Member state representatives responsible for cross-border health care in the EU [7].

4.2.3. **Authentication & Authorization Module (Auth Module)**

This module should provide standards-based authentication and authorization mechanism for all clients of the FAIR4Health system. The following standards and profiles can be used during the design of this module:

- Authentication and authorization should follow well-established international standards. **OpenID Connect**\(^5\) is a simple identity layer on top of the OAuth 2.0\(^6\) protocol and seems to be a good candidate to be integrated in this module. It allows clients to verify the identity of the end-user based on the authentication performed.

---

\(^2\) https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/tutorials/oid.pdf

\(^3\) https://wiki.ihe.net/index.php/Patient_Identifier_Cross-Referencing

\(^4\) https://www.hl7.org/fhir/operation-patient-match.html

\(^5\) https://openid.net/connect/

\(^6\) https://oauth.net/2/
by an Authorization Server, as well as to obtain basic profile information about the End-User in an interoperable and REST-like manner. This functionality should interact with the Identity Management Module for the user/client profiles.

- **SMART Health IT** is an open, standards-based technology platform that enables innovators to create apps that seamlessly and securely run across the healthcare system. Using an electronic health record (EHR) system or data warehouse that supports the SMART standard, patients, doctors, and healthcare practitioners can draw on this library of apps to improve clinical care, research, and public health.

- **SMART on FHIR** is a set of open specifications to integrate apps with Electronic Health Records, portals, Health Information Exchanges, and other Health IT systems. This set of specifications exposes an authorization mechanism named as SMART App Authorization Guide to authorize the applications within the ecosystem. This specification is lately published by HL7 itself with the name of SMART App Launch Framework.

### 4.2.4. Audit Module

Logging requirement can be considered in two different levels. One for the **software logging** that needs to be done in different levels within the source code to track the activities with info messages, inform the associated parties with error messages and debug the system when necessary. On the other hand, on the application integration layer, each **transaction** (i.e. data access, query request) is very important and needs to be logged explicitly for further user inspection. The following standards and profiles can be used while designing this module:

- **IHE’s The Audit Trail and Node Authentication (ATNA) Integration Profile** establishes security measures which, together with the Security Policy and Procedures, provide patient information confidentiality, data integrity and user accountability. ATNA contributes to **access control** by limiting network access between nodes and limiting access to each node to authorized users. This profile not only serves to the logging mechanism, it also provides facilities for user and connection authentication. The **logging facility** is provided by the Audit Trail. The Audit Trail needs to allow a security officer in an institution to audit activities, to assess compliance with a domain’s security policies, to detect instances of non-compliant behaviour, and to facilitate detection of improper creation, access, modification and deletion of Protected Health Information (PHI).

- **HL7 FHIR (Audit Event)** provides a means to maintain security logs. FHIR repository to be used as the audit trail repository, the audit messages are designed to be FHIR resources with this standard. A record of an event made for purposes of

---

7 [https://smarthealthit.org/](https://smarthealthit.org/)
8 [http://docs.smarthealthit.org/](http://docs.smarthealthit.org/)
10 [https://wiki.ihe.net/index.php/Audit_Trail_and_Node_Authentication](https://wiki.ihe.net/index.php/Audit_Trail_and_Node_Authentication)
11 [https://www.hl7.org/fhir/auditevent.html](https://www.hl7.org/fhir/auditevent.html)
maintaining a security log. Typical uses include detection of intrusion attempts and monitoring for inappropriate usage.

4.2.5. PPDDM Security Module

Existing PPDDM approaches can be categorized under two concepts when the security is in question:

❖ De-identified data leaves the data source so that the PPDDM engine on the platform can process whole data for predictive modelling. **Only software components access data.** Once the processing completes, data is deleted permanently. Communication is performed through end-to-end encryption.

❖ Encrypted data leaves the data source. PPDDM engine processes the encrypted data and tries to build the models by using the encrypted data. No client can decrypt the data except the data owner. This is also called **homomorphic encryption** in literature. This approach is rather new in theory and no practical implementation exists today for real world use.

**Blockchain** concept can be analysed as a candidate to be utilized during the design and implementation of the PPDDM security module. **Smart contracts** can reflect and link the agreements made among the peers to the transactions occurring through PPDDM.

4.2.6. End-to-end Security Module

**Asymmetric encryption** algorithms should be used to maximize the level of encryption of the transmissions among components. On the other hand, FAIR4Health is about sharing data between computer systems. Hence, respective developments in the secure data sharing arena should be explored. The following standards and profiles can be used while designing this module:

❖ Public-key cryptography or **asymmetric cryptography**, is a cryptographic system that uses pairs of keys: public keys which may be disseminated widely, and private keys which are known only to the owner. In such a system, any client can encrypt a message using the receiver’s public key, but that encrypted message can only be decrypted with the receiver’s private key.

❖ **Dat**\(^{12}\) is a secure peer-to-peer protocol for sharing data between computers. Dat can be considered as a candidate to be used as the base protocol while sharing data between remote FAIR4Health components.

4.3. Utilization of the Cybersecurity modules within the scenarios

This section provides the placement of the cybersecurity modules in each identified scenario. The modules should provide the identified security functionalities:

\(^{12}\) [https://datproject.org/](https://datproject.org/)
Improving Health Research in EU through FAIR Data

- **Identity Management Module**
  - Identities of all clients are managed.
  - Trust is ensured.
- **Authentication (Auth) Module**
  - Clients are authenticated (in line with the identity management)
  - Authorization is handled through roles, scopes and authorization rules
- **Audit Module**
  - All transactions are audited into an audit repository.
- **End-to-end Security Module**
  - All transactions are encrypted using asymmetric encryption algorithms.
- **PPDDM Security Module**
  - Based on the selected data mining method, this module should ensure that only allowed components process the FAIR data.

4.3.1.1. Scenario 1: Local access to raw data

![Security Modules attached to Scenario 1.](image)

- Clients are authenticated (in line with identity management)
- Authorization is handled through roles, scopes and authorization rules
- All transactions are encrypted.

**Figure 6.** Security Modules attached to Scenario 1.
4.3.1.2. Scenario 2: Local access to FAIR data

- All transactions are audited.
- All transactions are encrypted.
- Identities of all clients are managed.
- Trust is ensured.
- Clients are authenticated (in line with identity management)
- Authorization is handled through roles, scopes and authorization

**Figure 7.** Security Modules attached to Scenario 2.

4.3.1.3. Scenario 3: PPDDM access to FAIR data

- All transactions are audited.
- All transactions are encrypted.
- Identities of all clients are managed.
- Trust is ensured.
- Clients are authenticated (in line with identity management)
- Authorization is handled through roles, scopes and authorization

**Figure 8.** Security Modules attached to Scenario 3.
4.3.1.4. **Scenario 4: P2P access to FAIR data**

![Diagram of Security Modules attached to Scenario 4](image)

- All transactions are audited.
- All transactions are encrypted.
- Identities of all clients are managed.
- Trust is ensured.

**Figure 9:** Security Modules attached to Scenario 4.

4.3.1.5. **Scenario 5: Access to eHealth services**

![Diagram of Security Modules attached to Scenario 5](image)

- All transactions are audited.
- All transactions are encrypted.
- Identities of all clients are managed.
- Trust is ensured.

**Figure 10.** Security Modules attached to Scenario 5.
5. Summary: Overcoming Strategies for Technical Barriers

This document builds upon other deliverables in FAIR4Health project, such as D2.2 and D2.3, to present technical barriers identified for FAIR data policy implementation in EU Health research as well as security requirements and proposed overcoming strategies. The proposed overcoming strategies presented below will provide input to WP3 and WP4 which will design and implement the FAIR4Health platform and agents.

Proposed strategies for building trustworthiness in the use of FAIR data for health research:
- To encourage public institutions to deploy FAIR demonstrators with health data
- To reach a harmonization of FAIR metrics
- To implement sound data provenance methods
- To boost the availability of trusted data repositories
- To develop a FAIR code of conduct
- To encourage the development of a sustainability plan beyond the research project
- To engage software providers to make data FAIR

Proposed overcoming strategies for health data management life cycle barrier:
- To raise awareness and provide training on the use and management of metadata
- To emphasize the use of community-based standards and ontologies

Proposed strategies to advance interoperability:
- To set-up a best practice guide for the use of health information exchange standards
- To address alignment and harmonization of metadata in HL7 FHIR IGs.
- To democratize tools for mapping (meta) data with related standards and ontologies

6. References


