FAIR 4 Health

D2.3. Guidelines for implementing FAIR open data policy in health research

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824666

European Commission

HORIZON 2020

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824666

Document Information

GRANT AGREEMENT NUMBER 824666			ACRONYM: FAIR4Health	
Full title	Improving Health Research in EU through FAIR Data			
Horizon 2020 Call	SwafS-04-2018: Encouraging the re-use of research data generated by publicly funded research projects			
Type of action	Research and Innovation Action			
Start Date	1st December 2018	**Duration**	36 months	
Website	www.fair4health.eu			
Project Officer	Raluca Iagher			
Project Coordinator	Carlos Luis Parra Calderón, Andalusian Health Service (P1–SAS)			
Deliverable	D2.3. Guidelines for implementing FAIR Open Data policy in health research			
Work Package	WP2. Comprehensive Analysis for FAIR Data Policy Implementation in Health Research			
Date of delivery	**Contractual**	M06	**Actual**	May 2019
Nature	Report	**Dissemination Level**	Public	
Lead beneficiary	P3-UC3M			
Responsible Authors	Tony Hernández-Pérez, Eva Méndez Rodríguez			
e-mail	tony@bib.uc3m.es	**Phone number**	+34916249252	
Reviewers	Francisco J. Núñez-Benjumea (SAS), Catherine Chronaki (HL7), Devika Madalli (External Scientific Advisory Board)			
Keywords	Guidelines, FAIRdata, ethical issues, legal issues, cultural issues			

Document history

Version	Issue Date	Stage	Changes	Contributor(s)
0.1	24/04/2019	Draft	Scope and outline	E. Méndez (UC3M)
0.2	25/05/19	Draft	Final draft for review	T. Hernández (UC3M), E. Méndez (UC3M)
0.3	27/05/19	Draft	Review	F. Núñez (SAS), C. Chronaki (HL7), D. Madalli (ESAB)
1.0	31/05/19	Final	Final document	T. Hernández (UC3M), E. Méndez (UC3M)

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824666

Table of Contents

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824666

List of figures

List of tables

List of acronyms

API	Application Programming Interface
BoF	Birds of a Feather
CDR	Clinical Data Repositories
CIM	Clinical Information Models
CSQ	Client Satisfaction Questionnaire
CTS	CoreTrustSeal
DOI	Digital Object Identifier
DMP	Data Management Plan
DPO	Data Protection Officer
EC	European Commission
EHR	Electronic Health Record
EOSC	European Open Science Cloud

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824666

EU	European Union
EUOSPP	European Open Science Policy Platform
FAIR	Findable, Accessible, Interoperable and Reusable
FDP	FAIR Data Point
FDP-API	FAIR Data Point API
FHIR	Fast Healthcare Interoperability Resources
GDPR	General Data Protection Regulation
HLEG	High Level Expert Group
HRPO	Health Research Performing Organization
ICD	International Classification of Disease
ICU	Intensive Care Unit
IEC	International Electrotechnical Commission
IG	Interest Group
IFDS	Internet of FAIR Data & Services
INB-ISCIII	Spanish National Bioinformatics Institute
ISO	International Organization for Standardization
LOINC	Logical Observation Identifier Names and Codes
MDR	Meta Data Repository
OSF	Open Science Framework
ORD	Open Research Data
OSSE	Open Source Registry for Rare Diseases
PCI	Practical Commitment for Implementation
PDCA	Plan, Do, Check and Act
PDF	Portable Document Format
PE	Public Engagement
PHT	Personal Health Train
PID	Persistent IDentifier
PPDDM	Privacy-preserving Distributed Data Mining
PSI	Public Sector Information
SNOMED CT	Systematized Nomenclature of Medicine – Clinical Terms
SRDC	Software Research and Development Consultancy
RDA	Research Data Alliance

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824666

RDF	Resource Description Framework
RDM	Research Data Management
RFO	Research Funding Organization
ROO	Radiation Oncology Ontology
RPO	Research Performing Organization
SO	Specific Objective
SUS	System Usability Scale
TAB	Technical Advisory Board
WG	Working Group
WHO	World Health Organization
WP	Work Package

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824666

1. Executive summary

This deliverable gathers an **analytical overview** of all the considerations addressed through WP2 - to identify, report and overcome all the barriers that could prevent Health Research Performing Organizations (HRPOs) from opening, sharing and FAIRifying their research data. The document describes:

- The scenario and the results of the comprehensive analysis (section 2).
- The methodology established to tackle the analysis (section 3);
- The results of the complete study in the form of key elements to determine a FAIR/Open Data Policy in Health research (section 4);
- A set of guidelines addressed to Health Research Performing Organizations (section 5);
- The mechanisms and further discussion to convert those guidelines in a primer for HRPO policies on research data management in Health research (section 6).

2. Introduction

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

In the framework of the FAIR4Health project, the main objective of WP2 "Comprehensive analysis for FAIR data policy implementation in health research" was to elucidate 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 Health Research Performing Organisations (HRPOs) to implement a FAIR data policy.

This deliverable D2.3 "Guidelines for implementing FAIR data policy in health research" is part of the WP2 and represents a first draft of these Guidelines 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 are:

- The set of **technical recommendations** (D2.1) that would facilitate the implementation of these processes and functionalities considered in the FAIRification workflow.
- 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.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824666

2.2. Objective and target audience of this document

The guidelines aim at serving as a primer for all those Research Performing Organizations (**RPOs**) dealing with **health research data** and willing to: a) Manage, curate and disseminate their dataset as a principal research outcome. b) Fulfill the Open Science requirements of the public funders granting their research projects. c) Convert their research data to become **FAIR** (Findable, Accessible, Interoperable and Reusable).

This document reflects all the analyses done under WP2 to come up with a first draft of the guidelines (recommendations, instructions, suggestion and concrete advice) for Health and Medical Research Institutions and their managers. They are **the starting point** to open the discussion and further support in the Research Data Alliance (RDA) context. If successful, the guidelines will be an RDA Recommendation, to help health research institutions to state their own **policies in data management,** aligned with FAIR data principles and with funder's policies, but taking into account all the legal, ethical, technical and cultural issues that FAIR4Health have investigated.

2.3. FAIR DATA does not equal to Open Data

The preamble to the proposal for the new *EU Directive on Open Data and the Re-use of Public Sector Information* (**PSI directive**, already approved by the European Parliament), states that open data "as a concept is generally understood to denote data in an open format that can be freely used, re-used and shared by anyone for any purpose". It also encourages Member States to promote the creation of data based on the principle of "open by design and by default" ensuring the protection of personal data, including where information in an individual dataset may not present a risk of identifying a natural person, but when that information were to be combined with other available information, it could entail such a risk.

When talking about Open Data people mainly refer to Open Government Data, Open Data produced by the public sector. But there is also Open Data produced and released by private companies. Netflix, RTVE, Uber, Lyft, BBVA and more companies have opened part of their data. For example, Netflix opened a part of its data for a competition wherein the participants could create the best algorithm for suggesting films to Neflix viewers. Lyft and Uber have opened a part of their data to allow studying the effects of transport companies on traffic. But, besides such data, we must consider also the data produced by public or private companies while conducting their research. It may not be possible for these research data to always be open for legitimate commercial interests or for security or privacy reasons, but they can and should always be FAIR.

The new **EU PSI directive** defines Research data as "documents in a digital form, other than scientific publications, which are collected or produced in the course of scientific research activities and are used as evidence in the research process, or are commonly accepted in the research community as necessary to validate research findings and results". It includes statistics, results of experiments, measurements, observations resulting from fieldwork, survey results, interview recordings and images. It also includes metadata,

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824666

specifications and other digital objects, like software code or research notes. For the first time, research data have a dedicated article (Art. 10) on the new PSI directive:

> **Article 10. Research data**: "1. Member States shall support the availability of research data by adopting national policies and relevant actions aiming at making publicly funded research data openly available ('open access policies'), following the principle of "open by default" and compatible with the FAIR principles. In that context, concerns relating to intellectual property rights, personal data protection and confidentiality, security and legitimate commercial interests, shall be taken into account in accordance with the principle of "as open as possible, as closed as necessary". Those open access policies shall be addressed to research performing organisations and research funding organisations".
>
> "2. Without prejudice to point (c) of Article 1(2), research data shall be re-usable for commercial or non-commercial purposes in accordance with Chapters III and IV, insofar as they are publicly funded and researchers, research performing organisations or research funding organisations have already made them publicly available through an institutional or subject-based repository. In that context, legitimate commercial interests, knowledge transfer activities and pre-existing intellectual property rights shall be taken into account".

Not all open data, specially research data, are or must be completely "open" or "free", but it should at least be **FAIR**. That is what is implicitly recognized in the wording of the article 10.1 when it mentions "as open as possible, as closed as necessary". There are two main types of reasons for keep research data "closed":

1. Legal and ethical reasons (Cf. section 4.2):
    a. **Privacy**: research involving human beings, especially in the health sector but also in sociologist studies, may include data that could permit the identification of individual subjects. That is a legitimate reason not to make the data completely open. But it doesn't mean that research data could not be FAIR. It may be Findable, Accessible, Interoperable and Reusable — findable through metadata and accessible and reusable if some requirements imposed by the data owner are met (e.g. asking for permission or getting consent). What is needed to be FAIR is that the dataset can be findable, that once found the researcher may be able to study that data without restrictions, how it was generated and by whom (provenance), under what conditions it could be accessible and what type of reuse is permitted, all under a technical and semantically interoperable environment, in order to computational agents may be capable of interoperate.
    b. **Confidentiality, Intellectual Property and Security**: Confidentiality, established as an agreement between the researcher and the research subjects, is about how the gathered data will be managed and stored, and to who or under what conditions will be made accessible. Research data may affect the reputation of a brand, of individuals or simply that data should not be used out of certain contexts. Sometimes research data may be part of a collection of documents with copyright (secondary data) so reuse of them for a new analysis may require a new consent or agreement. Or simply, it is related to legitimate commercial or national security interests. We must not think only in terms of defense or public security but also in the protection of population groups that may be exposed to marginalization due to health reasons, as may be the case with the protection of environmentally sensitive places or cultural sites.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824666

2. Economic reasons
    a. **Costs**: research data are often too big to be curated, stored or easily transmitted. The storage and transmission of such volume of data may be too expensive. Although the new PSI directive urges that access to open data must be free of charge to avoid entry barriers to markets, especially to SMEs, it also recognizes that institutions can charge marginal costs. It is not feasible to offer all the clinical data that a hospital can have and to afford the costs (time, human resources, etc.), not to mention privacy issues. Again, we can find cases where research data is not completely open but FAIR. "FAIR only speaks to the need to describe a process – mechanized or manual – for accessing discovered data; a requirement to openly and richly describe the context within which those data were generated, to enable evaluation of its utility; to explicitly define the conditions under which they may be reused; and to provide clear instructions on how they should be cited when reused. None of these principles need data being "open" or "free" [1].
    b. **Disincentives for researchers:** for researchers, especially early career researchers, it is needed to take into account that in some fields the process of collecting data could be a hard and complicated process (move to a remote place to get the data, long time for convincing someone for being interviewed, etc.). Gathering research data is not always made at developed countries and many times data (e.g. Ebola crisis) is far away from the desk. Forcing researchers to always open their research data immediately may disincentive them from collecting good data. They may feel there their hard work may not be available to benefit them first and worse, someone else may make use of their data. To avoid these "data parasites," as they are named, it may be reasonable to set an embargo period on the data in such circumstances [2].
    c. **Disincentives for private companies collaborating with public sector.** A similar situation exists when private companies collaborate with the public sector. While the aim of such public-private collaborations is to promote new research, releasing research data immediately could discourage the private partners as they may feel that they would lose advantage over their competitors. This is especially true in the health sector where many projects are funded not only government funded or funded by charities but also by the pharmaceutical industry.

As said in the report *Turning FAIR data into reality*: "Data can be FAIR or Open, both or neither. The greatest benefits come when data are both FAIR and Open, as the lack of restrictions supports the widest possible reuse, and reuse at scale" [3]. The FAIR Guiding Principles express it also very clearly when talking about accessibility: "Once the user finds the required data, she/he needs to know how they can be accessed, possibly including authentication and authorization". Even more, research data with FAIR properties (after a FAIRification process) can be accessible, interoperable and reusable even through computational agents.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824666

2.4. FAIR data principles and Health Research in the context of Open Science

The **FAIR Data Principles** were first published in 2016 [4]. FAIR seeks the reuse of data and other digital research output and objects (algorithms, tools and workflows that led to that data) making them Findable, Accessible, Interoperable, and Reusable. The principles consider applications and computational agents as stakeholders with the capacity to find, access, interoperate and reuse data with none or minimal human intervention and recognize the importance of automated process to do that because humans increasingly rely on computational support to deal with intensive data processes.

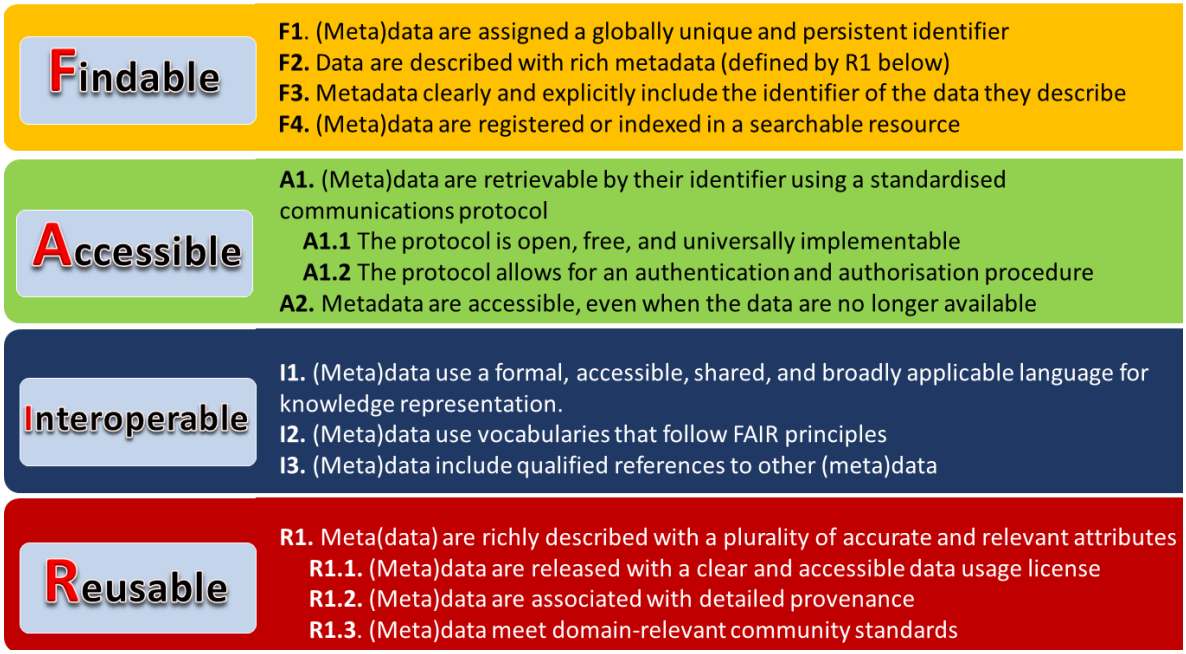


Figure 1: FAIR Data Principles. Summary

The first draft of the FAIR Data Principles was born in January 2014 at the Lorentz Center in Leiden, Netherlands, by a community of scholars, librarians, archivists, publishers and research funders as a part of FORCE11[1]. Since then, a lot of organizations have adopted the FAIR principles. As early as july 2016 the European Union published the Guidelines on FAIR Data Management in Horizon 2020[2]. The principles are also explicitly mentioned in the new Open data and reusable PSI directive, and the European Open Science Cloud focuses on enabling FAIR data and principles. In the United States the National Institutes of Health[3], also support the FAIR principles and it can be said that most important research funding agencies and international organizations support or have adopted these principles.

FAIR data is **necessary for Open Science**. And for a real Open Science scholarly articles and research objects associated (data, software, workflows, algorithms, etc.) should be

[1] https://www.force11.org/
[2] http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf
[3] https://commonfund.nih.gov/data

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824666

open and available without barriers. FAIR may not be open, but open has to be FAIR. As recognized in a recent study from National Academies of Sciences, Engineering, and Medicine:

> "Most data in repositories today are not available under FAIR principles, and the complexities of realizing this will entail significant costs. Making data FAIR is a difficult task for investigators, and substantial public investment is going to be required to change the current situation. Making data "findable" is going to require better standards for metadata; new ontologies for the vast majority of scientific disciplines, which currently lack standardized, granular terms that can be used by data search engines; and new tools to enable investigators and curators to author scientific metadata that are sufficiently comprehensive and standardized so that search engines can locate appropriate datasets with adequate precision and recall"[5].

An Open Science ecosystem ideally strives for researchers and citizens to have **immediate access to published articles and data**, software and other research products under FAIR principles, ideally without cost and with the possibility of reusing everything as deemed convenient. For a real open science there are some barriers that still remain to be brought down. Sharing data, code and other research objects is becoming common but not in all disciplines.

Data sharing and data **stewardship** practices are not uncommon in **health research**, but they are very inefficient. The 2017 *Data Scientist Report*, from Figure Eight[4], claims that a data scientist spends around 53% of their time collecting, labeling, cleaning and organizing data. Others raise that percentage to 80% [6]. This is a big waste of research time caused by many reasons: lack of standards for metadata creation, lack of common vocabularies and ontologies, lack of trust in data integrity, or privacy concerns. This lack of good data management and sharing practices has brought a crisis of reproducibility [7]. Only recently journals and funders have begun to ask for the data that underpin the results of a research.

FAIR Data Principles have had a wide and excellent acceptance by **funding agencies, publishers and scientific communities** for two main reasons: a) Because its intention of to make all scholarly output should be Findable, Accessible, Interoperable, and Reusable means to go one step further than merely sharing data and it implies good data management. b) Because the principles emphasize that processes to find, access, interoperate and reuse must be automated and carried out by machines and software agents. This makes it credible and approachable in this era of data deluge.

[4] https://www.figure-eight.com/download-2017-data-scientist-report/

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824666

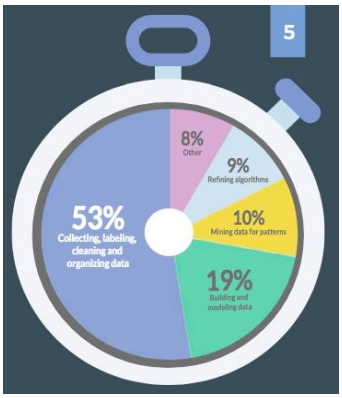


Figure 2: For Data Scientist. What activity takes up most of your time (Data Scientist Report: 2017)

Figure 3: Roadmap for a better data management (Elsevier)

The drivers of the first definition of the FAIR principles came largely from the field of life sciences, mostly biotechnology and, therefore, with a strong relationship with the world of technologies. The enormous attention and attractiveness of the FAIR principles are enabling the development of an **Internet of FAIR Data & Services (IFDS)** with the main focus on early developments in the European Open Science Cloud (EOSC), led by the GO FAIR[5] initiative.

In the field of health, beyond the articles that call for the awareness of the FAIR principles, the design of policies to achieve FAIR datasets and other scholarly objects (codes, algorithms, images, etc.) or metrics of Fairness, much of the research is focused on infrastructure and semantic interoperability. For example, in The Netherlands there is a project "exploring the relationship between the development of diabetes and socio-economic factors such as lifestyle and health care utilization" [8]. It means to extract information using access-restricted data from different sources and institutions. The aim of the technical part of the project is to develop a computational framework to facilitate access, reuse, and combining data from at least two different national agencies in a secure environment, essentially by addressing the FAIR principles. They develop an infrastructure based on the **Personal Health Train** (PHT).

PHT is probably the most innovative concept and implementation of health data initiatives following FAIR principles. The main goal of PHT is to provide a general-purpose infrastructure where many different questions can be asked of multiple data owners such as the hospitals or even the patients themselves. Some of the advantages of PHT are that it eliminates the need to make multiple copies of the data. "The PHT architecture comprises algorithm 'trains' that visit 'stations', which check algorithmic processing credentials and provide access to data it is authorized to release" [8]. A key concept in PHT is to **bring algorithms to the data** where they happen to be, **rather than bringing all data to a central place**. PHT is designed to give controlled access to heterogeneous data sources while ensuring privacy protection and maximum engagement of individual patients and citizens. As a prerequisite, health data are made FAIR (Findable, Accessible, Interoperable

[5] https://www.go-fair.org/

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824666

and Reusable). Stations containing FAIR data may be controlled by individuals (general) physicians, biobanks, hospitals and public or private data repositories [9].

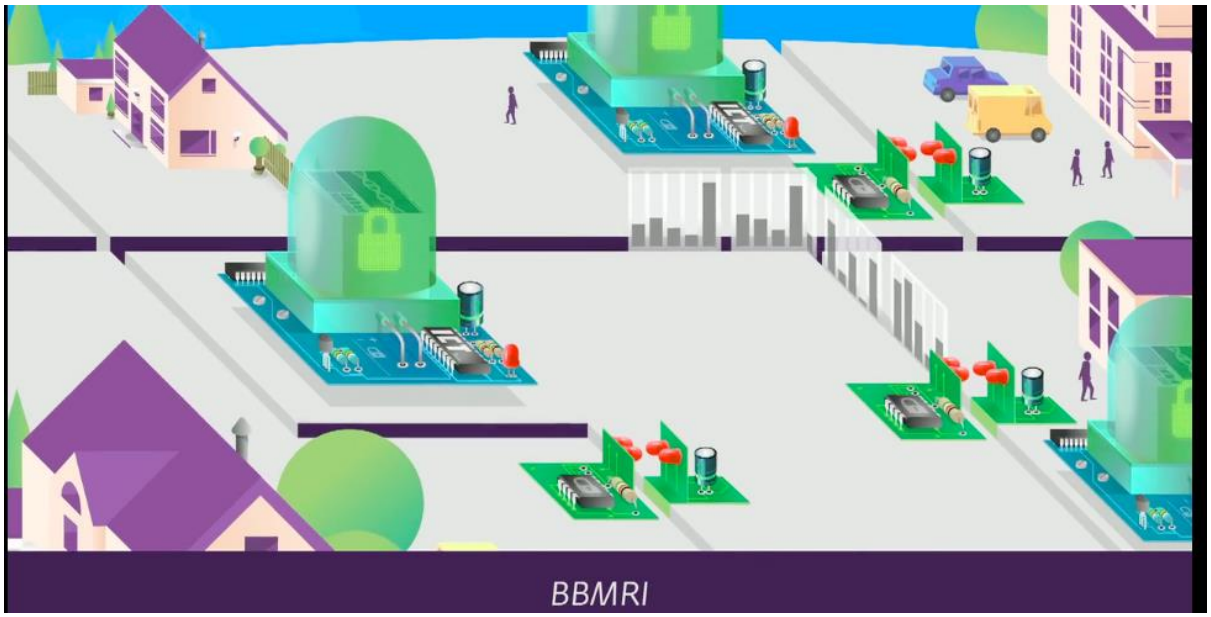


Figure 4: Personal Health Train video. (Source: https://vimeo.com/143245835)

The **Open Source Registry for Rare Diseases** (OSSE) has built a first prototype implementing FAIR Data Principles by adding a FAIR Data Point (FDP) to his architecture. The OSSE is basically software for the management of registries for patients with rare diseases with a metadata repository (MDR) that contains data elements. As said by Schaff et al. "the main goal of the FDP is to provide meta information about the data in the registry. Which means that only the descriptions of the medical data elements are available, no patient-identifying and medical data is provided. For metadata description the FAIR Data Point API (FDP-API) is the most important interface for sharing this information with other registry systems". [10]

Although Mons [1] warns that "FAIR is not equal to RDF, Linked Data, or the Semantic Web [...] and FAIR Principles explicitly do not prescribe the use of RDF or any other Semantic Web framework or technology", the reality is that some of the most relevant advances in the field of health are occurring in or are related to these technologies. Indeed, the biopharmaceutical industry perceives as a technical barrier to the implementation of FAIR principles the lack of agreement for the representation of data in a common way and the agreement on standards, for example, ontologies [11].

Researchers from Maastricht University Medical Centre [12] have developed a radiation oncology ontology (ROO) from a clinical database of 80 oncological patients with diagnosed rectal cancer from a trial. The database contains information from different sources, combining demographic and clinical outcomes. The ROO uses entities from other ontologies and tools like the International Classification of Disease (ICD) and contains 1,183 classes covering almost all concepts in radiation oncology (cancer diseases, treatments, etc.). This is a first step toward transforming different clinical databases into FAIR and linked data.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824666

3. Methodology

The purpose of this section is to describe the methods and procedures carried out to achieve the aim of this deliverable: identify the elements that might be a part of consistent guidelines for implementing FAIR open data policy in health research. The methodology followed in this deliverable gathers several approaches that we might describe in two main **phases** that encompass **different methodological approaches**:

Phase 1: **Literature review**. FAIR4Health performed a comprehensive analysis of both scholarly publications and web-based publications about FAIR in general and specifically in health data management, including the current legal documents (GDPR, national regulation from Italy, Serbia, Switzerland and Spain) and key reports. The analysis resulted in an overview of the current status of the research on FAIR data, and specifically identified studies dealing with research data in the health domain. For the purpose of this review, two key studies have been chosen, one in public health [13], and other from the biomedical perspective [14]. This is in addition to other papers and reports featuring barriers related to a broader array of subjects and practical challenges (legal issues, ethical approaches, technical features, etc.), as well as those addressing a wider number of researchers and disciplines [15,16]. (See complete list of references).

Other key documents and resources taken into account for this deliverable were:

- *Turning FAIR into reality* (2018)[6], the report of the EC High Level Expert Group on FAIR Data, pointing out the 27 recommendations to implement FAIR data, where some of them are targeting institutions and research communities.
- European and National regulations related with personal data protection, like the GDPR[7] or the regulation concerning Italy, Spain, Serbia and Switzerland.
- The reflections around the new PSI Directive[8] that agrees to make publicly-funded research data open by default.
- The information resources from the Research Data Alliance[9] (RDA) particularly its recommendations and the work done by the Health Data Interest Group (IG).
- The report of *Global diffusion of eHealth*[10] published by the World Health Organization (WHO) that shows the growing interaction of patients and citizens with Digital health systems, where patients are shown as Health research data users.

Phase 2: **Asking stakeholders and pop-up research.** In this phase we practiced direct observation of stakeholders and users. It included three main methodologies: qualitative short interview with experts recorded (Korsakow film), general open surveys and focus groups.

[6] https://ec.europa.eu/info/sites/info/files/turning_fair_into_reality_1.pdf
[7] https://eugdpr.org/
[8] https://oeil.secure.europarl.europa.eu/oeil/popups/ficheprocedure.do?reference=2018/0111(COD)
[9] https://rd-alliance.org/
[10] https://apps.who.int/iris/bitstream/handle/10665/252529/9789241511780-eng.pdf%3Bjsessionid=D42FEE119D337087F31396649A9B36DB?sequence=1

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824666

a) **Korsakow film.** The creation of this video-material, known as Korsakow[11] has different roles in FAIR4Health project: 1) As part of the outreach strategy to create audiovisual content and raise awareness around the problems tackled by FAIR4Health. 2) As a mean of engaging experts in our project and involving them in FAIR4Health. 3) As part of the methodology (pop-up research) of WP2. Pop-up research helps us to intercept intended people (stakeholders in health sciences research, in our case) in context for short interviews.

   In this Korsakow film we recorded experts during the 13th RDA plenary held in Philadelphia in April 2019. That way we guaranteed that the experts were in the right context to think about FAIR data. The **interviewees** were:

   - **Leslie McIntosh**. CEO of Ripeta. Executive Director of RDA US.
   - **Oya Beyan**. Biomedical Informatics. RWTHaachen University. FAIRplus project
   - **Kevin Ashley**. Director of the Digital Curation Center.
   - **Jane Greenberg**. Professor Drexel University. Metadata Research Center.
   - **Ville Tenhunen**. University of Helsinki. University Library.
   - **John Graybeal**. Technical Program Manager at Stanford University's School of Medicine.
   - **Angela Murillo**. Indiana University.
   - **Edit Herczog**. Vision & Values / RDA Council.
   - **Devika Madalli.** Indian Statistical Institute. RDA Technical Advisory Board (TAB)

   In addition, we also interview two Drexel University students — Kelin Baldridge, and Yuvraj Sharma — as they are potential researchers in the field of data management.

   The **questions** posed to the participants were:

   - Enumerate the main problems you think health researchers face regarding research data.
   - Name the main social, technical and cultural barriers that prevent progress in the FAIRification of health data.
   - Can you identify the benefits of FAIRfying health research data? Can you identify the benefits of OPENing health research data?

For each question we generated an individual film where the videos are selected randomly making a new film each time. The FAIR4Health first Korsakow films can be accessed here: http://galan.uc3m.es:7088/korsakow_uc3m/login.jsp (password: F4HUC3M). The videos will be also available in the FAIR4Health Youtube channel and linked and integrated from FAIR4Health website and disclosed through social networks (Cfr. D7.3).

b) **Open surveys** and data collection. In order to gather stakeholders and public at large views, we launched three open surveys (Figure 5), adapted to an online format using the Google Forms® tool:

[11] http://korsakow.tv/formats/korsakow-film/

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824666

1. *Open survey on ethical implications of reusing FAIR data for health research*.[12] The purpose of this survey was to gather feedback from all involved stakeholders regarding their perceived importance of putting in place mechanisms to guarantee the compliance of Health Research Performing Organizations (HRPOs) with the ethical considerations.

2. *Open Survey on Boosting Citizen Science in Health Research*[13], targeting in this case, citizens at large, but also patient associations since they bring together citizens motivated to learning and solving shared health problems. The purpose of this survey was to collect feedback from citizens about the suitability of several Public Engagement (PE) strategies to be applied for boosting citizen science in EU health research.

These two surveys (ethical implications, citizen science) were released in all the languages of the consortium: English, Spanish, French, Italian, Portuguese, German, Dutch, Serbian and Turkish.

3. The third survey was the *Open Survey on cultural barriers using FAIR data for health research*[14]. The purpose of this survey was to gather opinions and attitudes about sharing of health data among health research practitioners and senior researchers. In this case, the survey was only published in English, and we used snowball sampling, staring for 18 key institutions, projects and people, that received the survey and they pass it to their colleagues. Similar surveys have taken place over the last years. One of the biggest has been the study of Springer-Nature with 7700 respondents [17]. Some of the main barriers highlighted in the study were organizing data in a presentable and useful way (46%); concerns about copyright and licensing (37%); not knowing which repository to use (33%); lack of time to deposit data (26%) and costs of sharing data (19%).

The results of all these surveys are discussed under section 4 and have also impacted deliverable 2.2 to identify functional requirements based on the perspective of the researchers.

[12] https://osf.io/sczpd/
[13] https://osf.io/czbmj/
[14] https://osf.io/vb6sk/

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824666

Figure 5: QR codes pointing to the 3 Open surveys.

c) Focus groups. A focus group is a qualitative methodology that uses a small group of people whose reactions to new challenges or political endeavors are studied. This methodology was used to explore ethical issues as well as technological barriers to sharing data in the context of Health research. It was also the initial methodology to explore among stakeholders. After the survey on ethical issues and technological barriers, two focus groups were convened.

In the focus group on ethics, the following members of the **External Ethics Advisory Board** of the FAIR4health project participated:

- **Ricard Martínez:** Assistant Professor of Constitutional Law at the University of Valencia. Director of an Institutional Chair about privacy and data protection. Dr. Martinez works at the Spanish Data Protection Agency and is the former president of the Spanish Data Professional Association, and a Data Protection Officer (DPO) of BigMedylitics, a Horizon2020 project.
- **Carolin Reuter:** Medical Doctor and PhD candidate at the department of Medical Ethics and History of Medicine at the University of Göttingen, Germany.

In the focus group on technological barriers, the following members of the FAIR4Health consortium and from the **External Scientific Advisory Board participated**:

- **Alfonso Valencia:** Life Sciences Department Director at the Barcelona Supercomputing Center (BSC), Formerly Director of the Spanish National Bioinformatics Institute (INB-ISCIII).
- **Anil Sinaci:** Senior researcher and project manager at SRDC (Software Research and Development Consultancy).
- **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.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824666

- **Marcos Martínez:** Senior research software engineer, MED/BMIR at the Stanford Center for Biomedical Informatics Research.

d) We also organized a **BoF (Birds of a Feather)** at the **13th RDA Plenary meeting** to indirectly solicit the opinions of experts and stakeholders. A BoF is a typical modality in RDA plenaries, conceived as an informal meeting where the attendees gather together in groups based on a shared interest and carry out discussions to share ideas and approaches around a topic of common interest. In our case the accepted RDA BoF was: *Assessing FAIR Data Policy Implementation in Health Research*[15] where we initially discussed the landscape, particularly in terms of ethical challenges and cultural and behavioral barriers for FAIR open data policy implementation in general and in the health research domain in particular. We discussed questions of the korsakow film openly and we identified potential international collaborators pursuing similar activities so the European work may be set in a broader global context (See Collaborative Notes[16]). We agreed with the group to share the FAIR4Health Guidelines included here with the RDA Health Data IG and create a specific working group (WG) to refine and adopt them by a broader community (see section 6 below).

4. Elements of a FAIR Open Data policy in Health Research (Guidelines)

4.1. General considerations

Research **Data Management practices** such as the creation of Data Management Plans (DMP), making datasets openly available, the deposition of data in repositories, and the application of FAIR data principles to research outcomes are becoming increasingly common as they are required by funder mandates such as in Horizon 2020 and Horizon Europe, as well as because of legislative push factors such as the newly revised EU Directive on Public Sector Information (PSI).

Over the last years, under the umbrella of **Open Science agenda for Europe**, different communities, stakeholders and institutions have been producing their guidelines, recommendations or other kind of documents tackling research data management (including FAIR data and/or Open Data approaches). For example:

- *Practical Guide to the International alignment of Research Data Management alignment*[17] addressed to Research Funding Organizations (RFOs), launched by Science Europa in January 2019.

[15] https://rd-alliance.org/bof-assessing-fair-data-policy-implementation-health-research-rda-13th-plenary-meeting
[16] https://docs.google.com/document/u/1/d/1HtjHcWzknrmtD2ck99tLXwTgtq3edrK1Gv2NqQeulJE/edit
[17] https://www.scienceeurope.org/wp-content/uploads/2018/12/SE_RDM_Practical_Guide_Final.pdf

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824666

- *Recommendations on managing research data addressed to Researchers*[18], by Maredata Spanish Research network of Open Research Data[19].
- Or even the *Guidelines on FAIR Data Management in Horizon2020*[20] by the European Commission addressed to EU funding applicants and beneficiaries to help them to make the data coming out from their research projects Findable, Accessible, Interoperable and Reusable.

Along with all these documents, we must cite again the EC HLEG report on FAIR data (*Turning FAIR into reality*[21]), giving recommendations to different stakeholders to implement the FAIR action plan, especially within the European Open Science Cloud (EOSC).

However, there is not a clear guideline addressing Research Performing Organizations (**RPOs**) to be proactive on creating real policies to implement FAIR data within institutions. Furthermore, different disciplines, communities and research domains, have different issues to undertake a proper FAIR data policy that many times relies on the particular nature of the datasets produced within the domain. Health research is probably one of the most sensitive and complex domains to make its research data open[22] and FAIR.

On the other hand, following the EUOSPP (European Open Science Policy Platform[23]) we fully believe that beyond recommendations, principles and declaration for an open data-driven research, we need **Practical Commitments for Implementation** (PCI). A PCI is a realistic commitment that a stakeholder might adopt to implement Open Science in a practical way. Implementation of Open Science, in any of its challenges and dimensions (e.g. FAIR data), is only possible when the stakeholder has jurisdiction, and the HRPO has the capability to create a policy at institutional level.

In this section, we will reflect on the findings under **WP2, facing all the issues** (legal, ethical, cultural, technical, etc.) to help define common frameworks for health research data policy allowing for different levels of commitment and requirements at institutional level (HRPO) with a special focus on de-identification and other health data specific issues. This section will help us build up the foundations of the intended *Guidelines for implementing FAIR open data policy in health research*, as one of the important outcomes of FAIR4Health project, and hopefully, in the future, an endorsed outcome at the RDA community level.

4.2. Legal framework in EU

The analysis of the legal framework of Fair4Health outlined below is based on the General Data Protection Regulation **(GDPR)** since it will be applicable to all EU-based research

[18] https://digital.csic.es/handle/10261/173801
[19] One of the Maredata (https://maredata.net) members (UC3M) is also member of FARI4Health.
[20] http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf
[21] https://ec.europa.eu/info/sites/info/files/turning_fair_into_reality_1.pdf
[22] See the workshop held in Brussels in April 2018 about this topic: Open Research Data to Support Sustainable Health Initiatives.
[23] https://ec.europa.eu/research/openscience/index.cfm?pg=open-science-policy-platform

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824666

projects. Consequently, if the national legislation contradicts the GDPR, the GDPR will prevail. A table reflecting the national regulations of the selected EU countries is added at the end of this section (Table 1). A more detailed report about this legal framework is available at FAIR4Health virtual research environment: https://osf.io/h6auf

4.2.1. Health data under Fair4Health Project

According to the GDPR it is necessary to differentiate between:

- 'Data concerning health': "means personal data related to the physical or mental health of a natural person, including the provision of health care services, which reveal information about his or her health status". (Article 4.15 GDPR[24]).
- 'Genetic data': "means personal data relating to the inherited or acquired genetic characteristics of a natural person, which give unique information about the physiology or the health of that natural person and which result, in particular, from an analysis of a biological sample from the natural person in question" (Article 4.13 GDPR[25]).

However, for the purposes of this document, both will be referred jointly as "**health data**". We should also consider that "health data" is a very broader term that might include clinical, research data or both that might have different conditions from a legal point of view, under the privacy general regulation in Europe.

As technology progresses, it is more and more difficult to completely anonymize data. But, in the other hand, better technology can both, make data anonymous as well as re-identify it. But even when data have been anonymized, they still have the condition of **personal data** and its processing is therefore subject to **data protection regulations**. If, at some point it is possible to achieve total anonymization that would guarantee the absolute impossibility of re-identifying the data subject, anonymized data would cease to have the status of personal data. It would therefore be possible to process such data without having to comply with the data protection requirements.

Article 89 of GDPR[26] provides that to ensure **the principle of data minimization** (only the minimum necessary data for the purpose of research must be gathered) the data controller must apply technical and organizational measures that may include pseudonymization. As such, GDPR envisions the possibility of resorting to pseudonymization, but also to other measures that could help achieve minimization of personal data collection in such a way that data subjects cannot be identified.

[24] http://www.privacy-regulation.eu/en/article-4-definitions-GDPR.htm
[25] http://www.privacy-regulation.eu/en/article-4-definitions-GDPR.htm
[26] http://www.privacy-regulation.eu/en/article-89-safeguards-and-derogations-relating-to-processing-for-archiving-purposes-the-public-interest-scientific-or-hi-GDPR.htm

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824666

4.2.2. Legal implications for FAIR data policies

4.2.2.1 Prohibitions and exemptions

The GDPR creates a distinction between personal data and **special categories of personal data**. Article 9[27] establishes the general principle of the **prohibition** of processing **health data.** So: processing health data, as personal health data, is prohibited in the new EU data protection landscape. However, there are **exceptions** to this principle, listed in Article 9 as follows:

> Article 9.2 GDPR:(a)"data subject has given explicit consent to the processing of those personal data for one or more specified purposes, except where Union or Member State law provide that the prohibition referred to in paragraph 1 may not be lifted by the data subject"; and (j)"processing is necessary for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes in accordance with Article 89(1) based on Union or Member State law which shall be proportionate to the aim pursued, respect the essence of the right to data protection and provide for suitable and specific measures to safeguard the fundamental rights and the interests of the data subject".

4.2.2.2 Consent

The first essential requirement to process health data is the **consent** of the data subject. Consent should be given by a clear affirmative act establishing a freely given, specific, informed and unambiguous indication of the data subject's agreement to the processing of personal data relating to her or him, such as by a written statement, including by electronic means, or an oral statement. Therefore, the data subject must consent to the use of human biological samples for scientific research purposes as well as to the use of clinical data. Ticking a box when filling out a form at the time an analysis or medical test is conducted is admitted as consent.

Consent will always be **required**, irrespective of whether or not it is anonymized, provided that the resulting dataset would still be personal data. In certain circumstances, coded or identified samples may be processed for biomedical research purposes without the consent of the data subject where it is possible to re-use them.

Consent may be generic or specific. As far as possible, consent should preferably be generic in order to enable platform users to make use of the data in different types of research. However, data subject may also give her or his consent more specifically: for a specific research or for a type of research (e.g. consent for the processing of data in the context of cancer research).

In the context of **scientific research** and whenever it is necessary in the **absence of a reason of public interest** that makes consent unnecessary, the data subject may withdraw

[27] http://www.privacy-regulation.eu/en/article-9-processing-of-special-categories-of-personal-data-GDPR.htm

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824666

her/his consent at any time, without affecting the lawfulness of the processing based on consent.

In any case, consent is something that might be taken into account at institutional level when creating FAIR data or Open Data policies.

4.2.2.3. Right of information

It is important to point out that any policy around FAIR data in Health Research will have to consider the information rights of data subjects contained in Articles 13 and 14 GDPR. Where **personal data** are directly obtained from the data subject, obligations of Article 13 GDPR[28] must be considered. According to Article 14.5, this obligation is not applicable where the data subject already has the information, or the provision of such information proves impossible or would involve a disproportionate effort. This is particularly true in the case of **processing for archiving purposes** when conducting scientific research, subject to the conditions and safeguards set forth in Article 89(1) GDPR. It is likely to render impossible or seriously impair the achievement of the objectives of that processing. In such cases, the institutional responsible (the controller) shall take appropriate measures to protect the data subject's rights and freedoms and legitimate interests, including making the information publicly available.

4.2.2.4 Transfer

The objective of the EU with the **GDPR is to guarantee the protection of personal data** and the **free circulation** of such data. One of the requirements to be met by the responsible **data controller** at institutional level is to ensure the transfer of the data securely and in accordance with the implementing rules such as consent, information and transfer according to guarantees and adequacy.

In our case the guidelines that we are going to work in inside FAIR4Health are not limited to the European Union, as non-European countries and international organizations may also participate both, in the creation of the guidelines (in the **global context** of RDA) as well as in the adoption of them by HRPOs in any part of the world.

4.2.2.5 Reuse

According to FAIR principles, it must be possible/desirable to **reuse the data for different investigations** without altering its essential content. In order to do so, the data subject must have given her/his prior consent to the reuse of the personal data.

The scope of the consent given by the data subject determines whether or not such data may be reused. This means that when the processing must be based on consent, the data

[28] http://www.privacy-regulation.eu/en/article-13-information-to-be-provided-where-personal-data-are-collected-from-the-data-subject-GDPR.htm

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824666

subject should be clearly and unequivocally aware of the purposes for which the processing will be carried out.

Personal data may be **reused** in certain circumstances on the basis of:

a) The type of consent given: generic or specific;
b) Prior/previous processing of the data for the purposes for which consent was given;
c) The exercise of the rights of opposition, rectification and erasure; and
d) Withdrawal of consent

Article 159 of GDPR[29] states that "the processing of personal data for **scientific research purposes** should be interpreted in a broad manner". The EC its "Guidelines on consent under Regulation 2016/679"[30] (wp259rev.01) considers that: "the notion may not be stretched beyond its common meaning and understands that 'scientific research' in this context means **a research project** set up in accordance with relevant sector-related methodological and ethical standards, in conformity with good practice". Therefore, consent is required if the data is going to be processed as part of a different project for which consent was not initially given. This situation might happen very often in HRPOs where more than one project about the same disease is running at the same time or correlatively in time. Any policy about FAIR data should take this restriction into account.

Once the circumstances of the research institution have been addressed, it is necessary to analyze the **possible scenarios** detailed below:

a) The **data are reusable**:

- Where the data subject has given generic consent for the purpose of scientific and/or biomedical research;
- In the case of further processing of data for which consent has been given for purposes other than those for which they are intended to be re-used, in which case they may be reused within the same scientific research despite the fact that the purposes may be different.

b) The **data are not reusable**:

- Where the data subject has given her or his consent for specific purposes and the data have not been processed;
- Where the data subject has withdrawn consent, even though the data may not have been deleted [1][31];
- Where the data subject has exercised her/his rights of objection or withdrawal;

[29] http://www.privacy-regulation.eu/en/recital-159-GDPR.htm
[30] https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=623051
[31] Whereas 65 considers the exception to the exercise of the right of suppression when it states that "the further retention of the personal data should be lawful where it is necessary, for exercising the right of freedom of expression and information, for compliance with a legal obligation, for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller, on the grounds of public interest in the area of public health, for archiving purposes in the public interest, scientific or *historical research purposes or statistical purposes, or for the establishment, exercise or defence of legal claims*".

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824666

- Where the data have been previously processed
- When data processing is planned outside the research project for which consent was originally given.

By virtue of the principle of **data minimization**, and the guarantees that it establishes, the reuse of personal data is possible for different purposes provided that it forms part of the same scientific research project for which the consent was collected.

Whereas article 157[32] broadens the scope of the investigation taking into account the possible collection of health data from registries. It thus recalls that "by coupling information from registries, researchers can obtain new knowledge of great value with regard to widespread medical conditions such as cardiovascular disease, cancer and depression. On the basis of registries, research results can be enhanced, as they draw on a larger population. Within social science, research on the basis of registries enables researchers to obtain essential knowledge about the long-term correlation of a number of social conditions such as unemployment and education with other life conditions. Research results obtained through registries provide solid, high-quality knowledge which can provide the basis for the formulation and implementation of knowledge-based policy, improve the quality of life for a number of people and improve the efficiency of social services. In order to facilitate scientific research, personal data can be processed for scientific research purposes, subject to appropriate conditions and safeguards set out in Union or Member State law."

4.2.2.6 Other rights and exceptions of the data subject

According to the GDPR, **data subjects** have the following **rights**: (a) right of access (Article 15 GDPR[33]); (b) right to rectification (Article 16 GDPR[34]); (c) right to erasure (Article 17 GDPR[35]); (d) right to restriction of processing (Article 18 GDPR[36]); (e) right to data portability (Article 20 GDPR[37]); and (f) right to object (Article 21[38]).

As an **exception**, the right to erase will not apply where it is necessary for exercising the right of freedom of expression and information, for compliance with a legal obligation, for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller, **on the grounds of public interest in the area of public health** or for **archiving scientific research** purposes among others (Article 17.3 GDPR[39]).

There is also an exception to **the right to object**, when the personal data is being processed for scientific research purposes pursuant to Article 89(1), on grounds relating to the data

[32] http://www.privacy-regulation.eu/en/recital-157-GDPR.htm
[33] http://www.privacy-regulation.eu/en/article-15-right-of-access-by-the-data-subject-GDPR.htm
[34] https://gdpr-info.eu/art-16-gdpr/
[35] http://www.privacy-regulation.eu/en/article-17-right-to-erasure-'right-to-be-forgotten'-GDPR.htm
[36] http://www.privacy-regulation.eu/en/article-18-right-to-restriction-of-processing-GDPR.htm
[37] http://www.privacy-regulation.eu/en/article-20-right-to-data-portability-GDPR.htm
[38] http://www.privacy-regulation.eu/en/article-21-right-to-object-GDPR.htm
[39] http://www.privacy-regulation.eu/en/article-17-right-to-erasure-'right-to-be-forgotten'-GDPR.htm

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824666

subject's particular situation, unless the processing is necessary for the performance of a task carried out for reasons of public interest according to Article 21.6 GDPR[40].

Requirement Country	Consent is mandatory for				Consent's Withdrawal	Right of information	Anonymi-sation and/or Pseudony-misation
	Processing	Transfer	International transfer	Reuse			
Italy	NO Scientific and biomedical research: Article 110 Code for the protection of personal data	YES Article 2 sexies Code for the protection of personal data	YES Article: 9 Authorisa- tion n. 9/2016	YES Article 110-bis Code for the protection of personal data	YES Article: 4.5.1 Measure identifying the provisions contained in General Authorisations no. 1/2016, 3/2016, 6/2016, 8/2016 and 9/2016	Articles 77 to 82 Code for the protection of personal data	No preference
Serbia	YES Articles 4.12), 12, 15, 17 Law on Personal Data Protection	YES Articles 11 and 36 Law on Personal Data Protection	YES Article 69 Law on Personal Data Protection	YES Articles 6 and 31 Law on Personal Data Protection	YES Articles 15 and 30.2) Law on Personal Data Protection	Articles 23 and 24 Law on Personal Data Protection	Pseudonymisation and encryption Articles 4.6), 6.5), 42, 50 and 92 Law on Personal Data Protection
Spain	YES (Express and written) Sixteenth additional provision. 2.a) Organic Law 3/2018 -------- General: Article 4 and 13 Genetic analysis: Article 48 Biomedical research: Article 58.2 Law 14/2007	YES (Express and written) Article 5.2 Law 14/2007	Ref. GDPR Article 40 Organic Law 3/2018	YES, if for different means (Express and written) Sixteenth additional provision 2.c) Sixth transitional provision Organic Law 3/2018 -------- Article 5.3 Law 14/2007	YES Article 4.3 Law 14/2007	Article 11 Organic Law 3/2018 -------- Articles 4.5, 26 and 27 Obligation: Article 59 Law 14/2007	Pseudony-misation Sixteenth additional provision. 2.d) Organic Law 3/2018
Switzerland	YES Article 17 Federal Act on Data Protection -------- Article 7 and 16 Human Research Act	YES Article 10a and 19 Federal Act on Data Protection -------- Articles 41 and 59 Human Research Act	YES Article 10a 6.b) Federal Act on Data Protection -------- Article 42 and 60 Human Research Act	YES Article 17 and 32 Human Research Act	YES Article 19.1.b) Federal Act on Data Protection -------- Article 34 Human Research Act	Article 8, 9, 14, 18a and 18b Federal Act on Data Protection -------- Article 8, 16 and 18 Human Research Act	Anonymi-sation Article 22 Federal Act on Data Protection -------- Coded Article 32 Human Research Act

Table 1: Executive summary of the National Regulations related with data protection

4.3. Ethical implications

This section focuses on the ethical implications of reusing FAIR data for health research in order to consider them when HRPOs implement FAIR data policies. It is based on the

[40] http://www.privacy-regulation.eu/en/article-21-right-to-object-GDPR.htm

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824666

answers gathered through the open survey on ethics and the focus group discussion over these results, as explained in the methodology section. The report about ethical implications is available at https://osf.io/bn5cs/

1. Health Research Performing Organizations are responsible for **performing research** under a **legal and ethical framework** acknowledged by the **scientific community** and the **public community at large**. They must implement governance systems in keeping with this legal and ethical framework, considering the principles stated in the GDPR, a research integrity and good practices code of conduct, compliance with data ethics principles and, as much as possible, stand for producing ethical outcomes for science and society.
2. Ethical questions dealing with the reuse of health data for research are very complex, even for experts in the field. There is an understanding gap that may hinder the validity of the responses gathered from lay citizens/patients. Therefore, there is a **strong need to train all the stakeholders** involved in this field, from patients to health researchers, in the **ethical and legal issues** within research.
3. **Patients/citizens and researchers** may have different views about the meaning of honesty and fairness regarding the openness of the research process. In addition, it seems that patients should be engaged in the research process beyond the signature of an informed consent. Therefore, a deeper and structured dialogue between both groups should be encouraged to **avoid unrealistic expectations from both sides** and achieve common objectives. This can be done by developing public participation events such as joint workshops between patient associations and research groups, for instance.

4.4. Security and data privacy issues

This section provides an overview of **relevant security** requirements to inform policies to be adopted by HRPOs. The conceptual map below summarizes related topics on security requirements that need to be considered in the proposed approach. The report on cybersecurity is available at https://osf.io/268f3/

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824666

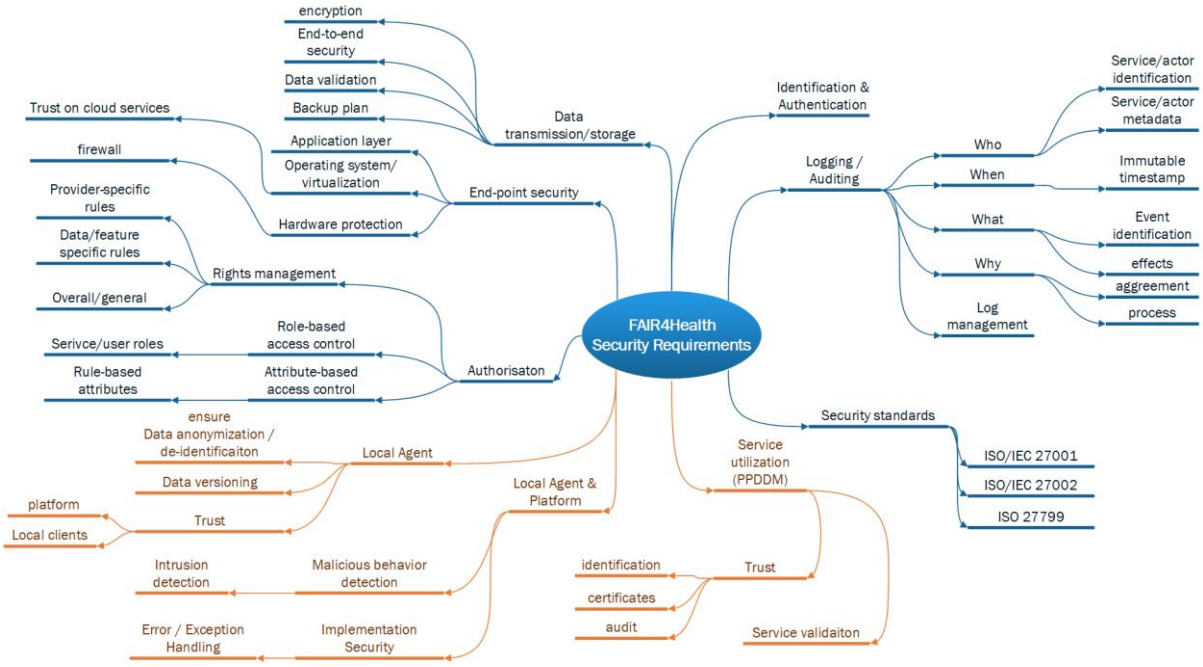


Figure 6: Conceptual map of general security requirements (blue) and specific security requirements (orange) for FAIR4Health (by mindmap)

The stated security requirements were derived according to general security concepts established for shared, cloud platforms. Nonetheless, specific security requirements have been developed and released in the deliverable D2.2, while technical requirements derived from these functionalities have been added to the deliverable D2.1.

4.4.1. General Security requirements

The following general security requirements are relevant to all identified FAIR4Health scenarios and the overall idea of the FAIR4Health approach:

Identification & authentication: Each actor and each component (data sources, PPDDM services, etc.) needs to be unambiguously identified within the FAIR4Health environment. The identification needs to be unambiguous, immutable and transparent.

Auditing or logging: is a fundamental feature for enabling confidentiality and security. For traceable actions, each triggered event within a software system should be logged with related information [18]. Considering the FAIR4Health approach, the following parts are considered crucial for logging:

- Who triggered the event? Identification of the **actor/service** who triggered the event. Requirements comprise a unique identity for each actor/service and related metadata.
- When was the event triggered? The exact **date and time** when the event happened. This can be realized through an immutable timestamp, accessible from all actors.
- What caused the event to trigger? The **description** of the event unambiguously identifying the task/action that triggered the event. In this case it is necessary to

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824666

determine a suitable way to denominate and identify events (e.g. severity levels similar to syslog, meaningful messages, for relevant cases even interdependency among events). It might be useful to state any effect on the further process caused by the triggered event, e.g. deny access, or depending on its severity, send an email notification to administrators or other responsible actors.

- Why was an event triggered? Specifies the **context** of the current event and describes its intention. E.g. an agreement between data providing party and data retrieving party is necessary depending on the actual access on the data. A log event might be related to such an agreement for its completeness. Further the log might link interrelated events or create a history sequence on events based on the perspective (e.g. events triggered for the process of accessing **FAIR data** on the platform for one user). Beside the proactive detection of security vulnerabilities it might also be used for intrusion detection. It might be further used to verify access.

Apart from the actual logging task, it is necessary to **manage the gathered logs** for secondary use (debugging) and ensure access authority as well as data integrity. Responsible roles or policies need to ensure confidentiality.

Data transmission and storage: Data transmission can occur over a point-to-point or multipoint channel. This raises the risk of data loss and manipulation during access on FAIR data. Data transmission over the internet needs state of the art protection for both, the endpoints as well as the transmission channel. The following requirements need to be met:

- Encryption of data.
- Secure end-to-end communication enables that only the communicating participants can read the content of the messages.
- Data validation should be performed before transmission and after receiving it to avoid the transmission jeopardizing data integrity.
- A suitable backup plan needs to be part of the system including the data emerging throughout the system life cycle.

Endpoint-security: The underlying and dependent components of the platform need to be kept in a secure state as well. This comprises the security management of the different system levels like application layer (e.g. malware protection), operating system or virtualization (e.g. trust on cloud services/systems) and hardware protection (e.g. firewalls).

Authorization is the required function of specifying access rights/privileges to data. It is necessary to secure 1) the intellectual property of the data set or the data mining algorithm and 2) to avoid data loss and privacy risks. In order to establish authorization mechanism, two major concepts are well known: The first is called attribute-based access control and defines policies, which describe rules based on (user-)attributes for access control (Binary "If-Then"-Logic). The second concept is called role-based access control and defines user roles with specific access rules (e.g. user-role "healthcare professional" is allowed to access metadata of all data sets stored in the FAIR4Health platform). The management of authorization and access control needs to be well defined and mandatory for all

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824666

participating actors. The rules need to be elaborated in a joint-effort and transparently communicated according to the used concepts:

- It is required, to define/manage/enforce **Service and User roles**.
- It is required, to define/manage/enforce access **rules and policies**.
- **Authorisation** needs to support different levels of granularity and security management.
  - General right management defines rules which are valid for the platform in general.
  - Feature specific right management defines rules for concrete specific elements of the FAIR data platform (e.g. special access rules for birth dates within a dataset).

Security Standards are highly relevant for the specific security requirements, yet general concepts contained therein need to be addressed. Well-known security standards include ISO/IEC 27001; ISO/IEC 27002 and ISO 27799. They not only describe security requirements limited to IT, but also deal with the integration of information security into the organizational structure and processes. The standard **ISO/IEC 27001** also advocates the proven concept of the PDCA-cycle, comprising the steps Plan, Do, Check and Act to manage IT security for a continuous, frequent and quality-oriented improvement of information security.

4.5. Policies to facilitate a cultural change towards FAIR data implementation

The task 2.5 of WP2 *(Cultural and behavioral barriers in EU for FAIR open data policy implementation in health research and overcoming mechanisms)* was to do an analysis of current cultural and behavioral barriers in EU that could hinder the FAIR open data policy implementation. The analysis distinguishes between FAIR data and Open data, identifying the **main barriers to opening or sharing research data** among European health scientists. One of the subtasks and methodologies was to design and distribute an open survey to relevant stakeholders (researchers in Health Sciences, but also extremely motivated citizen scientists). The next section describes the results of that survey. The review of current key studies and bibliography, the contextualization of research data in the new EU Directive of Public Sector Information (PSI) and the pop-up research associated to this task may be found in sections 2.3 and 2.4 of this deliverable.

4.5.1 Open survey on cultural barriers using FAIR data for health research

The Open Survey on cultural barriers using FAIR data for health research was released in April 2019 and disseminated among all stakeholders involved in the health data workflow, ranging from patients to medical doctors and researchers. The full survey is also available in PDF format in the FAIR4Health virtual research environment (OSF): https://osf.io/vb6sk In order to facilitate its dissemination and collecting responses, the survey was adapted to an online format making use of the Google Forms® tool.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824666

The main purpose of this survey was to gather opinions and attitudes on sharing health data among the stakeholders involved in the health data workflow and to assess their degree of knowledge about basic FAIR concepts. There were 99 respondents with a little over a third from Spain.

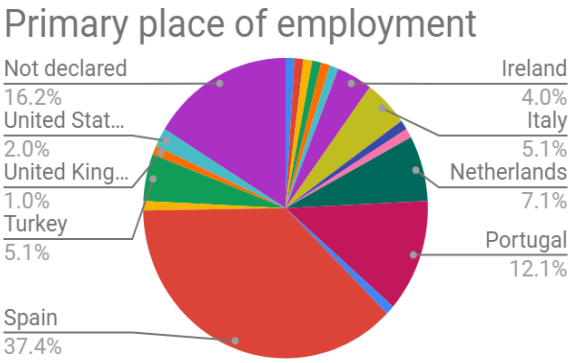


Figure 7: Primary place of employment of respondents. N=99

Most respondents (48,5%) identified themselves as researchers, 39,4% as clinical researchers, medical doctors or health professionals, while "others" category was 12,1%. With regards to attitudes on data sharing, most of the respondents showed a great degree of agreement with the options: "I don't usually share my data but have no problems to do it if someone or a funding agency asks for it" (3.117 ± 1.366)[41]. And "I only share my data upon request (from other researchers, agencies)" (2.947 ± 1.389), confirming what other studies says: while there is no problem with data sharing as a concept, most people usually don't do it. And when they do it, they prefer to share on request.

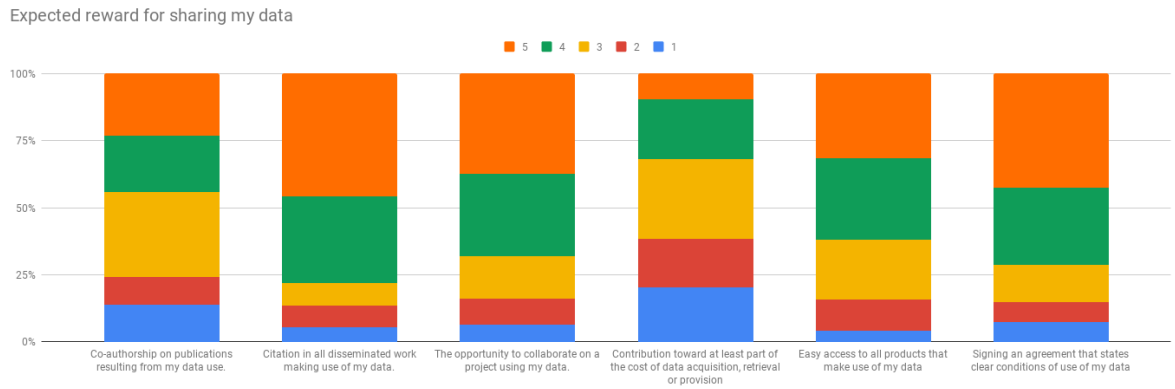


Figure 8: Expected reward for sharing my data

[41] Results are shown in terms of Average ± Standard Deviation. Answers can range from 1 (totally disagree) to 5 (totally agree).

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824666

Expected reward for sharing my data	1	2	3	4	5	N
Co-authorship on publications resulting from my data use.	13	10	30	20	22	95
Citation in all disseminated work making use of my data.	5	8	8	31	44	96
The opportunity to collaborate on a project using my data.	6	9	15	29	35	94
Contribution toward at least part of the cost of data acquisition, retrieval or provision	19	17	28	21	9	94
Easy access to all products that make use of my data	4	11	21	29	30	95
Signing an agreement that states clear conditions of use of my data	7	7	13	27	40	94

Table 2: Expected reward for sharing research data (by researchers)

Respondents perceived their lack of knowledge about which repository to use (2.968 ± 1.548) or which license to apply to their (meta)data (3.179 ± 1.465) as big barriers to data sharing. They all expect assistance from their institutions to manage research data (3.510 ± 1.267), citations from other researchers making use of their data (4.052 ± 1.229), the opening of opportunities for new collaborations (3.830 ± 1.271) and signing clear conditions of use of their data (3.915 ± 1.298). Most of them value more accessibility (4.358 ± 0.998) or re-usability (4.206 ± 1.052) over findability (4.021 ± 1.205) or interoperability (3.863 ± 1.188).

4.5.2 Towards a cultural change for FAIR Data implementation

The results of the Open Survey cited here confirm the main barriers for researchers in health sciences:

- Concern about privacy and confidentiality of data;
- Lack of knowledge about which licenses to apply to datasets and other digital research objects;
- Concern about how, when and where to deposit;
- Lack of training, on how to package, manage and share data in an appropriate way.

The **findings are consistent** with other similar studies such as the Wellcome Trust [19] report in which population and public health researchers declared more barriers to data sharing because they confront challenges such as the effort involved in data preparation, or concerns about privacy, or the clinical researchers worrying about misuse of data, as well as a lack of time curating data before it may be deposited in a repository..

Data sharing culture is very dependent on the researcher's discipline. For example, genetic and molecular sciences as well as infection and immunobiology researchers practice data sharing very well and don't pay much attention to the kind of barriers that researchers from other fields consider signficant — fear that data will be misused or misinterpreted, and fear that sharing data may jeopardise future publication opportunities [19].

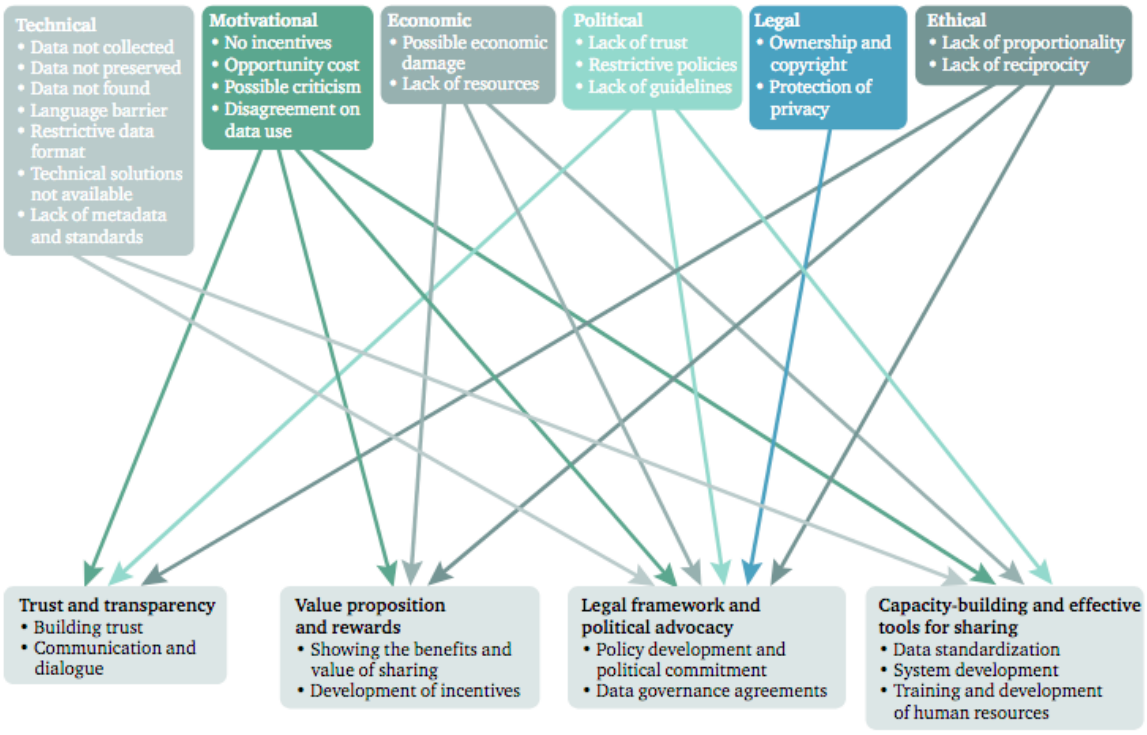
The main motivations identified [20] for **overcoming barriers to** data sharing among researchers include:

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824666

- direct career benefits deriving from sharing;
- following the norms of the discipline;
- funders or publishers requiring sharing; or
- when data sharing is a fundamental part of the process.

Sane and Edelstein [21] propose the following solutions for public health grouped in four categories:

1. Trust and transparency.
2. Value proposition and rewards.
3. Legal framework and political advocacy.
4. Capacity-building and effective tools for sharing.



*Barrier themes adapted from van Panhuis et al. A systematic review of barriers to data sharing in public health. *BMC Public Health*. 2014;14:1144.

Figure 9: Solutions to data sharing in public health. Source:(Sane & Edelstein, 2015) adapting barriers identified by (van Panhuis et al., 2014)

Promoting cultural change in a scientific discipline is not an easy task. Some of the most important solutions proposed depend more on Research Funding Organisations (RFOs) and publishers than on Research Performance Organisations (RPOs). Overcoming barriers to data sharing sometimes involves promoting stick and carrot policies.

From the point of view of the RPOs the main initiatives should be **aimed at building trust** by developing systems that facilitate the researcher's tasks, allowing them to have confidence and discharge responsibilities by complying with certain protocols that are easy to activate, and offering training on data management throughout their life cycle so that they can benefit from the incentives and rewards provided by the RFO.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824666

Building trust

- Systems (FAIR4Health platform in this case) should provide electronic workflows among researchers and Ethical Committee or IRB facilitating communication with researchers, inside the institution, and from outside institution requesting access. This should feed a kind of knowledge base for internal researchers and Ethical Committee.
- Systems must provide automated deanonymization processes, discharging researchers from this responsibility. Researchers should take part of the final decision for considering datasets finished.
- Systems must guarantee to researchers somehow that all data they're working on have "the appropriate consent".
- Systems must help user to the FAIRification process of the datasets.
- Use licenses of data to share should be easy to add to datasets. Systems must provide a closed list with suggestions of which licenses could apply. Ethical Committee should be involved in the definition and redefinition of these licenses. This is part of the FAIRification process.
- Systems can help providing suggestions of trusted repositories where to deposit the data.
- Most of the stages of the processes should provide the names of the people to contact in case of doubts.
- Systems must provide folders, directories or some kind of website with other datasets shared before as exemplary cases (practising by example).

Value propositions: incentives and rewards

- Appraise journals and data citations as a part of personal and/or departmental assessment process
- Track and audit reuse of FAIR datasets shared through data citations.
- Provide training in data management and FAIR principles
- Funding FAIR and data management related projects
- Audit and improving FAIRification processes and fairness of datasets.
- Promote and stimulate co-authorship and institutional collaborations among researchers using and aggregating datasets from different sources. Recognising authorship for sharing the data.

From the point of view of the RFOs, it seems that mandating the use of a Data Management Plan (DMP) and also requiring data deposit in journals or repositories that meet certain conditions increases the awareness of researchers about the need to share data. Of course, these policies must be accompanied by: funding for data sharing; new assessment mechanisms, both for DMPs and data sharing deposited in repositories, as well as certain rewards for those who comply with these policies.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824666

(H)RFOs	Increase Awareness of FAIR practices	Mandating DMPs	
		Mandating FAIR data sharing	*Only Journals and services (repos) complying FAIR principles. Clear policies, promote data standards*
		Assessing DMPs	*Good DMPs assessment improve possibilities of extra funding for tracking or scholarship for data management.*
		Assessing FAIR data sharing	*Data Citation counts for assessment. Increase possibilities of get funding or extra funding*
	Promote use of journals and services (repositories) complying FAIR principles		
	Funding FAIR practices		

Figure 10: Role of research funding organizations (including Health research) to incentivize FAIR principles

4.6. Public engagement and citizen science in health research

The task T2.7 of WP2 *(Boosting citizen science for FAIR data generation in health research)* addresses the **perceived suitability of public engagement mechanisms** and strategies that may **leverage citizen participation** in health research based on the report "Innovative Public Engagement: A Conceptual Model of Public Engagement in Dynamic and Responsible Governance of Research and Innovation" from the PE2020 project. These Public Engagement methods have been analysed making use of an open survey about their perceived suitability for health research. Additionally, an analysis has been performed of the effectiveness of mobile health-sensor data recording by individuals via smartphone applications (mHealth apps) as a suitable method for boosting citizen science in health research.

4.6.1. Open survey on boosting citizen science

The open survey on boosting citizen science in EU health research was released in April 2019 and disseminated among all stakeholders involved in the health data workflow, ranging from citizen and patients to medical doctors and researchers. The survey was released in nine languages: English, Spanish, French, Italian, Portuguese, German, Dutch, Serbian and Turkish. None of the questions were mandatory, that is, it was possible to complete the survey without filling out any or all the questions.

The full survey in English is available in PDF format in FAIR4Health-OSF: https://osf.io/czbmj. In order to facilitate its dissemination and collecting of responses, the survey was adapted to an online format making use of the Google Forms® tool.

The main findings and conclusions of the survey are presented below. The complete results can also be found at the OSF repository in the following link: https://osf.io/kfd48/

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824666

In total, 182 respondents completed the survey. Most of them chose Spanish (n=80, 43.96%) and Portuguese (n=43, 23.63%) as the language of the survey, and about 90% of them were citizens of the European Union. The vast majority of respondents had completed at least a Bachelors' or equivalent educational level (n=163, 95.33%), and the most prevalent profiles were: scientific researchers (n=63, 34.62%), healthcare professionals (n=47, 25.82%) and general audience (n=46, 25.27%). Please note that this question could be answered with more than one option (multiple choice).

According to the results of this open survey, the most suitable public engagement mechanisms for boosting citizen science in health research is Public Deliberation (3.572 ± 1.375) closely followed by Public Participation (3.544 ± 1.415), while the least suitable one is Public Consultation (3.024 ± 1.457).

4.6.2. Literature review on the use of mHealth apps as an effective method for citizen science

In this section, the main findings and conclusions extracted from the literature review are presented. A comprehensive report on the use of mHealth apps as an effective method for citizen science can also be found in the OSF repository at: https://osf.io/3n6q2/

In recent years, there has been **a change in health care systems** due to an increased use of technology. **Digital health** owes its rapid expansion to an increase in access to patients´ clinical history. **Citizens are the driving force** behind this rapid expansion of digital health utilization.

Additionally, the surge of **mobile applications related to health** in recent years promises to provide citizens the ability to better understand their health and achieve their overall goals related to their well-being. Having said that, it comes of no surprise that of the more than one million applications available in the global apps market, over **325,000 are health related** [22].

A particular literature review [22] aimed to identify and explore the current methods related to the **usability of health applications**. As with other digital instruments, ease of use is one of the key factors in successful implementation.

A randomized clinical trial was conducted on patients after discharge from Intensive Care Unit (ICU) due to respiratory failure, in order to **assess the feasibility, acceptability, and usability of a mobile, self-directed mindfulness training app** in comparison to both used methods: a therapist-led telephone-based mindfulness program as well as a web-based critical illness education program. The study concluded that a majority of users preferred the mobile app as a method of delivery. More importantly, mobile mindfulness performed similarly to therapist-led mindfulness training program and generally better than an education program [23].

Innovations in digital health face several **ethical and political challenges**. We have argued that in order for digital health products and applications to produce tangible innovation and health impacts, either at individual or population level, four conditions must be met [24]:

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824666

- First, data are of paramount importance for digital health: **access** to sufficient amounts of data is thus a primary requirement for the development of innovative diagnostic and therapeutic.
- Second, alignment with existing **legal provisions** regarding data protection, data security and privacy are key to digital health innovation. Legal frameworks can thus have a major impact in facilitating or hindering progress in this field. Nonetheless, legal provisions do not address the full range of ethical issues in data processing. Nor do they cover the full spectrum of legitimate concerns of data subjects.
- Third, robust and transparent **accountability** mechanisms should ensure the precise identification of responsibility for data uses and their consequences on individuals, families and communities. What is more, accountability also sets up mechanisms for communicating health relevant information to data subjects.
- Finally, fourth, evidence of **safety and efficacy** is a significant condition for the success of digital health. Licensed digital health products and applications will have to go through extensive assessment processes.

We conclude that health systems are slowly moving towards digital health. Research and support for health information technologies related to patient participation for its potential benefits are important. In this sense, **it seems reasonable and advisable to include accessible, legally compliant, accountable and safe mobile technologies** as an effective method for boosting citizen science in health research.

4.7. Technical considerations for the implementation of a FAIR data policy in health research

There are several **essential components** needed for the data to accomplish FAIR principles**.** The technical consideration to implement the strategy relies on the use of well-defined tools to support the use of Persistent Identifiers (PIDs), policies, metadata, standards, vocabularies and certified and trustworthy repositories.

FAIRification is the process "*to translate raw (meta) data into Findable, Accessible, Interoperable and Reusable (meta) data according to the FAIR data guiding principles*". The minimum entities resulting from the FAIRification process are **FAIR Digital Objects which can represent data, source code, workflows, models and other resources**. (Cf. FAIRification workflow in section 5 of D2.2)

But turning data and code into **FAIR Digital Objects** is only the first stage of the process**.** They need to be deployed inside a FAIR ecosystem of services in which the data can be findable, accessible, interoperable and reusable. The objects need Persistent Identifiers (PIDs) and metadata to be discoverable. [3]

To implement FAIR data policies HRPOs must define useful practices for data sharing and common eHealth standards. FAIR4Health framework needs to be **based on eHealth standards and vocabularies** together with **semantic technologies** over distributed repositories, such as in the Personal Health Train approach [9].

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824666

4.7.1. FAIR technical ecosystem

As was mentioned earlier, the central part of the FAIR ecosystem is the FAIR Digital Object, which is composed of the following layers one on top of each other [3].

- The Digital Object itself is the basic element of data, code or other resources. It is the information that the HRPOs decide to make available. This includes expertise to curate and steward such objects.
- Identifiers that are persistent, global and unique (PIDs) are needed to identify the Digital Object unambiguously. They can be a DOI or URN, among others.
- Standards & Code that are well documented and open. In the case of FAIR4Health, semantic interoperability can be achieved with the use of Clinical Information Models (CIM) that allow organizing the information inside an EHR repository or for EHR communication and controlled medical terminologies and vocabularies and the mappings between them:
  - HL7 FHIR resources[42] [1], a collection of 120 definitions of data structures to be exchanged in a health data interoperability scenario.
  - Some terminologies are designed to serve specific purposes in a specific medical field. For example, Logical Observation Identifier Names and Codes[43] (LOINC) is designed for laboratory result encoding, whereas other general-purpose terminologies such as SNOMED CT[44] or International Classification of Diseases (ICD), can be used to clarify the clinical meaning of data.
  - Terminology mapping, engine for the normalization of local term to standard terminologies.
- Reusability and discoverability can only be achieved if the data do include rich metadata (how, when and by whom the objects were created). The use of metadata standards adopted by the given research communities to enable interoperability and reuse is a must. The metadata also need to have the appropriate documentation about dependencies and licensing. Semantic technologies allow finding the FAIR Digital Objects with the help of the metadata.

4.7.2. Technical issues for a FAIR data policy from FAIR4Health perspective

The implementation and deployment of the **FAIR data infrastructure, repository, or even a FAIR data ecosystem** must cover the suitable treatment of the FAIR Digital Objects during the FAIRification workflow from raw data. It will also include services, applications, tools, and algorithms that allow using the clinical data for research. Different architectures should be considered and reflected in the institutional policy (e.g. considering federation when integrating data across distributed repositories from the same RPO). Local and PPDDM agents need to access the information and data taking into consideration that the infrastructure should work for humans and machines.

[42] https://www.hl7.org/fhir/
[43] http://www.loinc.org/
[44] http://www.snomed.org/

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824666

Interoperability is the hardest FAIR principle to fulfil [25]. Multiple standards and frameworks can be applied but FAIR4Health challenge is to address all the aspects to solve the technological interoperability problem, along with legal interoperability and other issues regarding standards**.** All this technical standardization approaches should be also stated at institutional level when crating a FAIR data policy.

4.7.3. Application and Tools

As an example, the FAIR4Health ecosystem will enable researchers, health professionals and stakeholders to use FAIR Digital Objects by making them findable, accessible, interoperable and reusable. In the FAIR4Health platform, applications and tools will be implemented to execute the steps of the FAIRification workflow so the datasets are transformed and annotated with suitable metadata. Semantic web technologies are used to automatically discover and search the FAIR artefacts adequate to their use case.

The services must also implement metadata specifications, standards and ontologies following FAIR principles. This means making them discoverable, identifiable and registered in catalogues. These services should provide the ability to curate the artefacts based on security mechanisms such as expiration, organization authorization and so on (Cfr. D2.2 – 5.2 - Data Curation). FAIR4Health will be ready to use distributed, trusted and certified (**CoreTrustSeal (CTS)**) repositories, so FAIR4Health agents are forced to use common protocols independent of RPOs. Moreover, the services in which FAIR Digital Objects are managed should also be certified by a trusted entity, which poses further needs to explore this potential role for CTS repositories.

4.7.4. Repositories and registries

FAIR4Health ecosystem must only include certified and trusted repositories with the following criteria:

- Persistent unique identifiers (PIDs)
- Metadata to enable data discovery by machines
- Suitable licenses
- Long-term preservation to ensure data set persistence and repository sustainability [3]

4.7.5. Storage and infrastructure

FAIR4Health ecosystem needs to take into account the distributed nature of the Clinical Data Repositories (CDR) spread across different countries with different legislations. The system must be "privacy by design and by default" in order to be fully compliant with GDPR. Service catalogues must be offered, and workflows can be run automatically over the data. Local and PPDDM agents must be easily connected with the FAIR services. The distributed repositories must be federated to integrate clinical data or results and provide the users a unified vision over distributed environments.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824666

Proposed technical considerations for the requirements identified in order to implement a useful FAIR data ecosystem for Health research extracted from the report Turning FAIR Data into Reality [3]

- ❖ Rec. 2: Implement a model for FAIR Digital Objects, compatible with eHealth standards and well-known terminologies and vocabularies.
- ❖ Rec. 3: Develop components of a FAIR ecosystem, application and tools on top of the FAIR Digital Objects to be searchable and interoperable.
- ❖ Rec. 4: Develop the FAIR4Health interoperability framework.
- ❖ Rec. 7: Support semantic technologies, to automatically discover the datasets, facilitating the automated processing.
- ❖ Rec. 9: Develop assessment frameworks to certify FAIR services.

5. Guidelines for implementing FAIR/Open data policy for Health Research Performing Organizations

This guideline gathers **5 principles** and **10 steps** that a Health Research Performing Organization (HRPO) should follow to implement a FAIR/Open data policy. They take into account all the legal, ethical cultural and technical reflections that FAIR4Health has done as if the project was itself a HRPO.

PRINCIPLES

1.	To implement a FAIR/Open data policy implies **to manage a complex change**. The policy needs **strategical vision** and **leadership**.
2.	There are not policies without **resources** and **incentives** supported by necessary Infrastructure.
3.	It is necessary to count with the right **knowledge** and **skills** about FAIR data and research data management in the current Open Science landscape.
4.	Before implementing any Open/FAIR data policy, the institution needs a clear **action plan** identifying the main **actors** and a credible **timeline**.
5.	The policy must be **written** down and **approved** by the institution.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824666

STEPS

1.

Define the **vision** and **objectives** of the policy.

This should include:

- The **description of the current research** performed in the institution: research funders and their policies regarding research data; Types of research outcomes that the institution produced in the last 5 years; Kinds of data used/reused by the researchers; Current infrastructures used at institutional level (VREs, data repositories, etc.)
- The **definition** of the **objectives** of the policy, such as: comply with funders' requirements, improve transparency and reproducibility, include datasets in Health research infrastructures (EOSC) etc.
- The **adoption** of a comprehensive approach to overcoming **mechanisms** that can enhance data sharing.

2.

Identify/name a **responsible person/unit** for the Open/FAIR data policy as well as the envisaged **team** to put it in practice.

Design an **action plan**, state **resources** to be committed for the Research Data Management and a feasible **timeline.**

3.

Raise awareness among researchers, as well as providing adequate training and **assistance** (data stewards, data scientist, or alike in the institution).

Definition of needed **skills** and a **training programme**.

4.

Identify and describe current and needed **data infrastructures.** This includes data storage and architecture definition but also the provision of tools to make data description and formatting easy and affordable for researchers.

All the research data should be deposit in a **trustworthy** repository.

5.

Establish **responsible Research Data Management** practices within the institution and **define FAIR for implementation**. This should include:

- Analyse the scope of the FAIR principles, that includes concepts like: data selection/curation, long-term stewardship, legal interoperability and the timeliness of sharing
- Creation of a standard institutional template for **Data Management Plans.**

6.

Determine the agreed level of **openness**, **transparency** and **re-usability** for research data produced in the HRPO, including **licensing** and **provenance** as well as the intended mechanisms for personal data **protection** (GDPR compliance).

7.

Technical decisions and **standards** adoption. FAIR data implies, among many other technical issues, **metadata** and **Persistent IDs**. So, the policy must include at least decision, at institutional level, around technical standards:

- PIDs policy and control for the identification of data, publications and other outcomes, as well as researchers.
- Metadata policy for data accessibility and interoperability: Selection of metadata vocabularies, best practices on metadata completeness, etc.

These decisions might include a technical analysis in domain relevant (Health) standards including metadata schemas, data modeling and vocabularies.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824666

8.	**Devise** credit and **reward mechanisms** in order to ensure researchers consider it is worth allocating time and energy to data management/sharing.
9.	**Write** the policy in Open/FAIR research data management → Submit it to approval in your **governance** bodies ⇒ Disseminate the policy inside the HRPO.
10.	Create mechanisms for **FAIR data assessment** within the institution. **Re-align** and **consolidate** the policy with the funders to guarantee that publicly-funded research data are made FAIR and Open, except for legitimate restrictions.

6. Guidelines refinement: discussion and next steps

The guidelines stated above are only the first attempt, in the context of FAIR4Health project, to come up with general and broadly endorsed guidelines in the Health Research Data domain. In this sense, this first version of the guidelines is the starting point to create a Working Group (WG) in the Research Data Alliance. It is intended to validate the draft guidelines in the international community. RDA is deemed as a good place to get the attention of international community. Based on the feedback and interest of the wider international community, the guidelines may be further developed in the context of different stakeholders including developed, middle income and least developed countries. Perhaps such an endeavor may evolve as a working group of the RDA in the future.

As we have already mentioned, **FAIR4Health** had already approached RDA community organizing a BoF "Assessing FAIR data policy implementation in Health Research[45]" during the 13th Plenary in April 2019. During this session, we discuss the methodology, surveys, etc. to be applied within FAIR4Health project in this first step. We decided to keep the Community informed and look for potential collaborators internationally within the Health Data IG, about these possible and needed guidelines. Therefore, the next steps will be:

1. Organize another BoF during the 14th RDA plenary in Helsinki in October 2019 specifically about *"Practical Commitments for Implementation of Open/FAIR data in Health Research Performing Organizations: Guidelines and collective engagement".*
2. Announce the BoF and the envisaged work around the guidelines through Health Data IG list, which currently (May 2019) has 186 members.
3. Formalize the creation of the RDA WG for the Guidelines for Open/FAIR data policies in HRPOs. To create an RDA WG requires commitment and a series of steps (Figure 11):
   - ❖ Definition of a Case Statement that describe the Recommendation that the group will produce, as well as the value proposition and work plan.
   - ❖ Identification of international memberships. We must include members from 3 or more continents.

[45] https://rd-alliance.org/bof-assessing-fair-data-policy-implementation-health-research-rda-13th-plenary-meeting

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824666

- Name 2-4 co-chairs leading the initiative. In our case two people from FAIR4Health project and at least one non-European member.
- Propose work, outcomes, deliverables and the action plan to come up with the acknowledged guidelines before 12-18 months.
- Publication of the Case Statement in a request for comment mode in the RDA homepage. Review of the Case statement by the RDA Technical Advisory Board (TAB) (4-6 weeks).
- If the TAB approves the Case statement, RDA Council will review it as is, or with recommendations or subject to specific revisions.

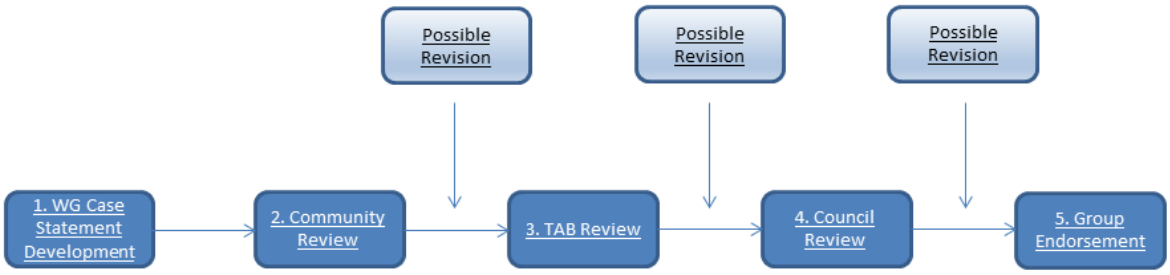


Figure 11: Process to formalize a RDA WG. Case statement's review process

4. Work within the WG: Teleconferences, F2F meeting/workshop at RDA 15 and 16, mailing lists, Open discussion though twitter polls or other simple mechanisms.
5. (Hopefully) Publication of the Guidelines as RDA Recommendations.
6. Dissemination of the recommendations among Health RPOs worldwide.
7. Closing out the specific Working Group in RDA and come back to the discussion within the Heath Data IG to follow up implementation and use cases.

7. References

[1] Mons B, Neylon C, Velterop J, Dumontier M, da Silva Santos LOB, Wilkinson MD. Cloudy, increasingly FAIR; revisiting the FAIR Data guiding principles for the European Open Science Cloud. Information Services & Use. 2017;37(1):49-56. doi:10.3233/ISU-170824

[2] Martone ME, Garcia-Castro A, VandenBos GR. Data sharing in psychology. American Psychologist. 2018;73(2):111-125. doi:10.1037/amp0000242

[3] European Commission. Turning FAIR Data into Reality: Final Report and Action Plan from the European Commission Expert Group on FAIR Data.; 2018. https://publications.europa.eu/en/publication-detail/-/publication/7769a148-f1f6-11e8-9982-01aa75ed71a1/language-en/format-PDF. Accessed December 6, 2018.

[4] Wilkinson MD, Dumontier M, Aalbersberg IjJ, et al. The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data. 2016;3:160018. https://doi.org/10.1038/sdata.2016.18.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824666

[5] National Academies of Sciences E. Open Science by Design: Realizing a Vision for 21st Century Research.; 2018. doi:10.17226/25116

[6] Press G. Cleaning Big Data: Most Time-Consuming, Least Enjoyable Data Science Task, Survey Says. Forbes. March 2016. https://www.forbes.com/sites/gilpress/2016/03/23/data-preparation-most-time-consuming-least-enjoyable-data-science-task-survey-says/. Accessed May 18, 2019.

[7] National Academies of Sciences E. Reproducibility and Replicability in Science. Washington, DC: The National Academies Press; 2019. https://www.nap.edu/catalog/25303/reproducibility-and-replicability-in-science.

[8] Sun C, Ippel L, Wouters B, et al. Analyzing Partitioned FAIR Health Data Responsibly. arXiv preprint arXiv:181200991. 2018. https://arxiv.org/pdf/1812.00991.

[9] Dutch Techcentre for Life Sciences. Manifesto of the Personal Health Train Consortium.; 2018. https://www.dtls.nl/fair-data/personal-health-train/manifesto-personal-health-train-consortium/. Accessed May 19, 2019.

[10] Schaaf J, Kadioglu D, Goebel J, et al. OSSE Goes FAIR - Implementation of the FAIR Data Principles for an Open-Source Registry for Rare Diseases. Stud Health Technol Inform. 2018;253:209-213. http://ebooks.iospress.nl/publication/50058.

[11] Wise J, de Barron AG, Splendiani A, et al. Implementation and relevance of FAIR data principles in biopharmaceutical R&D. Drug Discovery Today. 2019;24(4):933-938. doi:10.1016/j.drudis.2019.01.008

[12] Traverso A, van Soest J, Wee L, Dekker A. The radiation oncology ontology (ROO): Publishing linked data in radiation oncology using semantic web and ontology techniques. Med Phys. 2018;45(10):E854-E862. doi:10.1002/mp.12879

[13] van Panhuis WG, Paul P, Emerson C, et al. A systematic review of barriers to data sharing in public health. BMC Public Health. 2014;14(1):1144. doi:10.1186/1471-2458-14-1144

[14] Federer LM, Lu Y-L, Joubert DJ, Welsh J, Brandys B. Biomedical data sharing and reuse: Attitudes and practices of clinical and scientific research staff. PloS one. 2015;10(6):e0129506. http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0129506.

[15] Tenopir C, Allard S, Douglass K, et al. Data Sharing by Scientists: Practices and Perceptions. PLoS ONE. 2011;6(6). doi:10.1371/journal.pone.0021101

[16] Tenopir C, Dalton ED, Allard S, et al. Changes in Data Sharing and Data Reuse Practices and Perceptions among Scientists Worldwide. van den Besselaar P, ed. PLOS ONE. 2015;10(8):e0134826. doi:10.1371/journal.pone.0134826

[17] Stuart D, Baynes G, Hrynaszkiewicz I, et al. Whitepaper: Practical challenges for researchers in data sharing. March 2018. doi:10.6084/m9.figshare.5975011.v1

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824666

[18] Zeng L, Xiao Y, Chen H, Sun B, Han W. Computer operating system logging and security issues: a survey. Security and Communication Networks. 2016;9(17):4804-4821. doi:10.1002/sec.1677

[19] Eynden VV den, Knight G, Vlad A, et al. Survey of Wellcome researchers and their attitudes to open research. October 2016. https://wellcome.figshare.com/articles/Survey_of_Wellcome_researchers_and_their_attitudes_to_open_research/4055448. Accessed May 10, 2019.

[20] Eynden VV den, Bishop L. Incentives and Motivations for Sharing Research Data, a Researcher's Perspective: A Knowledge-Expert Report. University of Essex; 2014:48. repository.jisc.ac.uk/5662/1/KE_report-incentives-for-sharing-researchdata.pdf.

[21] Sane J, Edelstein M. Overcoming Barriers to Data Sharing in Public Health.; 2015:26. https://www.chathamhouse.org/publication/overcoming-barriers-data-sharing-public-health-global-perspective/20150417OvercomingBarriersDataSharingPublicHealthSaneEdelstein.pdf.

[22] Maramba I, Chatterjee A, Newman C. Methods of usability testing in the development of eHealth applications: A scoping review. International Journal of Medical Informatics. 2019;126:95-104. doi:10.1016/j.ijmedinf.2019.03.018

[23] Cox CE, Hough CL, Jones DM, et al. Effects of mindfulness training programmes delivered by a self-directed mobile app and by telephone compared with an education programme for survivors of critical illness: a pilot randomised clinical trial. Thorax. 2019;74(1):33-42. doi:10.1136/thoraxjnl-2017-211264

[24] Vayena E, Haeusermann T, Adjekum A, Blasimme A. Digital health: meeting the ethical and policy challenges. Swiss medical weekly. 2018;148:w14571.

[25] Wilkinson MD, Verborgh R, Santos LOB da S, et al. Interoperability and FAIRness through a novel combination of Web technologies. PeerJ Comput Sci. 2017;3:e110. doi:10.7717/peerj-cs.110

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824666