D6.1 Expected users' interactions

VERSION 1.1

Document Information

Contract Number	965345
Project Website	http://www.healthycloud.eu/
Contractual Deadline	M15, May 2022
Dissemination Level	PU
Nature	R
Author(s)	Laura Portell Silva (BSC) Lidia López Cuesta (BSC) Salvador Capella-Gutiérrez (BSC)
Contributor(s)	Alicia Martinez (SAS) Annika Jacobsen (LUMC) Celia Alvarez (SAS) Christian Ohmann (ECRIN) Esmeralda Ruiz (UB) Eva García (BBMRI-ERIC) Helena Lodenius (CSC) Kurt Majcen (BBMRI-ERIC) Maria Panagiotopoulou (ECRIN) Peter Holub (BBMRI-ERIC)
Reviewer(s)	Adrian Thorogood (UNILU) Anamika Chatterjee (EMBL-EBI) Christian Ohmann (ECRIN) Maria Panagiotopoulou (ECRIN)

Keywords	User profiles, Health data portal, User interactions, Personae

Change Log

Version	Author	Date	Description of Change
v0.1	Laura Portell-Silva	28.03.2022	Initial table of contents
v0.2	Laura Portell-Silva Salvador Capella-Gutiérrez Lidia López Cuesta	06.05.2022	First content sent to reviewers
v0.3	Maria Panagiotopoulou Christian Ohmann Anamika Chatterjee Adrian Thorogood	23.05.2022	Review finished
v0.4	Laura Portell-Silva Salvador Capella-Gutiérrez Lidia López Cuesta	30.05.2022	Improvements considering reviewers feedback
V0.9	Laura Portell-Silva Salvador Capella-Gutiérrez Lidia López Cuesta	03.06.2022	Final version sent to coordinators
v1.0	Salvador Capella-Gutiérrez	06.06.2022	Final version submitted
V1.1	Laura Portell-Silva	27.09.2022	Additional minor changes after EC feedback

Table of contents

Executive Summary

HealthyCloud WP6 is focused on defining the reference architecture for a FAIR health data portal. This portal is conceived as an access gateway for existing resources and a place for providing references to different users. The first step to reach this goal is to define the different user profiles that will interact with the portal. Indeed, different users such as citizens, researchers or infrastructure providers have different expectations from the portal and what it can offer. This deliverable seeks to define the needs and objectives of the user profiles in order to later on detect the functionalities that the portal should have based on their needs.

The user profiles are divided into fourteen categories, which are grouped into the six orthogonal concepts, designed to capture relevant aspects for the users' profiles when interacting with the FAIR health data portal. These categories include legal roles, e.g. data controller and data processor according to the European General Data Protection Regulation (GDPR), a temporal component on the use and impact of expected behaviours from prototypical users, among others.

In the context of this report, eight different profiles are considered to define the user interactions with the FAIR data portal. Importantly, these users' profiles build on existing efforts and bring in specific profiles associated with the secondary use of health-related data for research, innovation and policy-making purposes. In some cases, sub-profiles are considered to reflect the broad diversity of users that can demand different functionalities for the FAIR health data portal.

In addition to the categorization of the user profiles, user personae are also defined as archetypical users whose goals and challenges represent the needs for a larger group of users. The definitions of the user personae include their skills, goals, challenges, needs and expectations from the FAIR data portal. The definition of these user personae helps to better understand future users of the FAIR health data portal.

This effort will support the designing of the reference architecture for the FAIR health data portal. Such a portal is aimed at facilitating the interoperability of existing resources by defining the minimal information needed to interconnect them. Identifying data oriented resources is a concerted effort done in the HealthyCloud WP3 and WP4 and will be relevant to other tasks in WP6. All efforts in HealthyCloud aim to provide the relevant context for the definition of the strategic agenda of the future FAIR health data portal, including the users that will make use of it.

1 Introduction

A detailed user profiles definition and their corresponding challenges, goals and skills is needed to translate the vision of the FAIR health data portal into reality. This deliverable reports the outcome of Task 6.1 "Specifications for end users' interactions with the FAIR health data portal", which aims to identify all relevant end-users expected to interact with the future FAIR health data portal. This work will be used for capturing what functionality and services the portal should have.

One of the key objectives of this report is to properly model the expectations and needs of each user profile. Such an effort should contribute towards the engagement of future users and drive the generation of guidelines and recommendations for the portal. The guidelines and recommendations include expected interactions - within the portal itself and beyond - with both the data and computational components of the FAIR health data portal. Interestingly, these users' profiles will facilitate the work performed by other work packages and the stakeholder mapping. Final outcomes will be part of the HealthyCloud Strategic Agenda outlining the implementation of the future FAIR health data portal.

2 Working methodology

The work for this deliverable was developed with the collaboration of all HealthyCloud WP6 members during the monthly meetings and dedicated workshops scheduled specifically for the user profiles definition. From September 2021 to March 2022, six workshops of two hours were organized, and focused in four main areas that form the basis of this report:

- Definition of the user profiles categories that can be part of the FAIR health data portal.
- Development of the different user profiles and sub-profiles that fall into each category.
- Defining prototypic personae for each of the profiles for later defining the functionalities of the FAIR health data portal.
- Definition of the interactions of the users with the FAIR health data portal.

During the dedicated workshops, the user profiles were characterised based on the identified categories using a 5-likert scale, being 1 the lowest influence of the category into the profile and 5 the highest influence. In some cases, it was not possible to define a single type of user for a profile and the definition includes sub-profiles. Annex A includes the resulting data from the workshops and the figures found throughout this report were created based on this data, considering the input of all workshop participants.

In addition, it is important to mention that the definition of the users' profiles was based on previous work done by similar projects and initiatives:

- "The Digital Skills for FAIR and open science" report by the European Open Science Cloud (EOSC) executive board.[1]
- "Building digital workforce capacity and skills for data-intensive science" report by the Organisation for Economic Cooperation and Development (OECD) Global Science Forum.[2]
- "Final report of analysis of the training needs of different EOSC-Life users" by EOSC-Life.[3]

---

[1] https://doi.org/10.2777/59065
[2] https://doi.org/10.1787/e08aa3bb-en
[3] https://doi.org/10.5281/zenodo.3692068

3 Users' profiles categories

During this work, fourteen categories have been identified for defining the FAIR health data portal users' profiles. To facilitate specific discussions, these categories are grouped around six major orthogonal groups. Categories are designed to capture relevant aspects for the users' profiles when interacting with the FAIR health data portal. These categories include legal roles, e.g. data controller and data processor according to the European General Data Protection Regulation (GDPR), a temporal component on the use and impact of expected behaviours from prototypical users, among others. Categories and their orthogonal groups are described below.

Categories related to data generation and usage

1 **Data provider**: person or entity that generates health-related data either as product of their activity or as mandated by another individual or organization.

2 **Data user**: person or entity that makes use of health-related data that might be under their direct control or not, for which they might need to request access to.

Categories related to legal roles of health-related infrastructures

3 **Data controller**: Under Regulation (EU) 2018/1725, as well as under the GDPR, the data controller is the party that, alone or jointly with others, determines the purposes and means of the processing of personal data. The actual processing may be delegated to another party, called the data processor. The controller is responsible for the lawfulness of the processing, for the protection of the data, and respecting the rights of the data subject. The controller is also the entity that receives requests from data subjects to exercise their rights.

4 **Data processor**: According to Article 3 (12) of Regulation (EU) 2018/1725, a processor shall mean "a natural or legal person, public authority, agency or other body which processes personal data on behalf of the controller." The essential element is therefore that the processor only acts "on behalf of the controller" and thus only subject to its instructions. In some cases, the processor may choose not to process the data by itself, but may have recourse to a subcontractor who processes the data on its behalf. In practice, this will depend upon the processor agreement entered into with the controller.

Categories related to career development of a given professional

5 **Beginner**: limited experience specialist, e.g. software developers, researchers. They are usually starting out in their field of activity and that are often supervised by middle-level or senior professionals.

6 **Intermediate**: experienced specialist, e.g. researchers and software developers, with a balance work between technical developments and decision-making processes.

7 **Advanced**: highly experienced individual oriented mainly to decision-making processes and potentially the supervision of other team members.

Categories related to the intended use of the health-related data

8 **Technology-oriented**: individuals or organizations whose primary focus is the use of technological solutions for managing health-related data as part of their daily activities. Such users could also be involved in the development of new technological solutions.

9   **Usage-oriented**: individuals or organizations which are focused on the consumption and usage of the health-related data content for their daily activities.

Categories related to the professional sector of the main activities of a given professional

10   **Academia**: individuals and entities who are primarily affiliated to a university, research centre, or similar type of organization, mostly funded by public resources.

11   **Clinical**: individuals and entities who are primarily affiliated to clinical settings or health research performing organizations.

12   **Industry**: individuals and entities whose primary affiliation is the industrial sector, mostly funded with private capital.

Categories related to the temporal scale of the activities considered

In order to capture the temporal scale of the activities that the different roles have these two categories were added even though they don't refer to persons or institutions as the previous ones described.

13   **Long term**: it focuses on the use and impact of any action over a long period of time. The meaning of this temporal scale is specific for each of the analysed users' profiles, e.g. long-term impact of policies put forward by policy-makers, long-term users of specific software, long-term implications of making available high-quality health-related data, etc.

14   **Short term**: it is oriented towards the immediate use and impact of any action. This term can be used as well to define few uses of a given technology, e.g. one-time software users.

4 Users' profiles definition

In the context of the FAIR health data portal, eight different profiles have been considered to define its user interactions. In some cases, sub-profiles are taken into account to reflect the broad diversity of users that can fall into the same category. These profiles are configured after the fourteen categories mentioned in the previous chapter. The definition and categorization of these profiles serve to model the needs and goals of each individual profile when interacting with the FAIR health data portal. Those interactions should serve to define the expected portal's functionalities.
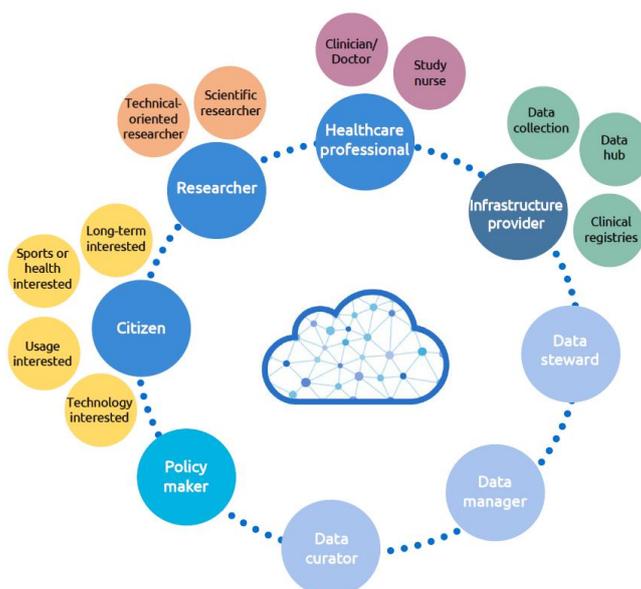

¡Error! Argumento de modificador desconocido.: FAIR health data portal user profiles and sub-profiles.

As mentioned before, the work done in this deliverable is based on previous efforts done in the context of the EOSC and the OECD Global Forum. Existing efforts revolve around the design and deployment of reference FAIR data portals, which require defining the profiles of users interacting with them. However, there are some differences when defining the users that interact with these two portals. In the context of the FAIR health data portal, two new user profiles - healthcare professionals and infrastructure providers - were considered relevant that were not taken into account for EOSC nor OECD. Working with health-related data that by definition is of sensitive nature requires paying special attention of the context where either the data is generated (health professionals) for secondary use in research or shared (infrastructure provider) to authorized users. To facilitate having a reasonable number of profiles, the data scientist and software engineer profiles - explicitly modelled in EOSC - were included in the category of researchers in HealthyCloud. Other profiles like EOSC enablers, data RI support and trainers and technology enablers are not specific to the future FAIR health data portal. Thus, their specific profiles will be taken directly from the definitions made in the broader context of the EOSC and its cluster project for Medical and Life Sciences, EOSC-Life.

4.1 Citizen

Citizen, individual who wants either to get information from one or more scientific biomedical disciplines or to contribute to a citizen science initiative. They might also consent to share their health information in a de-identified manner for research purposes. Importantly, sub-profiles were considered to reflect the broad diversity of citizens regarding the health-related data generation and usage.

As seen in Figure 2, citizens can be divided into four different categories. Some of the citizens will be more technology interested than others, e.g. how their data can be processed to give them some particular insights on their health. On the other hand, citizens can be more usage interested when they want only to look for specific disease information or population-level aggregated patterns through the FAIR health data portal. The two last sub-profiles are divided into sports and health interested citizens, when they produce information from their wearables that they would like to share with the scientific community; and diseases interested, when they want to find or share information about their chronic condition.
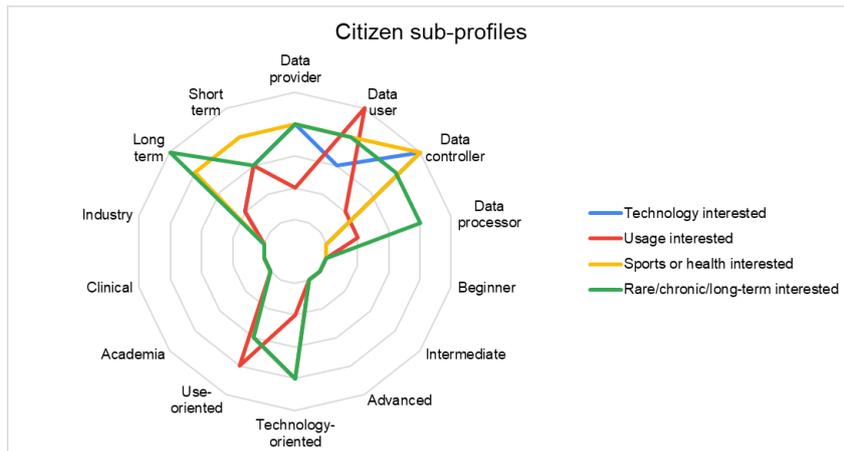
¡Error! Argumento de modificador desconocido.: Visual representation of 4 different sub-profiles according to the pre-defined categories for citizens.

4.2 Researcher

Researchers are those individuals that will interact with the future FAIR health data portal to obtain, process, produce, analyse, deposit or share research data and its potentially associated outcomes.
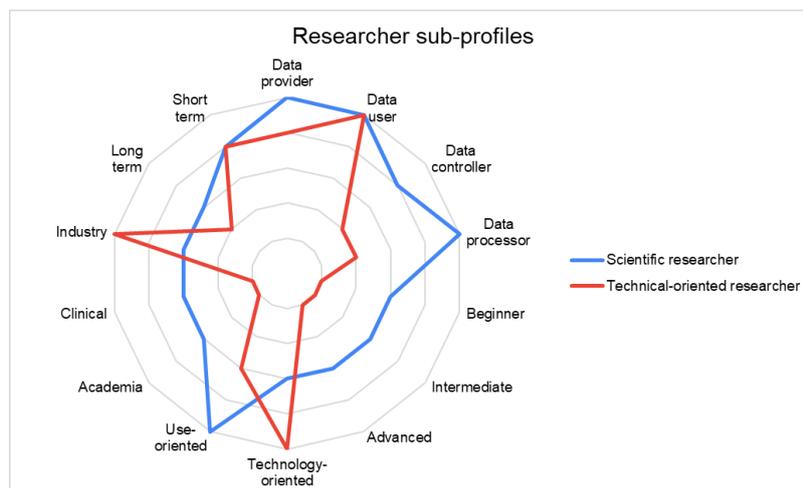

Figure 3: Visual representation of the different sub-profiles according to the pre-defined categories when considering different researchers' profiles.

As seen in Figure 3, researchers can be divided into two categories depending on their level of technical expertise.

- They can be **scientific researchers**, which include clinical researchers, biomedical researchers and bioinformaticians.
- **Technical-oriented researchers** with a higher technical expertise, which include for instance software engineers and data scientists. These users mainly process data – especially when defining input and output data types or creating AI-based models – but they do not own or control the data.

The two profiles mainly differentiate on how they would want to find information in the portal; and the user interface they would find more useful for accessing data, metadata and relevant information for their scientific activities.

4.3 Policy and decision maker

Policy makers gather information through consultation and research. Policy makers develop a policy or strategic framework based on those conclusions and the problem at hand. Therefore, it is likely that they would consume aggregated data rather than provide it, and they can be part of the industrial, clinical or academic sectors.

Figure 4: Visual representation of policy maker profile according to the pre-defined categories.

4.4 Healthcare professional

A **healthcare professional** is a person that works in the healthcare sector and has an active role in providing health-related data, which can eventually have a second use for research purposes. They are classified as clinicians/doctors and study nurses. The secondary use of the data generated by these profiles is very valuable for the FAIR health data portal and has to be considered as one of the ecosystem's main assets. Indeed, this profile sets a clear difference with other user profiles modelling efforts in the context of EOSC. The healthcare professional can also act as data users of the FAIR health data portal, since they can be interested in the research being done in specific areas or in general population trends.
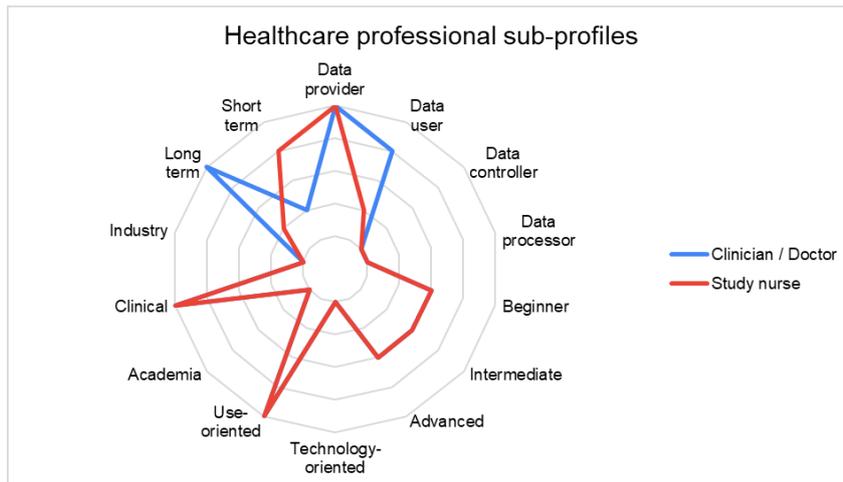
Figure 5: Visual representation of the different sub-profiles according to the pre-defined categories when considering the healthcare professional profile.

4.5 Data management profiles

4.5.1 Data curator

Data curators are responsible for the quality and FAIRness of health-related data, and to make sure data is discoverable and accessible. This profile also considers the possibility of enriching data when increasing its quality. Importantly, data curators can be data users while working for data providers or platforms with a processor role.

¡Error! No se encuentra el origen de la referencia.
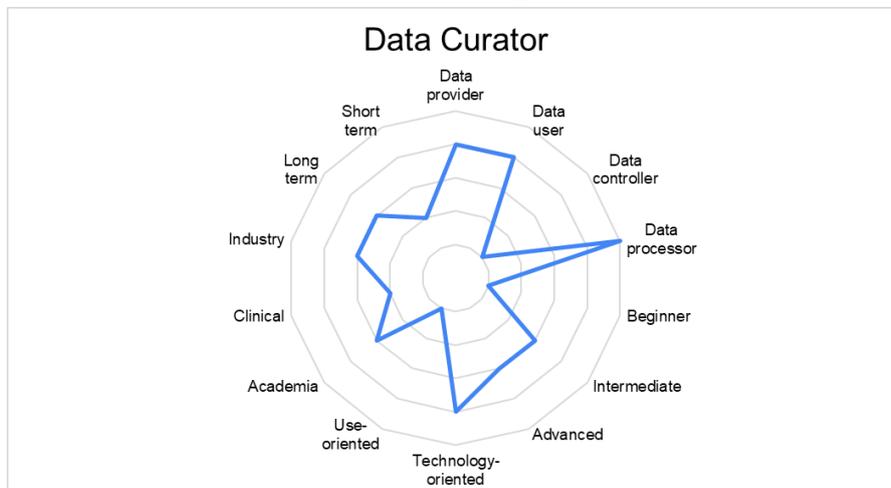
Figure 6: Visual representation of the data curator profile according to the pre-defined categories.

4.5.2 Data steward

Data stewards have an administrative role; they do not really consume the data to analyse it, but to understand it. They create guidelines to make data FAIR and advice on how to do it. Stewards may actually process data and have direct responsibility on the data at hand or not.
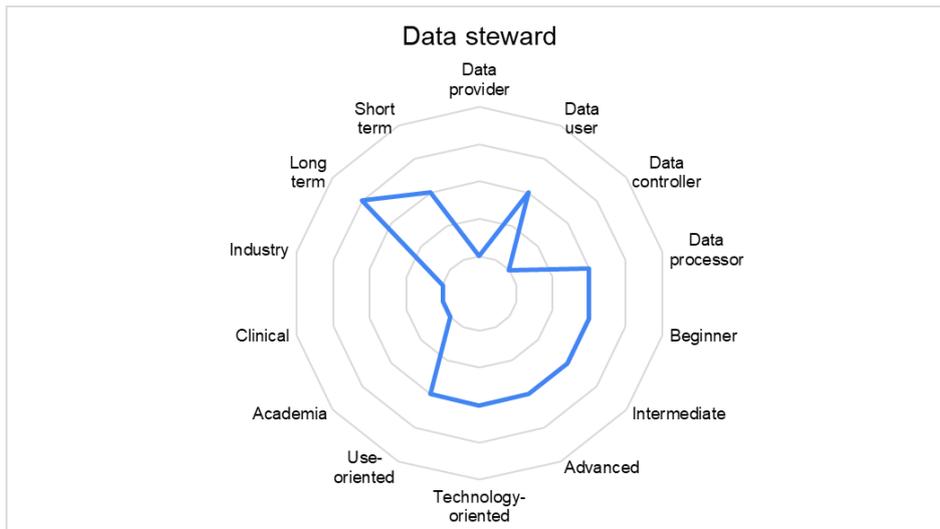
Figure 7: Visual representation of the data steward profile according to the pre-defined categories.

4.5.3 Data manager

Data managers ensure a correct flow of the data, which implies a holistic approach to how data is collected, used, re-used and potentially shared. In their managerial capacity, they play an integral role in processing, and in creating Data Management Plans (DMPs) for their projects. They can be part of both data processor and data controller organizations.
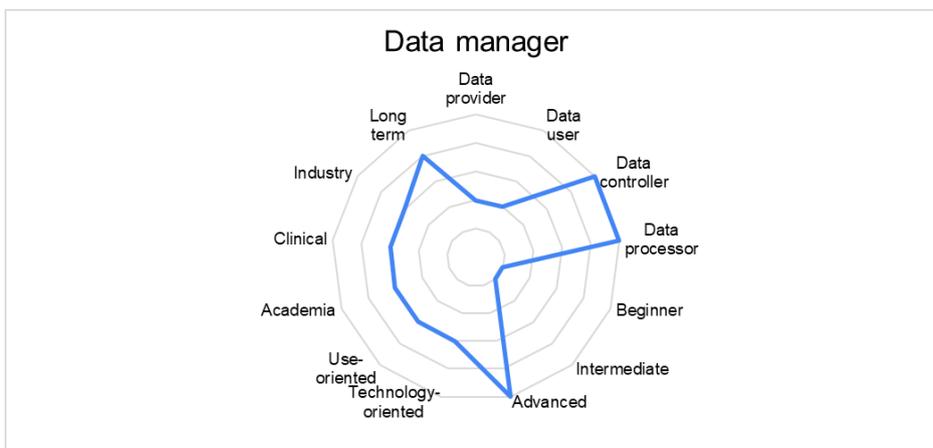
Figure 8: Visual representation of the data manager profile according to the pre-defined categories.

4.5.4 Data management profiles considerations

Data management is a complex and continuous process that requires **domain experts**, usually understood as *data curators*, which care about the content of the dataset and make sure that the data can be used, it is up-to date and potentially enriched. It also requires **technology professionals** to make sure data is used efficiently, e.g. *data managers and stewards* which guarantee usability of the data across the data life cycle and make sure that the data has the appropriate FAIRness level, respectively. Thus, a combination of these profiles should be part of any users' profiles definition when designing relevant interfaces for the FAIR health data portal. As expected, those profiles share important similarities when it comes to the overall management of data.
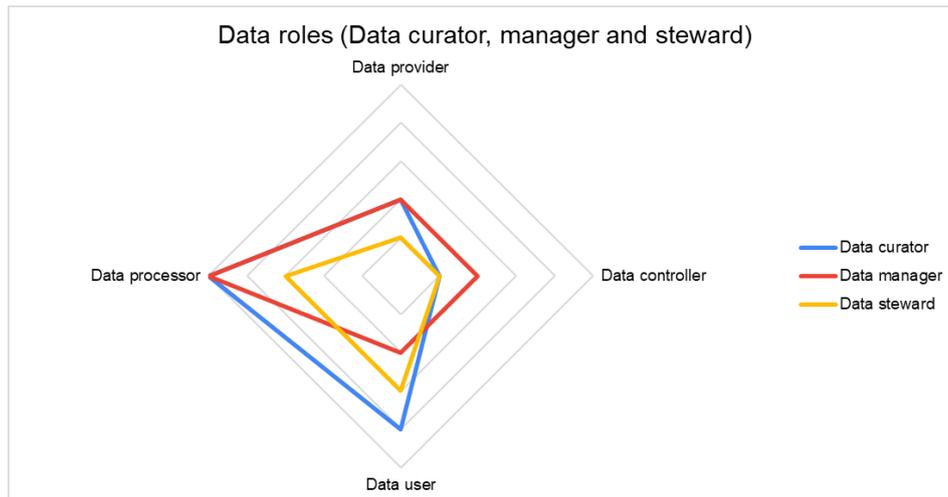
Figure 9: Data-related categories for data management profile.

As seen in Figure 9 and 10, data management profiles mainly consume data to process it. In most cases, they do not act as data controllers, which implies they cannot share the data with the FAIR health data portal. However, they can potentially share metadata to make data discoverable and also information on how to access it.
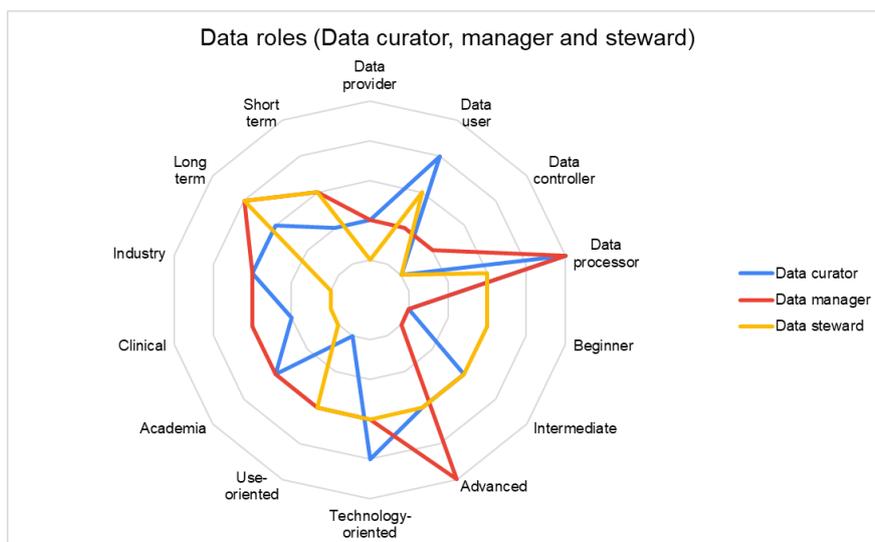
Figure 10: Visual representation of the different sub-profiles according to the pre-defined categories when considering all professional profiles related to data management.

4.6   Infrastructure provider

Infrastructure provider is the responsible organization to support the physical and digital management of health-related data following existing regulations. As seen in Figure 11, three different infrastructure provider profiles are identified as part of the FAIR health data portal: data collection, data hub and data repository or clinical registry. These infrastructures will receive the data from the primary data generators in order to make it available for the scientific community for secondary use. They can have mix roles as data controllers and processors. In the latter case, data can be shared when authorized by the data controller. Infrastructure providers would also include secure processing environments which are used for secure data analysis of the data and HPC.
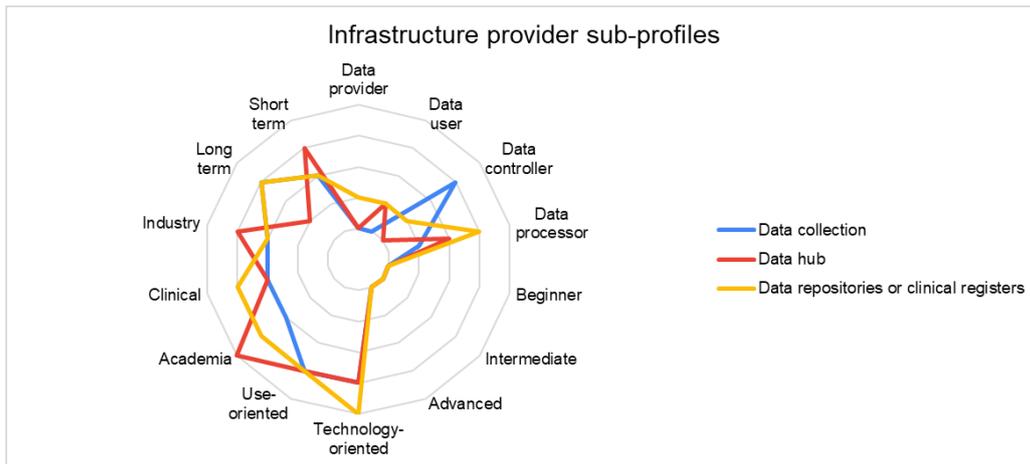
Figure 11: Visual representation of the different sub-profiles according to the pre-defined categories when considering different infrastructure provider profile.

4.7   Profiles relationships

To understand better how data is used in the FAIR health data portal and how it gets to the FAIR health data portal, Figure 12 illustrates the different users and how the data flows among them. It is important to mention that even if the data flows in one way, the data users also influence the way the data providers work. For instance, the policies created by the policy makers will influence the organization and management of the infrastructure providers, especially when working with sensitive data.
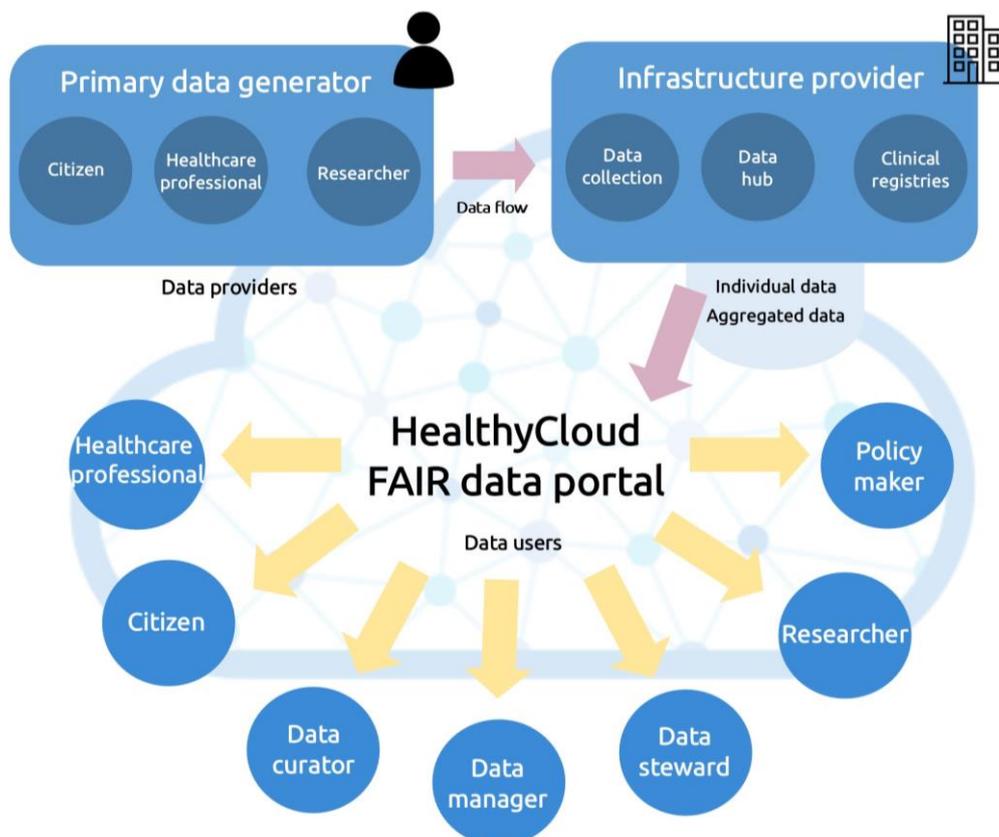
Figure 12: Data flow among the different user profiles and the FAIR health data portal. On the top, the data providers are described and at the bottom the data user who will consume the data from the portal.

5   Personae definition from user profiles

To later define the functionalities of the FAIR health data portal, user personae are defined as *archetypical users whose goals and challenges represent the needs for a larger group of users*. For this report, hey are used a mean to deeper understand the diversity of needs and requirements that a given user profile can have. The definitions of the user personae include their skills, goals, challenges, needs and expectations from the FAIR health data portal.

The definition of these user personae helps to understand better and get a similar perspective to the future users of the FAIR health data portal, which will be of help when designing a reference architecture.

5.1 Citizen

MARIA
CITIZEN

Age: 25
Occupation: Science teacher
Interests: Research

"I want to give my data for research, innovation and policy making purposes."

SKILLS
- Basic knowledge on the importance of contributing data for research
- Use of the online portal

GOALS
- Want to share their data for research, innovation and policy making
- Want to know what research is conducted for their medical condition and how the data is used
- Want to have access to summarized information about healthcare trends in the general and/or disease-specific population

CHALLENGES
- Do not know if there is research done for their condition/disease
- Do not know how to share their data for research, innovation and policy making
- Do not know if it is secure to share their data

NEEDS
- Find what research is being done for their condition/disease
- Where their data is being used and how
- Understand how secure the ecosystem is (data and communication)
- Know who is accessing the data (researchers or companies)

EXPECTATIONS
- The whole system is secure and simple to be used (for both technical and non-technical users)
- Easy-to-use (easy search mechanism) and easy-to-learn

DATA GENERATION AND USAGE
Data provider
Data user

CAREER STAGE
Beginner
Intermediate
Advanced

PROFESSIONAL SECTOR
Academia
Clinical
Industry

LEGAL BASIS
Data controller
Data processor

INTENDED USE OF DATA
Technology-oriented
Usage-oriented

TEMPORAL SCALE OF ACTIVITIES
Short term
Long term

5.2   Researcher

BEN
RESEARCHER

Age: 32
Occupation: PhD
Interests: Immunology

"I want to use health data for my research."

SKILLS
* General knowledge of sensitive data providers ecosystem
* Knowledge on existing services for enabling research activities
* Knowledge on study design and statistics
* Use of statistical tools

GOALS
* Want to find data about the same topic as their research
* Want to share their research outcomes
* Want to do analysis with the data
* Want to have access to summarized information about healthcare trends in the general and/or disease-specific population

NEEDS
* A reference repository for sharing their raw and/or processed data
* Find available health-related data for its further re-use.
* A reference portal where to find information about the infrastructure providers
* A reference place where I can build virtual cohorts (on top of existing data sources)
* Best practices, guidelines and tools for working with sensitive data

CHALLENGES
* Do not know where to find data about a specific topic
* Do not know how to ask permission to use the data found through the FAIR health data portal
* Waiting too long to get access to the data
* Programmatic access to health-related data may be too cumbersome
* Complex user-interfaces might be overwhelming
* Do not have time to get complex training on how to use analytical tools

EXPECTATIONS
* easy-to-find health-related data mechanism
* Straightforward user interfaces for achieving their goals, either gathering further information for their research activities, better understanding how to share/access to data, or to get to know how to process their data

DATA GENERATION AND USAGE
Data provider
Data user

CAREER STAGE
Beginner
Intermediate
Advanced

PROFESSIONAL SECTOR
Academia
Clinical
Industry

LEGAL BASIS
Data controller
Data processor

INTENDED USE OF DATA
Technology-oriented
Usage-oriented
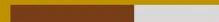
TEMPORAL SCALE OF ACTIVITIES
Short term
Long term

5.2.1 Technical-oriented researcher

Since it is expected that researchers and technical-oriented researchers will have different interactions with the FAIR health data portal, both of them have been defined as separated user personae.

5.3 Policy and decision maker

PHILIP
POLICY AND DECISION MAKER

Age: 50
Occupation: Policy officer
Interests: Politics

"I want to advance on policies for a trustworthy sensitive data usage"

SKILLS
- Know how to combine information from different sources to promote new policies and adjust existing ones
- Understand the challenges for a legit use of health-related data for research purposes.

GOALS
- Want to promote policies enabling health-related data research in a ethically sound, legally compliant and socially aware manner
- Want to understand the up-take of current policies and existing barriers preventing their implementation
- Want to promote best practices for translating data-driven research results into good clinical practice guidelines
- Want to identify and extract information about general and/or disease-specific population trends to implement healthcare plans
- Want to increase the impact of the policies they are promoting
- Want to to have access to aggregated information for data usage patterns from different sources and/or domains

NEEDS
- Access to summarized information about health-related trends in the general and/or disease-specific population
- A reference place to gain access to heterogeneous health-related data sources, including aggregated information about specific healthcare aspects or data usage patterns
- Measurement of impact of the policies they put forward

CHALLENGES
- Do not know where to find health-related general population or disease-specific trends
- Access to heterogeneous data sources, which might be geographically distributed and may fall under different legal frameworks
- Justify investments and health-related decisions to tax-payers and voters

EXPECTATIONS
- Easy-to-filter trends and visualizations to enable the policy formulation and decision-making process
- A portfolio of indicators for measuring health-related data usage impact

DATA GENERATION AND USAGE
- Data provider
- Data user

LEGAL BASIS
- Data controller
- Data processor

CAREER STAGE
- Beginner
- Intermediate
- Advanced

INTENDED USE OF DATA
- Technology-oriented
- Usage-oriented

PROFESSIONAL SECTOR
- Academia
- Clinical
- Industry

TEMPORAL SCALE OF ACTIVITIES
- Short term
- Long term
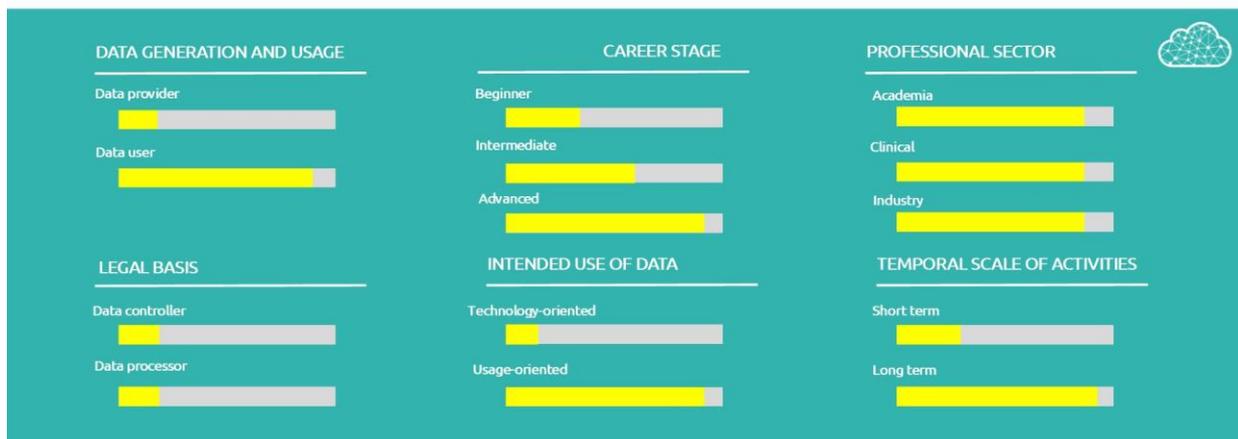
5.4 Healthcare professional

RACHEL
HEALTHCARE PROFESSIONAL

Age: 43
Occupation: Doctor
Interests: Animal care

"I want to have a reference place where to find relevant data for my professional activities"

SKILLS
* Provide health-related data for secondary use
* Formulate data-driven hypothesis based on real-world evidence

GOALS
* Want to identify if a given patient can be associated to a particular observational research cohort
* Want to identify best practices for developing and validating data-driven protocols for preventing, diagnosing and treating individual diseases
* Want to corroborate current observed patterns with existing data-driven research projects

CHALLENGES
* Lack of time and references in the standardization of the health-related data collection
* Do not know how to access to health-related data for conducting their research activities
* Do not know how to identity if an individual patient's observation match existing cohorts.

NEEDS
* A reference place for their professional activities
* Information about how to validate data-driven protocols for treating individual diseases
* A user-friendly experience for identifying if a given patient's observation match existing cohorts

EXPECTATIONS
* Reference place for identifying existing cohorts and creating new (virtual) ones
* A reference place for best practices for developing and validating data-driven protocols for preventing, diagnosing and treating individual diseases

DATA GENERATION AND USAGE
Data provider
Data user

CAREER STAGE
Beginner
Intermediate
Advanced

PROFESSIONAL SECTOR
Academia
Clinical
Industry

LEGAL BASIS
Data controller
Data processor

INTENDED USE OF DATA
Technology-oriented
Usage-oriented

TEMPORAL SCALE OF ACTIVITIES
Short term
Long term

5.5   Data curator

CHRIS
DATA CURATOR

Age: 38
Occupation:
Interests:

"I want to contribute to richly described high-quality data"

SKILLS
- Ability to validate that data is compliant with policies and regulations.
- High understanding of the domain
- Deep understanding of the data to be curated
- Knowledge of using controlled terminology for annotating data and metadata

GOALS
- Want to ensure data complies with the community's practices and uses the right community's formats and standards
- Want to ensure data is compliant with policies and regulations
- Want to make sure that data can be used by someone else in their own research activities
- Want to make sure users understand the purpose of the data collection
- Want to have access to aggregated information for data usage patterns from different sources and/or domains

NEEDS
- Provide the context where data would be primary used. Reference guidelines and mechanisms - either manual or automated - for controlling for data quality
- Reference guidelines for working with domain-specific controlled vocabularies and ontologies

CHALLENGES
- Low secondary-use data quality (see FAIR-Health principles).
- The lack of accepted - either formal or de facto - standards for the domain of the data

EXPECTATIONS
- Guidelines space for data quality and domain-specific controlled vocabularies and ontologies
- Data quality visualization
- easy-to-contact with the primary data providers for better understanding how data has been collected and generated

DATA GENERATION AND USAGE	CAREER STAGE	PROFESSIONAL SECTOR
Data provider	Beginner	Academia
Data user	Intermediate	Clinical
	Advanced	Industry

LEGAL BASIS	INTENDED USE OF DATA	TEMPORAL SCALE OF ACTIVITIES
Data controller	Technology-oriented	Short term
Data processor	Usage-oriented	Long term
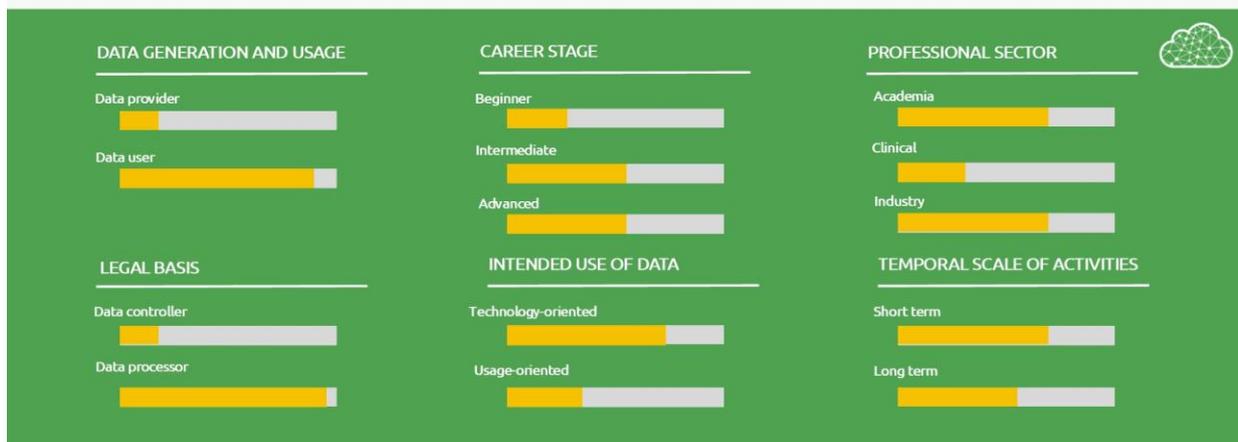
5.6   Data steward

PETER
DATA STEWARD

Age: 32
Occupation: Research assistant
Interests: Nature

"I want to follow the FAIR principles when working with health data."

SKILLS
* Deep knowledge about the FAIR data principles and application criteria
* Understanding of the sensitive data implications for properly FAIRifying data and its associated metadata
* Ability to provide recommendations on how to generate FAIR data by design and increase the FAIRness level of data already generated

GOALS
* Want to provide guidelines for researchers using health-related data
* Want to know if researchers follow the FAIR data principles for identifying major bottlenecks when adopting those principles
* Want to have access to aggregated information for data usage patterns from different sources and/or domains

CHALLENGES
* Do not find specific guidelines for adopting the FAIR principles when working with health data
* No standard approach for measuring health data FAIRness level

NEEDS
* Mechanisms for evaluating health-related data FAIRness
* Aggregated information about adoption of the FAIR data principles
* How to work with domain-specific controlled vocabularies and ontologies to increase data interoperability - one of the most difficult aspects of the FAIR data principles

EXPECTATIONS
* Aggregated information about the FAIRness level of the different data providers connected with the portal
* Dedicate space offering guidelines to learn about the FAIR data principles, especially regarding its implemetationw with health-related data, and how to measure data FAIRness

DATA GENERATION AND USAGE	CAREER STAGE	PROFESSIONAL SECTOR
Data provider	Beginner	Academia
Data user	Intermediate	Clinical
	Advanced	Industry

LEGAL BASIS	INTENDED USE OF DATA	TEMPORAL SCALE OF ACTIVITIES
Data controller	Technology-oriented	Short term
Data processor	Usage-oriented	Long term
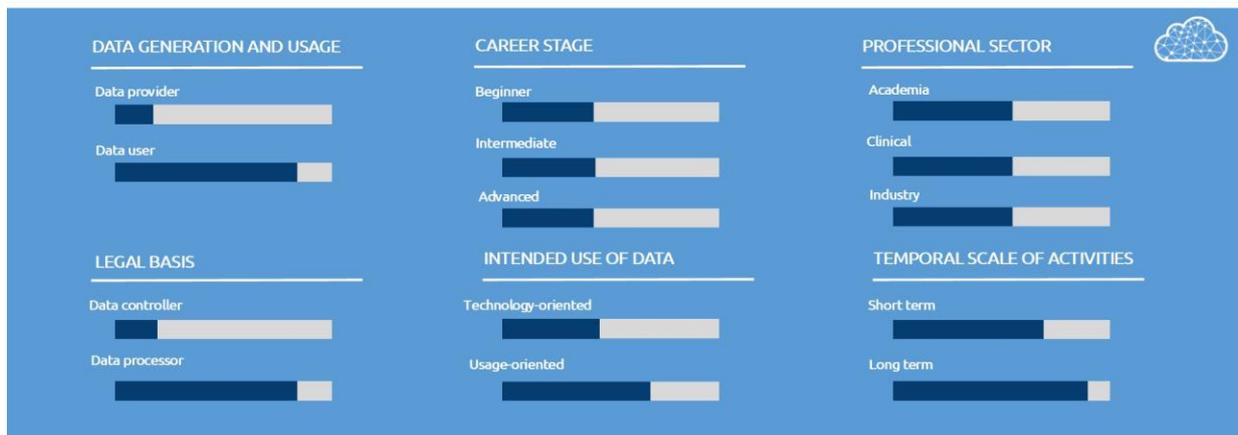
5.7   Data manager

HELENA
DATA MANAGER

Age: 34
Occupation: Operations support
Interests: Sports

"I want to properly manage data across the whole data life cycle"

SKILLS
* Deep knowledge of the different steps of the research data life cycle
* Knowledge of data management planning
* Knowledge of the implications of working with health-related data

GOALS
* Want to ensure high data quality in all steps of the data life cycle
* Want to find a template to create a DMP
* Want to have access to aggregated information for data usage patterns from different sources and/or domain
* Want to access to description on data quality validation mechanisms
* Provide guidance to researchers for proper management of health-related data, including the implications of data access and sharing

NEEDS
* Have access to the latest tools and resources to do proper data management of health-related data for its secondary use in research activities
* Find guidelines on good data management practices for health data usage
* Best practices, guidelines and community-driven common data models
* Best practices, guidelines and tools for creating and maintaining up-to-date data management plans

CHALLENGES
* Low understanding by researchers of the full data life cycle
* Poor understanding by researchers of the implications of working with health-related data
* Lack of references for preparing and keep up-to-date data management plans for health-related data

EXPECTATIONS
* Exemplary use-cases on how to manage data in the different steps of the data life cycle
* Community efforts for good data management practices
* Registry of tools and resources for managing health data
* Dedicated space in the portal for finding reference information for and tools for creating and maintaining up-to-date research data management plans

DATA GENERATION AND USAGE
Data provider
Data user

CAREER STAGE
Beginner
Intermediate
Advanced

PROFESSIONAL SECTOR
Academia
Clinical
Industry

LEGAL BASIS
Data controller
Data processor

INTENDED USE OF DATA
Technology-oriented
Usage-oriented

TEMPORAL SCALE OF ACTIVITIES
Short term
Long term
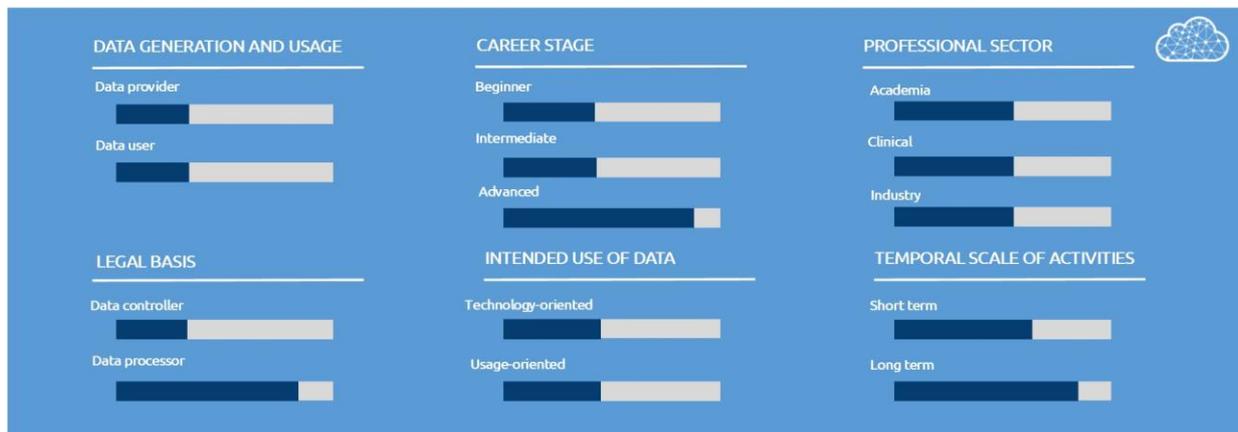
5.8 Infrastructure provider

JENN
INFRASTRUCTURE PROVIDER

Age: 45
Occupation: Data hub
Interests: HPC
Computational facilities

"I want to facilitate access to the data that is under my responsibility to those who have permission for it."

GOALS

* Want to facilitate data deposition and management
* Want to promote visibility of the available data
* Want to work towards the recognition of the data providers
* Want to facilitate data access to those who has the rights for it
* Want to enable access control to data providers to manage access to available data
* Provide access to aggregated information for data usage patterns from different sources and/or domains
* Want to have access to description on data quality validation mechanisms

NEEDS

* Increase visibility across the research community, making sure that data providers understand the legal and ethical implications of sharing data
* Be part of existing registries and catalogues for infrastructure providers to share best practices and raise general awareness
* Best practices, guidelines and community-driven common data models
* Easy-to-follow instructions for submitting data and associated metadata to the infrastructure provider
* Best practices, guidelines and tools for creating and maintaining up-to-date research data management plans

SKILLS

* Technical skills to facilitate different mechanisms for data access and sharing
* A deep understanding on the legal basis for data sharing
* Knowledge of the existing ontologies and controlled vocabularies for describing metadata associated to the managed data.

CHALLENGES

* Lack of visibility across the research community
* Low data management practices, which impacts on data description (metadata) and quality
* Lack of sustainable funding for the infrastructure
* Limited awareness on the existing mechanisms for trustworthy and secure data access and sharing by data providers

EXPECTATIONS

* Data is findable through the portal
* easy-to-register as data provider into the portal, including sharing data models and other metadata descriptors
* Single sign-on mechanisms available through the portal for facilitating users recognition and authorization on the connected data providers
* Reference space in the portal for visualizing indicators on the impact of health-related data usage

DATA GENERATION AND USAGE

Data producer
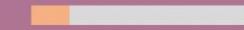
Data consumer

CAREER STAGE

Junior

Middle-level

Senior
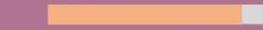
PROFESSIONAL SECTOR

Academia

Clinical

Industry

LEGAL BASIS

Data controller

Data processor

INTENDED USE OF DATA

Technology-oriented

Usage-oriented

TEMPORAL SCALE OF ACTIVITIES

Short term

Long term

6 Users' interactions with the FAIR data portal

Based on the user profiles definitions, the following table contains a summary of the exemplary user interactions with the FAIR health data portal.

Table 1. Expected interactions of the different user profiles with the future FAIR health data portal.

	Citizen	Researcher	Policy and decision maker	Healthcare professional	Data curator	Data steward	Data manager	Infrastructure provider
A place to share their data in a secure environment, including easy-to-follow instruction on how to do it	x	x		x				x
Find research outcomes about a specific topic	x	x		x				
Do analysis with the data		x						
Reference place for identifying existing cohorts and creating new (virtual) ones, including documentation		x		x				
Effective programmatic means to discover/access/process data		x						
Access summarized information about healthcare trends in the general and/or disease-specific population	x	x	x					
Easy-to-combine aggregated information from different sources and/or domains			x					
Aggregated information for data usage patterns from different sources and/or domains			x		x	x	x	x
Reference place for identifying geographically distributed infrastructure providers for data management and analysis, including documentation		x						x
Reference place for best practices, guidelines and tools for working with sensitive data, including data quality		x			x		x	
Reference place for best practices and guidelines for developing and validating data-driven protocols for preventing, diagnosing and treating individual diseases				x				
Reference place for best practices, guidelines and tools to evaluate the FAIRness of datasets and contribute towards their FAIRification						x		
Access to description on data quality validation mechanisms, ideally driven by community standards					x		x	x
Reference place for best practices, guidelines and tools to work with domain-specific controlled vocabularies and ontologies					x	x		x

Reference place for best practices, guidelines and community-driven common data models							x	x
Reference place for best practices, guidelines and tools for creating and maintaining up-to-date Data Management plans (DMPs)							x	x

7 Next steps

The current effort will be a key contribution for developing the architecture of the FAIR health data portal (deliverable 6.2) and beyond. For instance, it will serve to ensure that all relevant stakeholders have been identified in the context of the project. Regarding the portal itself, this document contains the definition of the users' profiles interacting with it. Indeed, prototypical personae were defined to represent specific users and their goals, challenges, and needs when using this portal. Taking all together has allowed us to define the first set of users' interactions with the FAIR health data portal. Expected interactions are an important asset for defining the reference architecture of the health portal.

In addition, the work done on this deliverable will also be an important step towards establishing agreed definitions of commonly used terms to harmonise the work in HealthyCloud. For that same reason, part of the nomenclature used for the user profiles categories definition were adapted from the HealthyCloud glossary[4]. The HealthyCloud glossary is a project level effort to define core concepts across the project. These included the data producer category which was changed to data provider, and the data consumer category, which was changed to data user. Also, the infrastructure provider profile was added to the HealthyCloud glossary as a parent definition of other infrastructures that were already part of the glossary like data hub, data collection and secure processing environments.

---

[4] https://zenodo.org/record/5998128#.YzLEiXZByUk

Annex A - User profiles data

This annex contains the results of the user profiles dedicated workshops described in the Methodology section (Section 0

Citizen

	Data provider	Data user	Data controller	Data processor	Beginner	Intermediate	Advanced	Technology-oriented	Use-oriented	Academia	Clinical	Industry	Long term	Short term
Technology interested	4	3	5	1	1	1	1	4	3	1	1	1	2	3
Usage interested	2	5	2	2	1	1	1	2	4	1	1	1	2	3
Sports or health interested	4	4	5	1	1	1	1	4	3	1	1	1	4	4
Rare/chronic/long-term interested	4	4	4	4	1	1	1	4	3	1	1	1	5	3
Average	3,50	4,00	4,00	2,00	1,00	1,00	1,00	3,50	3,25	1,00	1,00	1,00	3,25	3,25
Median	4	4	5	1	1	1	1	4	3	1	1	1	2	3

Researcher

	Data provider	Data user	Data controller	Data processor	Beginner	Intermediate	Advanced	Technology-oriented	Use-oriented	Academia	Clinical	Industry	Long term	Short term
Scientific researcher	5	5	4	5	3	3	3	3	5	3	3	3	3	4
Technical-oriented researcher	4	5	2	2	1	1	1	5	3	1	1	5	2	4
Average	3,75	5,00	3,50	2,75	1,50	1,50	1,50	3,50	4,50	2,50	3,00	2,50	2,25	4,00
Median	3	5	4	2	1	1	1	3	5	1	3	1	2	4

Policy and decision maker

	Data provider	Data user	Data controller	Data processor	Beginner	Intermediate	Advanced	Technology-oriented	Use-oriented	Academia	Clinical	Industry	Long term	Short term
	1	5	1	1	1	2	4	1	5	5	5	5	5	3

Healthcare professional

	Data provider	Data user	Data controller	Data processor	Beginner	Intermediate	Advanced	Technology-oriented	Use-oriented	Academia	Clinical	Industry	Long term	Short term
Clinician / Doctor	5	4	1	1	3	3	3	1	5	1	5	1	5	2
Study nurse	5	2	1	1	3	3	3	1	5	1	5	1	2	4
Average	5,00	3,00	1,00	1,00	3,00	3,00	3,00	1,00	5,00	1,00	5,00	1,00	3,50	3,00
Mode	5		1	1	3	3	3	1	5	1	5	1		

Data curator

	Data provider	Data user	Data controller	Data processor	Beginner	Intermediate	Advanced	Technology-oriented	Use-oriented	Academia	Clinical	Industry	Long term	Short term
Data Curator	4	4	1	5	1	3	3	4	1	3	2	3	3	2

Data steward

	Data provider	Data user	Data controller	Data processor	Beginner	Intermediate	Advanced	Technology-oriented	Use-oriented	Academia	Clinical	Industry	Long term	Short term
Data steward	1	3	1	3	3	3	3	3	3	1	1	1	4	3

Data manager

	Data provider	Data user	Data controller	Data processor	Beginner	Intermediate	Advanced	Technology-oriented	Use-oriented	Academia	Clinical	Industry	Long term	Short term
Data Manager	2	2	5	5	1	1	5	3	3	3	3	3	4	3

Infrastructure provider

	Data provider	Data user	Data controller	Data processor	Beginner	Intermediate	Advanced	Technology-oriented	Use-oriented	Academia	Clinical	Industry	Long term	Short term
Data collection	1	1	4	2	1	1	1	4	4	3	3	3	4	3
Data hub	1	2	1	3	1	1	1	4	4	5	3	4	2	4
Data repositories or clinical registers	2	2	2	4	1	1	1	5	4	4	4	3	4	3
Average	1,33	1,67	2,33	3,00	1,00	1,00	1,00	4,33	4,00	4,00	3,33	3,33	3,33	3,33
Median	1	2			1	1	1	4	4		3	3	4	3