FAIR4Health: Improving Health Research in EU through FAIR Data

Vasiliki Foufi
Division of Medical Information Sciences
Geneva University Hospitals
University of Geneva
Geneva, Switzerland
vasiliki.foufi@unige.ch

Jean-Philippe Goldman
Division of Medical Information Sciences
Geneva University Hospitals
University of Geneva
Geneva, Switzerland
jean-philippe.goldman@unige.ch

Jessica Rochat
Division of Medical Information Sciences
Geneva University Hospitals
University of Geneva
Geneva, Switzerland
jessica.rochat@hcuge.ch

Carlos Luis Parra Calderón
Andalusian Health Service (SAS)
Seville, Spain
carlos.parra.sspa@juntadeandalucia.es

Christophe Gaudet-Blavignac
Division of Medical Information Sciences
Geneva University Hospitals
University of Geneva
Geneva, Switzerland
christophe.gaudet-blavignac@hcuge.ch

Christian Lovis
Division of Medical Information Sciences
Geneva University Hospitals
University of Geneva
Geneva, Switzerland
christian.lovis@hcuge.ch

***Abstract*—*This paper presents an overview of the FAIR4Health European project (2019-2021) which aims at encouraging the reuse of research data generated by publicly funded research projects. The project is coordinated by the Virgen del Rocío University Hospital, Andalusian Health Service (SAS) and the consortium consists of 17 partners from 11 EU and non-EU countries. A technological platform and tools are being developed for data FAIRification and data mining tasks. To test the feasibility of the technological solutions on real data, two pathfinder case studies will be performed.*

Keywords*—*FAIR data, interoperability, data sharing, data mining

I. Introduction

Sharing meaningful data is an important challenge in the field of personalized medicine. Besides legal and ethical aspects, there are technical challenges to be met so as to handle massive amounts of multimodal and heterogeneous distributed data, and also semantic challenges in order to build an interoperable framework. The H2020 FAIR4Health (F4H) project (2019-2021) aims at facilitating and encouraging the European Union (EU) Health Research community to FAIRify, share and reuse datasets derived from publicly funded research initiatives. A user-centered FAIR4Health platform and F4H agents are being developed to enable the translation from raw (meta) data to FAIR (meta) data. To validate the F4H platform and demonstrate the feasibility of FAIRifying medical research data, two pathfinder case studies prototypes will be performed on real medical data. The F4H consortium consists of 17 partners from both public and private entities. (*FAIR4Health - Consortium*, n.d.)

II. The FAIR4Health Project

A. Objectives

The F4H EU project, coordinated by the Virgen del Rocío University Hospital, Andalusian Health Service (SAS), consists of:

- 6 health research organizations;
- 2 universities, experts in data management;
- 4 academic partners with strong background on medical informatics;
- 5 business actors.

The ultimate goal of the project is to facilitate and encourage the EU Health Research community to FAIRify, i.e. make findable, accessible, interoperable and reusable, share and reuse their datasets derived from publicly funded research initiatives. The specific objectives are:

1. To design and implement an effective outreach strategy at EU level.

2. To produce a set of guidelines to set the foundations for a FAIR data certification roadmap.

3. To develop and validate an intuitive, user-centered F4H platform and F4H agents.

4. To demonstrate the potential impact in health research and health outcomes.(*FAIR4Health - Project*, n.d.-a)

B. Challenges and technological solutions

To meet these objectives, various aspects should be taken into account: legal and ethical aspects, technical aspects as well as semantic aspects. To address these challenges, various technological solutions are being developed in the framework of the F4H project.

First of all, FAIRification tools (*FAIR4Health - Newsletter*, n.d.) are being built to enable users transform the raw data into FAIR datasets. These tools are composed of:

- a repository created based on the Health Level Seven International (HL7) Fast Healthcare Interoperability Resources (FHIR) standard; (*Overview - FHIR v4.0.1*, n.d.)
- a Health Digital Terminology for concept translation and mappings (SNOMED CT (*SNOMED Home Page*, n.d.), LOINC (*Home*, n.d.), ICD (*Classification of Diseases (ICD)*, n.d.), etc.);
- a Data Curation & Validation tool to connect health data sources and migrate data into the HL7 FHIR Repository;
- a Data Privacy tool for privacy challenges on sensitive health data via de-identification and pseudonymization techniques.

The technological solution that will support this project will be based on two main entities: The F4H platform and F4H agents. The F4H agents, which will be located at the data owner's premises, will enable the FAIRification of local datasets. At the end of this process, datasets will be normalized, curated and mapped to domain vocabularies and ontologies, acting like Data FAIRports. The agents will also host instances of the Privacy-Preserving Distributed Data Mining (PPDDM) services so they could be run locally without the need of hosting these datasets outside the owner's premises. (*FAIR4Health - Newsletter*, n.d.)

Figure 1 displays the main architecture of the F4H platform and the interaction among the F4H community

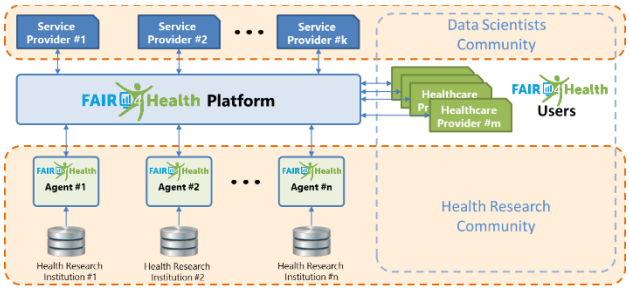

Fig. 1. FAIR4Health Open Community. (FAIR4Health - Project, n.d.-b)

As shown in the figure, the health researchers will be the main contributors in the FAIRification tasks. In turn, they will be able to browse and access to FAIRified datasets from their peers, always in compliance with the data owners' policies in terms of licenses, an approval from the local ethics committee, the regulatory framework, etc.

C. Use Cases

To validate the F4H platform and demonstrate the feasibility of FAIRifying medical research data, two pathfinder case studies prototypes will be performed on real medical data, both retrospective and prospective.

1. The first use case will focus on the identification of multimorbidity patterns and polypharmacy correlation on the risk of mortality in elderly patients via a multicentric observational study on datasets from 5 different European cohorts.

2. In the second use case, an early prediction service for 30-days readmission risk in patients with Chronic Obstructive Pulmonary Disease (COPD) will be developed. For this goal, both a retrospective and a prospective observational study will be carried out. (*FAIR4Health - Newsletter*, n.d.)

III. Conclusion

The F4H project goes beyond the health research domain and addresses the beneficial impact that the FAIR data strategy may have in health outcomes as well. There is a triple win behind this community:

1. Health researchers could access large datasets and accelerate the discovery of knowledge while avoiding bias due to local datasets;

2. eHealth services providers could develop and exploit innovative services in the EU Digital Single Market;

3. Healthcare providers could have access to this eHealth services portfolio to improve their quality of care. (*FAIR4Health - Project*, n.d.-a)

Interested parties who would like to benefit from the FAIRification workflow and tools developed by FAIR4Health, and the knowledge gained, can join the FAIR4Health community via the website https://www.fair4health.eu/en/membership.

Acknowledgment

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

References

Classification of Diseases (ICD). (n.d.). Retrieved May 5, 2021, from https://www.who.int/standards/classifications/classification-of-diseases

FAIR4Health—Consortium. (n.d.). Retrieved February 11, 2021, from https://www.fair4health.eu/en/partners

FAIR4Health—Newsletter. (n.d.). Retrieved February 11, 2021, from https://www.fair4health.eu/en/newsletter

FAIR4Health—Project. (n.d.-a). Retrieved February 11, 2021, from https://www.fair4health.eu/en/project#

FAIR4Health—Project. (n.d.-b). Retrieved February 11, 2021, from https://www.fair4health.eu/en/project#

Home. (n.d.). LOINC. Retrieved May 5, 2021, from https://loinc.org/

Overview—FHIR v4.0.1. (n.d.). Retrieved May 5, 2021, from https://www.hl7.org/fhir/overview.html

SNOMED Home page. (n.d.). SNOMED. Retrieved May 5, 2021, from /