Approaches to the integration of TRUST and FAIR principles

Jaime Delgado [1], Celia Alvarez-Romero [2], Alicia Martínez-García [2], and Carlos Luis Parra-Calderón [2]

[1] *Universitat Politècnica de Catalunya (UPC BarcelonaTECH), Campus Nord, C/ Jordi Girona, 1-3, Barcelona, 08034, Spain*
[2] *Group of Research and Innovation in Biomedical Informatics, Biomedical Engineering and Health Economy, Institute of Biomedicine of Seville, IBiS / Virgen del Rocío University Hospital / CSIC / University of Seville, Avda. Manuel Siurot s/n, Seville, 41013, Spain*

Abstract

FAIR4Health, an EU's funded project, promotes the application of FAIR principles (Findable, Accessible, Interoperable, Reusable) in data derived from publicly funded health research initiatives to share and reuse them in the European Union Health Research community, defining an effective EU-wide strategy for the use of FAIR in Health.

This paper analyses how to apply FAIR principles to "trust" in the context of the open-source software development of FAIR4Health. The FAIRification process and TRUST principles are discussed and related. The paper tries to open a new view on the trustworthiness of data access using open-source software.

Keywords

FAIR Principles, Health Data, TRUST Principles

1. Introduction

The FAIR Data principles [1] aim to ensure that data are shared in a way that enables and enhances reuse by humans and machines. Although FAIR (Findable, Accessible, Interoperable, Reusable) emerged from a workshop for the life science community, the principles are intended to be applied to data and metadata from all disciplines.

Since the formal release via the FORCE11 community [1], the FAIR data principles have been adopted by several funders and governments worldwide. The European Commission data management guidelines were updated in 2017 to introduce the notion of FAIR. On the other hand, the EOSC Declaration [2] launched following the Summit in June 2017 and the recent Staff Working Document proposing an Implementation Roadmap for the European Open Science Cloud [2] both emphasize the central role of FAIR data.

FAIR is being adopted by a diverse range of research disciplines. Several groups have been assessing uptake to date and the challenges being encountered. FAIR4Health [3] and other projects add to the State-of-the-Art by documenting good practices and applying them to other domains where possible.

2. FAIR principles and its software implementation

FAIR4Health has designed and implemented a workflow [4] to apply the FAIR principles to health research data, based on the FAIRification process of GO FAIR [5], but addressing the ethical, legal and

technical aspects that health data include due to their nature. These aspects have been analysed for sensitive data, and FAIRification tools, based on the use of the HL7 FHIR standard [6], have been developed to obtain FAIR data from data resulting from biomedical research.

FAIRification tools are standalone, open-source desktop applications developed by the FAIR4Health project. They include:

- Data Curation Tool: used to connect the health data sources, which can be in various formats (Excel files, CSV files, SQL databases), and migrate data into an HL7 FHIR Repository. The tool shows the available FHIR profiles to the user so that he/she can perform mappings appropriately. The tool can also contact a Terminology Server (which is actually another HL7 FHIR Repository) so that data fields can be annotated if coding schemes such as ICD10 [7] or SNOMED-CT [8] are in use.
- Data Privacy Tool: aims at handling the privacy challenges exposed by the sensitive health data. It is designed to work on an HL7 FHIR API so that it can be used on top of any standard FHIR Repository as a data de-identification, anonymization, and related actions toolset. The tool accesses FHIR resources, presents metadata to the user, guides the user about the configuration to be applied and then outputs the processed FHIR resources.

The architecture of the FAIR4Health ecosystem, as depicted in Figure 1, shows the high-level process and interoperability of different components.
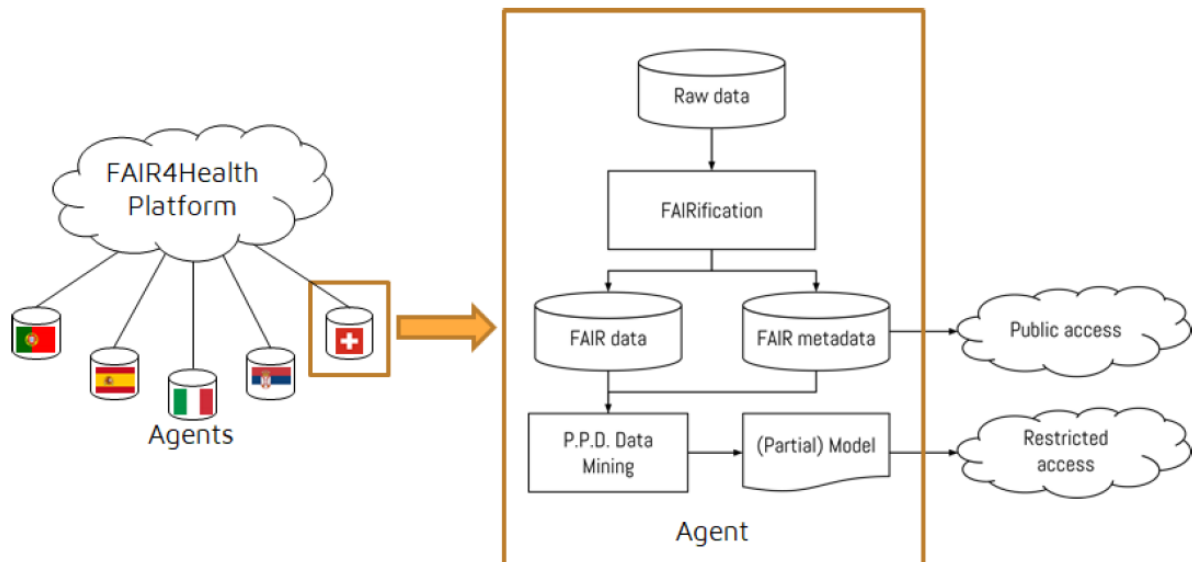
Figure 1: Overview of the architecture and components of the FAIR4Health ecosystem (FAIR4Health agents and FAIR4Health platform)

The FAIR4Health agents, located at the data owner's facilities, enable the application of the FAIRification process to local datasets through their user-driven Extraction and Loading Transformation functionalities. At the end of this process, the datasets are standardized, curated and mapped to domain vocabularies and ontologies. The agents also host instances of Privacy Preserving Distributed Data Mining services so that they can operate locally without the need to host these datasets outside the owner's facilities.

Furthermore, the FAIR4Health platform functions as a facilitator to deploy innovative data-based services, hosting the repository of actionable data mining models that could be executed. The exchange of datasets is properly secured on an end-to-end basis to minimize the risk of information disclosure.

3. TRUST principles approach

This section analyses the concepts related to the TRUST principles and then maps them to the FAIRification process.

3.1. TRUST principles concepts

One of the possible approaches to "trust" is through the so-called "TRUST principles". We first consider the approach taken at the eTRICKS project [9]. This project developed a "Recommended Data sharing model" for which a specific version of TRUST principles was defined.

The defined TRUST principles by eTRICKS are:

- Transparency: data subjects are informed of data users' requests, if they wish, and data breaches, when required, according to what the GDPR establishes [10].
- Reciprocity and reward: the contribution in a study of data subjects, data providers, and data users is acknowledged or rewarded.
- Universality: the use of data is open to any registered data user, provided that use is authorized by a national law and/or a data subject.
- Security: data are processed in a controlled environment. Data users and their requested processes are recorded for auditing purposes.
- Tiered data use: the authorization of data use depends on the data type, the analysis purpose, the data user's profile, the analytical algorithm that a data user wants to use, and the data subject's will.

However, the TRUST principles defined by eTRICKS is not the only approach. Another relevant definition of TRUST principles [11] comes from the RDA (Research Data Alliance) [12]. The defined TRUST principles by the RDA are:

- Transparency: to provide publicly accessible evidence of the services offered by a repository.
- Responsibility: a commitment to provide reliable data services and authenticity and integrity of data.
- User focus: to ensure the implementation and enforcement of the standards and data management norms of a specific user community.
- Sustainability: the capability to support long-term data preservation and use.
- Technology: the infrastructure and capabilities to support secure, persistent, and reliable services.

These 5 principles by the RDA have several associated key concepts, such as data provenance and data curation for *Transparency*, data quality and sharing of *Responsibility* for the 2_{nd} principle, engagement and responsiveness for a broad *User focus*, implication of all the partners for *Sustainability*, and features such as reliability, flexibility, scalability, transferability, security and agility for *Technology*.

In the context of FAIR principles, we try to combine and integrate the two different approaches to TRUST principles. One possibility is the following:

- Transparency: data subjects informed (users' requests, data breaches, publicly accessible evidence of the services).
- Responsibility & Reciprocity: data subjects, providers, users; Commitment to provide reliable data services.
- Universality and standards: access for registered users (if authorized) and use of (implement and enforce) standards.
- Sustainability: capability to support long-term data preservation and use.
- Technology: infrastructure and capabilities to support the repository operations, including Security (controlled data processing and auditing) and Tiered data use.

3.2. TRUST principles and FAIRification

In order to select the specific TRUST principles concepts to further analyse them in the context of software development, we need to consider other issues, such as the FAIRification workflow, so we could map TRUST concepts to workflow steps. As mentioned before, FAIR4Health has adopted its own workflow.

The different steps of the FAIR4Health's FAIRification workflow [4] could be summarized as:
- 1. Raw data analysis
- 2. Data curation & validation
- 3. Data de-identification / anonymization
- 4. Semantic modeling
- 5. Make data linkable
- 6. License attribution
- 7. Data versioning
- 8. (Meta)data aggregation
- 9. Archiving

An example of mapping between TRUST principles (adapted to the FAIR context, as described before at the end of subsection 3.1) and FAIRification workflow steps could be:
- Transparency: steps 1 and 9.
- Responsibility & Reciprocity: steps 1, 6 and 8.
- Universality and standards: steps 2, 3, 5 and 7.
- Sustainability: steps 2, 7 and 9.
- Technology: steps 3, 4, 5 and 6.

Other mapping options could be possible depending on the specific use cases.

Nevertheless, mapping TRUST principles to the FAIRification steps is not the only option we consider in order to develop a methodology for trustable open-source software implementations. It is also interesting to consider the steps related to the process of accessing information offered through FAIR principles, which are in any case related to FAIRification. A simple example of steps could be Data Acquisition, Data Processing/normalization and Data Request (by patient, health professional or researcher).

4. Conclusions

The FAIR principles open a new approach to the sharing of information in different areas of research. The FAIR4Health project focuses in promoting and implementing these principles in the context of health data. The software being developed is open-source.

The paper rises a new issue in the consideration of trustworthiness, not only for the software but also for the information to share. Our approach to trust is based on the so-called TRUST principles, to which there is no full agreement yet. Therefore, we are trying to refine these TRUST principles in the context of the FAIRification process. The result will be a methodology to keep trust and FAIR principles while developing open-source software for providing access to health data. We still need more work to produce something formal, but we have tried to define the approach to follow.

Finally, concerning future work, apart from the continuation of the analysis started here, we are working on another methodology for adding security and privacy to FAIR-oriented systems and to the different associated steps.

5. Acknowledgements

6. References

[1] The FAIR Data Principles. URL: https://www.force11.org/group/fairgroup/fairprinciples.
[2] The EU's open science policy. URL: https://ec.europa.eu/info/research-and-innovation/strategy/goals-research-and-innovation-policy/open-science_en.
[3] FAIR4Health Project. URL: https://www.fair4health.eu.
[4] Sinaci, A. A., Núñez-Benjumea, F. J., Gencturk, M., Jauer, M. L., Deserno, T., Chronaki, C., ... & Parra-Calderón, C. L. "From raw data to FAIR data: the FAIRification workflow for health research". Methods of Information in Medicine, 59(S 01), e21-e32 (2020).
[5] GO FAIR initiative. URL: https://www.go-fair.org.
[6] HL7 FHIR website. URL: http://hl7.org/fhir.
[7] World Health Organization. Classification of Diseases (ICD). URL: https://www.who.int/classifications/icd/en.
[8] Systematized Nomenclature of Medicine -- Clinical Terms (SNOMED CT). URL: http://www.snomed.org.
[9] eTRIKS (European Translational Information and Knowledge Management Services). URL: https://www.etriks.org.
[10] GDPR (General Data Protection Regulation): Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016, Official Journal of the European Union, L 119, 4 May 2016. URL: https://gdpr-info.eu.
[11] Lin, D. (2019). The TRUST Principles for Trustworthy Data Repositories – An Update. RDA/WDS Repository Certification IG. In: RDA/WDS Certification of Digital Repositories IG. URL: https://www.rd-alliance.org/groups/rdawds-certification-digital-repositories-ig.html.
[12] Research Data Alliance. URL: https://www.rd-alliance.org.