### Overview

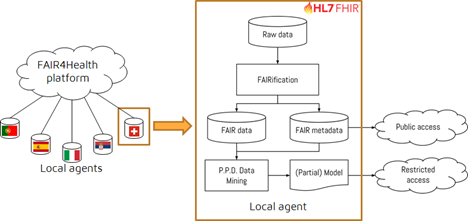
[FAIR4Health](https://www.fair4health.eu/) is a project funded by the European Union Horizon 2020. It is carried out by 17 partners from 11 different countries and lasts 36 months, bringing together expertise from different domains (health research, data managers, medical informatics, software developers, standards and lawyers).

The overall objective of FAIR4Health is to facilitate and encourage the European Union **health** research community to **FAIRify**, that is, to augment, share and reuse datasets derived from publicly funded research initiatives, demonstrating the FAIR strategy’s potential impact on health outcomes and health research. It focuses on two main use cases:

1. FAIR4Health Pathfinder Use case 1 (P1): to support the discovery of disease onset triggers and disease association patterns in comorbid patients and demonstrate the reproducibility of research

2. FAIR4Health Pathfinder Use case 2 (P2): to develop and pilot a prediction service for 30-days readmission risk in patients with COPD

This project has developed an intuitive, user-centered **technological platform** to enable the translation from raw (meta)data to FAIR (meta)data, and the use of this FAIRified information to execute Privacy-preserving Distributed Data Mining (PPDDM) algorithms.



Raw data extracted from real healthcare institution EHRs across Europe, are FAIRified and converted to HL7 FHIR resources complying to a common data model.

The studies are made accessible through the FAIR4Health portal (<https://portal.fair4health.eu/>) describing the covered use cases, the number of patients involved to generate the models in each case, and the number of contributing sites. Computable metadata concerning the health data sets published by each site is also given.

Published health datasets cannot be accessed directly, but they can be reused by authorized researchers in the context of specific use cases through algorithmic access via FAIR4Health Agents.

### Using HL7 FHIR

The raw healthcare data considered for the purpose of the project at the subject level have been:

* Data related to patients (pseudoanonimized identifier, gender, age, country,...)
* Information from encounters of these patients: identifier, start and end date, type of encounter (planned vs unplanned), ICD-10 codes, drugs prescription at discharge (identifier, code).
* Specific observations, such as: smoking status, institutionalized status, mortality status, domiciliary oxygen prescribed, result of the modified Medical Research Council (mMRC) Dyspnea Scale.
* Results of laboratory tests: identifier, code, quantity, unit, date.

These data have been represented in HL7 FHIR by [Patient](https://www.hl7.org/fhir/2021Mar/patient.html), [Encounter](https://www.hl7.org/fhir/2021Mar/encounter.html), [Condition](https://www.hl7.org/fhir/2021Mar/condition.html), [MedicationStatement](https://www.hl7.org/fhir/2021Mar/medicationstatement.html), [Observation](https://www.hl7.org/fhir/2021Mar/observation.html) resources, in accordance with the agreed FAIR4Health FHIR profiles. These profiles also indicate the commonly agreed vocabularies to be used.

For each set of resources generated by the FAIR4Health FAIRification process, a [Provenance](https://www.hl7.org/fhir/2021Mar/provenance.html) resource is created to document such a transformation; a [DocumentManifest](https://www.hl7.org/fhir/2021Mar/documentmanifest.html) resource is used to provide minimal documentation about data set metadata including the licence of use and the [CapabilityStatement](https://www.hl7.org/fhir/2021Mar/capabilitystatement.html) is used to describe the capabilities offered by each site.

### Lessons learned

#### Metadata

The practical application of the requirements about distinct, identifiable and rich metadata is not so straightforward: this requires a common understanding on what ‘metadata’ is (easy in theory, less easy in practice) and an agreement on what is the minimal set of information defining the metadata, decision that needs to take in consideration the actual capability to provide and represent this minimal set.

In this initial phase the project has chosen to consider a minimal starting set of information defined only for the project and study level. No distinct subject level metadata have been implemented. Some of these metadata have been published for human readability (case study - project level) through the welcome page of the FAIR4Health portal; others, enabling machine processing, have been defined by using FHIR Provenance, DocumentManifest and CapabilityStatement resources and published for the time being in a GitHub repository.

On this topic a possible future improvement will be the increase of the quantity of information provided as machine processable metadata, considering the adoption of alternative resources such as the FHIR Library or the Citation to better describe richer study level metadata and their publication through FHIR API. Moreover, a better formalization of what is intended as rich metadata in the associated FHIR IG.

#### FHIR data, FAIR by design

Even though the Data Curation tool developed by the project to enable the transformation of the raw data made available by the pilot sites into FHIR resources, facilitated the realization of interoperable and reusable data. Having native FHIR health data, designed to fulfill FAIR principles, is the way to overcome all the experienced shortages, potentially more and more critical when the domain of application will be extended.

Added value of FHIR

Even if not all the capabilities offered by HL7 FHIR have been used at this stage by the project, the adoption of HL7 FHIR gave a strong contribution in terms of data reusability and interoperability. Allowing to agree, formalize, share and adopt a common health data model for the shared data within the project community.