

FAIR Software as a Service: Combining tools to implement FAIR for clinical data platforms

Introduction

Semantic Web technology is a common implementation choice for software services that help meet the machine actionability requirement of the FAIR principles. IT departments and Data Competency Centres (DCCs) have strong operational expertise in deploying software services, but lack expertise in using Semantic Web technology. An example is the DCAT-based FAIR Data Point (FDP) [1]. Although local deployment is lightweight for Semantic Web experts and non-invasive for institutional infrastructure, adoption is hampered if engineers are not yet familiar with the underlying technologies of FAIR services. This poses a risk for delivering the FAIR foundation of a federated health data infrastructure that is ready for federated analytics, AI, and machine learning.

We present our first results of a project that aims to make deployment of FAIR services by institutional DCCs cost-effective, scalable, and sustainable by delivering 'FAIR Software as a Service' (FAIR-SaaS). FAIR-SaaS is open-source software that is preconfigured and updated by FAIR specialists for institutional DCCs that can be run locally or in the (SURF) cloud. The project starts with (i) common FAIR Data Point (FDP) configurations that DCCs can deploy to describe their resources for machines, (ii) configuration files using commonly used formats, such as yaml, facilitating DCCs to generate common APIs from FAIR metadata.

An early target is to simplify the processes for DCCs to contribute FDPs for data and compute capacity to the Dutch national health infrastructure, and comply with multiple existing and future common APIs. FAIR SaaS are straightforward to update for FAIR experts (e.g. grlc SaaS requires simple queries on FAIR metadata) who can guarantee convergence to international standards. The hypothesis is that FAIR SaaS provided by SURF, the Dutch national research hosting organisation, which requires a minimal number of Semantic Web experts for maintenance, has a significant impact on the applicability, scalability, and sustainability of a health data infrastructure, demonstrating how a critical mass of FAIR (meta)data can be achieved for enhancing health care and life sciences.

From concept to development

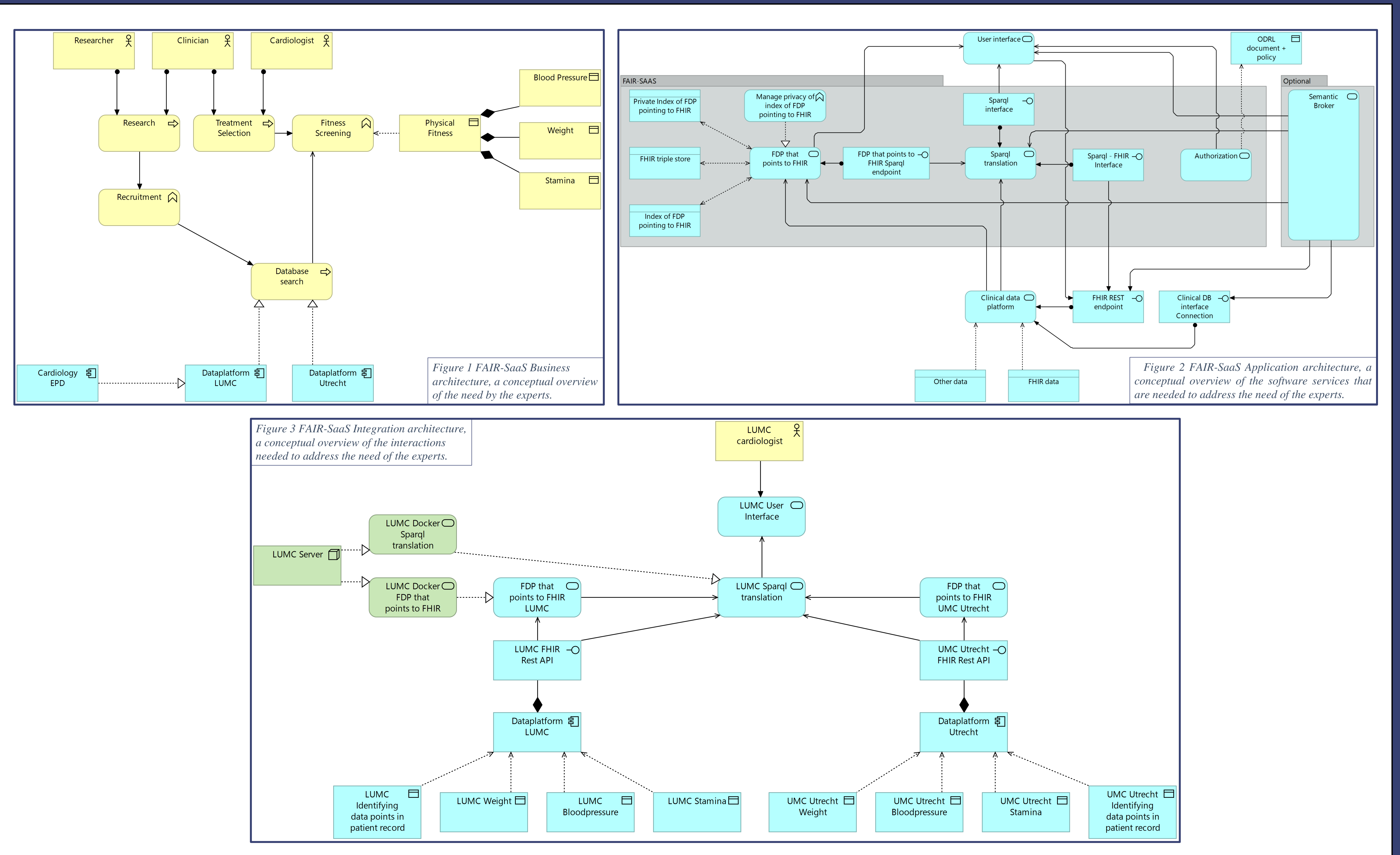
A use-case has been defined with local health data project leaders from the LUMC, UMC Utrecht and Erasmus MC, to develop the overall architecture and concept of the service. The use-case is:

As a cardiologist I would like to see information on physical fitness (weight, stamina, blood pressure, etc.) for patients planned for thorax surgery across two hospitals, such that I can screen patients based on their physical fitness for surgery and take decisions on timing of surgery and capacity planning. Patients that are not selected must undergo a fitness program.

The use-case was used to create the business architecture that has been defined in Figure 1, which resulted in the application architecture defined in Figure 2, and an integration architecture defined in Figure 3. From a FAIR mindset, we have defined that these services do not need to be created by us but should be re-used where available and adjusted if necessary.

To lower the amount of knowledge about Semantic Web technology needed by IT personnel we concluded that the first services will be delivered in a docker container and that we need to set up a Virtual Machine (VM) image with these dockers pre-installed, as shown in Figure 3. The VM is optional if docker knowledge is present at the specific IT department. We are currently combining existing tools to develop these features.

Architecture of FAIR-SaaS



References

[1] Luiz Olavo Bonino da Silva Santos, Kees Burger, Rajaram Kaliyaperumal, Mark D. Wilkinson; FAIR Data Point: A FAIR-Oriented Approach for Metadata Publication. Data Intelligence 2023; 5 (1): 163-183. doi: <https://doi.org/10.1162/dint.a.90300>

Acknowledgements

We acknowledge SURF and NWO for support and funding, Health RI, the European Joint Programme Rare Diseases and EOSC-Life for developments leading to this project (H2020 N°82557, 824087).