# Overview of a suite of middle-ware services for implementing FAIR data principles

Mark Thompson[1], Luiz Bonino[6], Mark D Wilkinson[2], Rajaram Kaliyaperumal[1], Kees Burger[6], Shamanou van Leeuwen[6], Annika Jacobsen[1], Claudio Carta[3], Erik Schultes[6], David van Enckevort[4], Richard Finkers[5], Mascha Jansen[6], Barend Mons[1], and Marco Roos[1]

[1] Leiden University Medical Centre, The Netherlands
{m.roos,m.thompson,r.kaliyaperumal,a.jacobsen}@lumc.nl
[2] Universidad Politcnica de Madrid, Spain
markw@illuminae.com
[3] Istituto Superiore di Sanitá, Italy
claudio.carta@iss.it
[4] University Medical Center Groningen, The Netherlands
david.van.enckevort@umcg.nl
[5] Wageningen Plant Research, The Netherlands
richard.finkers@wur.nl
[6] Dutch Techcentre for Life Sciences, The Netherlands
{luiz.bonino,erik.schultes,mascha.jansen}@dtls.nl

## 1 Introduction

Principles of Findable, Accessible, Interoperable, and Reusable data for humans and computers (FAIR)[1]are widely endorsed by organizations such as the European Open Science Cloud, the life science data infrastructure ELIXIR, the NIH via its commons program, the biobanking infrastructure consortium BBMRI-ERIC, the G20 and the G7. Implementing a data ecosystem based on FAIR principles requires guidelines, tools, and training, and FAIR data stewards to help apply them. The principles as such do not recommend any particular implementation: user communities will have to decide the most appropriate implementation for their domain. Here, we demonstrate the use of a suite of Semantic Web-based middle-ware services that help communities implement FAIR data principles[2]. Aiming to facilitate adoption, the services are made to complement existing data infrastructures, including local and centralised data resources, and thus establish a robust, federated ecosystem of FAIR resources. The services are also particularly suited for training data stewards. We demonstrate the application of the services by rare disease and plant breeding communities where the combination of Ontologies, Linked Data, and light-weight FAIR services are being explored as the means to implement FAIR data principles.

[1] Wilkinson et al. doi:10.1038/sdata.2016.18
[2] https://github.com/DTL-FAIRData

## 2 FAIR services

We present the following middle-ware services and tools to implement FAIR principles:

- F: a FAIR Data Search Engine based on harvesting metadata from FAIR Data Points using widely adopted metadata structures and standards (see "A" below). The FDP web interface also exposes (bio)schema.org-compatible metadata for use by third-party search engines.
- A: the FAIR Data Point RESTful API that uses the Data Catalogue Vocabulary (DCAT[3]) and Datacites Registry of Research Data Repositories (RE3Data[4]) to provide high level metadata descriptors about data deposits, and to provide instructions to access various distributions of data sets (such as both an original CSV file and its fully interoperable RDF representation).
- I: a FAIRifier tool that is based on OpenRefine and its RDF plug-in and used to convert tabular data into ontology-grounded RDF; this tool is actively used for events such as the bring your own data workshops (BYODs) , and other FAIR data training courses.
- R: for reusability we provide the FAIR Metadata Editor to support richer (meta)data to optimise future reuse of FAIR data sources, and early prototypes that apply machine readable metadata to govern access, such as license, consent, and privacy preservation.

## 3 Conclusion

In the context of our contribution to cross-national infrastructure for data stewardship in communities such as the rare disease community and the plant breeding community, we consider that tools described above, and their implementation based on Semantic Web technologies to help the adoption of the FAIR approach. The services we present are light-weight and still allow communities sufficient freedom to make design decisions together with FAIR data stewards. Training data stewards is therefore an important objective for further adoption and we found the service suite valuable for training events and BYODs. We also work with software service providers in user communities. Turning for instance patient registry software and biobank cataloguing software such as MOLGENIS into FAIR data generating tools will substantially lower the burden of FAIRification.

**Acknowledgments.** We acknowledge the participants of the BYODs, FAIR workshops and hackathons for their kind contributions and feedback, and FAIRdICT, ODEX4All, ELIXIR and ELIXIR-EXCELERATE, BBMRI-ERIC and BBMRI-NL, RD-Connect, Istituto Superiore di Sanitá in Italy, the Dutch Techcentre for Life Science (representative of ELIXIR-NL), Ministerio de Economa y Competitividad (Spain) grant number TIN2014-55993-RM, and the NBDC/ DBCLS BioHackathon series.

[3] https://www.w3.org/TR/vocab-dcat
[4] http://www.re3data.org/