German Medical Data Sciences: A Learning Healthcare System
U. Hübner et al. (Eds.)

doi:10.3233/978-1-61499-896-9-209

OSSE Goes FAIR – Implementation of the FAIR Data Principles for an Open-Source Registry for Rare Diseases

Jannik SCHAAF[a,1], Dennis KADIOGLU[b], Jens GOEBEL[a], Christian-Alexander BEHRENDT[b], Marco ROOS[c], David van ENCKEVORT[d], Frank ÜCKERT[e], Fatlume SADIKU[e], Thomas O.F. WAGNER[f] and Holger STORF[a]

[a] *Medical Informatics Group, University Hospital Frankfurt, Frankfurt, Germany*
[b] *University Medical Center Hamburg-Eppendorf, Hamburg, Germany*
[c] *Dutch Tech Centre for Life Sciences, Leiden, Netherlands*
[d] *University Medical Center Groningen, University of Groningen, Groningen, Netherlands*
[e] *Division of Medical Informatics for Translational Oncology, German Cancer Research Center, Heidelberg, Germany*
[f] *Frankfurt Reference Center for Rare Diseases, University Hospital Frankfurt, Frankfurt, Germany*

Abstract. The Open Source Registry for Rare Diseases (OSSE) provides a concept and a software for the management of registries for patients with rare diseases. A disease is defined as rare if less than 5 out of 10,000 people are affected. Up to date, approximately 6,000 rare diseases are catalogued. Networking and data exchange for research purposes remains challenging due to the paucity of interoperability and due to the fact that small data stocks are stored locally. The so called "Findable, Accessible, Interoperable, Reusable" (FAIR) Data Principles have been developed to improve research in the field of rare diseases. Subsequently, the OSSE architecture was adapted to implement the FAIR Data Principles. Therefore, the so-called FAIR Data Point was integrated into OSSE to provide a description of metadata in a FAIR manner. OSSE relies on the existing metadata repository (MDR), which is used in to define data elements in the system. This is an important step towards unified documentation across multiple registries. The integration and use of new procedures to improve interoperability plays an important role in the context of registries for rare diseases.

Keywords. OSSE, FAIR, interoperability, rare diseases, registries

1. Introduction

Approximately 30 million people with rare diseases live in the European Union (EU) [1]. According to the World Health Organization (WHO), a disease is rare if it affects less than 5 out of 10,000 people. Although, there are about 6,000 different recognized rare diseases [2]. Large-scaled collection of registry data has evolved to an important tool in

[1] Corresponding Author, Jannik Schaaf, Medical Informatics Group, University Hospital Frankfurt, Theodor-Stern-Kai 7, 60590 Frankfurt am Main, Germany; E-mail: schaaf@med.uni-frankfurt.de.

health services research. Especially in the field of research on rare events or diseases, registries or comparable real-world data sources have substantial advantages [3,4]. Registries are generally considered as a form of medical documentation. In this context, health relevant data, e.g. information on a certain disease, are collected. In addition to providing support in clinical research, they serve quality assurance and quality improvement and the description of epidemiological connections [4].

1.1. Background & Motivation: OSSE – Open-Source Registry for Rare Diseases

OSSE was developed as part of the National Action Plan for People with Rare Diseases, promoted by the German Federal Ministry of Health in the years of 2013 and 2015 [5].

OSSE provides technical and legal concepts and a software-system as a registry toolbox to collect longitudinal and medical data. Data elements are defined and stored within the metadata repository (MDR) following the International Organization for Standardization (ISO) 11179. For each data element, its data type, value range, and unit of measure are defined [6].

However, OSSE does not provide a limited interface to communicate data with registries of other software solutions. An open approach, with standardized vocabulary and ontologies would be desirable in order to facilitate a simple connection. The FAIR Data Principles have been developed for this purpose, which is explained in the next chapter. This paper describes the first steps towards the architecture extension and implementation of the FAIR Data Principles in OSSE. The focus is to build a first prototype.

1.2 The FAIR Data Principles

Interoperability of registries plays a crucial role in increasing data communication in the field of rare diseases. At present, OSSE offers the possibility to integrate other registry solutions into a registry landscape through the OSSE bridgehead, but the access is limited to the participating registries. In order to see which registry includes what kind of metadata, is currently only available via the MDR. However, only owner of an OSSE registry gain access to the MDR, which prevents other research institutes from retrieving the metadata. The FAIR Data Principles are taking this aspect into account. An important aspect of FAIR is to provide existing and new records in a interoperable format, which is legible by computers and humans. Through the semantic annotation of data and metadata, computer systems can automatically combine different data sources, resulting in greater knowledge discovery. The main reason to adopt FAIR in OSSE is to provide a standardized interface to communicate with registries of other software solutions and registries. The criteria of FAIR are examined in more detail below [7].

(1) Findable: By describing metadata, people and computers can interact with the data to search for specific records. (2) Accessible: Data is stored long-term, with defined license and access conditions, both at the level of metadata as well as the level of the data. (3) Interoperable: Data sets can be combined with other data sets. (4) Reusable: Data can be used for further research using computational methods [7].

1.3 Expose Datasets FAIR: The FAIR Data Point

The FAIR Data Point (FDP) is a software component that allows data owners to store data in a FAIR manner and, on the other hand, retrieve data records via metadata. In

addition, retrieving the data is only possible if permitted by the license conditions. FDP are used in some domains, but in this paper we focus solely on patient registries. The FDP has the following objectives: Publish datasets in a FAIR compliant way, allow to retrieve data usage information of the FDP, provide information on the available records, allow data consumers to access the data and make an interaction via GUI and API available. The API (Application Programming Interface) is used by software systems to integrate data directly into a system while the GUI (Graphic User Interface) is designed to provide users with a simple interface to access metadata [8]. In summary, the FDP has two purposes: First, it can be used as a stand-alone web application, where data owners give access to their datasets in a FAIR manner and second, it can be integrated in a larger data interoperability system, like a FAIR port, where many FAIR Data Points of different systems are available. The Dutch Tech Centre of Life Sciences (DTL) is working on a "DTL Data FAIRport" to make FAIR Data Points available to connect different systems in life sciences [8].

2. Methods

To implement the FDP, the existing OSSE architecture needs to be adapted. OSSE already has a component that is ideally suited to connect a FAIR Data Point. This includes the MDR, which provides the data of the registry as metadata. Specifically, the FDP is implemented as a new, standalone component with connections to the electronic data capture of OSSE (OSSE.EDC) and the MDR, as shown in Figure 1. The OSSE.EDC is responsible for managing and storing the data in the registry and thus contains the medical data, while the MDR provides information about the data, stored in the registry. Both components are needed to get access to the registry data and provide metadata information for the FDP. The MDR provides a Representational State Transfer (REST) interface to retrieve the metadata of the registry and it can be used for communication between systems [9]. Also, the OSSE.EDC retrieves the registry data of the OSSE.Store component, which is responsible for the database access. This component also provides a REST interface.

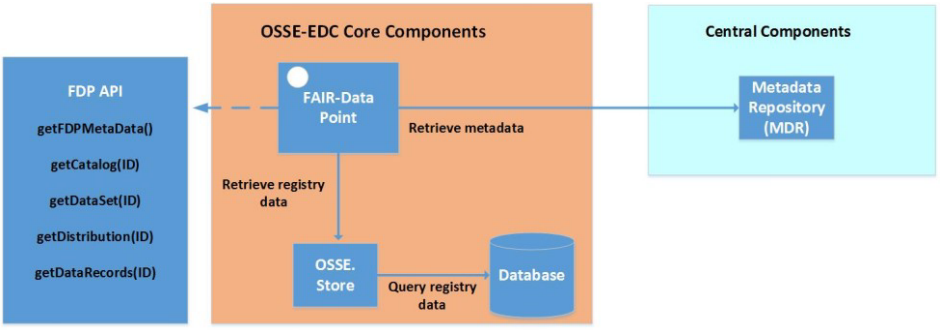

Figure 1. Extended OSSE-Architecture.

3. Results

The main goal of the FDP is to provide meta information about the data in the registry. Which means that only the descriptions of the medical data elements are available, no patient-identifying and medical data is provided. For metadata description the FAIR Data Point API (FDP-API) is the most important interface for sharing this information with other registry systems (shown in Figure 1). The FDP-API contains the Metadata Provider API. This component includes a public service for retrieving the data, which includes several layers to retrieve data. Each layer defines a set of (minimally required) metadata and recommends ontologies for that metadata. The FDP layer describes the FDP as a basic data information source (Data Repository). For example, information about the title, description, a unique ID and information about the data owner of the FDP can be obtained. The description is based on R3Data, "the Metadata Scheme for the Description of Research Repositories" [10]. Each layer points to the respective lower layer. For example, the FDP layer contains a reference to one or more catalogs. A catalog is a collection of metadata about datasets.

The Catalog, Dataset and Distribution layers are based on the Data Catalog Vocabulary (DCAT) published by the W3C in 2014. The main goal of DCAT is to improve the data catalogues interoperability and make applications easy consume metadata from multiple catalogues [11]. The catalog layer contains metadata from the catalog of the FDP. Again, a unique ID, title and description is provided. There are also references to the dataset metadata. The dataset layer gives information about each of the offered datasets. The distribution layer gives information about how the dataset is distributed. Information about the access URL, download URL, format and media type is given here. The bottom layer, data record layer, describes a collection of data, which is available for access or download in one or more formats. In OSSE, each of the described layers can be accessed via a RESTful API. To access the available catalogs, datasets, distributions and data records, only the identification of the respective element is needed. The individual layers were implemented as functions in the OSSE FDP REST API. For example, the *getFDPMetadata* function returns appropriate information about the FDP, so for every layer a function to get metadata out of the layer is available.

A corresponding RDF document is provided, which is created with the Apache Jena framework, which allows the writing and reading of RDF [12]. The concrete metadata uses the FDP according to the function *getUserRootElements* of the MDR-REST API, which returns all metadata for the specific register. In addition, OSSE offers a configuration web page for the FDP to set individual parameters. This includes, for example, an indication of the website or the name of the institution that publishes the FDP.

4. Discussion

The increasing number of FAIR infrastructures is evident. Various collaborations and research institutions as the European Reference Network for Rare Multisystemic Vascular Diseases (VASCern) are currently promoting the utilization of FAIR Data Principles. It is planned to implement a decentralized registry infrastructure with OSSE registries and the FAIR approach in the L-ACMAG (Longitudinal Study Registry of Aortic, Myocardial, Arterial and Genetics in aortic diseases) [13]. L-ACMAG aims to implement a data privacy compliant research collaboration among fifteen German

Marfan Reference Centres to improve the diagnosis and treatment of patients with rare genetic vascular diseases. To meet the changing requirements in the field of digital health care, the European Commission proposed a comprehensive reform of data protection rules in the EU. The novel regulation will apply from 25 May 2018. Time will show if registries can remain FAIR and data privacy compliant at the same time.

5. Conclusion

In the field of rare diseases, registries have evolved to an important research tool. The FAIR Data Principles have been developed to improve data exchange and networking of registries. However, semantic interoperability is not considered in FAIR. Therefore, the OSSE-MDR offers a possibility to uniformly describe and define data elements. The MDR can thus be used as an additional tool for creating semantic interoperability in a FAIR infrastructure.

6. Conflicts of Interest

The authors declare that there is no conflict of interest.

References

[1] Bundesministerium für Gesundheit Seltene Erkrankungen. Seltene Erkrankungen. Retrieved May 9, 2018, from http://www.bmg.bund.de/themen/praevention/gesundheitsgefahren/seltene-erkrankungen.html
[2] World Health Organization. Coming together to combat rare diseases. Retrieved May 9, 2018, from http://www.who.int/bulletin/volumes/90/6/12-020612/en
[3] C.A. Behrendt, F. Heidemann, H.C. Riess, K. Stoberock, S.E, Debus, Registry and health insurance claims data in vascular research and quality improvement. *Vasa* **46** (2017), 11-5.
[4] TMF. IT-Infrastrukturen in der patientenorientierten Forschung –aktueller Stand und Handlungsbedarf. Retrieved May 9, 2018, from https://www.toolpool-gesundheitsforschung.de/produkte/it-report>
[5] Geschäftsstelle des Nationalen Aktionsbündnisses für Menschen mit Seltenen Erkrankungen. Nationaler Aktionsplan für Menschen mit Seltenen Erkrankungen. Handlungsfelder, Empfehlungen Und Maßnahmenvorschläge. Retrieved May 9, 2018, from https://www.bundesgesundheitsministerium.de/fileadmin/Dateien/3_Downloads/N/NAMSE/Nationaler_Aktionsplan_fuer_Menschen_mit_Seltenen_Erkrankungen_Handlungsfelder__Empfehlungen_und_Massnahmenvorschlaege.pdf.
[6] M. Muscholl, M. Lablans, T. Wagner, F. Ückert, OSSE – open source registry software solution. *Orphanet Journal of Rare Diseases* **9(1)** (2014), 09.
[7] Dutch Tech Centre of Life Sciences. FAIR-DATA. Retrieved May 9, 2018, from https://www.dtls.nl/fair-data/fair-data/>
[8] Dutch Tech Centre of Life Sciences. FAIR Data Point Software Specification. Retrieved May 9, from https://dtl-fair.atlassian.net/wiki/spaces/FDP/pages/6127622/FAIR+Data+Point+Software+Specification.
[9] SearchMicrosservices. REST (Representational state transfer). Retrieved May 9, 2018, from http://searchmicroservices.techtarget.com/definition/REST-representational-state-transfer.
[10] Karlsruhe Institute of Technology. Registry of Research Data Repositories. Retrieved May 9, 2018 from http://www.re3data.org/>.
[11] World Wide Web Consortium (2014). Data Catalog Vocabulary (DCAT). Retrieved May 9, 2018, from https://www.w3.org/TR/vocab-dcat/
[12] The Apache Software Foundation. *Apache Jena*. Retrieved May 9, 2018, from https://jena.apache.org/
[13] Arbeitsgruppe GermanVasc Universitätsklinikum Hamburg-Eppendorf (2017). Retrieved May 9, 2018, from http://rarevasc.org/