Applying the FAIR Data Principles to the Registry of Vascular Anomalies (VASCA)

Bruna DOS SANTOS VIEIRA[a,1], Karlijn GROENEN[a], P.A.C. 'T HOEN[a], Annika JACOBSEN[b], Marco ROOS[b], Rajaram KALIYAPERUMAL[b], Martijn KERSLOOT[c,d], Ronald CORNET[c] and Leo SCHULTZE KOOL[a,e]

[a] *Radboud University Medical Center, Nijmegen, Netherlands*
[b] *Leiden University Medical Center, Leiden, Netherlands*
[c] *Amsterdam University Medical Center, Netherlands*
[d] *Castor EDC, Amsterdam, Netherlands*
[e] *VASCERN VASCA European Reference Centre*

Abstract. Background: Connecting currently existing, heterogeneous rare disease (RD) registries would greatly facilitate epidemiological and clinical research. To increase their interoperability, the European Union developed a set of Common Data Elements (CDEs) for RD registries. Objectives: To implement the CDEs and the FAIR data principles in the Registry of Vascular Anomalies (VASCA). Methods: We created a semantic model for the CDE and transformed this into a Resource Description Framework (RDF) template. The electronic case report forms (eCRF) were mapped to the RDF template and published in a FAIR Data Point (FDP). Results: The FAIR VASCA registry was successfully implemented using Castor EDC (Electronic Data Capture) software. Conclusion: FAIR technology allows researchers to query and combine data from different registries in real-time.

Keywords. registries, rare diseases, health information interoperability

1. Introduction

Registries of rare disease (RD) patients are extremely valuable for establishing genotype-phenotype relationships, natural history studies and patient's selection for research. Usually, analyses across different RD registries are difficult because each registry is set up differently and data may not be accessed nor collected uniformly. Therefore, the FAIRification of registries (e.g., making the data Findable, Accessible, Interoperable, and Reusable) might benefit researchers, for they can pool and analyse data from different registries using the same syntax[1]. Aiming at improving interoperability among registries, the Joint Research Centre produced a set of Common Data Elements (CDEs) to be adopted by RD registries[2]. This paper describes how the CDEs and the FAIR principles were implemented in the Registry of Vascular Anomalies (VASCA). VASCA is a working group of the Rare Multisystemic Vascular Diseases European Reference Network (VASCERN)[3]. Currently, the registry collects the CDEs and does not share patient summary and prescription information, as considered by the European eHealth Connecting Europe Facility (CEF) Digital Service Infrastructure (DSI).

[1] Corresponding Author: Bruna dos Santos Vieira, Radboud university medical centre, PO box 9101 6500 HB Nijmegen, E-Mail: bruna.dossantsovieira@radboudumc.nl

2. Methods

The CDEs were implemented in an electronic Case Report Form (eCRF) in Castor EDC (Electronic Data Capture)[4]. Then, a semantic model for the CDEs, which describes the relationships between CDEs and maps them to ontologies, was created and peer reviewed[5].Next, an application was developed to feed a Twig template[6], that converts the eCRF collected data into a Resource Description Framework (RDF)[7]. The generated RDF was exposed in a FAIR data point, supporting federated SPARQL queries[8,9]. Among the stakeholders involved were a patient organization, ethical committee, disease specialists, data-, and ontology experts.

3. Results

The FAIR RD registry was successfully set up, allowing for real-time data query and re-use. Various documents including manuals, approved patient information leaflet and informed consent forms, were produced. VASCA is published in the EU directory of registries (ERDRI.dor) and metadata repository (ERDRI.mdr)[10].

4. Discussion

Implementing the FAIR principles brings sustainability to the research environment. The templates and infrastructure can be reused by other registries and is facilitated by the rich documentation produced. However, their implementation needs to be adapted to local clinical workflows. VASCA's next steps entail data collection in the pilot institute, scale-up to participating centres, quality checks, federated query tests, and disease-specific data elements addition.

This work is partially supported by the funding from the European Union's Horizon 2020 research and innovation programme under the EJP RD COFUND-EJP N° 825575. One author of this publication is member of the Vascular Anomalies Working Group (VASCA WG) of the European Reference Network for Rare Multisystemic Vascular Diseases (VASCERN) - Project ID: 769036.

References

[1] A. Jacobsen, R. Kaliyaperumal, L.O. Bonino da Silva Santos, B. Mons, E. Schultes, M. Roos & M. Thompson, A generic workflow for the data FAIRification process, Data Intelligence 2 (2020), 56–65, doi: 10.1162/dint_a_00028.
[2] JRC, CDEs, https://eu-rd-platform.jrc.ec.europa.eu/set-of-common-data-elements, last access: 17.1.2020.
[3] VASCERN, Vasca-WG, https://vascern.eu/expertise/rare-diseases-wgs/vasca-wg/, last access: 17.1.2020.
[4] Castor EDC, Electronic Data Capture Solution, https://www.castoredc.com/electronic-data-capture-solution/, last access: 17.1.2020.
[5] EJP RD, ERN common data elements, https://github.com/ejp-rd-vp/ERN-common-data-elements, last access: 19.3.2020.
[6] Twig, Twig, https://github.com/twigphp, last access: 17.1.2020.
[7] W3C, RDF, https://www.w3.org/RDF/, last access: 17.1.2020.
[8] DTL, FAIR Data Tools, https://www.dtls.nl/fair-data/find-fair-data-tools/#fourth, last access: 17.1.2020.
[9] W3C, SPARQL, https://www.w3.org/2001/sw/wiki/SPARQL, last access: 17.1.2020.
[10] JRC, Erdri description, https://eu-rd-platform.jrc.ec.europa.eu/erdri-description, last access: 20.1.2020.