

The FAIRification of data and the potential of FAIR resources demonstrated in practice at the Rome Bring Your Own Data workshop

Claudio Carta¹, Marco Roos², Annika Jacobsen², Rajaram Kaliyaperumal², Mark Thompson², Mark D. Wilkinson³, Ronald Cornet⁴, Andra Waagmeester⁵, David van Enckevort⁶, Mascha Jansen⁷, Luana Licata⁸, Allegra Via⁹, and Domenica Taruscio¹

¹ Istituto Superiore di Sanità, Italy

{claudio.carta,domenica.taruscio}@iss.it

² Leiden University Medical Centre, The Netherlands

m.roos@lumc.nl, annika.jacobsen.86@gmail.com, rajireturn@gmail.com, mark9630@gmail.com

³ Centro de Biotecnología y Genómica de Plantas UPM-INIA, Spain

mark.wilkinson@upm.es

⁴ Academic Medical Center - University of Amsterdam, The Netherlands

r.cornet@amc.uva.nl

⁵ Micelio, Belgium

andra@micelio.be

⁶ University Medical Centre Groningen, The Netherlands

david.van.enckevort@umcg.nl

⁷ Dutch Techcentre for Life Sciences, The Netherlands

mascha.jansen@dtls.nl

⁸ University of Rome Tor Vergata, Italy

luana.licata@gmail.com

⁹ National Research Council, Italy

allegra.via@cnr.it

Abstract. It is widely agreed that rare disease patient registries should be international and follow the guiding principles of Findable, Accessible, Interoperable, Reusable (FAIR) for humans and computers. Furthermore, the procedures to collect and exchange data should be harmonised. Since 2014, the Bring Your Own Data (BYOD) annual workshop has been organised by and held at the National Centre for Rare Diseases - Istituto Superiore di Sanità (CNMR-ISS), Rome, Italy with the aim to promote the establishment of FAIR rare disease registries in compliance with IRDiRC and EU recommendations. The event has been arranged with the support of RD-Connect and ELIXIR, in particular, the Dutch Techcentre for Life Sciences representative of ELIXIR-NL. The general roadmap of the BYOD workshop contains at least a preparatory phase, an execution phase, and a follow-up phase to foster the results of the workshop by surveying and having phone conferences with participants. At the 4th edition of the BYOD this year, held in September, there were twenty eight participants selected on the basis of their background, registry role, and involvement in European Reference Networks

(ERNs). This edition took place in collaboration with or with the support of several additional organisations. This year the focus was on (i) the FAIRification process, with partially pre-prepared semantic models, (ii) FAIR data tools, and (iii) FAIR data management. A new element was a semantic model sketching exercise with the registry managers. At the beginning of the workshop, an overview of the FAIRification process and an introduction to four selected, scrambled, and anonymous sample datasets were provided. In the following sessions, the BYOD became highly interactive and the participants, split into four groups, went through the process of FAIRification step by step with break-out sessions alternated to plenary sessions in which participants presented the results of their group work. In the first step, each group discussed and drew a conceptual model of their assigned dataset, followed by a plenary session to discuss commonalities and complementarities. In the second step, the groups created an ontological model based on the conceptual model from the previous step which required ontology search. In the third step, the OpenRefine-FAIRifier tool and the metadata editor were used to aid in the creation of the machine readable format of the selected data. Then, to show the potential of data linkage, a cross resource question was executed over the four FAIRified sample datasets. The final part of the workshop was focused on a discussion about FAIR data management. Based on our experience from the previous editions of the workshop, the new elements were: (i) in the preparatory phase only four heterogeneous datasets among those received were selected, (ii) a breakout and plenary session at the end of each FAIRification step was introduced, (iii) a time slot was allocated for the self sketching the semantic model with pen and paper by the registry managers, (iv) a time slot was allocated to discuss FAIR data management and FAIR project planning. We concluded that the main objectives of the BYOD for registry managers were achieved. The participants were able to see the potential of a FAIR registry and this allowed them to get quicker responses to cross-resource questions improving the use of available information on rare diseases and accelerates research. In addition, participants had the chance to get further into the FAIRification process and understand the importance of data management planning when setting up a registry. As a final note, it is important to stress that the friendly environment and the high interactivity among the different players, i.e., data managers, researchers, clinicians, patient representatives, and IT-trainers, strongly contributed in making the BYOD workshop a success.

Keywords: Rare Disease Registries, FAIR data principles, BYOD

Acknowledgments. We greatly thank speakers, trainers, FAIR data experts, organisers, participants, patient associations and their representatives for their contribution to the BYOD workshop. This edition took place in collaboration with or with the support of CNMR-ISS, RD-Connect, ELIXIR (in particular Dutch Techcentre for Life Sciences representative of ELIXIR-NL and ELIXIR-IIB), RD-Action, EURORDIS, ERNs, EpiRare, and ICORD.