

Surveyed common data access policies preferences amongst European Reference Networks

Alberto Cámara^{1,*}, Nirupama Benis^{2,3}, César H. Bernabé⁴, Inés D.O. Coelho⁵, Clémence M. A. Le Cornec⁶, Aylin Demir⁷, Bruna D.S. Vieira^{5,8}, Jose A. Ramírez⁶, K. Joeri van der Velde⁹, Shuxin Zhang^{2,3}, Ronald Cornet^{2,3}, Annika Jacobsen⁴, Marco Roos⁴, Franz Schaefer⁶, Morris A. Swertz⁹ and Mark D. Wilkinson¹

¹*Departamento de Biotecnología-Biología Vegetal, Escuela Técnica Superior de Ingeniería Agronómica, Alimentaria y de Biosistemas, Centro de Biotecnología y Genómica de Plantas. Universidad Politécnica de Madrid (UPM) - Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria-CSIC (INIA-CSIC). Campus Montegancedo 28223 Pozuelo de Alarcón (Madrid), Spain.*

²*Department of Medical Informatics, Amsterdam UMC location University of Amsterdam, Meibergdreef 9, Amsterdam, The Netherlands.*

³*Amsterdam Public Health, Digital Health Methodology, Amsterdam, The Netherlands.*

⁴*Department of Human Genetics, Leiden University Medical Center, Leiden, The Netherlands.*

⁵*Centre for Molecular and Biomolecular Informatics, Radboud University Medical Center, Nijmegen, The Netherlands.*

⁶*Division of Paediatric Nephrology, Center for Paediatrics and Adolescent Medicine, University of Heidelberg, Heidelberg, Germany.*

⁷*Institute of Medical Informatics, Goethe University Frankfurt, University Hospital Frankfurt, Frankfurt am Main, Germany.*

⁸*Department of Medical Imaging, Radboud University Medical Center, Nijmegen, The Netherlands.*

⁹*Genomics Coordination Center, University of Groningen and University Medical Center, Groningen, The Netherlands.*

Abstract

Background: Data sharing amongst existing Rare Disease (RD) registries, even though being a process that presents multiple barriers, would enrich and ease research, as well as facilitate interoperability between the registries themselves. Methods: To understand their preferences on sharing data, we surveyed 24 European Reference Networks (ERNs) from the RD Domain. Results: The answers show that most ERNs are willing to share a set of Common Data Elements for free with authenticated users at an aggregated or pseudonymized level the moment the data is collected. The one exception is the industry sector, to which ERNs prefer to ask for a fee. Objective: Our aim is to create a reference for how most RD registries are willing to share their data, improving the ability of other stakeholders to make informed decisions to make their data interoperable.

Keywords

Rare diseases, patient registries, FAIR data sharing

14th International SWAT4HCLS Conference, February 13–16, 2023, Basel, Switzerland

*Corresponding author.

✉ alberto.camara-ballesteros@ejprd-project.eu (A. Cámara)

ORCID 0000-0001-5613-9704 (A. Cámara); 0000-0002-2101-6154 (N. Benis); 0000-0003-1795-5930 (C. H. Bernabé); 0000-0002-0756-2722 (I. D.O. Coelho); 0000-0001-7893-0505 (B. D.S. Vieira); 0000-0003-0942-4371 (J. A. Ramírez); 0000-0002-0934-8375 (K. J. v. d. Velde); 0000-0003-4715-9070 (S. Zhang); 0000-0002-1704-5980 (R. Cornet); 0000-0003-4818-2360 (A. Jacobsen); 0000-0002-8691-772X (M. Roos); 0000-0001-7564-9937 (F. Schaefer); 0000-0002-0979-3401 (M. A. Swertz); 0000-0001-6960-357X (M. D. Wilkinson)



© 2023 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

1. Introduction

Amongst the clinical domains, the Rare Disease (RD) domain is one of the most convoluted regarding data sharing. This is the result of numerous factors, such as the intrinsic low prevalence of RDs that causes data to be siloed, scarce, and heterogeneous (as it ranges from documents updated by patients or their caretakers to clinical registries operated by data stewards and managing institutions). Another relevant aspect of RD data is that it usually is highly distributed [1], as specialised institutions usually focus on one RD or a group of RDs. Perhaps the most crucial limitation is that the data from the RD domain is personally identifiable by nature (as a consequence of the small number of people affected by these diseases) [2]. These circumstances, together with general problems related to information sharing, such as the absence of a common standardised vocabulary to describe it, cause great impediments to research activities [3].

To alleviate some of these issues, the Joint Research Centre¹ composed a set of Common Data Elements (CDEs) [4] that are considered essential for further research on RDs. This set of CDEs collects information that is prevalent for all European Registry Networks (ERNs), such as diagnosis, sex, status, phenotype, etc. The CDEs are modeled by the European Joint Programme on Rare Diseases (EJP-RD)² as the CDE Semantic Model, which recommends a collection of widespread ontologies that serve as a common group of terminologies to represent knowledge. By implementing the CDE Semantic Model, ERNs can increase their data sharing capabilities amongst themselves and with external resources. With the help of EJP-RD, most registries are undergoing a FAIRification process and implementation of the CDEs, which are crucial steps towards improving interoperability and data sharing within the RD community. [5] shows the procedure for applying the CDE Semantic Model to a registry of vascular anomalies, which entails the use of standardised terms from widespread ontologies, applying a transformation layer from whatever kind of data existed in the registry to a Resource Description Framework (RDF), as a way to make the data machine-interpretable and provide semantics amongst other steps. The EJP RD is currently working on the Virtual Platform (VP), a place where ERNs and other RD-related resources can connect to become discoverable (requiring the creation and collection of a Data Catalog Vocabulary (DCAT)-based minimal set of metadata about the registry and its contents, which is being designed by the experts at the EJP-RD) and share their data. The objective of this paper is to inventorize the ERNs' preferences when sharing their data, such as exactly what data to share, how to share it, whom to give access to it, and its time of publication.

2. Methods and results

All 24 ERNs were invited to complete a survey in which they describe which stakeholder is allowed for which level of data sharing. Nine types of stakeholders were considered: Contributing researcher (CR), Non-contributing researcher (NCR), Industry (IND), National Health Authority (NHA), Regulatory authority (RA), Health technology Assessment/Payors (HTA), European Health Data Space (EHDS). Patient Organisation (PO), and Non-Governmental Organization

¹https://joint-research-centre.ec.europa.eu/index_en

²<https://www.ejprarediseases.org/>

(NGO). Five aspects of data sharing are considered: Data to be shared (1), Highest data level to be shared (2), Access modalities(3), Timing (4), and Need for Data Access Committee (DAC) (5). The options presented for each aspect, as well as the results from the survey can be found on their respective tables.

Table 1

Data to be shared. *CDEs plus selected disease history/intervention/outcome data.

Stakeholder	None	CDEs	CDEs + extras*	All available data
CR	0.0%	21.4%	28.6%	50.0%
NCR	0.0%	21.4%	35.7%	42.9%
IND	0.0%	45.5%	45.5%	9.1%
NHA	0.0%	63.6%	27.3%	9.1%
RA	0.0%	54.5%	36.4%	9.1%
HTA	0.0%	60.0%	30.0%	10.0%
EHDS	0.0%	60.0%	30.0%	10.0%
PO	0.0%	41.7%	33.3%	25.0%
NGO	0.0%	44.4%	44.4%	11.2%
Average	0.0%	45.8%	38.9%	19.6%

To help with visualisation, the responses were plotted, they can be found here [6]. You can find the original responses from the ERNs that agreed to share them here [7].

3. Discussion

The results from the survey indicate that the paramount priority for ERNs is to protect their patient's privacy, as they all want to share their data, but in a controlled and safe manner. That is why no ERN answered that they did not want to share any of their data, and most of them are willing to give access to the CDEs by themselves (45.8% on average) or with some extras (38.9% on average) to all stakeholders. Most ERNs want to share data that is pseudonymised (37.0% on average) or aggregated (42.1% on average), as it allows for research to be done without risking the privacy of their patients. The majority of ERNs want to share their data for free (42.8% on average), but a considerable amount of them require the user to be authenticated (42.5% on average). Regarding the data sharing timing, the overwhelming majority are willing to share it immediately upon collection (72.2% on average), probably because that way it becomes available for research as soon as possible, and almost all registries require stakeholders trying to access their data need to submit a request to their DAC - the next point of the survey. This is no surprise, as that way they can carefully examine exactly who accesses their data, what data they share and exactly when to share it. Two stakeholders seem to be polar opposites in the ERNs perspective: researchers and industry. For researchers, ERNs want to share all their data at pseudonymised patient level (57.1%), and for free (some of them with (50.0%) and some of them without authentication (42.9%)). These answers further support the point that the RD community is willing to assist researchers as much as possible while protecting patients' privacy. This might be the cause for the trend towards industry, one of the stakeholders with

Table 2
Highest data level to be shared

Stakeholder	None	Yes/No	Counts	Aggregated data	Anonymised data	Pseudonimised data
CR	0.0%	0.0%	7.1%	28.6%	7.1%	57.1%
NCR	0.0%	0.0%	7.1%	35.7%	21.4%	35.7%
IND	0.0%	0.0%	7.1%	57.1%	7.1%	28.6%
NHA	0.0%	0.0%	8.3%	41.7%	8.3%	41.7%
RA	0.0%	0.0%	16.7%	25.0%	25.0%	33.3%
HTA	0.0%	0.0%	16.7%	41.7%	8.3%	33.3%
EHDS	0.0%	0.0%	8.3%	33.3%	16.7%	41.7%
PO	0.0%	0.0%	14.3%	57.1%	0%	28.6%
NGO	0.0%	0.0%	8.3%	58.3%	0%	33.3%
Average	0.0%	0.0%	10.4%	42.1%	10.4%	37.0%

Table 3
Access modalities

Stakeholder	Free	Fees	Authenticated	Authenticated and fees
CR	42.9%	0.0%	50.0%	7.1%
NCR	42.9%	0.0%	57.1%	0.0%
IND	23.1%	15.4%	7.7%	53.8%
NHA	50.0%	0.0%	41.7%	8.3%
RA	41.7%	0.0%	50.0%	8.3%
HTA	50.0%	8.3%	33.3%	8.3%
EHDS	41.7%	0.0%	58.3%	0.0%
PO	42.9%	0%	42.9%	14.3%
NGO	50.0%	0%	41.7%	8.3%
Average	42.8%	2.6%	42.5%	12.0%

which ERNs have the lowest percentage of willingness to share all available data (9.3%), the highest willingness to share aggregated data (57.1%), and the highest willingness to have access with both authentication and fees (53.8%). Remarkably, researchers seem to favor data sharing with the industry over other stakeholders like patient organisations, which might suggest that researchers are faced with maintaining balance between protecting their patients and finding in the industry a source for funding that increases the sustainability of the registry itself.

Acknowledgments

This work was supported by the European Joint Programme on Rare Diseases, ERICA and the following European Reference Networks: BOND, ERKNeT, Endo-ERN, ERNICA, EURACAN, eUROGEN, EURONMD, GUARD-Heart, LUNG, MetabERN, PaedCan, RARE-LIVER, RITA, VASCERN.

Table 4

Timing

Stakeholder	Never	Immediately after collection	X years after collection	Only published data
CR	0.0%	91.7%	8.3%	0.0%
NCR	0.0%	54.5%	45.5%	0.0%
IND	0.0%	80.0%	10.0%	10.0%
NHA	0.0%	70.0%	30.0%	0.0%
RA	0.0%	80.0%	20.0%	0.0%
HTA	0.0%	66.7%	22.2%	11.1%
EHDS	0.0%	77.8%	22.2%	0.0%
PO	0.0%	54.5%	45.5%	0.0%
NGO	0.0%	75.0%	25.0%	0.0%
Average	0.0%	72.2%	25.4%	2.3%

Table 5

Need for DAC

Stakeholder	Yes	No
CR	100.0%	0.0%
NCR	100.0%	0.0%
IND	100.0%	0.0%
NHA	75.0%	25.0%
RA	100.0%	0.0%
HTA	75.0%	25.0%
EHDS	100.0%	0.0%
PO	75.0%	25.0%
NGO	75.0%	25.0%

References

- [1] H. J. S. Dawkins, et al., Progress in rare diseases research 2010–2016: An irdirc perspective, *Clinical and Translational Science* 11 (2018) 11–20. doi:10.1111/cts.12501.
- [2] M. G. Hansson, et al., The risk of re-identification versus the need to identify individuals in rare disease research, *European Journal of Human Genetics* 24 (2016) 1553–1558. doi:10.1038/ejhg.2016.52.
- [3] A. Schiepati, et al., Why rare diseases are an important medical and social issue, *The Lancet (British edition)* 371 (2008) 2039–2041. doi:10.1016/S0140-6736(08)60872-7.
- [4] EJP-RD, Cde semantic model, 2020. URL: <https://github.com/ejp-rd-vp/CDE-semantic-model>, last accessed: 2 Nov 2022.
- [5] B. D. S. Vieira, et al., Applying the fair data principles to the registry of vascular anomalies (vasca), *Studies in health technology and informatics* 271 (2020) 115–116. doi:10.3233/SHTI200085.
- [6] A. Cámara, Plots for data sharing policies survey, 2022. doi:10.5281/zenodo.7274752.
- [7] A. Cámara, Responses to the data access policies survey, 2022. doi:10.5281/zenodo.7535291.