Digital Health and Informatics Innovations for Sustainable Health Care Systems
J. Mantas et al. (Eds.)

doi:10.3233/SHTI240672

Review of Key Elements in Developing a Common Data Model for Rare Diseases: Identifying Common Success Factors

Adam S L GRAEFE[a,b,1], Filip REHBURG [a],
Miriam HÜBNER [a], Sylvia THUN[a] and Oya BEYAN [b]
[a] *Berlin Institute of Health at Charité – Universitätsmedizin Berlin, Core Unit Digital Medicine and Interoperability, Germany*
[b] *Institute for Biomedical Informatics – University Hospital Cologne, Germany*
ORCiD ID: Adam S L Graefe https://orcid.org/0009-0004-8124-8864

Abstract. This paper explores key success factors for the development and implementation of a Common Data Model (CDM) for Rare Diseases (RDs) focusing on the European context. Several challenges hinder RD care and research in diagnosis, treatment, and research, including data fragmentation, lack of standardisation, and Interoperability (IOP) issues within healthcare information systems. We identify key issues and recommendations for an RD-CDM, drawing on international guidelines and existing infrastructure, to address organisational, consensus, interoperability, usage, and secondary use challenges. Based on these, we analyse the importance of balancing the scope and IOP of a CDM to cater to the unique requirements of RDs while ensuring effective data exchange and usage across systems. In conclusion, a well-designed RD-CDM can bridge gaps in RD care and research, enhance patient care and facilitate international collaborations.

Keywords. Rare Diseases, Interoperability, Common Data Model, Digital Medicine, European Health Data Space, Registry, Artificial Intelligence

1. Introduction

Despite their name, Rare Diseases (RDs) affect more than 17 million individuals in the European Union (EU) and over 260 million worldwide [1]. In the EU, RDs are defined as having a prevalence of less than 5 per 10000 individuals. Also known as Orphan Diseases, at least 70 % of RDs have a genetic origin, with the majority manifesting during childhood [1].

RDs are systematically underrepresented in routine care, leading to challenges for rare disease (RD) care and research [2]. Diagnoses can be complicated by classification problems with symptoms and common diseases and lack of awareness in non-specific care [3, 4]. Delayed diagnoses impede the application of suitable treatments and research into novel therapies [5,6], which are presently scarce [7]. Further, precise clinical data is limited, hindering cross-institutional care, research, and exchange [2,8].

High-quality data collected in routine clinical care is quintessential for effective RD care and research [6,8]. However, a substantial knowledge gap persists between data

[1] Corresponding Author: Adam S L Graefe; E-mail: adam.graefe@charite.de.

captured in routine care and machine-readable data within healthcare information systems (HIS) [2]. Heterogeneous systems often lack specification and standardisation, thus, precision, widening this gap [2]. Interoperability (IOP) is indispensable in bridging this divide by ensuring the unambiguous interpretation and seamless exchange of medical data across diverse systems [8].

Considering the vast number of over 6000 different RDs [1], each with unique requirements for research and care, the necessity for consensus data becomes apparent to bundle knowledge, expertise, and research. A data model defines how data within a data set is stored and its elements, data types, and interrelations within a database. Thus, a common data model (CDM) is designed to serve as a common denominator for multiple application scenarios [9].

Our paper defines key issues and recommendations for an RD-CDM addressed by existing RD infrastructure, studies and data models. This work supports developing RD-CDMs based on IOP standards, addressing the lack of RD data processed for secondary use in many countries. We delineate and analyse two main factors: the scope and IOP of a CDM. Further, we discuss how these factors contribute to a CDM's main functionality of multiple secondary uses.

2. Background

Many organisations and projects analyse requirements and provide recommendations for digitalisation in RDs. We chose to investigate well-established sources with comprehensive RD coverage, including those addressing IOP and secondary use, while others were excluded for clarity. Among them, we can name the EU-funded Rare2030 Foresight Study [10] and the German National Action League for People with Rare Diseases (NAMSE) [11]. The European Union Committee of Experts on Rare Diseases (EUCERD) [12], the European Rare Disease Registry Infrastructure (ERDRI) [13], and the European Reference Networks (ERNs) [14] are central to the European RD infrastructure. Correspondingly, the ERDRI Common Data Set (ERDRI-CDS) [15], the French Minimal Data Set for Rare Diseases (F-MDS-RD) [16], and the Domain-specific Common Data Elements (DCDEs) [17] are established RD common data sets and models in the EU. We reviewed the named organisations' recommendations, publications and reports. Adding to this, we identified the following key factors related to common data sets and models: Organisational Grounds, International Consensus, Interoperability, Usage, and Secondary Use. Table 1 summarises and cites the corresponding recommendations addressed.

Table 1. Based on international recommendations (Rare2030, NAMSE, EUCERD), established common data sets (ERDRI-CDS, F-MDS-RD, DCDEs), and existing RD infrastructure (ERNs, ERDRI), Table 1 depicts key factors necessary to consider for the development, functionality, and success of an RD-CDM.

Key Factor	Recommendations
Organisational Grounds	➢ Organising specialised centres and international RD registries. [10, 11, 12 13, 14, 15, 16, 17] ➢ Targeted policies and legal frameworks. [10, 11, 12, 13, 14, 16, 17] ➢ Inclusion of patient organisations, expert groups, policymakers, standardisation entities and RD stakeholders. [10, 11, 12, 13, 14, 15, 16, 17]
International	➢ Establishing and publishing common, uniform & consensus data(sets) across the spectrum of RDs. [10, 11, 12, 13, 14, 15, 16, 17]

Consensus	➢ Define common semantic strategies for sharing RD datasets, the entire spectrum of RD data, and its analysis. [10, 11, 13, 14, 15, 17] ➢ Streamline data collection & utilisation into an RD-CDM. [10, 14, 16]
Interoperability	➢ Ensure semantic, syntactic, technical, and organisational IOP based on international standards, ontologies and terminologies. [10, 11, 12, 13, 14, 15, 16, 17] ➢ Provide recommendations of defined variables for IOP and the application of international standards. [10, 11, 12, 13, 14, 15, 16, 17] ➢ Enable IOP between all HISs, ERNs, Registries, expert centres & RD stakeholders. [10, 11, 12, 13, 14, 15, 16, 17]
Usage	➢ Usage in public, private, clinical and research data. [10, 11, 12, 13, 14, 15, 16, 17] ➢ Enable integration and data capture within all HIS. [10, 11, 12, 13, 14, 15, 16, 17] ➢ Enhance homogenous data collection across various research and care domains. [10, 11, 12, 13, 14, 15, 16, 17] ➢ Enhance the efficiency of data entry, automation, and analysis. [10, 11, 13, 14, 15, 16]
Secondary Use	➢ Connect international registries and ERNs. [10, 11, 12, 13, 14, 15, 16, 17] ➢ Enable international research exchange, and discoverability in research databases and reusable standard analysis formats. [10, 11, 12, 13, 14, 15, 16, 17] ➢ Ensure patient participation and accessibility to every European RD patient. [10, 11, 12, 13, 14, 15, 17] ➢ Ensure alignment with European Health Data Space (EHDS). [10, 11, 12, 13, 14, 15, 17]

3. Analysis

Based on our review, we identified two important aspects impacting the success of a CDM: its scope and IOP. This Section further explores these aspects, highlighting challenges, recommendations, and limitations to consider. Regarding the scope, the challenge lies in finding a common denominator for the unique clinical requirements of each RD while being limited in complexity and size. Data granularity denotes the degree of detail observable in clinical data, characterised by the precision of categorisation and depth of segmentation. Thus, the finer the data granularity, the more detail can be observed. However, data granularity can vary due to differing levels of detail, specifications, and scales across datasets, termed the spectrum of data granularity.

Navigating this spectrum presents a significant clinical challenge, balancing simplicity and comprehensiveness while remaining valid for all RDs. A CDM must remain simple for wide application and effective utilisation. Conversely, it must also be comprehensive, catering to the unique clinical nuances and detailed requirements to support meaningful research. For more complex questions, DCDEs provide support as a CDM may not suffice.

Each added element enhances the clinical potential of a CDM in particular scenarios, known as the marginal benefit, while excluding the element represents a deficit. Every additional element also requires more data collection effort, known as marginal costs. Initially, the marginal benefits of each added element rise, enhancing the model's potential. However, these benefits eventually begin to decrease. While marginal costs are consistent or may even increase, the extension of a CDM should cease once the marginal benefits fall below these costs. Since a CDM must meet complex technical and clinical requirements, IOP is at the core of its effectiveness. It enhances the seamless medical data capture and exchange across diverse HISs [9, 16]. To achieve IOP for a CDM, all data elements and value sets must be implemented and expressed with international standards. For instance, alignment with Health Level 7's Fast Healthcare Interoperability Resources (HL7 FHIR®) allows for compliance with semantic and syntactic IOP requirements on RDs [2,8,18]. Implementing technical and organisational IOP, along with a CDM's horizontal and vertical integration [19], necessitates the involvement of high-level authorities and legal policies to navigate legal and organisational complexities [10]. However, the limit of standardisation is the boundary of its dissemination. Moreover, existing data must be prepared for IOP, while previously missing data needs to be collected accordingly. Further, complex RD-specific clinical terms and expressions require extensive definitions to achieve IOP.

4. Discussion

Multiple application scenarios of a CDM include registry utilisation, patient engagement, and international research, alongside integration with ERNs and the EHDS. Navigating the data granularity spectrum and achieving IOP for a CDM is pivotal in harnessing its potential for secondary applications. Considering the inherently low prevalence of RDs, utilising patient data from multiple sources is indispensable. We recommend aligning a CDM with HL7 FHIR® resources and implementing it in routine care. This will increase the data available for secondary applications [8]. Multi-centred research and reusable analysis pipelines [20] for individual or multiple RDs are facilitated, potentially enhancing usability for clinicians. Also, the decreased data availability induced by the spectrum of data granularity is mitigated. This enhances Artificial Intelligence in RDs where large amounts of precise data are necessary [21]. Further, the operational flexibility of a CDM is enhanced for broad adaptability. This enables various operation modes and affects decentralised data management, including data privacy challenges, decentralised algorithms, and patient involvement.

5. Conclusions

This work delineates key success factors for developing an RD-CDM based on previous research, infrastructure, and work. Based on our review, we analysed two important values for the development of an RD-CDM: navigating the scope and IOP. We identified these two values as key to the CDM’s effectiveness as a common denominator for multiple application scenarios. The scope of a CDM includes navigating the granularity spectrum between simplicity and comprehensiveness, as well as marginal benefits and costs. In achieving IOP, alignment with international standards, such as those recommended by HL7 FHIR®, is indispensable. However, IOP limitations must be considered underlining the need for European-wide collaboration. The latter is quintessential regarding the secondary use of data for multiple purposes, such as ERNs, EHDS, international research, patient involvement, and registries. In conclusion, we recommend considering the aspects discussed for the ongoing effort to unite the European RD community in developing an RD-CDM.

References

[1] Wakap SN, Lambert D, Olry A, Rodwell C, Gueydan C, Lanneau V, et al. Estimating cumulative point prevalence of rare diseases: analysis of the Orphanet database. European Journal of Human Genetics. 2019 Sep 16;28(2):165–73. https://doi.org/10.1038/s41431-019-0508-0

[2] Schepers J, Fleck J, Schaaf J. The medical informatics initiative and rare diseases: next-generation routine data for diagnosis, therapy selection and research. Bundesgesundheitsblatt, Gesundheitsforschung, Gesundheitsschutz. 2022 Oct 28;65(11):1151-8. https://doi.org/10.1007/s00103-022-03606-y

[3] Vandeborne L, Van Overbeeke E, Dooms M, De Beleyr B, Huys I. Information needs of physicians regarding the diagnosis of rare diseases: a questionnaire-based study in Belgium. Orphanet Journal of Rare Diseases. 2019 May 4;14(1). https://doi.org/10.1186/s13023-019-1075-8

[4] Schaaf J, Sedlmayr M, Schaefer J, Storf H. Diagnosis of Rare Diseases: a scoping review of clinical decision support systems. Orphanet Journal of Rare Diseases. 2020 Sep 24;15(1). https://doi.org/10.1186/s13023-020-01536-z

[5] Feltmate K, Janiszewski PM, Gingerich S, Cloutier M. Delayed access to treatments for rare diseases: Who's to blame? Respirology. 2015 Feb 26;20(3):361–9. https://doi.org/10.1111/resp.12498

[6] Wright CF, FitzPatrick DR, Firth HV. Paediatric genomics: diagnosing rare disease in children. Nature Reviews Genetics. 2018 Feb 5;19(5):253–68. https://doi.org/10.1038/nrg.2017.116

[7] Tambuyzer E, Vandendriessche B, Austin CP, Brooks PJ, Larsson K, Needleman KIM, et al. Therapies for rare diseases: therapeutic modalities, progress and challenges ahead. Nature Reviews Drug Discovery. 2019 Dec 13;19(2):93–111. https://doi.org/10.1038/s41573-019-0049-9

[8] Lehne M, Saß J, Essenwanger A, Schepers J, Thun S. Why digital medicine depends on interoperability. Npj Digital Medicine. 2019 Aug 20;2(1). https://doi.org/10.1038/s41746-019-0158-1

[9] Ahmadi N, Zoch M, Guengoeze O, Facchinello C, Mondorf AW, Stratmann K, et al. How to customize Common Data Models for rare diseases: an OMOP-based implementation and lessons learned. Research Square (Research Square). 2023 Dec 8; https://doi.org/10.21203/rs.3.rs-3719430/v1

[10] Kole A et al, EURORDIS-Rare Diseases Europe, European Union Pilot Projects and Preparatory Actions Programme, European Commission, FIRPA International, Rare 2030 Young Citizens. Recommendations from the Rare 2030 Foresight Study: The future of rare diseases starts today. 2021. https://download2.eurordis.org/rare2030/Rare2030_recommendations_Exec_Summary.pdf

[11] Wessel T, Heuing K, Schlangen M, Schnieders B, Algermissen M. Rare diseases, digitization, and the National Action League for People with Rare Diseases (NAMSE). Bundesgesundheitsblatt, Gesundheitsforschung, Gesundheitsschutz. 2022 Oct 14;65(11):1119-25. https://doi.org/10.1007/s00103-022-03597-w

[12] Aymé S, Rodwell C. The European Union Committee of Experts on Rare Diseases: three productive years at the service of the rare disease community. Orphanet Journal of Rare Diseases. 2014 Jan 1;9(1):30. https://doi.org/10.1186/1750-1172-9-30

[13] Kölker S, Gleich F, Mütze U, Opladen T. Rare disease registries are key to Evidence-Based Personalized Medicine: highlighting the European experience. Frontiers in Endocrinology. 2022 Mar 4;13. https://doi.org/10.3389/fendo.2022.832063

[14] Tumienė B, Graeßner H, Mathijssen IMJ, Pereira AM, Schaefer F, Scarpa M, et al. European Reference Networks: challenges and opportunities. Journal of Community Genetics. 2021 Mar 17;12(2):217–29. https://doi.org/10.1007/s12687-021-00521-8

[15] EUROPEAN COMMISSION, JOINT RESEARCH CENTRE, Directorate F – Health and Food, Unit F.1 – Disease Prevention. SET OF COMMON DATA ELEMENTS FOR RARE DISEASES REGISTRATION. EUROPEAN PLATFORM ON RARE DISEASE REGISTRATION (EU RD Platform). https://eu-rd-platform.jrc.ec.europa.eu/sites/default/files/CDS/EU_RD_Platform_CDS_Final.pdf

[16] Choquet R, Maaroufi M, De Carrara A, Messiaen C, Luigi E, Landais P. A methodology for a minimum data set for rare diseases to support national centers of excellence for healthcare and research. Journal of the American Medical Informatics Association. 2014 Jul 18;22(1):76–85. https://doi.org/10.1136/amiajnl-2014-002794

[17] Abaza H, Kadioglu D, Martin SC, Papadopoulou-Bouraoui A, Viera BDS, Schaefer F, et al. Domain-Specific Common Data Elements for Rare Disease Registration: Conceptual approach of a European Joint Initiative toward Semantic Interoperability in Rare Disease research. JMIR Medical Informatics. 2022 May 20;10(5):e32158. https://doi.org/10.2196/32158

[18] Robinson PN, Graessner H. Rare-disease data standards. Bundesgesundheitsblatt, Gesundheitsforschung, Gesundheitsschutz. 2022;65(11):1126. Available from: https://doi.org/10.1007/s00103-022-03591-2

[19] Kohlmayer F, Praßer F, Eckert C, Kuhn KA. A flexible approach to distributed data anonymization. Journal of Biomedical Informatics. 2014 Aug 1;50:62–76. https://doi.org/10.1016/j.jbi.2013.12.002

[20] Jacobsen JOB, Baudis M, Baynam G, Beckmann JS, Beltrán S, Buske OJ, et al. The GA4GH Phenopacket schema defines a computable representation of clinical data. Nature Biotechnology. 2022 Jun 1;40(6):817–20. https://doi.org/10.1038/s41587-022-01357-4

[21] Schaefer J, Lehne M, Schepers J, Praßer F, Thun S. The use of machine learning in rare diseases: a scoping review. Orphanet Journal of Rare Diseases. 2020 Jun 9;15(1). https://doi.org/10.1186/s13023-020-01424-6