Published in final edited form as:

Contemp Clin Trials. 2015 May ; 42: 78–80. doi:10.1016/j.cct.2015.03.003.

NIH/NCATS/GRDR® Common Data Elements: A leading force for standardized data collection

Yaffa R. Rubinstein and Pamela McInnes

National Center for Advancing Translational Sciences (NCATS), National Institutes of Health

Abstract

The main goal of the NIH/NCATS GRDR® program is to serve as a central web-based global data repository to integrate de-identified patient clinical data from rare disease registries, EHR, clinical data and other data sources, in a standardized manner, to be available to researchers for conducting various biomedical studies, including clinical trials and to support analyses within and across diseases. The aim of the program is to advance research for many rare diseases and, by extension, common diseases as well.

One of the first tasks toward achieving this goal was the development of a set of Common Data Elements (CDEs), which are controlled terminologies that represent collected data. The use of CDEs facilitates the integration of patient information.

The GRDR CDEs have been the cornerstone of the GRDR repository, as well as of several other national and international patient registries. As a result many new opportunities for collaboration and networking are now available and being pursued. Most importantly, the establishment of the GRDR program has elevated the issue of data standardization and interoperability for rare disease patient registries, to international attention, resulting in a global dialog and significant change in the mindset of registry developers, patient advocacy groups, and other national and international organizations.

One of the main obstacles to advancing biomedical research is the inability to exchange and share data and knowledge. This is the result of: data collected using different terminologies, databases being established with lack of interoperability and with no linkage between them, negative results and lessons learned not being shared, and resources (including funding and patient population) being used in duplicated efforts with no coordination and collaboration. In the field of common diseases these obstacles may not stand in the way of making some significant progress, but with rare diseases, these issues are more acute and significant attention is needed.

Patients with rare diseases are scattered over large geographical areas around the world. Patient registries are a major source of patient data and essential to locate and identify these

© 2015 Published by Elsevier Inc.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

patients. Without them we cannot have a good estimate of the disease prevalence and enough data to conduct meaningful research to understand the pathogenesis of these diseases and to develop drugs and therapeutics to help the millions who are suffering from rare diseases.

It is estimated that out of many thousands of known rare diseases less than 20² percent have patient registries, of which some are duplications of the same disease, while others have data that are not openly available for all investigators, and many are established on different platforms using different terminology. In addition, registries are being established with no adequate long term strategy, lack of sufficient consideration of the needs of the patients, and unwillingness or inability to fully share and exchange the valuable data, experiences and knowledge. All of this contributes to hampering the efforts within the rare diseases community and the scientific community at large, to develop drugs and therapeutics and to improve the quality of life of millions of people around the globe.

Recognizing these challenges, and to bring a change in the manner in which patient data are collected, used and disseminated, the Office of Rare Diseases Research (ORDR) at the National Center for Advancing Translational Sciences (NCATS) first presented the concept of establishing the Global Rare Diseases Patient Registry Data Repository (GRDR) at an international workshop in 2010: “Advancing Rare Disease Research: The Intersection of Patient Registries”^{1,2}

The main goal of the NIH/NCATS GRDR[®] program³ is to serve as a central web-based global data repository to integrate de-identified patient clinical data from rare disease registries, EHR, clinical data and other data sources, in a standardized manner, to be available to researchers for conducting various biomedical studies, including clinical trials and to support analyses within and across diseases. The aim of the program is to advance research for many rare diseases and, by extension, common diseases as well.

One of the first tasks toward achieving this goal was the development of a set of Common Data Elements (CDEs)^{4,5}, which are controlled terminologies that represent collected data. The use of CDEs facilitates the integration of patient information and their clinical data from different sources and allows interoperability between databases. To develop the set of the GRDR CDEs, a national committee was established. This committee consisted of scientific expertise from disciplines representing all sectors of the community, including the National Institutes of Health and other federal agencies, academia, the private sector, health care providers, patient advocacy groups and patient organizations. The GRDR CDEs were tested, validated and implemented during a 2 year proof of concept period in the process of establishing new rare disease registries and mapping existing registries. For more details check <https://grdr.ncats.nih.gov>. Access to GRDR CDEs is freely available to download at: https://grdr.ncats.nih.gov/index.php?option=com_content&view=article&id=3&Itemid=5.

The 75 GRDR CDEs are organized into 10 different categories:

1. Current contact information
2. Socio-demographic information

3. Diagnosis
4. Family history
5. Anthropometric information
6. Patient-reported outcome
7. Medications, devices, and health services
8. Clinical research participation and biospecimen
9. Communication and preferences
10. Administrative

For each element the following information is provided:

Item#, Item concept, Question text, Comments, Response category, Variable structure, Reference categories, Reference Categories Link (if applicable) and Recommended Degree of Requirement

The GRDR CDEs are composed of 5 main components:

1. CDEs that are recommended for any group planning to establish a patient registry (that includes all the listed CDEs)
2. CDEs that represent data collected by the patient and/or on behalf of the patient
3. CDEs that represent data that are generated only by the registry (or the GRDR program) and is considered administrative information and labeled N/A (not applicable) for the patient
4. CDEs that represent data that will be aggregated into the GRDR (Labeled GRDR)
5. CDEs that represent data that are needed to assign a Global Unique Identifier (labeled GUID)

Since the GRDR integrates only de-identified patient information, CDEs for data collected without identifiable patient information were labeled with "GRDR" and designated by a GRDR number e.g. GRDR001, GRDR002, etc.. CDEs with identifiable information meant to be used only by the individual patient registries collecting the data from their community are not accepted to the GRDR. In addition, elements that are required in order to assign the GRDR Global Unique Identifier (GUID) are designated as "GUID". The GUID is an individual patient ID, for which the code resides at, and is known only to, the registry providing the data. The GUID is created using personally identifiable information such as a subject name, date of birth and city of birth. This personally identifying information is never transferred to GRDR nor stored within the GRDR database. This allows data from a de-identified individual subject to be aggregated, tracked and linked across projects, time, databases, and biospecimens, so that when studying one aspect of a patient's presentation, an investigator can relate it to other data from the same patient. (Additional information about the GRDR GUID can be found on <https://grdr.ncats.nih.gov/>)

The GRDR CDEs represent one of the first attempts to incorporate standard terminologies both prospectively and retrospectively⁶ into a new platform designed specifically to support long-term and follow up studies and data sharing across different sources at different locations. These CDEs are not meant to satisfy all the needs of rare disease-specific registries. Additional CDEs and disease specific CDEs will be developed as part of the GRDR[®] program.

The GRDR CDEs have been the cornerstone of the GRDR repository, as well as of several other national and international patient registries^{7,8}. As a result many new opportunities for collaboration and networking are now available and being pursued. Most importantly, the establishment of the GRDR program has elevated the issue of data standardization and interoperability for rare disease patient registries, to international attention, resulting in a global dialog and significant change in the mindset of registry developers, patient advocacy groups, and other national and international organizations. This change in mindset is evident by the adoption or modification of the GRDR CDEs, use of additional accepted and validated standards, and collaborations to develop additional disease specific standards and CDEs^{9,10,11}

The development and use of well-defined and validated terminologies and CDEs to support biomedical research is one of the goals of the NIH and other federal agencies. The NIH CDEs working group of the Trans-NIH BioMedical Informatics Coordinating Committee has established a website and a web-based database that includes sets of standards terminologies and CDEs covering many diseases. This resource is accessible to the public and hosted by the NLM at: <http://cde.nlm.nih.gov>.

In conclusion: Through the GRDR program, NCATS will continue to lead and engage the rare disease community, including the patients and the advocacy groups, in the dialog concerning the importance and the need to use data standards and CDEs. Additional tools and other resources developed through the GRDR program, to accelerate the rate of establishing high quality patient registries in a standardized manner, will be shared and disseminated. The aim is that through standardization, registries will be interoperable to enable exchange and sharing of data, knowledge, and experiences to improve the quality of life of patients with rare diseases as well as with common diseases.

References

1. Rubinstein YR, Groft SC, Bartek R, et al. Creating a global rare disease patient registry linked to a rare diseases biorepository database: rare disease-HUB (RD-HUB) *Contemp Clin Trials*. 2010; 31(5):394–404.
2. Forrest CB, Bartek RJ, Rubinstein Y, Groft SC. The case for a global rare-diseases registry *Lancet*. 2011; 377(9771):1057–1059.
3. GRDR website. <https://grdr.ncats.nih.gov/>
4. Jenders, RA.; McDonald, C.; Rubinstein, YR.; Groft, S. Applying Standards to Public Health: An Information Model for a Global Rare-Diseases Registry; American Medical Informatics Association symposium Proc.; 2011; p. 1819
5. GRDR CDEs. https://grdr.ncats.nih.gov/index.php?option=com_content&view=article&id=3&Itemid=5

6. Building a Common Pediatric Research Terminology for Accelerating Child Health Research. Kahn, Michael G., et al. PEDIATRICS. Mar.2014 133(3)
7. The NIH Office of Rare Diseases Research Patient Registry Standard: A Report from the University of New Mexico's Oculopharyngeal Muscular Dystrophy Patient Registry. Daneshvari PhD, Shamsi; Youssof MD, Sarah; Kroth, MD, MS, Philip J. Annu Symp Proc. 2013; 2013; p. 269-277.
8. Rémy, Choquet; Meriem, Maaroufi; de Carrara, Albane, et al. A methodology for a minimum data set for rare diseases to support national centers of excellence for healthcare and research. J Am Med Inform Assoc. 2014; 0:1–7.
9. Taruscio D, Mollo E, Gainotti S, Posada de la Paz M, et al. The EPIRARE proposal of a set of indicators and common data elements for the European platform for rare disease registration. Arch Public Health. 2014; 72(1):35. [PubMed: 25352985]
10. Swee Sung, Soon'; Gilberto, Lopes'; Hwee-Yong, Lim, et al. A for action to improve access to care and treatment for patients with rare diseases in the Asia-Pacific Region. Orphanet Journal of Rare Diseases. 2014; 9:137. [PubMed: 25223275]
11. AHRQ. A User's Guide: 3rd Edition. Registries for Evaluating Patient Outcomes. Research Report - Final – Apr. 30, 2014