Review

National information system for rare diseases with an approach to data architecture: A systematic review

Simin Derayeh, Alireza Kazemi, Reza Rabiei*, Azamossadat Hosseini, Hamid Moghaddasi

Department of Health Information Technology and Management, Faculty of Paramedical Sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran.

Summary	The study aims to systematically review literature on the rare diseases information system to identify architecture of this system from a data perspective. The search for relevant English language articles, based on keywords in title, abstract, Mesh and Emtree terms, was done in Pubmed and Embase (from 1980 to June 2017), Scopus, Science Direct and Cochran (from 1980 to July 2017). Articles were selected if they addressed data architecture of information systems with a focus on rare disease, and if at least one of their objectives dealt with design, implementation, and development of rare diseases information systems. Thirty-five studies met the inclusion criteria. The findings were categorized into six groups. This first group addressed organizations acting as data generators, data users, and data governors. The second group was related to data sources and databases. Datasets and data elements formed the third group of findings, including common datasets, specific datasets, and complementary datasets. The fourth group of findings was in relation to data standards. Data sharing and interactions among relevant bodies included the fifth group of the findings. The last group of findings was pertinent to procedures and criteria used for checking the quality of data, as cross review checking was a main procedure assessing the accuracy, consistency, and completeness of data. Design and development of an integrated information system for rare diseases considering data architecture principles in practice could help eliminating issues with management of rare diseases through facilitating sharing information and experiences.

Keywords: Rare disease, information system, data architecture, registry, data set

1. Introduction

A rare disease is a health condition that affects a small number of people in comparison with other common diseases (*1,2*). In most definitions of rare diseases suggested in various countries, a specific prevalence threshold is assigned for these diseases considering the population of countries and their requirements and policies (*2,3*). The World Health Organization (WHO) has suggested a frequency of less than 6.5 to 10 per 10,000 people to define rare diseases. In the European Union (EU), this definition is for less than 5 in 10,000 people (or 1 in 2,000). In the United States of America (USA), a number fewer than 200,000 people has been defined for rare diseases, but less than 50,000 in Japan and less than 2,000 people in Australia (*2,3*). The majority of countries have referred to the EU definition as their national strategy for rare diseases (*2*).

In addition to disease prevalence, other criteria common in rare diseases are considered in identifying these diseases (*2*). For example, these diseases are chronic, progressive, life threatening, body tissues degenerative and causing disability, and there is no curative and effective treatment for the majority of them (*2,4*). These diseases have genetic origins in 80% of the cases, 50-70% of the patients are children, 30% of patients die before they reach the age of five (*4,5*). So far, about 5,000 to 7,000 rare diseases have been identified and new rare diseases are regularly reported (*2,4*). Most of the known diseases are categorized

into several main groups including: metabolic disorders, neuromuscular disorders, blood disorders, cardiovascular and respiratory disorders, autoimmune diseases, skin diseases and rare neoplasms (*4*).

There are challenges with the management of rare diseases. These challenges include the geographical dispersion of patients and rare diseases specialists, limited number of specialists, the lack of consistency and integrity of studies, limited access to credible sources, and limited information and knowledge (*4,6*). Development of an integrated information system for rare diseases, for example, at a national level could help to eliminate part of the challenges experienced due to information restrictions (*7,8*).This issue was emphasized in the main recommendations of the European Union Council on rare diseases in 2013 (*9,10*). A well-developed information system for rare diseases requires addressing the principles of information systems architecture in practice (*7*). A key aspect of information systems architecture is data architecture (*11,12*), consisted of models, standards and methods that depict various data types and the methods for data collection, storage, processing, retrieval, and sharing (*12,13*). In relation to information system for rare diseases, the data architecture provides an overview of the data that should be available in the information system and provides a concrete infrastructure for data flow and sharing (*14*).

The main objective of this review is to identify the data architecture of rare diseases information system in the published studies addressing this system in terms of data content and data interactions. By presenting the current literature in rare diseases information systems, it is hoped that this review could contribute to a better understanding of the system data architecture.

2. Data Collection

The search for relevant English language articles, based on keywords in title, abstract, Mesh and Emtree terms, was done in Pubmed and Embase (from 1980 to June 2017), Scopus, Science Direct and Cochran (from 1980 to July 2017). Figure 1 shows the search strategy for identifying the relevant articles. The first part (Part A) included terms used for rare diseases. In the second part (Part B), the keywords related to the data management tools were used. The third part (Part C) contained terms in relation to studies on information system data architecture. The results of these three parts were combined using the Boolean operator "and". Searching was supplemented with checking the bibliographies of identified articles.

Three reviewers examined independently the title and abstracts of the identified articles. Articles were selected if they address data architecture of information systems with a focus on rare diseases, and if at least one of their objectives dealt with design, implementation, and development of rare diseases information systems, and if they studied the network, databases, and registries of these diseases at the regional, national or international level. All studies that merely focused on the design and evaluation of specific software, or data architecture with no focus on rare diseases were excluded. The process and the results of the literature search are illustrated in Figure 2.

3. Data Analysis

Searching the online databases resulted in 3520 articles from Pubmed, Embase, Scopus, Science Direct, and Cochrane after removing duplicates. Initial screening of titles and abstracts resulted in 145 articles, of which 113 articles were excluded because they did not address data architecture in relation to rare diseases. Three further articles were identified through checking the bibliographies, leading to a total of 35 articles for full text review (Figure 2).

Considering the geographical origins of the reviewed studies, 20 studies conducted in Europe (*8-10,12,15-30*), 7 studies in USA (*31-37*), 4 studies in Australia (*38-41*) and 4 studies were performed in Asia (*42-45*).
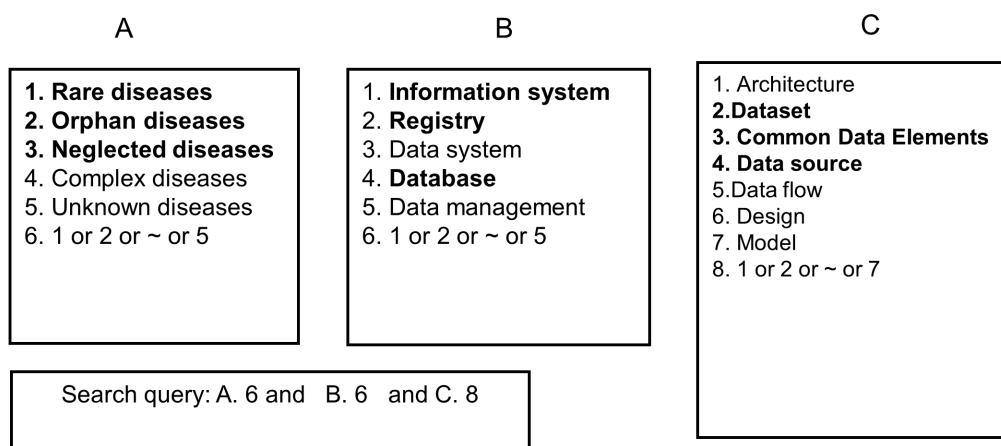
A

1. **Rare diseases**
2. **Orphan diseases**
3. **Neglected diseases**
4. Complex diseases
5. Unknown diseases
6. 1 or 2 or ~ or 5

B

1. **Information system**
2. **Registry**
3. Data system
4. **Database**
5. Data management
6. 1 or 2 or ~ or 5

C

1. Architecture
2. **Dataset**
3. **Common Data Elements**
4. **Data source**
5. Data flow
6. Design
7. Model
8. 1 or 2 or ~ or 7

Search query: A. 6 and B. 6 and C. 8

Figure 1. Keywords, Emtree and MeSH terms (bold terms) used in the search strategy.

```
┌──────────────┐ ┌──────────────┐ ┌──────────────┐ ┌──────────────┐ ┌──────────────┐
│ PubMed results│ │ Embase results│ │ Scopus results│ │ Science Direct│ │Cochran results│
│  (n= 235)    │ │  (n= 1761)   │ │  (n= 879)    │ │Results (n= 635)│ │  (n= 56)     │
└──────────────┘ └──────────────┘ └──────────────┘ └──────────────┘ └──────────────┘
```

Combined results
(*n*= 3566)

Excluded duplicates
(*n*= 46)

Identified records for title and
abstract review (*n*= 3520)

Excluded based on title & abstract review
(*n* = 3375)
Reasons: non English publications and
irrelevant to the selection criteria

Identified records for full-
text screening (*n*= 145)

Excluded (*n*= 113)
because of being merely about:
  - Software development
  - System evaluation
  - Data mining results
  - Bioinformatics subjects

Included (*n* = 32)
Criteria for analysis:
- Addressing data architecture aspects
- Focusing on rare disease information
system

Added by reviewing
bibliographies
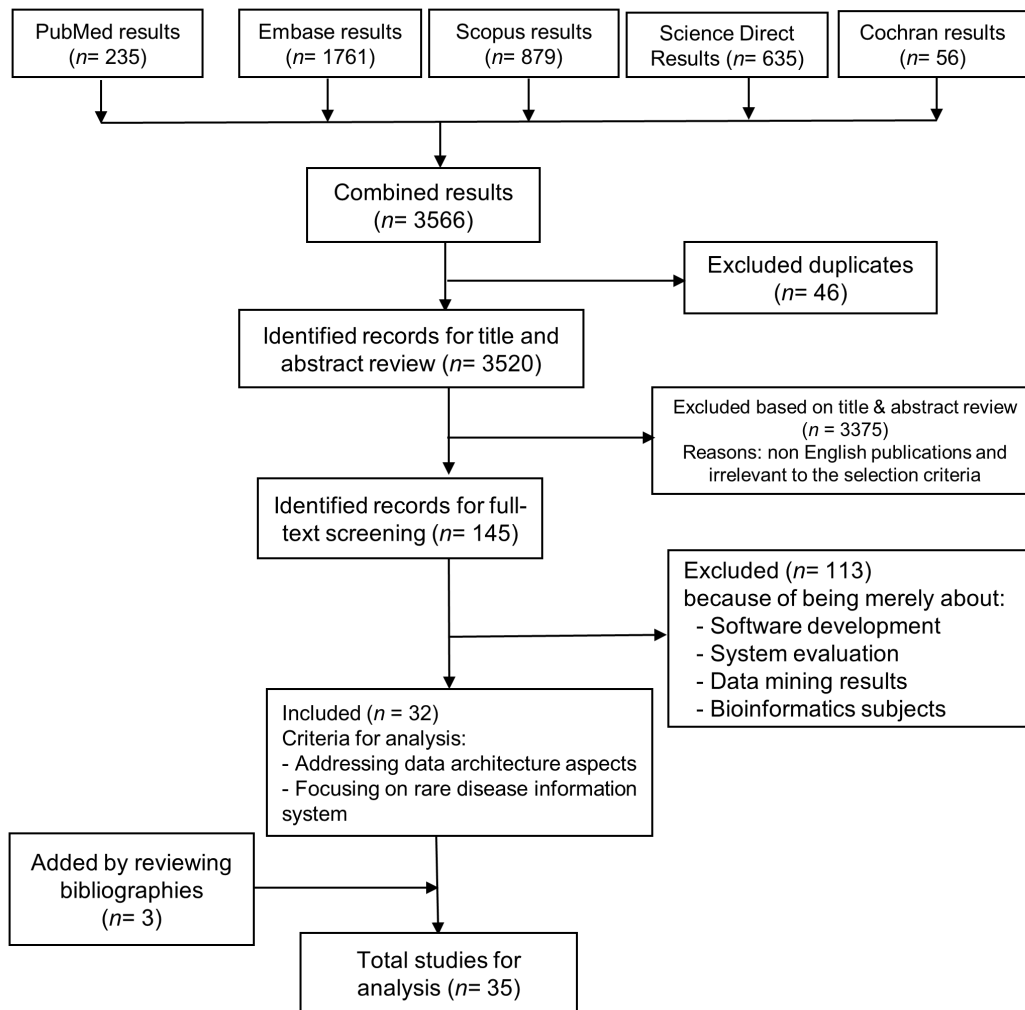(*n*= 3)

Total studies for
analysis (*n*= 35)

Figure 2. Search flow diagram.

The main findings from the analysis of the studies are categorized in the following six groups, including organization in charge of rare diseases data generation, adoption and governance (*10,15,17,18,20-24,27-41,43-45*); data sources (*8-10,15-18,20-24,26-36,39,40,43,45*), datasets, (*8-10,15-17,19,21-24,26,28-32,34,37-41*) data standards (*8,9,16,19-22,24,27,29,32,36,37,43,45*), data sharing (*15-18,20-22,27-29,31,33-35,38,40,43-45*) and data quality (*10,15-18,21-25,27,28,30-32,43,45*).

3.1. *Organizations involved in data generation, adoption, and governance*

Identifying the organizations that generate, adopt, and govern data could help to lay the foundation for concrete data architecture of information system for rare diseases. According to the studies, organizations dealing with rare disease information systems could be categorized into three main groups: (*i*) data-producers (*10,17,18,20-23,28,29,31-33,35,36,38-40,44,45*), (*ii*) data users (*15,17,18,21-23,27-29,31-35,37-39,41,43-45*), and (*iii*) data coordinators and governors (*10,15,17,18,20-24,27-31,33,35-37,45*).

Some of these organizations may belong to more than one group. The most important organizations producing rare disease data are clinics, reference centers for rare diseases, hospitals, research consortia, universities and academic centers (*21,31,35*).

Organizations considered as data-user included research centers, specialized scientific associations, physicians, universities, health managers and policy makers, insurance companies, pharmaceutical companies and medical equipment providers (*18,21,27,31,35,38*). The third group, i.e., coordinating and governing bodies, are in charge of implementing laws and standards and overseeing activities related to data generation, processing, storage and use to ensure that data are handled in safe and secure manner (*18*). The Office of Rare Diseases Research (ORDR) at the National Institutes of Health (NIH) in USA (*31,35*), and the NHS National Institute of Health Research (NIHR) in England (*18*) are examples of governing organizations.

3.2. *Data Sources and Databases*

The most significant data sources are the general and

Table 1. Identified common data sources in rare diseases information system and registration networks

Data source (*Ref.*)	Frequency of studies
Reference centers or clinics (clinical and epidemiological data) (*9,10,15-17,20-22,24,31,33,34,39,40,43*)	15
Research centers and consortia (research data) (*10,31,33,35,43,45*)	6
Patient self -reported data (*15-18,23,26,28,31-33,39*)	11
Laboratory, genetics and imaging data (*15,16,28,30,31,34,35,40,45*)	9
Specific databases or registries for diseases (*9,17,18,28,29,32,34,36,43*)	9
Data from cohort studies (*15,18,20,23,24,31*)	6
Medical Records (Electronic Medical Records, Electronic Health Records, Primary Care Records) (*8,9,17,18,20,24,32,34,36,43,45*)	11
Biobanks and bio-specimen (*29,32*)	2
Drug registry (*9,28,31,35*)	4
Birth registry (*8,9,21,22,40*)	5
Death registry (*9, 24*)	2

specialized rare diseases clinics and research centers, which are the most important producers of clinical and epidemiological data for these diseases. Furthermore, data can be captured from hospitals, research centers, bio-banks. In some systems, it is possible for patients or their relatives to enter data related to disease progress, the quality of their lives and other issues (*15,16,31,39*). A summary of the data sources is categorized in Table 1.

3.3. *Datasets and data elements*

Twenty three studies addressed datasets and their related data elements. The datasets are categorized into three main groups: a common dataset for rare diseases, specific datasets, and complementary or extended dataset. Common dataset and its data elements were mentioned in 22 studies (*8-10,15-17,19,21-24,26,28-32,37-41*). This dataset includes data elements applied to a wide range of rare diseases, such as patients' demographic data, different types of rare diseases, centers delivering services to patients, medication history and medical records of patients, family history, prenatal and neonatal records (Table 2). Eight studies indicated the data elements required for a specific dataset (*8,10,17,19,21,26,34,39*). This dataset includes data specific to a particular rare disease and is only applied for the same disease. The dataset for Cystinosis, Nephronophytosis, Turner Syndrome and Neonatal Diabetes (*21*) are examples of this type of dataset. The extended dataset includes data elements that are not available in the two aforementioned groups, but they add to, and improve, the knowledge about rare diseases. This category includes high encouraged data such as the experiences of patients and care givers, and data from studies addressed a particular aspect and conducted on an ad hoc basis (*8,10,15,16,24,26*).

3.4. *Data standards*

Fifteen studies showed the application of terminologies, nomenclatures, and international classifications of diseases including ICD 10, Systematized Nomenclature of Medicine-Clinical Terms (SNOMED-CT), and Logical Observation Identifiers Names and Codes (LOINC), as well as information exchange standards such as HL7 for data organization, developing datasets, and data sharing. The most commonly used standards are listed in Table 3. Classification systems such as ICD 10, Online Mendelian Inheritance in Man (OMIM) and ORPHA-codes were the most used systems.

3.5. *Data sharing*

Nineteen studies indicated data sharing among different bodies. Data sharing among parties such as scientific and academic communities, clinics, and the national database of rare diseases was reported in 12 studies (*15-17,21,22,27-29,31,33-35*). Data exchange of central registry with the local databases or registries were indicated in 12 studies (*20-22,27-29,31,33,38,40,43,45*).

Data interactions among academies, research centers and clinics were reported in 10 studies (*18,20-22,27,31,33,35,38,44*). In addition, seven studies indicated data sharing of patients or their relatives with registries of rare diseases (*17,18,29,31,33,35,40*).

3.6. *Data quality control*

The necessity of controlling the quality of data generated in the information systems was highlighted in 17 studies. In 14 studies, the data producers were in charge of data quality control (*15-17,21,22,24,25,27,28,30-32,43,45*); and in 8 studies, data coordinators and governors took this responsibility (*10,15,16,18,24,27,31,45*). In some studies, bodies such as quality control committees or departments (*10,16,23*) and independent experts (*10,17,25,30,43*) were in charge of checking the quality of data.

Table 4 demonstrates a summary of the main data quality control procedures together with criteria used in this respect. Ten studies referred to cross review checking as a routine for checking the quality of data through applying measures including accuracy, consistency, and completeness. Checking for duplications and missing data were among the other approaches for controlling data quality focusing mainly on the adequacy of data.

Table 2. The main categories of Common Data Elements in rare diseases registration and information system

Core Data category	Data Elements Instances	Frequency of Studies	*Ref.*
Demographic data	- Sex - Age (or date of birth) - Education level - Occupation - City and country of birth - City and country of residence - Contact details - Date of death	17	*(8,9,15-17,21-24,26,30,32,37-41)*
Diagnosis	- Date of current diagnosis - Status of diagnosis - Proof of diagnosis	15	*(8,9,15-17,19,21-24,26,30,37,40,41)*
Care pathway	- Type and name of treatment center - Referring physician - Center patient referred to - Documented visits - Date of contact	7	*(8,16,17,26,30,37,39)*
Medical history and medication	- Date of first treatment - Symptoms and date of onset of symptoms - Birth and reproductive history - Prenatal and neonatal information - Physical examination - Treatment strategy - Orphan drug treatment - Type of tests & results	20	*(8,9,15-17,19,21-24,26,28,30,32,36-41)*
Family history		9	*(8,15-17,19,24,36,37,40)*
Research and bio- banks data	- Patient participation in trials - Bio specimen storing in biobanks	6	*(8,16,24,26,28,43)*

Table 3. The most commonly used data standards

Category	Standard	Frequency of Studies	*Ref.*
Terminology and classification			
Code set	LOINC	3	*(8,36,37)*
Diagnostic Classification Systems	ICD 10	6	*(9,16,20,27,29,43)*
	ICD-9-CM or ICD 9	2	*(9,27)*
	Specific International Classification systems	9	*(9,16,20,27,29,32,36,37)*
Nomenclature	SNOMED-CT	7	*(8,9,19,27,32,36,37)*
Classification & Code Sets	ORPHA-codes	9	*(8,9,19-22,27,29,43)*
Information interchange	HL7	3	*(8,32,36)*
Gene codes set	OMIM	10	*(8,16,19-22,24,27,43,45)*

Table 4. The most common procedures conducted for controlling data quality

Quality Control Procedure	Data Quality Criteria	Frequency of Studies	*Ref.*
Predefined quality criteria for data	- Completeness - Accuracy	5	*(10,23,25,27,30)*
Cross review checking	- Accuracy - Consistency - Completeness	10	*(15-18,23-25,28,30,32)*
Checking for missing data	- Completeness - Adequacy	3	*(15,23,31)*
Checking for duplications	- Adequacy - Accuracy	5	*(21,22,25,27,45)*
Unification of data format	- Consistency	2	*(23,43)*
Double entry of partof all data	-Accuracy	2	*(23,25)*

Accuracy, as a data quality measure, was referred to in all of the studies that remarked the necessity of controlling data quality.

4. Discussion

Over the past two decades, there has been an increasing attention to rare diseases and managing these diseases (*6*). France, the United Kingdom (UK), and the United States are viewed as pioneer countries in terms of legislation, setting standards, developing of a network infrastructure for rare diseases, and creating information systems for these diseases (*6,10,18,21,27,29-31,35,37*).

4.1. *Organizations involved in data generation and governance*

Different organizations are dealing with rare diseases data. Continued monitoring and financial support by coordinators and governors could help data producer organizations to generate data considering data quality requirements. Data producer organizations, in turn, could feed governor organizations by high quality data to set and update relevant policies. In addition, data user organizations, such as research and care centers could have improved performance when using high quality data.

One of the main issues with management of rare diseases data is collecting data by different organizations in parallel resulted in data redundancy and duplication of efforts. Organizations such as research centers, clinics and hospitals may have their own exclusive databases for data collection. Developed countries have design and implemented a national integrated information network for managing rare diseases data, in which governing organizations play a coordinating role to avoid duplication of data collection (*31,32*).

4.2. *Data sources and datasets*

As the findings showed, there are a variety of data sources for a rare disease information system. Developing a national rare disease information system with the capability of data sharing with different data sources has a key role to play in efficient data management (*9,46*).

Due to the diversity of data sources, it is very challenging to determine a certain dataset for rare diseases. Common and specific datasets are two main datasets for rare diseases, in which the association between the common and specific datasets could be depicted in the form of a flower known as the Petal model (*17*). In this model, the common dataset constitutes the central core of the flower and the specific datasets of each rare disease make the petals and can be gradually added to the core part (*17,21*). In developed countries, work groups and specialized consortia are in charge of defining specific datasets for each of the rare

diseases groups (*8,26,35,47*).

The findings indicated the diversity of rare diseases data and data sources causing difficulties with data management. This issue could be eliminated through developing an integrated information system and the application of data standards.

4.3. *Data standards*

The use of standards could improve data quality and the interoperability of systems (*8*). In relation to the terminology and classification standards, these standards could be used to report the results of diagnostic tests (*8,21,22*), classify the conditions and provided services for statistical analysis, and reimburse of the costs (*20,27,29*). Since no standard for diagnostic codes is yet created for the classification of rare diseases, many countries have been using ICD-10. However, this disease classification lacks many of rare diseases and genetic disorders and does not fully support the detailed classification of these diseases (*21,22,48*). To overcome this issue, there have been global efforts coordinated by the World Health Organization to develop classification codes for rare diseases (*5,48*). Currently, a greater compatibility with the majority of genetic rare diseases is anticipated in ICD 11. Therefore, it is highly recommended to use ICD 11 and the ORPHA-codes simultaneously, as a valid ontology for rare diseases, although the alignment of ORPHA codes with SNOMED-CT, ICD-10 and OMIM is still in progress and requires further development (*5,17,48*). Some European countries such as France, use ORPHA codes rather than ICD-10 and SNOMED-CD (*21,22*), and this could be due the ORPHA codes being more specific and allowing better organization of rare diseases data.

4.4. *Data sharing*

With respect to data sharing, standards such as the HL7 support the exchange of data. This standard facilitates data sharing among different organizations, such as health and research centers, as well as policy making organizations (*8,31,35*). As findings indicated, different types of data could be shared among different bodies (*46*). Data sharing of health and research centers with the rare diseases registry is a main type of data interaction. Knowing that the most of bodies involved in data sharing are considered as both data producers and data users, their interactions with rare diseases registry is often two-dimensional, as they could send clinical and epidemiological data to the registry and receive information such as the relationship between phenotype and genotype characteristics of any rare disease (*18,27,46*). Another important type of data sharing could be seen between a central database or registry and other databases or specialized registries. In

this type of data sharing, the central registry acts as a hub and interacts with other registries (*20,43*).

4.5. *Data quality control*

The other key aspect of data architecture for rare diseases deals with laying the foundation for checking the quality of data, as the use of standards could improve the efficient use of the information system when data quality is addressed. Therefore, applying data quality procedures could help to ensure the quality of data.

5. Conclusion

To eliminate the challenges with the management of rare diseases, development of an integrated information system considering data architecture is crucial. Identifying data sources, defining datasets and data elements, and defining the interactions between data sources could help to design an integrated rare disease information system that could facilitate information sharing and provide the opportunity for sharing experiences at regional, national, and even transnational levels.

Acknowledgements

References

1. Gupta H, Badapanda C, Ghosh A, Chikara S. Rare DDB: An Integrated Catalog of Rare Disease Database. Clin Med Biochem. 2016; 2:111. doi:10.4172/2471-2663.1000111

2. Richter T, Nestler-Parr S, Babela R, Khan ZM, Tesoro T, Molsen E, Hughes DA; International Society for Pharmacoeconomics and Outcomes Research Rare Disease Special Interest Group. Rare Diseases Terminology and Definitions- A systematic Global Review: Report of the ISPOR Rare Disease Special Interest Group. Value Health. 2015; 18:906-914.

3. Aronson JK. Rare diseases and orphan drugs. Br J Clin Pharmacol. 2006; 61:243-245.

4. Zmazck V. EURORDIS Rare diseases Europ. What is a Rare disease. *http://www.mz.gov.si/fileadmin/mz.gov.si/pageuploads/kakovost/Redke_bolezni/prezentacija_2__Vlasta_Zmazek_EURORDIS.pdf* (accessed June 15, 2018)

5. The UK strategy for Rare Diseases. In: Department of Health. London: Crown, 2015. p. 1-35. *https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/260562/UK_Strategy_for_Rare_Diseases.pd* (accessed June 15, 2018)

6. Schieppati A, Henter JI, Daina E, Aperia A. Why rare diseases are an important medical and social issue. Lancet. 2008; 371:2039-2041.

7. Elze R, Fähnrich K P. The dispedia Framework: A Semantic Model for Medical Information Supply. ICONS International Conference on Systems. 2013; 59-63.

8. 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. J Am Med Inform Assoc. 2015; 22:76-85.

9. Zoni AC, Domínguez Berjón MF, Barceló E, Esteban Vasallo MD, Abaitua I, Jiménez Villa J, Margolles Martins M, Navarro C, Posada M, Ramos Aceitero JM, Vázquez Santos C, Zurriaga Llorens O, Astray Mochales J; Spain-RDR Group. Identifying data sources for a national population- based registry;the experience of the Spanish Rare diseases Registry. Public Health. 2015; 129:271-275.

10. Vittozzi L, Gainatti S, Mollo E, Donati C, Taruscio D. A Model for the European platform for Rare Disease Registries. Public Health Genomics. 2013; 16:299-304.

11. Winter A, Haux R, Ammenwerth E, Brigl B, Hellrung N, Jahn F. Health Information Systems Architecture and Strategies. 2nd ed. (Hannah K, Ball M, eds). Springer, London, UK, 2011; pp. 51-73.

12. Costechi N, Moisescu MM, Scacala S, Colin AM. Enterprise Architecture for e-Health System. The 4th IEEE International Conference on E-Health and Bioengineering-EHB. Popa Universityof Medicine and Pharmacy, Iasi, Romania., IEEE. 2013. DOI: 10.1109/EHB.2013.6707265

13. Cavalcanti JC. Effects of IT on Enterprise Architecture, Governance, and Growth. IGI Global, Hershey PA, USA, 2015; pp. 148-149.

14. Yilmaz S, Jonveaux P, Bicep C, Pierron L, Smail-Tabbone M, Devignes MD. Gene-disease relationship discovery based on model-driven data integration and database view definition. Bioinformatics. 2009; 25:230-236.

15. Evangelista T, Wood L, Torron RF, Williams M, Smith D, Lunt P, Hudson J, Norwood F, Orrel R, Willis T, Hilton-Jones D, Rafferty K, Guglieri M, Lochmuller H. Design, set-up and utility of the UK facioscapulohumeral muscular dystrophy patient registry. J Neurol. 2016; 263:1401-1408.

16. Farmer A, Aymé S, de Heredia ML, Maffei P, MacCafferty S, Mlynarski W, Nunes V, Parkinson K, Paquis-Flucklinger V, Rohayem J, Sinnott R,Tillmann V, Tranebjærg L, Barrett T.G. EURO-WABB: an EU rae diseases registry for wolfram syndrome, Alstrom syndrome and Bardet-Biedl syndrom. BMC Pediatrics. 2013; 13:130.

17. Kingswood JC, Bruzzi P, Curatolo P, *et al*. TOSCA-first international registry to address knowledge gaps in the natural history and management of tuberous sclerosis complex. Orphanet J Rare Dis. 2014; 9:182.

18. Javaid MK, Foretier-Zhang L, Watts L, *et al*. The RUDY study platform- a novel approach to patient driven research in rare musculoskeletal diseases. Orphanet J Rare Dis. 2016; 11:150.

19. Jaulent M-C, Assélé-Kama A, Savard S, Giavarini A, Touzé E, Jeunemaitre X, Ugon A, Plouin PF, Toubiana L. Building a Semantic Interoperability Framework for Care and Research in Fibromuscular Dysplasia. Stud Health Technol Inform. 2015; 216:217-221.

20. Toubiana L, Ugon A, Giavarini A, Riquier J, Charlet J, Jeunemaitre X, Plouin PF, Jaulent MC. A "Pivot" Model to Set up Large Scale Rare Diseases Information Systems: Application to the Fibromuscular Dysplasia Registry. Stud Health Technol Inform. 2015; 210:887-89.

21. Landais P, Messiaen C, Rath A, *et al*. CEMARA an

information system for rare diseases. Stud Health Technol Inform. 2010; 160:481-485.

22. Messian C, Le Mignot L, Rath A, *et al*. CEMARA: a Web Dynamic Application within a N-tier Architecture for Rare Diseases. Stud Health Technol Inform. 2008; 136:51-56.

23. D'Agnolo HM, Kievit W, Andrade RJ, Karlsen TH, Wedemeyer H, Drenth JP. Creating an effective clinical registry for rare diseases. United European Gastroenterol J. 2016; 4:333-338.

24. Nagel G, Unal H, Rosenbohm A, Ludolph AC, Rothenbacher D. Implementation of a population-based epidemiological rare disease registry: Study protocol of the amyotrophic lateral sclerosis (ALS)-registry Swabia. BMC Neurol. 2013; 13:22.

25. Lill CM, Bertram L. Developing the "Next Generation" of Genetic Association Databases for Complex diseases. Hum Mutat. 2012; 33:1366-1372.

26. Taruscio D, Mollo E, Gainotti S, Posada de la Paz M, Bianchi F, Vittozzi L. The EPIRARE proposal of a set of indicators and common data elements for the European platform for rare disease registration. Arch Public Health. 2014; 72:35.

27. Taruscio D, Gainotti S, Mollo E, Vittozzi L, Bianchi F, Ensini M, Posada M. The current Situation and Needs of Rare Disease Registries in Europe. Public Health Genomics. 2013; 16:288-298.

28. Pareyson D, Fratta P, Pradat PF, Soraru G, Finsterer J, Vissing J, Jokela M.E,Udd B, Ludolph A.C, Sagnelli A, Weydt P. Towards a European Registry and Biorepository for Patients with Spinal and Bulbar Muscular Atrophy. J Mol Neurosci. 2016; 58:394-400.

29. Thompson R, Johnston L, Taruscio D, *et al*. RD-Connect: An integrated platform connecting database, registries, biobanks and clinical bioinformatics for rare disease research. J Gen Intern Med. 2014; 29:s780-s787.

30. Taruscio D, Vittozzi L, Choquet R, Heimdal K, Iskrov G, Kodra Y, Landais P, Posada M, Stefanov R, Steinmueller C, Swinnen E,Van Oyen H. National Registries of Rare Diseases in Europe: An Overview of the Current Situation and Experiences. Public Health Genomics. 2015; 18:20-25.

31. Krischer JP, Gopal_Srivastava R, Groft SC, Eckstein DJ. The Rare diseases clinical Research Network's Organization and Approach to Observational Research and Health Outcomes Research. J Gen Intern Med. 2014; 29(suppl 3):739-744.

32. 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; 31:394-404.

33. Richesson RL, Sutphen R, Shereff D, Krischer JP. The Rare Diseases Clinical Research Network Contact Registry Update: Features and Functionality. Contemp clin Trials. 2012; 33:647-656.

34. Rubinstein YR, Mclnnes P. NIH/NCATS/GRDR Common Data Elements: A leading force for standardized data collection. Contemp clin Trials. 2015; 42:78-80.

35. Groft SC, Gopal_Srivastava R. A model for collaborative clinical research in rare diseases: Experience from the rare disease clinical research network program. Clin Invest.

2013; 3:1015-1021.

36. Richesson R, Shereff D, Andrews J. [RD] PRISM Library: Patient Registry Item Specifications and Metadata for Rare Diseases. J Libr Metadata. 2010; 10:119-135.

37. Daneshvari S, Youssof S, Kroth PJ. The NIH Office of Rare Diseases Research Patient Registry Standard: A Report from the University of New Mexico's Oculopharyngeal Muscular Dystrophy Patient Registry. AMIA Annu Symp proc. 2013: 269-277.

38. Bellgard MI, Macgregor A, Janon F, Harvey A, O'Leary P, Hunter A, Dawkins H. A Modular Approach to Disease Registry design: Successful adoption of an Internet-based Rare disease Registry. Hum Mutat. 2012; 33:E2356-E2366.

39. Bellgard MI, Napier KR, Bittles AH, Szer J, Fletcher S, Zeps N, Hunter AA, Goldblatt J. Design of a framework for the deployment of collaborative independent rare diseas-centric registries: Gaucher disease registry model. Blood Cells Mol Dis. 2018; 68:232-238.

40. Louise S, Fyfe S, Bebbington A, Bahi-Buisson N, Anderson A, Pineda M, Percy A, Zeev BB, Wu XR, Bao X, Leod PM, Armstrong J, Leonard H. InterRett, a model for international data collection in a rare genetic disorder. Res Autism Spectr Disord. 2009; 3:639-659.

41. Bellgard MI, Beroud C, Parkinson K, Harris T, Ayme S, Baynam G, Weeramanthri T,Dawkins H, Hunter A. Correction: Dispelling myths about rare disease registry system development. Source Code Biol Med. 2013; 8:21.

42. Song P, He J, Li F, Jin C. Innovative measures to combat rare diseases in China: The national rare diseases registry system, larger-scale clinical cohort studies, and studies in combination with precision medicine research. Interactable Rare Dis Res. 2017; 6:1-5.

43. Seo H, Kim D, Chae JH, Kang HG, Lim BC, Cheong HI, Kim JH. Development of Korean Rare disease knowledge base. Health Inform Res. 2012; 18:272-278.

44. Song P, Gao J, Inagaki Y, Kokudo N, Tang W. Intractable and rare diseases research in Asia. BioSci Trends. 2012; 6:48-51.

45. Sinha S, Black ML, Agarwal S, Das R, Bittles AH, Bellgard MI. Thallnd, α β -Thalassemia and Hemoglobinopathies Database for India: Defining a model country-specific and disease-centric bioinformatics resource. Hum Mutat. 2011; 32:887-893.

46. Kole A, Ensini M, Cam Yl. Proposed aims, scope, governance and sustainability options for a European platform for rare disease registries. In: European Organization for Rare Diseases, EURODIS, Paris; 2013: 1-20.

47. Minimum Data Set for Rare Disease Registries. In: EUCERD Joint Action. 2015. *http://www.eucerd.eu/ wp-content/uploads/2015/03/WP8_Registries_MDS.pdf* (accessed June 15, 2018).

48. Coding of rare diseases in health information systems with Orphacode. In: EUCERD Joint Action. 2014. Available from: *http://www.eucerd.eu/wp-content/uploads/2014/10/ WP5_WSReport_Ispra_Oct2014.pdf* (accessed June 15, 2018).