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

doi:10.3233/SHTI240424

A Framework to Support the Standardized Management of Real-World Data

Evangelia MINGA[a,1], Thomas CHATZIKONSTANTINOU[a], Theodore DALAMAGAS[b], Kostas STAMATOPOULOS[a] and Anastasia CHATZIDIMITRIOU[a]

[a] *Institute of Applied Biosciences, Center for Research and Technology Hellas, Thessaloniki, Greece*

[b] *Information Management Systems Institute, Athena Research Center, Athens, Greece*

Abstract. Relying on our experience on the development of data registration and management systems for clinical and biological data coming from patients with hematological malignancies, as well as on the design of strategies for data collection and analysis to support multi-center, clinical association studies, we designed a framework for the standardized collection and transformation of clinically relevant real-world data into evidence, to meet the challenges of gathering biomedical data collected during daily clinical practice in order to promote basic and clinical research.

Keywords. Real-World Data, Data Management, Knowledge Base

1. Introduction

Collection and analysis of Real-World Data (RWD) requires methodologies for data harmonization and quality assessment to achieve the effective integration of multi-originating biomedical data coming from different medical centers and laboratories daily practice towards enabling unified access to high-quality, clinically relevant data for research purposes. We used our expertise in the design of tools for the collection of clinical and biological data to support multi-center projects and clinical association studies in order to create a framework for the standardized management of RWD.

2. Methods

Given the proposal of a new clinical research study, our approach initially includes a close collaboration with domain experts in order to collect and analyze user requirements and define the data specifications. The data model is designed to express data requirements and complexity for each project and to effortlessly allow future domain expansions. The terminology is defined by the medical experts to meet the specific requirements. However, each term is also mapped to a code or concept retrieved from a standardized vocabulary or classification such as ICD-10. Data management applications are designed to provide simplicity, flexibility and easiness in use. To support

[1] Corresponding Author: Evangelia Minga, INAB|CERTH, 6th Km Charilaou-Thermi Rd, GR57001, Thessaloniki, Greece; E-mail: eva.minga@certh.gr.

retrospective data collection, we develop tools for the integration of data collected in project-based, specifically designed spreadsheets into a central database. For prospective studies, we prefer the design of user-friendly e-CRFs, organized and adjusted in a dynamic way to follow the project-based user scenarios. Data validation utilities are integrated in the data registration process and are developed based on a set of defined rules for allowed values, redundancy and consistency control, as well as domain-specific constraints to detect data discrepancies and assess the quality of data. Collected data are stored in a relational database, dedicated to the project. The database schema is generic and diagnosis-centric, designed for the organization of patient characteristics, diagnosis setting and disease course. Query and export tools are also developed and configured based on the project objectives, enabling the selection of anonymized, high-quality datasets, while statistical and visualization modules are integrated to facilitate data evaluation and analysis. Technical and organization measures are taken to ensure data privacy and data protection. Data specifications, validation rules and user scenarios are organized and stored in a structured way, creating a knowledge base (KB). All information obtained during the design of a new research study is used to enrich this KB. This knowledge-based approach is used to support the configuration of the tools for standardized data management.

3. Results

As a test case, we used Chronic Lymphocytic Leukemia (CLL), utilizing our medical expertise and our experience in data collection and analysis techniques in the domain [1]. We created a knowledge base representing CLL-based data entities and data relationships. Based on our previous work [2], we mapped the CLL concepts into standardized codes. We then utilized the knowledge base to standardize the development of a system with tools for CLL data collection and management.

4. Discussion and Conclusions

We focus on strategies to elaborate the acquisition process of RWD in a research context adjusting to different needs while adapting common policies for data harmonization and collection of high-quality datasets that can be re-used for research purposes. Our objective is to create a semantic framework, using knowledge-based approaches for data integration and quality assurance, that are continuously enriched and evaluated during the design and implementation of new projects.

References

[1] Chatzidimitriou A, Minga E, Chatzikonstantinou T, Moreno C, Stamatopoulos K, Ghia P. Challenges and solutions for collecting and analyzing real world data: the Eric CLL database as an illustrative example. Hemasphere. 2020 Sep 30;4(5):e425, doi: https://doi.org/10.1097/HS9.0000000000000425

[2] Minga E, Chamou D, Chatzikonstantinou T, Natsiavas P, Stamatopoulos K, Handakas E. Mapping data of patients with hematological malignancies to the OMOP Common Data Model: A Case Study of Chronic lymphocytic Leukemia. 2023 Jul 1-3; Rotterdam. OHDSI 2023 Europe Symposium.