Digital Health and Informatics Innovations for Sustainable Health Care Systems
J. Mantas et al. (Eds.)
© 2024 The Authors.
This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).
doi:10.3233/SHTI240418

Multi-Source Data ETL (Extract, Transform, Load) for a Genetic Epilepsy Diagnosis and Treatment Dashboard

Ariadna PÉREZ GARRIGA[a,1], Philipp HONRATH[b], Stefan WOLKING[b], Beatrice COLDEWEY[a], Susann A. BOZKIR[a], Patrick MAY[c], Yvonne WEBER[b], Rainer RÖHRIG[a] and Myriam LIPPRANDT[a]
[a]*Institute of Medical Informatics, Medical Faculty, RWTH Aachen University, Aachen, Germany*
[b]*Department of Epileptology and Neurology, Medical Faculty, RWTH Aachen University, Aachen, Germany*
[c]*Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-sur-Alzette, Luxembourg*
ORCiD ID: Ariadna Pérez Garriga https://orcid.org/0000-0001-8840-1770

Abstract. The growing number of genes identified in relation to epilepsy represents a major breakthrough in diagnosis and treatment, but experts face the challenge of efficiently accessing and consolidating the vast amount of genetic data available. Therefore, we present the process of transforming data from different sources and formats into an Entity-Attribute-Value (EAV) model database. Combined with the use of standard coding systems, this approach will provide a scalable and adaptable database to present the data in a comprehensive way to experts via a dashboard.

Keywords. Epilepsy, Genetics, Data Standardization, Common Data Model, ETL

1. Introduction

Epilepsy, a prevalent neurological disorder with a significant genetic component, poses challenges in the interpretation of genetic results due to the increasing number of known genes associated with the condition, which is particularly important for the diagnosis of rare epilepsies [1]. To address this, molecular boards, interdisciplinary groups dedicated to discussing treatment options for patients with complex genetic diseases, are invaluable [2]. TreatION-Project is currently developing a demonstrator for a therapeutic molecular board focused on genetic epilepsy.

2. Methods

The challenge is to consolidate various data sources [3], including hospitals' own REDCap-based questionnaires, as illustrated in Figure 1. Standardizing data formats is

[1] Corresponding Author: Ariadna Pérez Garriga, Pauwelsstraße 30, 52074 Aachen, Germany; E-mail: aperezgarrig@ukaachen.de.

crucial for storing and accessing information, including literature on genetic variants linked to epilepsy. A centralized dashboard will visualize data and provide access to clinical decision support tools for diagnosis and treatment of epileptic patients.

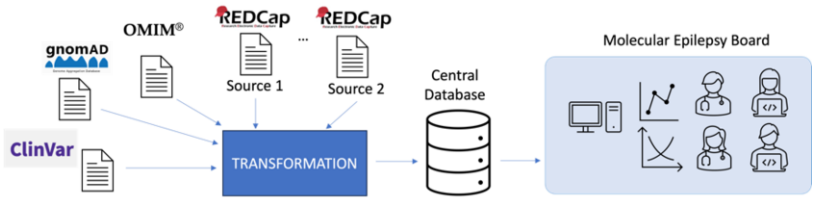


Figure 1. Architecture and data flow of the Extract, Transform, Load (ETL) process.

3. Results

The database will consist of three tables: "Patient", "Clinical Findings", and "Literature". The "Patient" table will contain immutable patient information such as birthdate, ethnicity, sex, and the data source. To accommodate clinical findings from various sources without constant modifications, the "Clinical Findings" table will use an Entity-Attribute-Value (EAV) model. REDCap questionnaire fields are mapped to clinical parameters, with answers corresponding to the values. Each field is considered an entity belonging to a patient. The coding or unit system used in the parameter is stored, allowing for different data types. A column with equivalent values in the standard coding system or preferred data type for the clinical finding is included. The “Literature” table contains necessary information regarding the literature of clinical findings, including epilepsy genetic diagnosis, also in an EAV structure.

4. Discussion and Conclusions

An EAV model speeds up data search without requiring a complex system of relational tables. This model allows for flexibility and storage of different data formats in the same table. Additionally, mapping to a standard coding system enables easy transformation to a common data model in the future. The use of standard coding systems and transformation into common data models should always be considered when creating medical survey databases.

References

[1] Symonds JD, Zuberi SM, Johnson MR. Advances in epilepsy gene discovery and implications for epilepsy diagnosis and treatment. Current Opinion in Neurology. 2017 Apr;30(2):193, doi: 10.1097/WCO.0000000000000433.

[2] Knowles JK, Helbig I, Metcalf CS, Lubbers LS, Isom LL, Demarest S, et al. Precision medicine for genetic epilepsy on the horizon: Recent advances, present challenges, and suggestions for continued progress. Epilepsia. 2022;63(10):2461–75, doi: 10.1111/epi.17332.

[3] López de Maturana E, Alonso L, Alarcón P, Martín-Antoniano IA, Pineda S, Piorno L, et al. Challenges in the Integration of Omics and Non-Omics Data. Genes. 2019 Mar;10(3):238, doi: 10.3390/genes10030238.