Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230257

Evaluation of Modeling Approaches for a Clinical Data Warehouse in a Highly Dynamic Environment

Stephan LORENZ[a,1], Richard GEBLER[a], Franziska BATHELT[a], Martin SEDLMAYR[a] and Ines REINECKE[a]
[a] *Institute for Medical Informatics and Biometry, Carl Gustav Carus Faculty of Medicine, Technische Universität Dresden, Dresden, Germany*

Abstract. The availability of clinical data for researchers is crucial for an improvement of healthcare and research. For this purpose, the integration, harmonization and standardization of healthcare-data from various sources in a clinical data warehouse (CDWH) is highly relevant. Our evaluation taking into account the general conditions and requirements of the project, led us to choose the Data Vault approach for the development of a clinical data warehouse at the University Hospital Dresden (UHD).

Keywords. Data Warehouse (DWH), Data Vault (DV), agile development

1. Introduction

A clinical data warehouse (CDWH) can be used to integrate and to harmonize medical data from various clinical sources in a central repository [1].The major requirements given by the Medical Informatics Initiative (MII) [2] and the general conditions of the implementing hospitals (i.e. available staff and budget) lead to the following question, addressed in this paper:

Which DWH modeling approach offers agile and flexible development of a CDWH according to volatile requirements of the MII for a small development team with limited budget, limited time and the aim to develop a framework for performant analytics of medical data?

2. Methods

We investigated the most suitable modeling approach for the CDWH based on the specific conditions at the University Hospital Dresden (UHD) and the requirements of the MII [2], conducting a selective literature research while considering the determining

[1] Corresponding Author: Stephan Lorenz, Institute for Medical Informatics and Biometry, Carl Gustav Carus Faculty of Medicine, Technische Universität Dresden, Dresden, Germany; E-mail: stephan.lorenz@mailbox.tu-dresden.de.

factors of the healthcare environment and the characteristics of heterogeneous medical data.

3. Results

A *flexible and agile development process* (Indicators for evaluation are marked *cursively*) is essential in order to make required adjustments caused by volatile requirements [2] and changing data sources without being limited by the necessity to redesign the entire data model.

According to Yessad and Labiod (2016) and Naame and Jovanovic (2016), the DV approach is the most suitable choice in terms of *philosophy*, *methodology* and *architecture*, *data integration* and *extraction, transform, and load (ETL)*, as well as *data modeling*. This approach is superior to the approaches of Inmon and Kimball in various aspects and can effectively address the challenges of *agile data-integration*, *-management*, and *-governance*. DV also follows an agile methodology, which results in faster *deployment times* and lower *costs of development*, compared to the other approaches, benefitting by its *flexibility* in handling *new analytical requirements* and *changes to the source data model(s)* [3,4].

4. Discussion and Conclusion

While the Data Vault approach offers many benefits such as flexibility, productivity, and scalability, it also has a higher level of complexity due to the high data standardization. In order to facilitate analysis and reporting, (dimensional) data marts (i.e. i2b2, OMOP CDM, MII core dataset, TriNetX) are used with the Data Vault approach in the CDWH of the UHD. These data marts need to adapt and reflect any changes in data sources through modification of the data delivery to the marts. It is important to consider this additional layer of complexity when implementing a CDWH using the DV approach. In order to identify and address potential bottlenecks in analytical performance, we regularly need to monitor and evaluate the performance and complexity of the CDWH.

References

[1] Hamoud A, Hashim AS, Awadh WA. Clinical data warehouse: a review. Iraqi Journal for Computers and Informatics. 2018;44(2).

[2] The Medical Informatics Initiative's core data set | Medical Informatics Initiative, (n.d.). https://www.medizininformatik-initiative.de/en/medical-informatics-initiatives-core-data-set (accessed January 3, 2023).

[3] Yessad L, Labiod A. Comparative study of data warehouses modeling approaches: Inmon, Kimball and Data Vault. In2016 International Conference on System Reliability and Science (ICSRS) 2016 Nov 15 (pp. 95-99). IEEE. doi:10.1109/ICSRS.2016.7815845.

[4] Naamane Z, Jovanovic V. Effectiveness of Data Vault compared to Dimensional Data Marts on Overall Performance of a Data Warehouse System. International Journal of Computer Science Issues (IJCSI). 2016 Jul 1;13(4):16. doi:10.20943/01201604.1631.