International Journal of Trend in Scientific Research and Development (IJTSRD)
Volume 4 Issue 4, June 2020 Available Online: www.ijtsrd.com e-ISSN: 2456 – 6470

Understanding the Need of Data Integration in E- Healthcare

Mrs. Shashi Rekha. H.[1], Dr. Chethana Prakash. M[2], Dr. K. Thippeswamy[3]

[1,3]Visvesvaraya Technological University, Mysuru, Karnataka, India
[2]B.I.E.T, Davangere, Karnataka, India

ABSTRACT

This paper discusses the current scenario of e- healthcare and different dimensions of Big data in healthcare and the importance of data integration in e- health care and the challenges associated with data integration and associated uses of data integration with respect to different use cases which might be helpful to physician's decision making because the data driven decision making involves combination of heterogeneous data which includes Electronic Health Record containing different types of data and connected healthcare organization in order to provide value based connected healthcare which would be useful to primary healthcare center located at different location because patients suddenly expect their healthcare experiences to be as exceptional and as transparent as those of retail or banking, and physician's have to scramble to adjust to these new expectations due to lack of data integrity.

KEYWORDS: *E- Healthcare, Big Data, Heterogeneous data, Data Driven Decision Making, Connected Health Care Organization*

How to cite this paper: Mrs. Shashi Rekha. H. | Dr. Chethana Prakash. M | Dr. K. Thippeswamy "Understanding the Need of Data Integration in E- Healthcare" Published in International Journal of Trend in Scientific Research and Development (ijtsrd), ISSN: 2456-6470, Volume-4 | Issue-4, June 2020, pp.665-667, URL: www.ijtsrd.com/papers/ijtsrd31007.pdf

IJTSRD31007

1. INTRODUCTION

Currently, computers dominate the world, and therefore the different data types are stored in computer warehouses. The biggest challenge in the big data era lies in collecting and storing data, the real challenge is the integration of all kinds of data sources to find successful big data value creation. This occurs, because many of these data sources have to be integrated with other data sources, because the data sources integrated or not often contain data variables that have to be further processed to create useful information. Data sources vary in terms of content, sources, and have to be present within the commercial data environment. There is a strong belief that instead of focusing on the "V" i.e volume of big data, the real challenge lies in addressing the "V" i.e variety of data sources, especially because every source adds a specific dimension to the healthcare data environment. Right now ,there is an existence of large number of autonomous and heterogeneous data storage repositories available on the worldwide information storage structure makes it impossible for users to be aware of the locations, structure/organization, and languages used in queries and semantics of the data stored in various repositories in different formats. So, there is a critical need to enhance current browsing, navigation and information retrieval techniques which focuses on information content sharing and semantics. In any strategy, which focuses on information content, the most difficult problem is usage of different vocabularies used to describe similar information across different domains. Big data dominates in various disciplines of real time applications such as E- governance, Smart Transportation, E- healthcare where the concept of data integration plays an important role because the data consists of heterogeneous data formats and online sharing of this data plays an important role in rapid decision making process [1]

2. E- healthcare:

In coming years, the healthcare field generates large amounts of data in different data formats. The value based treatment in hospitals and connected organization and digitization of data prefers to have the computerized view of data rather than hard copy format. The health care data includes Electronic Medical Record (EMR) of patient's data, clinical reports, doctor's prescription, diagnostic reports, medical images, pharmacy information, and health insurance related data. [2]. All these information collectively forms Big Data in Health care. By employing the analytics of big data in healthcare would produce the predicted results by understanding the data given to improve the health care and life time expectancy, proper treatment at early stages at low cost to the unreachable people in the mean time. The healthcare data is rich in information but poor in knowledge. There is a lot of data available within the healthcare systems which are rich in knowledge. But, there is a shortage of effective analysis tools to discover hidden relationships among associated types of data. An effective analysis tool helps to discover hidden relationships among different types of data which are heterogeneous data. [3]. Healthcare analytics is the systematic usage of healthcare data and

analysis tools to get associated and related insights by applying analytical methods ,e.g. such as mathematical statistical, contextual, quantitative, predictive, cognitive, and other techniques, to drive data driven fact -based decision making for Predicting ,Diagnosing ,Planning & Managing , and Learning in healthcare[4]. E-Health clouds are increasing popularly by facilitating the storage and sharing of big data in healthcare. But, adoption of such systems also brings about a series of challenges, especially related to the security and privacy of highly sensitive health data associated with the complete information of patient related data andintegration of heterogeneous data, data visualization and interpretation etc [5]. As per Dr. Martin Makary, professor of surgery and health policy at Johns Hopkins School of Medicine "People don't just die from heart attacks and cancer, they die from system wide failings and poorly coordinated care.[6]

3. Data Integration:

It combines the data obtained from many sources and sensors into a coherent data store, as in data warehousing. These sources may include multiple database, data cubes or flat files etc., store under a unified schema and that usually reside at a single site. Data integration is the problem of combining data residing at different sources in different formats and for providing the user with a unified view of these data [8, 9, 10]. The problem of designing data integration systems is important in current real time E-healthcare applications in rural healthcare system and is characterized by a number of issues that are interesting from a theoretical point of view. At higher level, data integration is defined as the combination of technical and business processes used to combine data from disparate sources into meaningful and valuable information. However, data integration can mean a lot of different things across different contexts [11]. Data integration involves gathering of data in different formats from different source repositories which can be on cloud or on-premises or both and putting it in unified form to be used in reporting and analysis. In short, data integration process centralizes data and makes it large which looks difficult to analyze if not integrated properly as shown below in the fig 1.

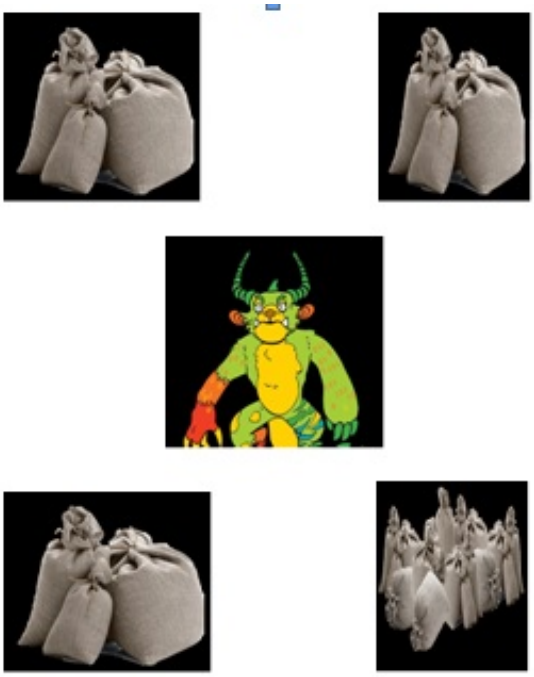

Figure 1: Healthcare Information Aggregation

4. Challenges of Data Integration:

Clinical data integration, however, has many challenges. Following is a discussion of the top existing challenges:

4.1. Clinical data rarely adheres to a schema, data dictionary, or data definition.

Discrete clinical data can be ingested and stored in databases by a wide range of electronic medical record (EMR) stored in hospitals .These hospitals store clinical data in proprietary schema structures, and the schema structures of a particular hospital may not be interoperable with those of other hospital .This creates a challenge, since understanding the data in its elemental form takes a lot of time and effort.

4.2. Standard formats like HL7 v2 and HL7 v3 (CCD, CCDA, etc.) exist, but vendors are not consistent in their implementation.

Most EMR systems rely on HL7 v2 for messaging, which is a very loose standard with a lot of variability in implementation. To ensure interoperability between different EMR systems, HL7 v3 was proposed as a semantic data representation, messaging, and document standard for patient medical information. However, apart from the document standard (CDA, CCDA), HL7 v3 has not had the adoption that was initially expected. Furthermore, even when EMR systems use HL7 v3, like for the CCDA document standard, loopholes can be exploited by vendors to produce CCDA-compliant documents that are still extremely inoperable between systems. These issues have rendered interoperability between EMR systems extremely challenging.

4.3. A huge amount of valuable clinical data exists as unstructured free text.

Most of the data recorded by physicians and providers is unstructured in nature, as physicians may not have the time or inclination to record data in structured formats. On the other hand, most health information technology systems use structured data to manage a wide range of healthcare processes that include care management, measuring performances such as clinical quality, cost re-imbursement, and reporting. If patient data elements for procedures, diagnosis, and medications are not recorded in a structured format, these healthcare processes for a patient are impacted since they cannot be tracked and measured properly.

A typical example is a patient's foot exam, for which the physician writes documentation, recording "pedal calluses" in an unstructured note. Even though the patient underwent a foot exam and was diagnosed with "pedal calluses," the quality measure does not record the fact that a foot exam for this patient was completed. Furthermore, since health information technology systems do not use unstructured data widely, the required interventions for managing "pedal calluses" are not triggered.

4.4. Reconciliation of clinical data with claims data is onerous.

As described earlier, claims and clinical data are meant for different purposes. Claims data is structured and is meant to provide a record of all the medical services that incur cost. Clinical data is often unstructured, and it records information about the medical care provided to the patient. Furthermore, clinical and claims data use different code sets to record diagnosis, procedure, and medication codes. These

differences between claims and clinical data make a straightforward reconciliation of information from the claims and clinical records laborious and time-consuming.

Healthcare data is quite complex and any text analytics/NLP solution needs to offer the following capabilities:

A. Recognizing synonyms of medical terms, and presenting the results accordingly. For example, "Hand pain" and "Finger pain" mean the same thing.

B. Processing large volumes of data in terms of millions of records.

C. Recognizing the context of healthcare terms, for example the terms "Family history of myocardial infarction," "No MI," and "MI ruled out" mean different things, so they must be handled accordingly to predict the disease. [11]

5. Benefits of Data Integration in E- healthcare:

The significance of data integration increases in hospitals & connected organizations because they are in use of more systems and applications. Organizations need the data stored in different data format in different systems to be brought together to achieve a comprehensive unified view. Then, doctors can use this integrated data to enhance up data driven computational intelligence to provide connected value based health care located at different locations [13].

The benefit of data integration includes the following:

- A single-view of the data available by keeping data synchronized across organizations and systems.
- A Comprehensive, informative view of patients EMR.
- Availability of data to doctors and staff across an organization in Unified view.
- Opportunities for analyzing and predicting data driven and data-based decision making based on high quality and complete data of the patient available.

6. Conclusion:

This paper discusses the significance of data integration in E-healthcare along with big data characteristics in E-healthcare comprising of heterogeneous data and challenges associated with data integration and the benefits of data integration when the data is collected from diversified sources to make data driven decision making in order to provide value based healthcare useful for the doctors and nurses while providing online services to E- healthcare system which would accelerate Universal Health Coverage and more and more organizations can realize the growing importance of data mobility which can increase care team productivity to deliver better patient outcomes located at different locations. The need of the hour is to provide Connected Value based healthcare to the unreachable people located in different places.

References

[1] Mohit Dayal∗ and Nanhay Singh Indian Health Care Analysis using Big Data Programming Tool," Twelfth International Multi-Conference on Information Processing-2016 (IMCIP-2016 "

[2] Muni kumar N, Manjula R,"Role of Big Data Analytics in Rural Health Care – A Step Towards Svasth Bharath", International Journal of Computer Science and Information Technologies, vol 5(6), pp 7172-7178,2014.

[3] Jyoti Soni Ujma Ansari Dipesh Sharma Predictive Data Mining for Medical Diagnosis: An Overview of Heart Disease Prediction, International Journal of Computer Applications (0975 – 8887) Volume 17– No.8, March 2011

[4] Cortada, J.W., Gordon, D., & Lenihan, B. (2012). The value of analytics in healthcare: from insights to outcomes. IBM Global Business Services, Life Sciences and Healthcare, Executive Report.

[5] A Qi Jiang· Muhammad Khurram Khan· Xiang Lu Jianfeng Ma· Debiao "Privacy preserving three-factor authentication protocol for e-Health clouds. J Super-computing(2016) 72:3826–3849 DOI 10.1007/s11227-015-1610-x

[6] http:// www.EHRintelligence .com

[7] Ragavi, B Srinidhi, V S Anitha Sofia "Data Mining Issues and Challenges: A Review International Journal of Advanced Research in Computer and Communication Engineering Vol. 7, Issue 11, November 2018.

[8] T. Kirk, A. Y. Levy, Y. Sagiv, and D. Srivastava. The Information Manifold. In *Proceedings of the AAAI 1995 Spring Symposium on Information Gathering from Hetero- generous, Distributed Environments*, pages 85–91, 1995.

[9] R. Hull. Managing semantic heterogeneity in databases: A theoretical perspective. In *Proc. of the 16th ACM SIGACT SIGMOD SIGART Symp. on Principles of Database Systems (PODS'97)*, 1997.

[10] J. D. Ullman. Information integration using logical views. In *Proc. of the 6th Int. Conf. on Database Theory (ICDT'97)*, volume 1186 of *Lecture Notes in Computer Science*, pages 19–40. Springer, 1997.

[11] http://dynamic.globalscape.com/files/whitepaper_5RingsDataIntegrationHell.pdf

[12] Clinical Data Integration for Health Plans: Top 5 Challenges and Solutions" ZEOMEGA Population Health Management Solution [White Paper].

[13] http://dynamic.globalscape.com/files/whitepaper_What is Data Integration.pdf.