Building a National Clinical Data Warehouse

Ustun Yildiz[1] and Ayse Keles[1,2]

[1]*Agency of Health Informatics, Ankara, Turkey*
[2]*Yildirim Beyazit University, Ankara, Turkey*
{ustun.yildiz, ayse.keles}@saglik.gov.tr

Keywords: EHR, Clinical Data Warehouse, Data Privacy, Health Transformation Program

Abstract: Turkey has created an e-health vision along with the Health Transformation Program. In this framework, standard coding systems such as standard definitions of the institutions, databank of healthcare providers, standard disease, drug and medical supplies classifications have been developed and a national clinical data warehouse (Health-NET) was established. Health-NET is an integrated, safe, fast and expandable health information system which aims to improve efficiency and quality of health services by collecting all kinds of data produced in the health institutions in line with the standards and generating information adequate for all stakeholders of the collected data.

1 INTRODUCTION

Access to healthcare is regarded as a fundamental element of social development. While classic theories of development explain development in terms of many socio-economic and cultural indicators such as GNP per capita, level of industrialization and employment, and consumption level of primary goods and services, today, it is explained primarily in terms of indicators related to the access to education and healthcare. At the end of 2002, the status of the Turkish health system made it necessary to undertake radical changes in many areas from service delivery to financing and from human labor to information systems. In order to improve the quality of service in these areas, Turkey has gone through a Health Transformation Program (HTP) started as early as 2003(OECD, 2014)(Akdag, 2008)(Akdag, 2011). As a result of the health infrastructure rehabilitation efforts, Turkey has observed a rapid decline in the Under-5 Mortality Rate (U5MR) from 29 in 2003 to 7.7 in 2011 per 1000 births (T.D., 2010) Similarly, the life expectancy at birth has increased to 75 in 2011 from 70 in 2004. Consequently, the satisfaction with the government provided health services was measured 75.9% in 2011 comparing to 39.5% in 2003(WHO, 2012). In the context of HTP, the effective collection and use of nation-wide Electronic Health Records (EHR)s became a primordial goal. In

parallel with the development of the relevant legislative framework, a national clinical warehouse (CDW) that collects EHR and other operational data from all health organizations in the country had been established in 2012 and became operational. As analogous to the worldwide approach of putting the health information at the center of decision processes along with the patient, the data collected from the healthcare facilities are used to structure and manage the new health infrastructure of the country (De Mul, 2012)(Yoo, 2014). Generally speaking, CDW is used to empower traditional application software in order to analyze public health behavior and support several different decision workflows such as clinical quality improvement (Weiner, 2012), pay-for-performance(Van Herck, 2010) and evidence-based medicine systems (Sacklett, 2000). The integration and use of diverse healthcare data from various sources into the same clinical repository is a challenging problem when implemented at a nation-size level. The problem becomes even more complex when Hospital Information Systems (HIS) that create the collected data are autonomous and implemented with different technologies. The current HIS market in Turkey includes more than 150 private firms that implement their customized software in more than 1500 healthcare facilities operating with more than 200.000 beds capacity (Kose, 2013). In this paper, we share our experience in implementing CDW and its use in the critical decision processes of the Ministry of

Health of Turkey. We summarize the design and operation processes, the software architecture, its use in Health Transformation Program and the lessons learned.

2 BACKGROUND

In this section, we describe some of the main aspects of a large-scale CDW that needs conceptual and technical considerations.

2.1 Data Collection

The health institution that provides CDW with data is composed by Level I Family Practitioners, Level II Public and Private Hospitals and Level II University Hospitals and Research Centers. It is obvious that the data collection might not be achieved with 100\% accuracy due to the complexity of data and the widespread use of the system. Data collection services are frequently updated due to the updates in the data packages definitions. This continuous process requires a tight coordination with Hospital Information System (HIS) providers that will implement client component for data upload to collection services. Due the difficulties in the integration process that needed continuous support, a help desk has been established by the Ministry of Health. The help desk provided regular data on the amount of data collected by CDW, the amount and type of upload errors to HIS firms. It has been recognized that the success of integration capabilities of institutions that belong to different levels are different. Level I institutions were 99% successful in uploading their data as described in the integration kits of CDW public website. Level II institutions were less successful in sharing their data. The main reason behind the successful integration of Level I institutions was based on the underlying business model which makes possible the calculation of practitioners salaries based on the collected data by CDW. While Level I data collection was successful, Level II data collection had not been as expected and it was below 60% average for the first year. The main reason for the lack of data was that Level II institutions were not subject to any business model implemented by CDW. The second year of the implementation of CDW, the data collection rate was increased to 77% as the parameters for Service Quality Standards were started to be calculated with CDW data. The integration of Level III research institutions were even less successful given that they were not managed by the Ministry but by independent universities. During the third year, the integration of Level II has improved by the increasing calculation of different healthcare service parameters using available CDW data.

2.2 Data Quality

It is important that we should not confuse the concept of 'data quality' to the aforementioned concept of 'clinical quality data'. While 'clinical quality data' is specific data based on clinical quality indicators which help to understand the clinical quality of the services provided to patients with specific diseases e.g. diabetes, stroke etc. On the other hand, 'data quality' is about the quality of any data collected by CDW and based on certain criteria sets e.g. complete, valid/correct, timely, without duplication (Kahn, 2012)(Arts, 2002). In this section, we detail the issues and our solutions to improve the 'data quality' of the 'clinical quality indicators' collected for CDW. The major issue that has been encountered in establishing the data quality was the data wrongly packed that do not fulfill data package acceptance rules. We identified the main reason was the difficulties associated with the establishment of data packages to be sent to CDW. On the other hand, we identified HIS users use different ICD-10 (WHO, 2004) codes for certain diagnosis and diseases because HIS require sophisticated data input interfaces for the latters. One particular aspects that needs special attentions is the geographic and temporal properties of the clinical data. It is observed that public health indicators could be misleading based on certain periods and on locations (e.g. Temporary Refuge Spaces), these problems are configured by expert knowledge.

2.3 Data Privacy

The privacy of EHR had been a high priority concern in the implementation of CDW and the tools that manage its data. In the collection process of EHR, the definition of data packages was mainly defined by the public health surveillance necessities and the establishment of personal health records. The idea behind the establishment of personal health records is to support the continuity of healthcare and prevent redundant services such as radiology. One other advantage of the involvement of patients in the structuring the records is the elimination of inconsistencies. Turkish citizens have a unique and publicly available 10 digit number. The use of this number considerably facilitates the consolidation and access of personal health records but in the same time could be a major privacy concern.
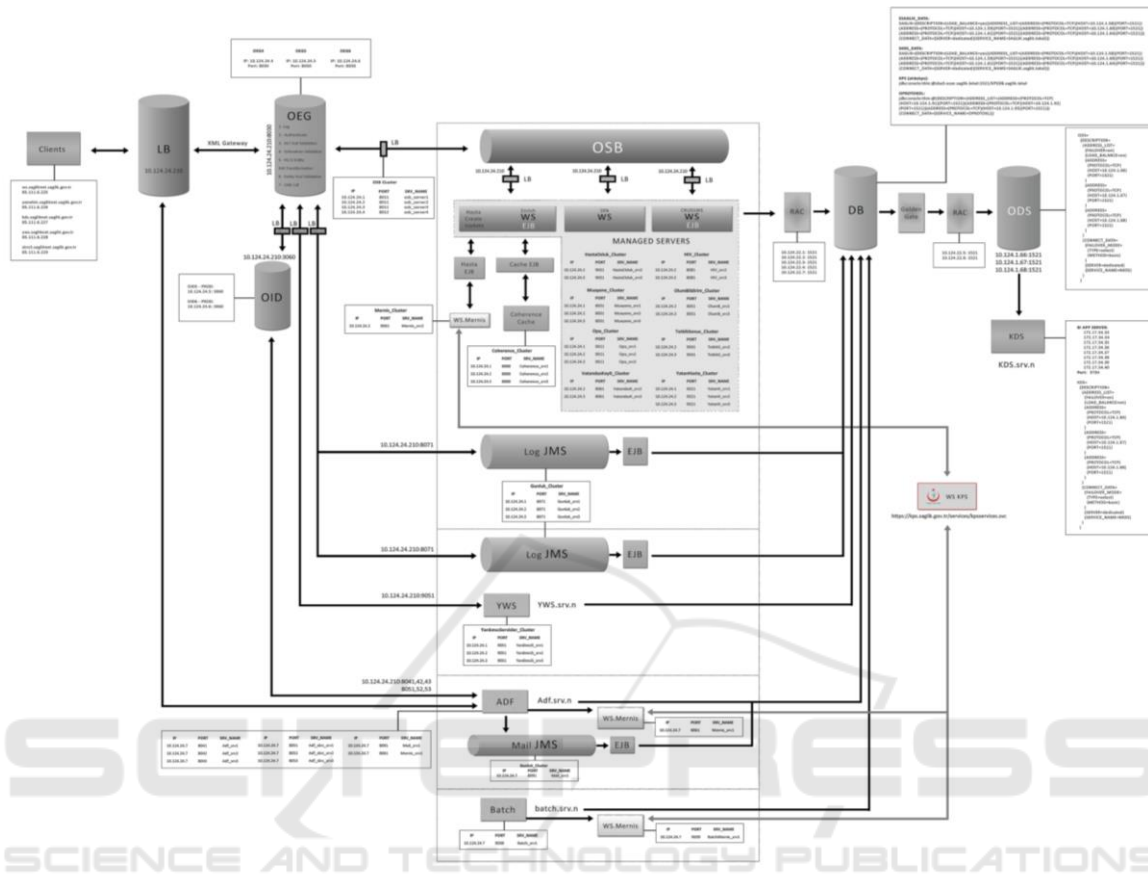
Figure 1: Overview of Architecture

In order to overcome with privacy concerns, an electronic consent form has been implemented in personal health record system (MoH, 2015). Patients can login and configure their preferences on the use of their records. The preferences are overwritten only in the case of emergency room services. It is recognized that the collection private health information and its computation after de-identification of data still can be a privacy concern as the identification of patient is possible with the combination of different queries (Fernandez, 2013) For certain diagnosis and diseases codes (e.g. HIV), we have decided to collect the records without identifiable information and a cryptic number without publicly available citizenship number

3 SOFTWARE ARCHITECTURE

Figure 1 describes the overview of the architecture of CDW with components related to data collection through web services, storage and data analyze modules. Messages coming from healthcare facilities are collected under HL7 form by Load Balancer component. Received messages are forwarded to an available XML Gateway. In parallel, HL7 messages are processed asynchronously by the JMS log queue. During its flow throughout the XML Gate way, the Authentication happens with LDAP on OID. Next, HL7 message is validated against XSD schema and Schematron processes business rules (mandatory and optional fields in data packages). After the validation, HL7 messages are transformed to a local data format to be saved in the database. The transformed messages are also validated against XSD Entity.
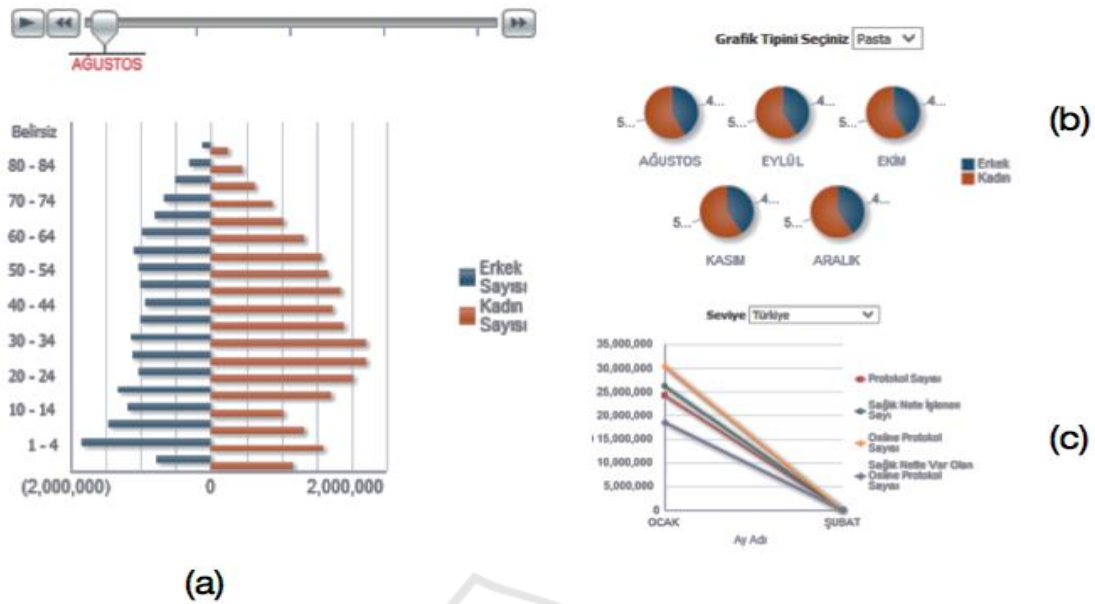
Figure 2: Functions available on dashboard

Collected messages are transferred to Service Bus. In the latter, the data is unpacked and the patient ID is verified through an external ID verification service (WS-Mernis). Figure 2 describes some of the functions available to users.

4 CONCLUSIONS

This paper has presented a small set of lessons learned from the establishment of a National Data Warehouse. The main research problems that we took the challenge on were: Data quality, data collection and data privacy issues. Although the patient count and basic computational problems were manageable with relation data models and systems, the advanced computation of patient records such as disease correlation analysis, organization of cohorts for evidence-based medicine applications require the use of big data solutions.

REFERENCES

OECD, Oecd reviews of health care quality: Turkey 2014doi:http://dx.doi.org/10.1787/9789264202054-en. URL /content/book/9789264202054-en

R. Akdag, Turkey health transformation program 2003-2010, Tech. Rep. 839, Re- public of Turkey, Ministry of Health (June 2011).

R. Akdag (Ed), Health transformation program in turkey and primary health care services, Tech. Rep. 770, Republic of Turkey, Ministry of Health (November 2008).

T. D. V. D. Lilia Jelamschi, Decline in the under-5 mortality rate (u5mr) in turkey: A case study, Tech. rep., UNICEF, http://www.unicef.org.tr/files/bilgimerkezi/doc/declin-in-the-under-5-mortality-rate-u5mr-in-turkye-a.pdf (March 2010).

W. H. O. R. O. for Europe, Successful health system reforms: The case of turkey, Tech. Rep. WHO LIS e96508, World Health Organization: Regional Office for Europe (May 2012).

M. De Mul, P. Alons, P. Van der Velde, I. Konings, J. Bakker, J. Hazelzet, Development of a clinical data warehouse from an intensive care clinical information system, Computer methods and programs in biomedicine 105 (1) (2012) 22–30.

S. Yoo, S. Kim, K.-H. Lee, C. W. Jeong, S. W. Youn, K. U. Park, S. Y. Moon, H. Hwang, Electronically implemented clinical indicators based on a data ware-house in a tertiary hospital: its clinical benefit and effectiveness, International journal of medical informatics 83 (7) (2014) 507–516.

J. P. Weiner, J. B. Fowles, K. S. Chan, New paradigms for measuring clinical performance using electronic health records, International Journal for Quality in Health Care 24 (3) (2012) 200–205.

P. Van Herck, D. De Smedt, L. Annemans, R. Remmen, M. B. Rosenthal, W. Sermeus, Systematic review: effects, design choices, and context of pay-for- performance in health care, BMC health services research 10 (1) (2010) 247.

D. L. Sackett, Evidence-based medicine, Wiley Online Library, 2000.

R. Kose, Health statistics year book, Yearbook 978-975-590 509-9, Agency of Health Research (2013).

M. G. Kahn, M. A. Raebel, J. M. Glanz, K. Riedlinger, J. F. Steiner, A pragmatic framework for single-site and multisite data quality assessment in electronic health record-based clinical research, Medical care 50.

D. G. Arts, N. F. De Keizer, G.-J. Scheffer, Defining and improving data quality in medical registries: a literature review, case study, and generic framework, Journal of the American Medical Informatics Association 9 (6) (2002) 600–611.

W. H. Organization, et al., International statistical classification of diseases and health related problems (the) icd-10, Ph.D. thesis, World Health Organization (2004).

T. M. of Health, E-nabiz personal health records system, , https://enabiz.gov.tr/Giris.aspx (2015).

J. L. Fernández-Alemán, I. C. Señor, P. Á. O. Lozoya, A. Toval, Security and privacy in electronic health records: A systematic literature review, Journal of biomedical informatics 46 (3) (2013) 541–562.