National Health Data Warehouse Bangladesh for Remote Health Monitoring: Features, Problems and Privacy Issues

Shahidul Islam Khan, Bangladesh University of Engineering and Technology
ABSTRACT

It is a great initiative by the Government of Bangladesh to establish the National Health Data Warehouse of Bangladesh. Currently, health data from different government organizations are being collected and summarized through two open-source software solutions: OpenMRS and OpenHIE. Experts manually mapped and integrated the contents of several available computerized databases and made them interoperable. The project was named as Data Management and Information System (DMIS). Over a two-year period, a team of experts manually mapped and integrated the contents of several separate databases of dissimilar platforms into a standard format. Currently, medical data from different healthcare organizations under Directorate General of Health Services (DGHS) of Bangladesh Government are being collected and summarized through two globally used open-source software solutions: DHIS2 and OpenMRS. The health warehouse also provides a platform for various reporting and data mining functions that could be used by the decision makers to monitor health progress at all levels down to Upazila health complexes [8], [9].

To discover effective knowledge from medical data sets, e.g., correlations among diseases, it is very essential to maintain linkage among health records. Record linkage is the process of identifying record pairs from different information systems which belong to the same real world entity. Record linkage is essential when joining datasets based on entities that may not always share a common identifier, such as national ID or social security number [10], [11]. On the other hand, medical record linkage has an inverse relation with patients’ privacy. If a health dataset preserves record linkage means privacy of individuals are at risk [12], [13]. Nowadays, a growing trend is observed of hacking medical records by cybercriminals. The objective of the hackers is to exploit personal information, which is a lucrative business for them.

Keywords
Health / Medical Data; Health Data Warehouse; Data Security; Privacy; Data Breach; DHIS2, OpenMRS;

1. INTRODUCTION

A data warehouse (DW) is a subject-oriented, integrated, non-volatile, and time-variant collection of data to support management decisions [1], [2]. It unites the data spread all over an organization into a distinct central structure. Main advantages of warehouses are standardizing data across organizations and improved turnaround time for analysis and reporting. Lengthy development time and high cost are treated as its major drawbacks [1] - [4].

A Health Data Warehouse (HDW) is a data repository, different from hospitals’ operational databases. HDW is used to analyze the consolidated historical health data [5]. Development of a Health DW contains two key phases. Firstly, a conceptual view of the DW is specified according to the user requirements in the configuration phase. Secondly, the allied data sources and the Extraction-Transform-Load (ETL) process are determined. After the initial load, during DW operation, data must be refreshed on a regular basis such that data stored in the DW reflect the current state of the operational systems [6], [7].

In the year 2009, Bangladesh government took initiative to develop National Health Data Warehouse (NHDW) with the help of German Donor GIZ. The objective of the warehouse was to build an electronic data repository which bridged the gaps between the various available computerized databases and made them interoperable. The project was named as Data Management and Information System (DMIS). Over a two-year period, a team of experts manually mapped and integrated the contents of several separate databases of dissimilar platforms into a standard format. Currently, medical data from different healthcare organizations under Directorate General of Health Services (DGHS) of Bangladesh Government are being collected and summarized through two globally used open-source software solutions: DHIS2 and OpenMRS. The health warehouse also provides a platform for various reporting and data mining functions that could be used by the decision makers to monitor health progress at all levels down to Upazila health complexes [8], [9].

To discover effective knowledge from medical dataset e.g., correlations among diseases, it is very essential to maintain linkage among health records. Record linkage is the process of identifying record pairs from different information systems which belong to the same real world entity. Record linkage is essential when joining datasets based on entities that may or may not share a common identifier such as national ID or social security number [10], [11]. On the other hand, medical record linkage has an inverse relation with patients’ privacy. If a health dataset preserves record linkage means privacy of individuals are at risk [12], [13]. Nowadays, a growing trend is observed of hacking medical records by cybercriminals. The objective of the hackers is to exploit personal information, which is a lucrative business for them.

In this paper we have presented the available features and limitations of NHDW, Bangladesh. We have also provided an overview of security and privacy risks of integrated healthcare information system. We have provided recommendations for the NHDW to improve the performance and security of the system.

2. NATIONAL HEALTH DATA WAREHOUSE (NHDW) BANGLADESH

Even with momentous health achievements in recent years, the absence of a well-functioning Health Information System (HIS) is the main barrier for the policy makers of Bangladesh to monitor health progress of mass people in real time. The HIS of Bangladesh is still highly fragmented. Data generated by private and public sector providers are not linked and, within the public sector, data from urban and rural areas, and from family planning programs, are handled independently. Numerous overlapping reporting systems result in heavy paperwork burdens and poor data quality [8].

Since 2009, the Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) has been supporting the Management Information System (MIS) unit of the Directorate General of Health Services (DGHS) to strengthen and modernize Bangladesh’s HIS. National Data Warehouse is established in

This research is supported by the ICT Division, Ministry of Posts, Telecommunications and Information Technology, Government of the People's Republic of Bangladesh.
Bangladesh which made existing electronic health datasets from different directorates and vertical programs interoperable for the first time. It is supported by two open source software products developed for the collection and management of health information in developing countries namely DHIS2 and OpenMRS [8], [9]. The first one is for aggregate data and reporting from national warehouse and the second one to incorporate individual patient records. They are described in the next sections.

2.1 The District Health Information Software (DHIS2)

DHIS2 is a tool for collection, validation, analysis, and presentation of aggregated healthcare statistical data [14]. It is generic software rather than a pre-configured database application. DHIS2 has an open meta-data model and a flexible user interface that allows the user to design the contents of a specific information system without extensive programming. It is a modular web-based software package built with free and open source Java frameworks. DHIS2 is open source software released under the BSD license and can be used at no cost. It runs on any platform with a Java Runtime Environment (JRE 6 or higher) installed.

The DHIS2 software is currently being used in more than 45 countries in Africa, Asia, and Latin America. The countries that have adopted DHIS2 as their nation-wide HIS software include Kenya, Tanzania, Uganda, Rwanda, Ghana, Liberia, and Bangladesh. A rapidly increasing number of countries and organizations are starting up new deployments [15]. The customized user interface of DHIS2 used by Bangladesh Government is shown in Fig.1 adopted from [16].

Fig. 1. DHIS2 user interface

The key features of DHIS2 are summarized bellow [14], [15]:

- Provide a comprehensive HIS solution based on data warehousing principles and a modular structure which can easily be customized to the different needs of the health systems - and supports the idea of an integrated HIS at all levels of the health hierarchy.
- Customization and local adaptation through the user interface. No programming required to start using DHIS2 in a new setting (country, region, district etc.).
- Provide data entry tools which can either be in the form of standard lists or tables, or can be customized to replicate paper forms.
- Provide different kinds of tools for data validation and improvement of data quality.
- Provide easy to use - one-click reports with charts and tables for selected indicators or summary reports using the design of the data collection tools.
- A user-specific dashboard for quick access to the relevant monitoring and evaluation tools including indicator charts and links to favorite reports, maps and other key resources in the system.
- User management module for passwords, security, and fine-grained access control (user roles).
- Messages can be sent to system users for feedback and notifications. Messages can also be delivered to email and SMS.
- Users can share and discuss their data in charts and reports using Interpretations, enabling an active information-driven user community.
- Functionalities of export-import of data and metadata, supporting synchronization of offline installations as well as interoperability with other applications.

In summary, DHIS2 provides a comprehensive HIS solution for the reporting and analysis needs of health information users at any level. It concentrated on aggregated health data.

2.2 OpenMRS

OpenMRS is a software platform and a reference application which enables design of a customized medical records system without programming knowledge [17]. It is a common platform upon which medical information system can be built in developing countries. The system is based on a conceptual database structure which is not dependent on the actual types of medical information required to be collected or on particular data collection forms and so can be customized for different uses.

OpenMRS is now in using around the world including Bangladesh, South Africa, Kenya, Rwanda, Zimbabwe, Uganda, Tanzania, India, China, United States, Pakistan, Philippines, and many other places. It is supported by many organizations including international and government aid groups, NGO’s, as well as for-profit and non-profit corporations [18]. The user interface of OpenMRS is shown in Fig.2 adopted from [19].

Fig. 2. OpenMRS user interface

The Key Features of OpenMRS are summarized bellow [16], [17]:

- **Patient repository**: Creation and maintenance of patient data, including demographics, clinical observations, encounter data, orders, etc.
• Data export: Data can be exported into a spreadsheet format for use in other tools (Excel, Access, etc.)
• Modular architecture: An OpenMRS Module can extend and add any type of functionality to the existing API and web apps.
• Patient workflows: An embedded patient workflow service allows patient to be put into programs (studies, treatment programs, etc.) and tracked through various states.
• Cohort management: The cohort builder allows creating groups of patients for data exports, reporting, etc.
• Support for complex data: Radiology images, sound files, etc. can be stored as “complex” observations.
• Reporting tools: Flexible reporting tools available.

2.3 Comparison of Aggregated Health Data and Individual Patient data for NHDW

Individual patient data is data related to a single patient, e.g., patient name, age, diagnosis, medical history etc. This data is typically based on a single patient-health care worker interaction. For example, when a patient visits a health center, a variety of details may be recorded, such as the patient’s temperature, their weight, and various blood tests. Whereas aggregated data is the consolidation of data relating to multiple patients, and therefore cannot be traced back to a specific patient. They are merely counts, such as incidences of Malaria, AIDS or other diseases. Typically, the routine data that a health facility deals with is this kind of aggregated statistics, and is used for the generation of routine reports and indicators. Aggregated data helps for strategic planning within a health system. Aggregate data cannot provide the type of detailed information which patient level data can, but is crucial for planning and guidance of the performance of health systems.

Individual medical records are still in the early stages of roll-out in Bangladesh. If maintained and used correctly, individual patient data have high potential to improve the quality of patient care. A technical issue which requires high attention is the confidentiality and privacy of individual patient data. Better measures need to be taken to protect the security of ever growing amounts of data of DHIS2 and OpenMRS, since the introduction of case-based records linked to identifiable individuals [8], [9].

2.4 Problems and Limitations of NHDW

Currently the national health warehouse is facing the following issues that need to be overcome soon to get the full benefits:

• Lack of skilled manpower to maintain the large scale and fast growing national data warehouse.
• 100% coverage of all the departments of Ministry of Health and Family Welfare (MoHFW) is still challenging.
• No automatic process to import data from different department of the ministry as because the source data is very much fragmented.
• Structure of MoHFW is very much scattered. Data redundancy is happening among few departments.
• Private sector data is missing and no working mechanism is available to include those into the data warehouse in near future.

• Internet infrastructure in the far most area of Bangladesh is still really poor. In most Upazilas, MODEMs are used with very slow internet connectivity.
• Data quality cannot be ensured yet. Sometimes unreal and practically impossible data are inputted as a consequence of data entry error.
• Reports are not standardized yet according to the international standards. Still there are many bugs in different reports.
• Lack of proper measure to protect data privacy for individual patients.

3. USE OF NHDW INFORMATION FOR REMOTE HEALTH MONITORING

The Government of Bangladesh has started to use the data available to NHDW for monitoring and decision making purposes. In this section, for better understanding, we have presented some graphical examples of using the information of the data warehouse for remote health monitoring. The figures are taken from MIS, DGHS public access web portal [20]. Following figures Fig. 3, Fig. 4, and Fig. 5 depicted the scenario of monitoring number of death by age groups in National level, Divisional level and District level.

![Fig. 3. Number of Death by age group: National view](Image)

![Fig. 4. Number of Death by age group: Divisional view (Chittagong)](Image)
From the following figures Fig. 6, Fig. 7, Fig 8, and Fig. 9, we can have the insight of percentage of Cesarean operations for child birth in National level, Divisional level, District level, and Upazila level.

The last figure Fig. 10 shows how the management of manpower is possible using the current NHWD. It illustrates the attendance situation of the healthcare staffs in different upazilas of Sirajganj district.

From the above figures we can say that, as the top level decision makers of the Bangladesh Government can have the statistical data of the whole country within minutes, the NHWD can play a vital role to significantly improve the healthcare situation in Bangladesh.
4. SECURITY AND PRIVACY ISSUES RELATED TO HEALTH DATA

People do not like their medical records to be revealed to others. All medical records are subject to stringent laws that address user access privileges. By law, security and authentication systems are often required for individuals that process and store medical records. Physicians in many cultures and countries have standards for doctor-patient relationships which include maintaining confidentiality. Patient data is highly confidential and therefore must be protected so that no one other than doctors can get it. When in paper, it must be properly stored in a secure place. For computers, patient data needs secure systems with passwords, restrained access and audit logs. Security concerns for aggregated data are not as crucial as for patient data, as it is usually impossible to identify a particular person to a aggregate statistic.

National health DW raises high risk to data security and privacy of individual people. After full deployment of warehouse, health service providers, and healthcare researchers can have access to private health data of millions of patients without bar [21]. It is not possible to guarantee that all doctors, health researchers, and health service providers will maintain patients’ privacy by themselves. So the Govt. should take proper measure so that individual patient cannot be identified from health database or warehouse and their privacy is safeguarded.

4.1 Health Data: Target of Hackers

There is a growing trend of hackers focusing into medical records. Their objective is to exploit personal information, which is a lucrative business to them. In U.S., a stolen Social Security number might sell for 25 cents in the underground market, and a credit card number for $1. Whereas sell value of a comprehensive medical record may varies from 10$ to $1,000 in black market [22], [23]. The data for sale includes names, birth dates, health policy numbers, diagnosis codes and billing information. Fraudsters use this data to create fake IDs to buy medical equipment or drugs that can be resold, or they combine a patient number with a false provider number and file made-up claims with insurers, according to experts who have investigated cyber attacks on healthcare organizations.

Health data on medical devices is on top position in the most wanted list of cybercriminals. For three years in a row, healthcare sector has reported the highest number of breaches and accounted for 42.5 percent of cyber attacks in 2014 [24]. Unlike credit cards, which can be quickly cancelled, medical identity theft is not immediately noticed by patients or providers. So criminals can use the information for longer periods to buy expensive medical equipment and drugs for resale, or file fraudulent insurance claims.

4.2 Health Data Privacy Laws in Bangladesh

There is no specific data protection law or privacy law that specifically applies to electronic health data in Bangladesh. In addition, still now, Bangladesh has not implemented specific laws that govern the use and disclosure of health and medical data in general similar to HIPAA in the United States. Certain protection for medical information is provided by the medical ethics and rules established by the Bangladesh Medical Association (BMA). The Medical Practitioners Act does not provide for any secrecy of information or data protection or communication, but physicians, accredited under BMA, are bound by its codes and declarations. Under such rules, patient-specific medical data may be disclosed only to the patient except in certain specified situations, such as where there is a possibility of adverse effect on the medical condition of a patient. [25].

4.3 Data Breaches of Health Information Systems

A data breach or leakage is defined as any incident involving the loss or exposure of digital personal records. Personal records are defined as data containing privileged information about an individual that cannot be readily obtained through other public means. This privileged information is only known by an individual or by an organization under the terms of a confidentiality agreement. According to 2015 Fifth Annual Benchmark Study on Privacy & Security of Healthcare Data which covered 90 healthcare organizations in the U.S., More than 90% of healthcare service providers had a data breach, and 40% had more than five data breaches over the past two years [24]. According to the report, for the first time, criminal attacks are the number one cause of healthcare data breaches. Criminal attacks on healthcare organizations are 125% higher compared to five years ago. During interview, 45% of healthcare organizations say the root cause of the data breach was a criminal attack and 12% say it was due to a malicious insider. According to the Fifth Annual Study on 2014, medical identity theft nearly doubled in five years, from 1.4 million adult victims to over 2.3 million in 2014 [26]. The types of healthcare data breach in the U.S. in year 2015 are illustrated in Fig. 11.

Hacking is also increasing in the Healthcare Servers at a shocking rate. From 21th October 2009 till data there are 1279 health data breach reported, among which number of server attacks is 190 that is average 14.85%. But up to 1st August 2015, in last 12 months $8 of 255 are server attack, which is 22.74%. We have uncovered this insight by analyzing the data provided by U.S. Department of Health and Human Services [27]. Hackers are increasingly targeted to the health servers which are very alarming to any national level health information system development.

![Type of health data breach experienced by health service providers in year 2015](image-url)

Fig. 11. Type of health data breach experienced by health service providers in year 2015

Prevention is better than cure. Security and privacy of healthcare data can never be afterthought. Though there is no reported case of healthcare data breach in Bangladesh till date, but any time cybercriminals can attack on the national health data warehouse. It is the proper time for the Government of Bangladesh to take initiatives regarding protection of this highly sensitive integrated data repository.
5. RECOMMENDATIONS

Followings are our recommendations to improve the usability, performance, and security of the NHDW of Bangladesh.

1. It is very important to include all healthcare providers, regardless public or private, in the NHDW soon. If the coverage of the national data repository reaches 100%, then the Govt. will be able to take more fruitful administrative and management decisions.

2. Data quality is very important because utilization of the data will improve the service delivery. Policy makers can be benefited by using quality data for taking better decisions. There is scope to improve the quality of the available data in the warehouse. For future data collection, data entry personnel at Upazila level should be given more training.

3. Currently there are more than 30 datasets (reporting forms) in the customized DHIS2 used by the Govt. These forms need standardization according to the international healthcare software standards.

4. Currently there is no institutional structure within the govt to support the NHDW. It is very essential to build the capacity of the Govt. in this regard.

5. For privacy reason, no medical record can be stored in any level, from diagnostic centers to NHDW, with personal identifiable attributes of the patients.

6. To facilitate knowledge discovery process of the Healthcare researchers, sufficient record-linkage data have to be kept in medical records by replacing personal identifiable attributes with unique code using suitable computer cryptographic technique.

7. A data-protection strategy have to be implemented that will cover data everywhere it is stored, and at every stages, from creation and processing, to storage, backup and transmission.

6. CONCLUSION

Health data warehouse development is a complex and time consuming process but is essential to deliver quality health services. The Government of Bangladesh has taken initiatives to improve the citizens’ health related services using ICT. One of the major initiatives is to develop National Health Data Warehouse (NHDW). Preserving record linkage by retaining identifiable attributes in health data warehouse is essential for effective data mining. On the contrary, identifiable health data have high risk to patients’ privacy and also increase the security risks of a health mining. On the contrary, identifiable health data have high risk to attributes in health data warehouse is essential for effective data mining. In this paper, we have briefly presented the available features and limitations of the NHDW. We have also provided an overview of security and privacy risks related to integrated healthcare information system. We have given some recommendations to improve the performance and security of the NHDW.

REFERENCES

[25] PATIENT PRIVACY IN A MOBILE WORLD http://www.trust.org/contentAsset/raw-data/03172bebf0f11-438e-94be-e029786de3036/file