INTEROPERABILITY MODEL FOR HEALTH INFORMATION SYSTEMS AND ASSOCIATED BENEFITS CHALLENGES AND APPROACH

Apurv Mehra

IIIT-D-MTech-CS-09-064

August 12, 2014

Indraprastha Institute of Information Technology
New Delhi

Thesis Committee
Amarjeet Singh (Chair)
Pushpendra Singh
Sundeep Sahay

Submitted in partial fulfillment of the requirements
for the Degree of M.Tech. in Computer Science.

This research was partially funded by DEITy and supported by NRHM-HP.

Keywords: Health Information System, Interoperability, OpenMRS, DHIS2, SDMX-HD, Weather, Air Pollution, Disease, Demographics

Certificate

This is to certify that the thesis titled **" Interoperability Model For Health Information System And Associated Benefits Challenges And Approach"** submitted by **Apurv Mehra** for the partial fulfillment of the requirements for the degree of *Master of Technology* in *Computer Science & Engineering* is a record of the bonafide work carried out by him under my guidance and supervision in the Mobile and Ubiquitous Computing group at Indraprastha Institute of Information Technology, Delhi. This work has not been submitted anywhere else for the reward of any other degree.

Dr. Amarjeet Singh
Indraprastha Institute of Information Technology, New Delhi

Abstract

Recent IT advancements have resulted digital collection of health related information. There exist diverse health information systems today, each addressing a specific need in the healthcare domain e.g. Hospital Information System collecting clinical, patient and administrative data; Health Management and Information Systems (HMIS) collecting health indicators across an area (such as mortality, births and vaccination) and systems for Telemedicine to provide access to healthcare for areas that are not easily accessible. Integrating the information collected across such diverse systems can be technically challenging, primarily due to the different data models and work flows being adopted to address specific requirements catered by such systems. However, such integration can provide significant benefits such as single point of access, detailed and real-time analysis and reduction in duplication of information.
In this work, we present a case study of integrating OpenMRS based hospital information systems and DHIS2 based HMIS deployed across the state of Himachal Pradesh (HP), India. OpenMRS based system is now deployed across 20 hospitals in the state while the DHIS2 based system is used for data collection from multiple reports under NRHM program. We used WHO backed interface for statistical data and metadata exchange in healthcare called SDMX-HD for data communication across the systems. We present the system architecture in detail and provide discussion on challenges and learning from development and deployment of such a system across the state of HP. We also discuss the data being collected at present and some analysis which can done to get useful insights from this data. We also looked into some external factors which might be influencing the healthcare setting in the state.

Acknowledgments

I would like to thank Dr.Amarjeet Singh for his constant guidance and support. Without your encouragement and advise my thesis could not have been materialized. Besides my advisor, I would like to thank my esteemed committee members Dr.Pushpendra Singh and Dr.Sundeep Sahay for agreeing to evaluate my thesis. I would like to acknowledge the support of Arunima Mukherjee, Arvind Chikne, Sagar Bele, Gaurav, Deepali Saha, Chahat Narula and several other people from HISP-India in gathering the user requirements and supporting the development and deployment of the system. Without their effort none of this work would have been possible. I would also like to thank Department of Electronic and Information Technology (DEITy), Government of India for partially funding the work done in this project. Last but not the least, I would like to thank all my family members and friends who encouraged and kept me motivated throughout the project.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The importance of collecting data to ensure better management of health care services is well understood in the present times. The need was first highlighted in a WHO study long back in 1986 which examined the role of Health Information Systems (HIS) in improving healthcare delivery to the people. The study emphasized that effective management and information support are some of the major obstacles in providing improved healthcare delivery to the people [12]. They further define a health information system as a set of components and procedures organised with the objective of generating information which will improve healthcare management decisions at all levels of the health system [12]. Thus HIS are not merely applications to collect relevant data but also should provide enough useful information which may be used to improve the efficiency and effectiveness of healthcare services.

Before the ubiquity of computers and mobiles based data collection, data was historically collected on paper in a set of prescribed forms. Across all steps in a medical procedure, from registration to diagnosis and lab reports, all required filling up of numerous forms and maintaining large volumes of paper based data. From data being collected on papers and notebooks and spreadsheets to applications with hundreds of modules and functionalities to generate reports along with visual aids, the health information systems have evolved to become a complex technology [8]. Today, Health Information Systems varies greatly in their scope and use. Hospital Information Systems, a sub-system of HIS are deployed in hospitals and clinics to collect clinical, demographic and administrative information. HIS are also used at the higher level, beyond the hospital (e.g. at the district, state or national level) to collect and aggregate statistical data for analysis and reporting. These systems typically aggregate data across multiple health programs such as immunization and maternal health. Data collected using these systems are increasingly being utilized by the administration for improved management of these programs and analyse their impact. Such integrated systems are typically called Health Management Information Systems (HMIS). Some of these also provide analytical tools to assist strategic decision making. Telemedicine, Telematics and eHealth systems are also being widely used as part of recent IT advancements impacting the Healthcare domain. Such systems are used in situations when all users are not present in the physical vicinity but can communicate over some network, usually internet [15]. With such a diversity of systems, seamless integration across these systems is the need of the hour that can allow for improved analytics and decision making.

Prior studies indicate numerous social and economic benefits that accrue from interoperability between different health information systems [6]. Investment in digitizing medical record and ensuring their interoperability can easily bring in the quick return of investment with benefits which eventually accumulate from the results [10]. One such study done in United States estimated that widespread adoption of interoperable EMR would improve overall efficiency of healthcare delivery and could result in savings of $142-$371 billion in United States alone. They calculated these figures based on many different parameters but primarily included savings from reduced adverse drug events, use of EMR for long term chronic disease management and use of EMR for short term preventive care [9]. Interoperability across the diverse components of HIS requires each of these systems to exchange information using a set of common and widely supported interoperability standards. As an example, consider two sub-components of HIS - first one a medical record system at a hospital level having clinical, demographic and administrative records and another as HMIS that aggregates various information regarding health indicators over time (e.g. vaccination, pregnancy, mortality rates etc.) essentially serving as a data warehouse for the district (or state or a country) that also encompasses multiple hospitals, each with their own medical record systems. The two systems are polemical in their use and functionality and collect a diverse set of information. Integrating two such systems will allow a single point of access to information of both systems. This would also facilitate the use of operational data coming from Hospitals to be captured instantaneously and be used for analysis rather than use of retrospective data. Further, such integration would also reduce duplication of information to some extent.

Research work in HIS integration has been studied in the past. One such work categorised the process of integration into information, service, process and user-oriented integration [16]. They mentioned the impossibility of finding a single technological solution set that can serve as a universal solution for HIS integration. Though largely accepted as beneficial and important, HIS integration has not become common because of the numerous challenges such as a diverse set of interfaces and technologies being used and lack of supported connectivity across the systems [17].This process is even more painstaking because various existing health information systems follow their own data models some of which are developed by private software vendors with proprietary data models. Hence integration is not feasible in all cases and seamless interoperability is practically hard to achieve.

In the first part I'll present a real world case study of integration of two open source health information systems deployed in the state of Himachal Pradesh, India. The government hospitals in Himachal Pradesh have OpenMRS as their hospital information system while the state government uses DHIS2, an open source health information management and analysis system. DHIS2 provides visualization, aggregation, validation and analysis of statistical information. The two systems were integrated using SDMX-HD which is WHO prescribed standard for exchange of statistical data and metadata in the healthcare domain. Since both the systems were previously working independently there was significant effort involved in making changes in both the systems to create a deployable integrated solution. Together with a team from our non-profit collaborators HISP-India, we developed an OpenMRS module called "Dhisreport" which is currently deployed in the district hospitals and connects with a central DHIS2 system of the state using SDMX-HD standard. Once we had a successful integration in 2013 a lot of data was collected at the state level which could be analyzed and used to gain useful insights about the statewide health services. In the second part we'll use some of the indicators that are being used in 'Dhisreport' module to show some vertical analysis that are possible at an individual hospital level and some which can be used to do horizontal analysis across multiple hospitals in the state.

Chapter 2

Background and Related Work on Health Information System

Rich literature exists in the field of interoperability and integration of health information systems. A HIXEn integration engine has been recently developed which sit between health information system applications that need to be integrated [19]. It has been designed to integrate patient level and aggregate level systems. It uses SPARQL and RDF extract to communicate data across systems. Since neither DHIS2 nor OpenMRS generate RDF it would be an unnecessary overload in Himachal case. Also the engine has not been tested for scaling on large datasets which is a basic requirement. A SDMX-HD based proof of concept model for Health Information Systems has been implemented in Sierra Leone [5]. Their use case involved DHIS2, implemented at a nationwide level while OpenMRS being used as an Electronic Medical Record (EMR) in the hospitals. In some cases mobile phones and PDAs were also being used for data collection and for interfacing with EMR. Since internet connectivity and other infrastructure were poor, additional flexibility had to be incorporated in their model for integration. Their model takes into account that adoption of the system at various levels of the organization will be uneven and thereby accommodates for the hybrid architecture by making use of any available data from systems that are plugged in the current deployment . The module deployed in Sierra Leone generates SDMX-HD reports which are then imported into the DHIS2. It also requires manually importing metadata generated from DHIS2 in SDMX-HD format into Hospital Information System. We have gone a step further to resolve these issues through automated synchronization between OpenMRS instances and DHIS2 instance.

The PlugIT project was carried out in Finland by health service providers, software vendors and research groups to understand integration specifications for health information systems [16]. Their study lays out various guidelines for carrying out large scale integration in health information systems. It also advocates the use of an incremental mode of development and addressing the primary integration concerns first thereby scaling the operation and features in later iterations. Another popular open source integration engine is mirth connect. Mirth Connect is cross platform integration engine which allows bi-directional message transfers, primarily HL7 among others, between systems over multiple protocols [18]. The project for integration of two major open source Health Information Systems undertaken by an NGO, HISP India, is first of its kind in India. The integration process has been finalised after extensive requirement study and the project has been executed in a bottom up manner. The deployment of DHIS2 was the first to be completed followed by deployment of OpenMRS in hospitals, which has now scaled to 20 hospitals. The development and This was followed by the deployment of integration module Dhisreport this year. Dhisreport will allow hospitals to easily submit reports to DHIS2 rather than filing paper based records and duplicating the effort. The collected data can be further used for horizontal analysis across hospitals, that was otherwise not possible without integration. The developed countries saw rapid proliferation of ICT in healthcare in 1990's and hence researches in statistical tools and methods for analysis and utilization of healthcare data were common way back in 1990's [7].The importance of analysis of healthcare data has been long established. The analysis is much easier for developed countries since there is abundant availability of surveillance data but developing countries do not have sufficient data and even the available data has not been sufficiently analysed [20]. Many researches have successfully shown how disease rates can be effectively correlated with various environmental and socio-economic factors. One such research in India showed how diarrhea is correlated with the number of healthcare centres, population growth rate, sex and region specific morbidity. They studied how diarrhea is correlated with number of above factors and how number of healthcare centres brings down the diarrhea prevalence in a region [20].

Chapter 3

Case Study: Himachal Pradesh Deployment

The requirements of state government of Himachal Pradesh in health care infrastructure are primarily met by use of two health information systems. At the state level there is a DHIS2 based system. The government of India has formulated some policies that require every state to keep track of some essential health indicators (disease, deaths, mortality rates etc.) in every state (NRHM, 2013). A report of these indicators is submitted to the central government periodically. The reports can be of varying periods weekly, monthly, quarterly or yearly. Apart from these requirements, the states are also under an obligation to generate some reports to keep track of their health care services. This data is collected on paper by various field workers working at various levels in the state. The data is then entered into the DHIS system usually at block/district level where users have internet connectivity. Every district also has district hospitals run by the state government. The district hospitals in Himachal Pradesh use OpenMRS as their Hospital Information System. Each district is composed of numerous blocks, which have many Community health centres (CHCs) and Primary Health Centres (PHCs). Here data is typically collected in registers.

There is another set of reports which are submitted by the hospitals to the state at regular intervals. These reports are built on data collected in the hospitals and various other indicators. For example a disease surveillance report is submitted by the hospitals to the state. The report consists of parameters such as Cholera, Malaria and Japanese encephalitis, among others. Figure 3.1 shows the flow of information and decision making process and typical infrastructure setting in the specified state.

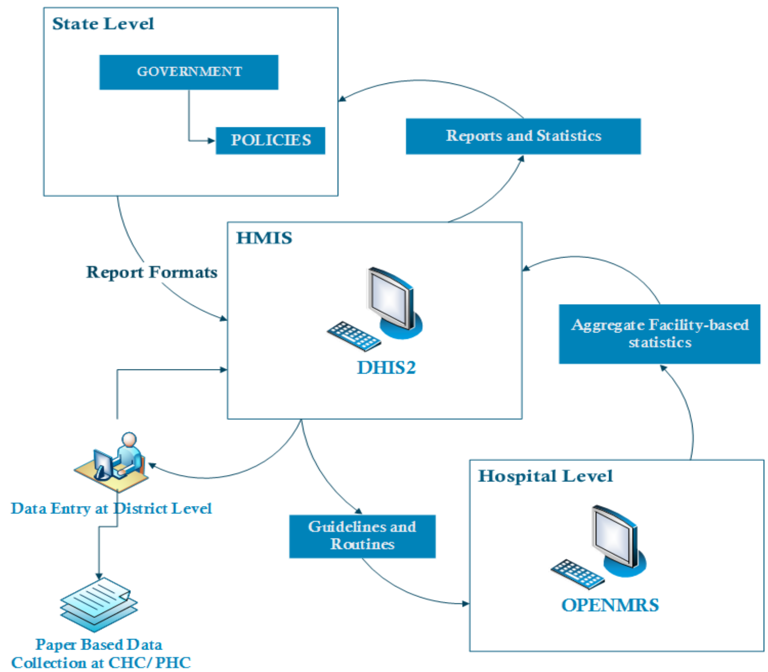


Figure 3.1: *Health ecosystem in Himachal Pradesh state. Various channels of information flow and collection which based on the policies and guidelines offered by the state is depicted.*

3.1 OpenMRS based Hospital Information System

OpenMRS is an open source electronic medical record system which has been built by a collaborative effort between teams at Regenstrief Institute in Indianapolis and Partners in Health (PIH), an NGO in Boston, USA. Today it is among the most popular open source EMR and is being used in varied domains. It has successfully been deployed in more than 25 developing countries [21]. The initial set up of OpenMRS provides a basic system which can be easily customized and scaled as per the needs of the organization. The most notable feature of OpenMRS is its patient centric data model. There are visits with one or more encounters for each interaction of the patient with the HIS. One or more "Observations" are generated during a single encounter which may involve one or more providers. These Observations record the actual health information and are mapped to a concept in the OpenMRS concept dictionary. The core of the data model is this concept dictionary which provides significant flexibility to the system by allowing addition of context dependent metadata.

The OpenMRS data model also conforms to the HL7 standards [14] which is an added benefit apart from being scalable. The model also provides room for internationalization. Easy to use APIs are provided and modules can be developed over the web framework to provide almost any required functionality. Over the years numerous modules for have been built for OpenMRS and are freely available for development and use.

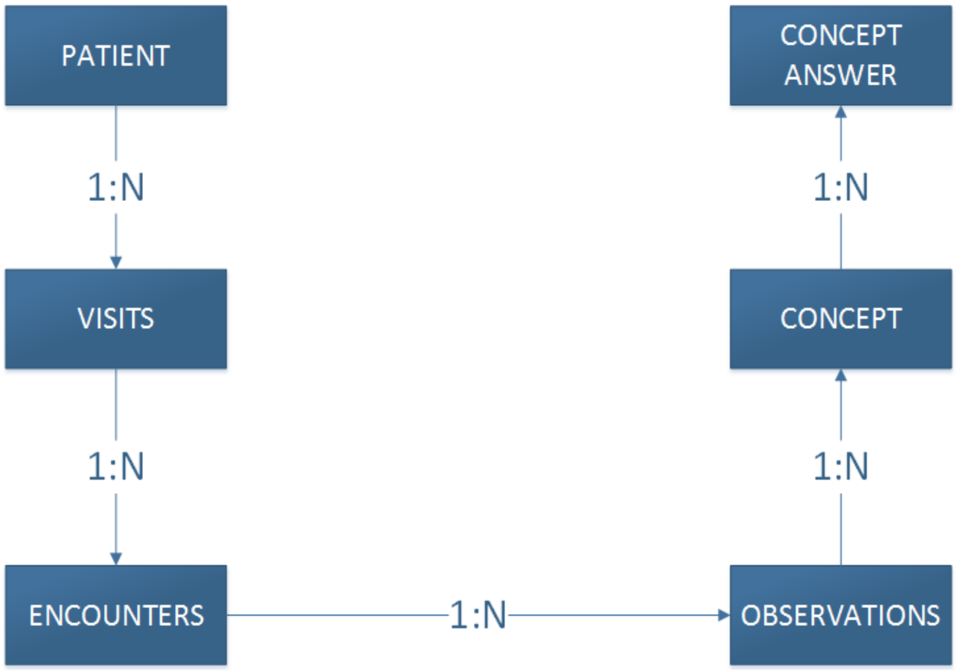


Figure 3.2: *Simplified OpenMRS data model showing how different items are linked with each other.*

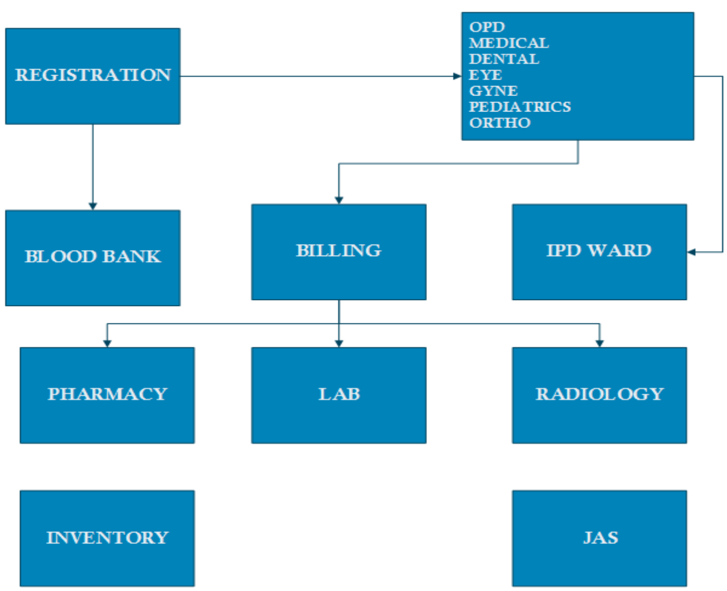


Figure 3.3: *Patient flow in hospital. The registration is the entry point for any patient and is compulsory.*

The system deployed in Himachal Pradesh has 10 major modules providing various functionalities. The modules are Registration, Billing, Laboratory, OPD, Inventory, Pharmacy, IPD, Blood Bank, Radiology and JAS (Jan Aushadi Store). The Registration and Billing modules were among the first to be implemented. The typical workflow of the patients in the hospital is shown in Figure 3.3. The entire system has been developed in a bottom up manner by HISP such that it is scalable and conforms to health standards [11].

Registration is mandatory for all patients. To use the pharmacy or get any medical tests done from the laboratory or the radiology department a patient is required to submit bills. Once the billing orders are generated patient can get the required service from the respective departments.

3.2 DHIS based Health Management Information System

The DHIS is an open source health information management tool and warehouse which has been developed by the HISP team. It was originally developed for use in South African districts but over the years its use has spread to more than 20 countries. The DHIS2 has tools for collection, validation, analysis and visualization of aggregate statistical data. The DHIS deployment in Himachal is responsible for aggregating and generating various GoI and State reports at periodic intervals. Data is entered and aggregated into the DHIS from the entire state to review various health indicators to understand performance of health infrastructure and personnel. For example, a district consolidated report is generated monthly to review indicators like maternal health, immunization and child health and malnutrition, among others. The DHIS deployed in Himachal Pradesh has been assigned a static public IP for making it accessible to all the hospitals.

Chapter 4

Interoperability between systems through Dhisreport Module

SDMX is a well-known and supported international standard for electronic exchange of statistical data and metadata information (www.sdmx.org). The SDMX standard is backed by numerous eminent organizations such as IMF, UN and World Bank. The primary aim of bringing in a new standard is to make processes more efficient and reduce duplication of effort. SDMX-HD is a WHOs implementation of the SDMX standard for health domain. It includes requirements for metadata profiles and support for non-enumerated data (www.sdmx-hd.org). It was intended to be used for exchange of indicators definitions and aggregate data and serve as an interoperability standard in the health domain.

The DHISREPORT is standalone module developed in the OpenMRS for sending aggregated patient records to the state DHIS of Himachal Pradesh. The module has four main components Report Definitions, Data Elements, Synchronization and Connection.

4.1 Report Definitions

Since the reports which are submitted to the state by the hospitals have well defined format their skeleton can be easily created in the module. Each such report is represent by a report definition. Each report definition has a unique uid and code. These essentially should be same across all the OpenMRS as well the DHIS implementation. This is very important in order to maintain consistency of reports and aggregated values at the HMIS end. The reports also have a pre-defined format and focused on data collection of specific indicators. Depending on the type of reports they need to be sent to DHIS on weekly, monthly or annual basis.

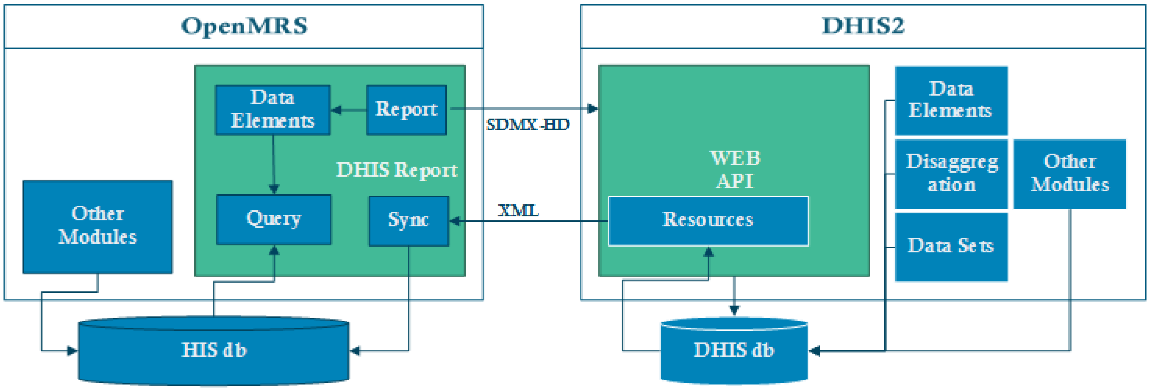


Figure 4.1: *Interoperability architecture in place. It shows the interactions between the DHIS and OpenMRS and how the data between the two systems is exchanged using Dhisreport module.*

4.2 Data Element

As described in 4.1, each report consists of indicators. Each such indictor is expressed as a data element in the module. Each data element also has its corresponding unique uid and code. Similar to the report definitions uid and code too need to be unique across all implementations. Since each data element has a well-defined context its value can be fetched from the database. To get the values currently each data element is mapped to a corresponding query. Due to the nature of the deployment in Himachal, which uses a common baseline database for every hospital thus having exactly the same Concepts mappings, the queries to fetch values for the data elements also remain the same across every hospital instance though this may not be necessarily true for all cases. Figure 4.1 explains the current interoperability architecture deployed in the state.

4.3 Synchronization

The purpose of this component is to facilitate synchronization between the DHIS and the OpenMRS instances. Currently this module synchronizes the report definitions, disaggregations and data elements. This is accomplished through the DHIS Web API . WEB API is a component of DHIS2 which allows systems to access and modify data stored in a DHIS2 instance (DHIS2, 2013). At any time instance, the current state or representations of all the above resources in DHIS can be accessed through its Web API. The resource can be retrieved in a variety of representation formats such as XML, HTML, PDF, JSON or EXCEL. As emphasized in the 4.1, consistent uids and code across all instances are important for proper functioning of the module. This consistency is maintained through the synchronization process. All data elements are synchronized with the DHIS on a weekly basis. The period has been chosen as weekly because it is the currently the minimum period for submitting a report to DHIS. This ensures that all the report definitions are consistent. Prior to the use of synchronizer, the consistency had to be manually checked. Further, if any update was to be made to any report at the DHIS instance the same had to be manually propagated to every OpenMRS instance. With current implementation, any update made at DHIS automatically propagates to all the OpenMRS instances. The user is notified of the last synchronization date and time as well as any missing queries to the existing reports after the latest synchronization. This above model worked during local testing but did not scale well on actual deployment in the initial stage. The synchronization process turned out to be very time consuming, taking more than five minutes in some cases. The problem was attributed to the fact that every data element in DHIS had to be checked and if found in OpenMRS then synchronization is to be carried out. This made the number of lookups proportional to the data elements present in DHIS. The Himachal Pradesh DHIS database has over a thousand data elements. This problem was resolved by using grouping feature available in DHIS. All the data elements which belong to the Hospital reports are grouped under a pre-specified data element group, in this case openmrsde. This greatly simplified the process of synchronization and also significantly reduced the time required for synchronization, bringing it to under half a minute.

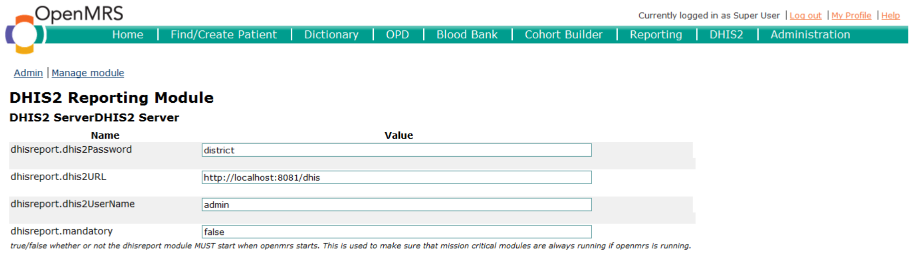
4.4 Connection and Automation

This component handles the connection information for communicating with DHIS. Currently it stores the URL, username and password of the DHIS instance. If required this can be easily changed to send data to a new server.

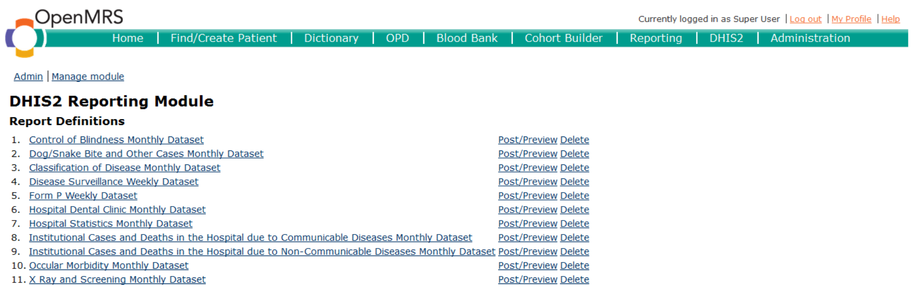
In the current model a user has to select a report and then its associated time period before pushing the data to the DHIS2 state server. This task has to be done weekly or monthly depending on the type of report. Although there are no daily reports as of now but there may very well be in the future. This process creates an unwarranted dependence on a user who is responsible for posting reports to the server. In an already understaffed and overworked hospital setting this might lead to problems of reports not being pushed on time. This problem was mitigated by use of the OpenMRS scheduler service. This service provides module API's to run a cron job or a scheduled task. We created a daily task which checks the report type and period. It checks if the report is weekly or monthly and then decides if it needs to be pushed to the DHIS2 server. The task accounts for a loss in connection and waits infinitely till it is online again and then publishes the pending reports. A major problem with this model is that the data gets compounded if the Internet downtime is large, extending to days.

4.5 Sample Use Case

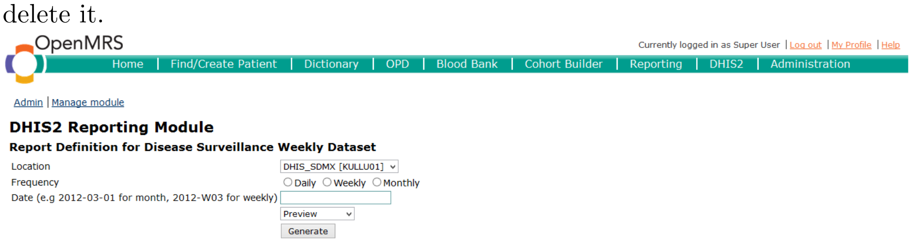
The module has a simple use case. On first installation there will be no reports definitions or data elements present in the module. The first step for the user is to add the DHIS2 connection information to the module.Once this step is completed the user can synchronize the module with DHIS2 server. The reports definition present in DHIS2 will appear in under module Home page.Queries need to be added for every data elements. This can be done using a pre-created



(a) The page to enter connection information for DHIS2. It is stored in database.



(b) The page showing list of reports. User can choose a report edit, send or delete it.



(c) The page where user can preview a report and send it to DHIS2.

Figure 4.2: A sample use case of the module. The module has been made simple to use but currently provides only limited options.

'XML' file if the number and types of queries is fixed. Once this is complete the report can be previewed and posted to the DHIS2 server after selected appropriate period for the report.On submission of report we get a success response from DHIS2. This finishes the process of submitting data to DHIS2.On subsequent submissions the user simply needs to synchronize the reports, add any missing queries and send the reports to DHIS2 in the usual way. Figure 4.2 shows the application screens.

Chapter 5

Data Collection

Data has been collected from various sources for study and in depth analysis. The entire database has patient level data from two hospital, weather data, ambient air pollution data and surface water pollution data of one city for the corresponding period.

5.1 Hospital Data

The primary source of data has been the HIS system deployed in the District Hospital, Shimla. The data that has been collected is from that date of deployment of the first module Registration in November 2010 to December 2013 when all the modules were successfully deployed and the hospital had a complete working system. The data comprises patient information of over six hundred thousand patients which resulted in around one million encounters over the period of more than three years. The state government needs to collect various parameters to monitor the state health services provided by the hospitals. To fetch these parameters over eight hundred queries were written to provide data from separate modules present in the database. For the purpose of analysis, some of the queries were selected. The data is available from mainly the following modules

- Registration: This module collects patient registration data. Apart from collecting patient demographic data such as age, gender, residence etc. it also collects information about patient categories which are mainly general, BPL and RSBY among others. This is important in context that BPL and RSBY patients receive free or subsidized health care services from all government hospitals and most private hospitals across country [1].
- Billing: This module records all the bills generated by different modules across the hospital. This includes Laboratory Tests , Radiology Tests , Ambulance services as well Inpatient and Outpatient charges. This also includes bills paid in form of tenders by the hospital.

[1] % of data available in the given period. Rounded to the nearest integer.

Table 5.1: Summary of Data

Hospital/Module	Frequency	Start Period	End Period	% Available [1]
District Hospital Shimla				
Registration	Daily	Nov-2010	Dec-2013	81
Billing	Daily	Nov-2010	Dec-2013	81
Laboratory	Daily	Jan-2011	Dec-2013	30
Radiology	Daily	Jan-2011	Dec-2013	80
IPD	Daily	Mar-2011	Dec-2013	43
District Hospital Solan				
Registration	Daily	Mar-2012	Nov-2013	89
Billing	Daily	Mar-2012	Nov-2013	89
Laboratory	Daily	Mar-2012	Nov-2013	26
Radiology	Daily	Mar-2012	Nov-2013	77
IPD	Daily	Mar-2012	Nov-2013	14
Weather				
Average Daily Temperature	Daily	Oct-2010	Dec-2013	100
Precipitation	Daily	Oct-2010	Dec-2013	100
Cloud Cover	Daily	Oct-2010	Dec-2013	100
Air Pollution				
NO_2	Weekly	Mar-2012	Mar-2013	100
SO_2	Weekly	Mar-2012	Mar-2013	100
SPM	Weekly	Mar-2012	Mar-2013	100
RSPM	Weekly	Mar-2012	Mar-2013	100
Surface Water Pollution				
Flourine	Fortnightly	Mar-2011	Dec-2012	90
Chlorine	Fortnightly	Mar-2011	Dec-2012	90
Calcium	Fortnightly	Mar-2011	Dec-2012	90
BOD	Fortnightly	Mar-2011	Dec-2012	90

- Inventory: This module stores the data of each and every goods and materials that are used inside the hospital. This module is responsible for recording the quantities purchases as well the quantities indented by individual hospital departments.

- OPD Module: This module records information associated with the out patient department. This includes patient attendance in the separate out patient departments as well their provincial diagnosis.

- IPD Module: This module records information related to the hospitals in patient department. This module collects information of admitted patients, their ward information as well what their admission outcome was.

- Radiology Module: This module records information related to the radiological tests. The information contains the type of test ordered for the patient. The date the test was carried out and when the results were returned to the patient. DDU Shimla hospital has facility for X-Ray, Ultrasound and Contrast Studies like CT scans.

- Laboratory Module: This module is similar to the radiology module but it records information regarding laboratory tests ordered for patients by doctors. DDU Shimla hospital has facility for BT/CT ,Serology, Cytology, Urine tests and various Biochemistry tests.

Table 5.2: Summary of Queries

Hospital	#Reports	#Queries	#Usable Queries
DH Shimla	12	817	145
DH Solan	11	567	49

We also have some data from District Hospital in Solan. Since the system was deployed on a later date we do not have data from all the modules. The data is mostly available from Registration, Billing and Laboratory Module. The other modules were deployed at a much later date which does not provide us with sufficient data points to do in depth analysis.

5.2 Weather Data

Historical weather data of Shimla City was collected from world weather online website [4]. This was collected in 24-hour intervals format. The API provided data for temperature, precipitation, wind speed, cloud cover, heat index, visibility etc. The data was collected for the interval October 2010 to December 2013 which corresponds to the availability of hospital data for the same period.

5.3 Ambient Air Pollution

The data for ambient air pollution was collected from the Himachal Pradesh government site of pollution control board [2]. The state pollution control authorities started monitoring air pollution March 2012 onwards. They recorded readings for NO2 , Residual Suspended Particulate Matter, SO2 and Suspended Particulate Matter. On an average around ten reading were taken every month from two separate sites in city Tekka Bench Residential Area and the residential area near Shimla bus stand. All the readings were in ug per m3.

5.4 Surface Water Pollution

The data for surface water pollution was also collected from the Himachal Pradesh government site of pollution control board [2]. The monitoring of water supplies was started in April 2011, but readings were available only till December 2012. The surface water was monitored and readings of more than ten important biological, chemical and physical testing parameters like Biological Oxygen Demand, Chemical Oxygen Demand, Calcium, Chlorine, Fluorine, Potassium, Magnesium, Nitrates, Total Suspended Solids, Temperature etc. parameters were recorded. In the period of over a year more than twenty observations were taken from four different sites.

Chapter 6

Data Cleaning and Preprocessing

As mentioned in chapter 5 the primary source of the data is the that collected from the hospitals information system. The EMR stores data in a MySQL database which has grown sufficiently large (1.5 GigaBytes) in three years of deployment. Since various modules were deployed during different times it resulted in a database where at any given time not all the data is available. For the purpose of analysis, it was important to understand the schema of the database as well as which modules contain how much information. Based on the availability of data we initially selected some six hundred queries which were pre-written for the dhisreport module. Each of the queries returned a single numerical value of a particular indicator given start date and end date. The indicators varied from being general census indicator like Attendance in Surgery OPD to diagnostic indicators like Total Malaria Cases. With the selected queries a time series database was built to record case value for each day from October 2010 to December 2013. Once we had the time series database of 'date' , 'dataelementid' and 'value' for all the indicators we selected some hundred queries out of available six hundred due to lack of quality data in terms of data points and possible data inconsistency.

6.1 Data Quality

Most of the queries which were not considered for analysis were due to the lack of sufficient data points. Almost five hundred queries did not have a single data point in the span of three years which raises questions on the whether all the modules are being extensively used by the hospital's staff. These queries had to be rejected due to the lack of data points to do any analysis. Also for the remaining queries p-value for each of the indicators was calculated and all indicators with p-value > 0.05 were not considered for any analysis.

6.2 Data Inconsistency

Some queries were rejected due to a possible of inconsistent values entry by the staff personnel. For example the database reported a total of 6 polio cases in the span of 3 years. Of these five in summer months of April - July 2011 and one in November 2012 but this seems inconsistent with the records of UNICEF which declares that India reported the last polio case on 11th January 2011 [3]. Also very recently after three years of no reported polio cases India has been declared polio free in February 2014. So it seems the data does collected in the state hospital does not agree with the national figures and UNICEF. So this and other similar cases have been overlooked. Also for the some of the selected query due to pilot study and deployment still in phase sometimes data was not available for modules for an extended period of time. For example data from radiology and laboratory module of total X-Rays and total lab test done is not available for seven months in the given period. Inclusion of these queries might necessarily result in inconsistent results. Similarly there are no eye disease cases reported from Eye OPD

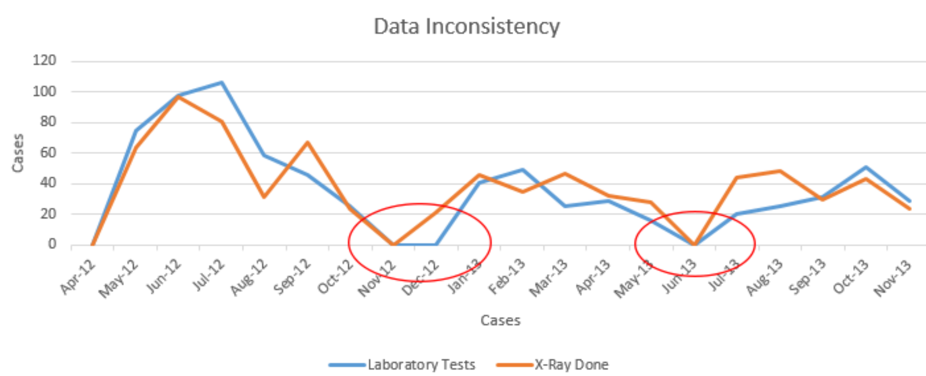


Figure 6.1: *The plot shows inconsistency in data values for two indicator. We have extended durations with 0 cases which is highly unlikely in practice.*

for three months which is highly improbable when an annual monthly average of reported cases is over 300. This indicates the module was either not working or not being used. To carry out better analysis the above inconsistencies had to be taken into account.

Chapter 7

Analysis and Observation

7.1 Methodology

The data collected was analyzed for information providing more insight into hospital working, general health index of people and looking for correlation between weather, air pollution and water pollution data with hospital statistical data. MS Office was used for various statistical analysis which includes finding p-value using chi test, correlation between different attributes and standard statistical functions like mean, standard deviation etc. For interactive visualization a Web based Django Application was built which allowed generation of interactive charts of multiple types to get a better understanding of the data from models. All streams which have been included for analysis have a p-value of less than 0.01 unless mentioned otherwise.

7.2 Observations

7.2.1 General Indicators

These indicators are mostly census based indicators which provide insight about the number of patients/cases reported to hospitals on a daily, weekly, monthly or annual basis. These can be used to understand and assess the workload on hospitals and individual departments. These can also be analysed to see if there are any recurring patterns so that the hospital can be better equipped and facilities may be better used. They can be used to quantify operational efficiency of the hospital.

Registration and Out Patient Department

Registration was one of the first deployed modules and hence is rich in terms of collected data. There is registration data of over six hundred thousand people in over 3 years.

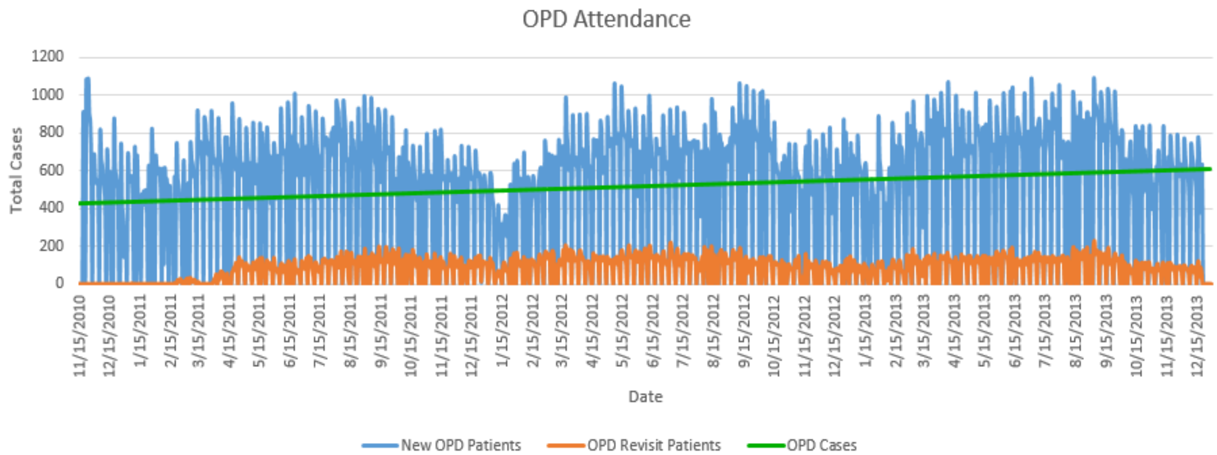


Figure 7.1: *The general trend of number of new patients and number of old patient revisiting the hospital Out Patient Department for check up. There has been a gradual annual increase incoming patients of more than 2% annually as indicated by the trend line.*

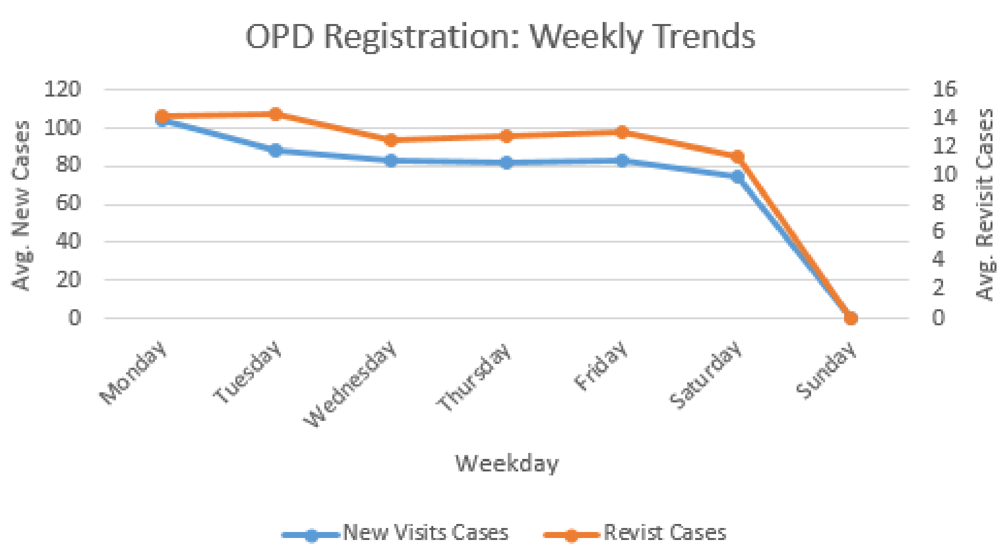


Figure 7.2: *The per week day average of number of patients arriving on average in a period of 3 years.*

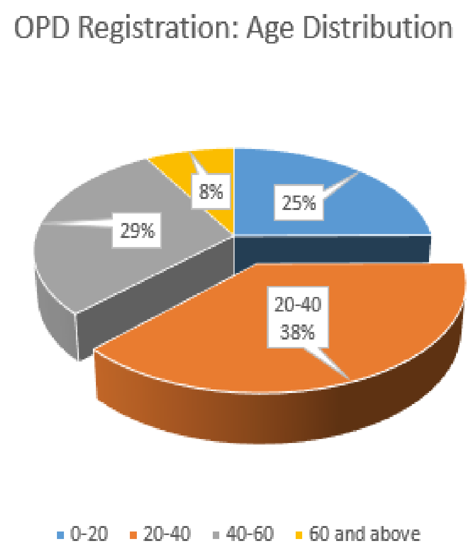


Figure 7.3: *Age-wise distribution of total OPD registration.*

It was also noted that the average OPD attendance was significantly higher on Mondays rather than other weekdays. Infact it was around 40% higher than weekly average and 20% higher than other working days average. Sunday is a non OPD day, patients which arrive are treated under emergency cases.
But on the contrast if we record the patients who come for revisits or second visit the visits are more equally distributed and Tuesday on an average has highest visits in the week indicating that they want to avoid the Monday rush. Figure 7.2 shows this trend.

Also when we do an age wise analysis we find that close to 40% of the patients belong to the age group 20-40 7.14b.

After this we compare the number of general patients and BPL or RSBY patients coming to the hospital. We get figure 7.4 from which it is evident that BPL patients form a very minuscule percentage of the total incoming patients. They account for less than 1 percent of the total patients coming in OPD.

We collate this data with all the individual out patient departments in the hospitals. As expected, we find highest number of cases on Monday in all OPDs except TB OPD. This is highest for Orthopedic department which has close to 30% higher average turn out on Mondays compared to the rest of the week output.

Laboratory

Highest laboratory tests are done on an average on Tuesdays. This value is 20% higher than the rest of the weeks working day average. This can be attributed to the fact that there is a high patient turn out on Mondays and therefore higher tests are ordered which get carried forward

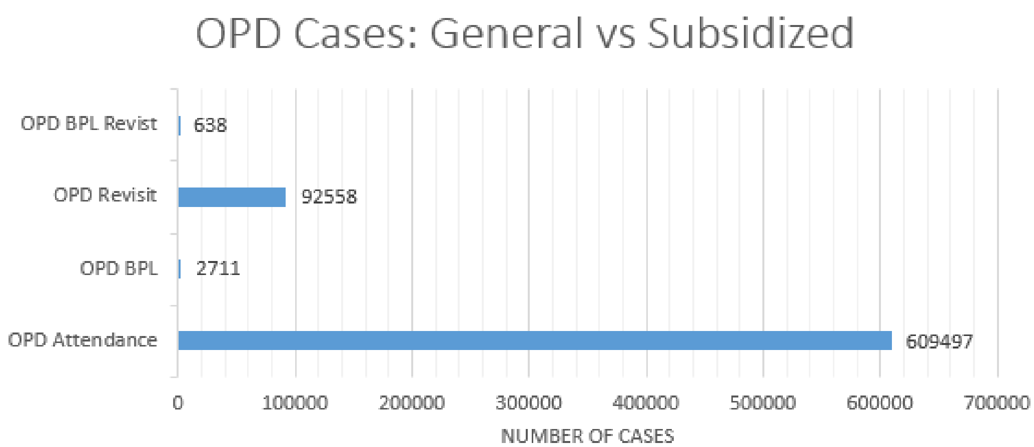


Figure 7.4: *General versus subsidized patients ratio of total OPD Cases. It is clearly evident that majority of the patients are general.*

till Tuesday. As expected there is a moderately high correlation of 0.34 between the number of lab test carried and patient attendance in out patient department. There is also particularly high correlation of 0.76 between ECG done and patient attendance.

Radiology

The radiological departments also show very similar results to the laboratory department. This is evident from the high correlation value of 0.67 between X-Rays done and OPD attendance in Orthopedic department. If we try to see the ratio of tests between general and BPL patients we find that though BPL patients are a minuscule percentage of total incoming patients the number of tests that they are prescribed are significantly higher.

Performance Indicators

There are numerous performance indicators values which can provide valuable insight into efficiency of hospital operations and help improve services offered. Some of such indicators are average patient waiting time for consultation, bed occupancy rate, waiting time for sample collection, cycle time for reporting, rate of rescheduling of laboratory or radiological tests, average revenue per patient etc. A few of these require information which is not present in the HIS database. For example for bed occupancy rate we need the information regarding the total number of available beds in each department as well as their occupancy information which is currently not captured by the system. Unfortunately most of the above indicators cannot be used for analysis purpose presently due to lack of sufficient data points. However, as more data is collected over time the calculations from above indicators will become feasible.

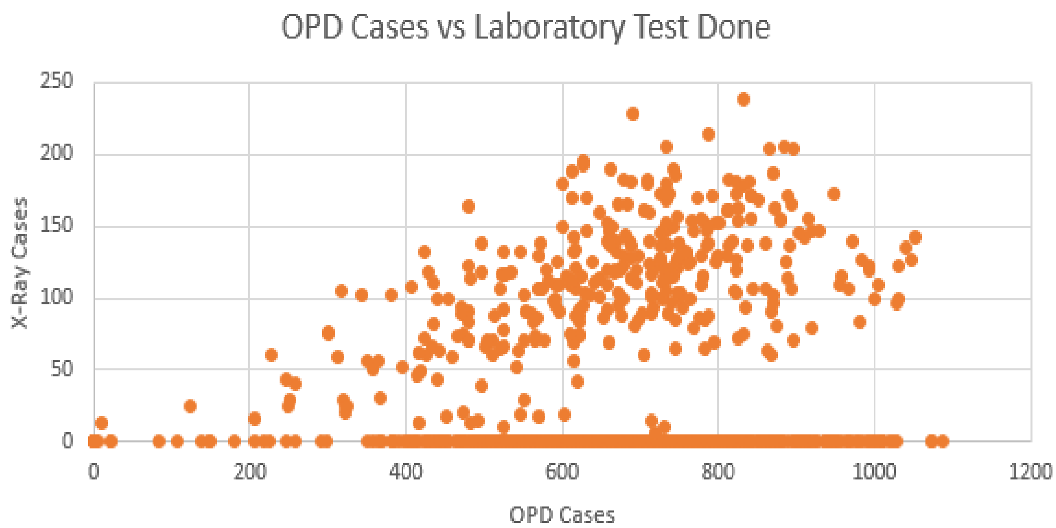


Figure 7.5: *Correlation between number of patients coming to OPD and number of Laboratory Tests done.*

Possible inferences

The set of indicators discussed above show trends in patient activity. For example the gradual increase in number of patients coming in for the OPD is probably a result of increase in population. Shimla district registered a growth rate of 12.67% from 2001 to 2011 (http://www.censusindia.gov.in). Next we consistently saw a rush on Mondays in all OPD's. This rush might be the resultant of cases which occur during the weekend. Since Saturday is a half day and OPD's remain closed on Sunday's, Monday see a much heavier patient influx which gradually settles down as the week progresses. This causes extra work overload for the laboratory and radiology department which consequently end up doing more tests on Tuesdays, likely compensation for the increased patient activity of Monday's. We show this by plotting daily laboratory cases and X-Ray cases with Orthopedic OPD and General OPD cases.
Next we made an interesting observation that nearly two-thirds of cases were from age group 20-60. This is in sync with the 2010 census stats pointing that 63.9% of the population of Himachal Pradesh lies in the age group of 15-59. After this we looked into category of patients coming for treatments. We have two categories here, one are the general patients and other are BPL or below poverty line patients. All facilities at the hospital are free for BPL patients. The percentage of BPL population in state is 4.33% but the percentage of patients coming to hospital from this category is less than 1%. This shows that people from BPL category are probably not using the health resources provided by the government, may be due unawareness.

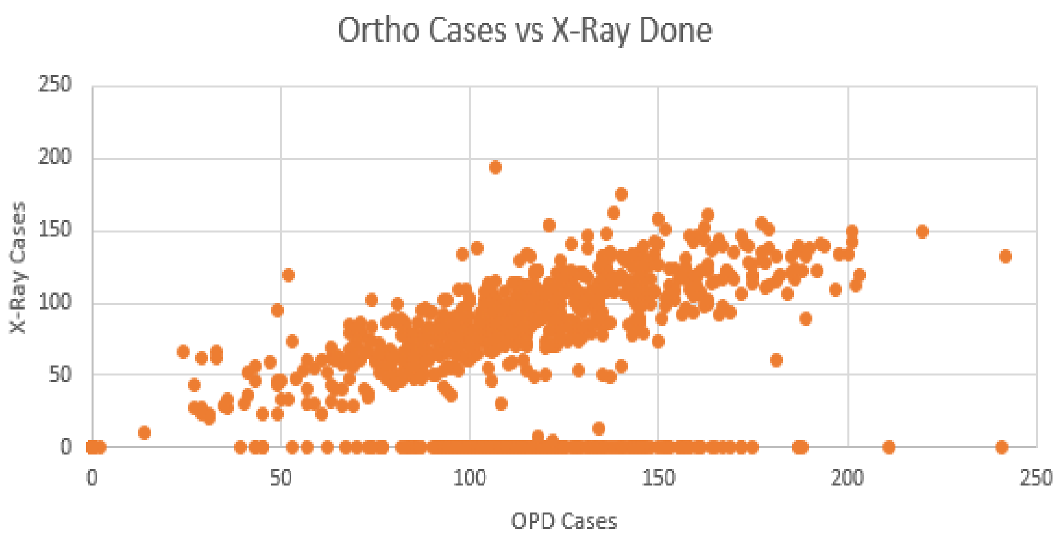


Figure 7.6: *Correlation between number of patients coming to orthopedic ward and number of X-Ray's done.*

7.2.2 Financial Indicators

This sections shows some financial indicators which have been computed from that data collected in the billing department.

Total Revenue

The hospital provides over two hundred and eighty standard billable services at highly subsidized rates apart from other miscellaneous services. Total billed amount in the period of study was of over 13 million rupees. Of this amount more than 5.5 million rupees was subsidized and only 7.9 million were the actual collections made.

As is evident from figure 7.7 that subsidy has increased over the years.

Revenue Across Patient Groups

As expected there is high correlation of 0.735 between billing amount and patient OPD attendance. The value could have been even higher if data from IPD was also available. To do a more accurate analysis on the amount hospital collected from each general patient vs the amount it subsidized for each BPL patient we a per patient analysis. On average in given period the amount collected per patient was rupees 19.1 while on the other hand average amount subsidized per BPL patient was just over 443 rupees.

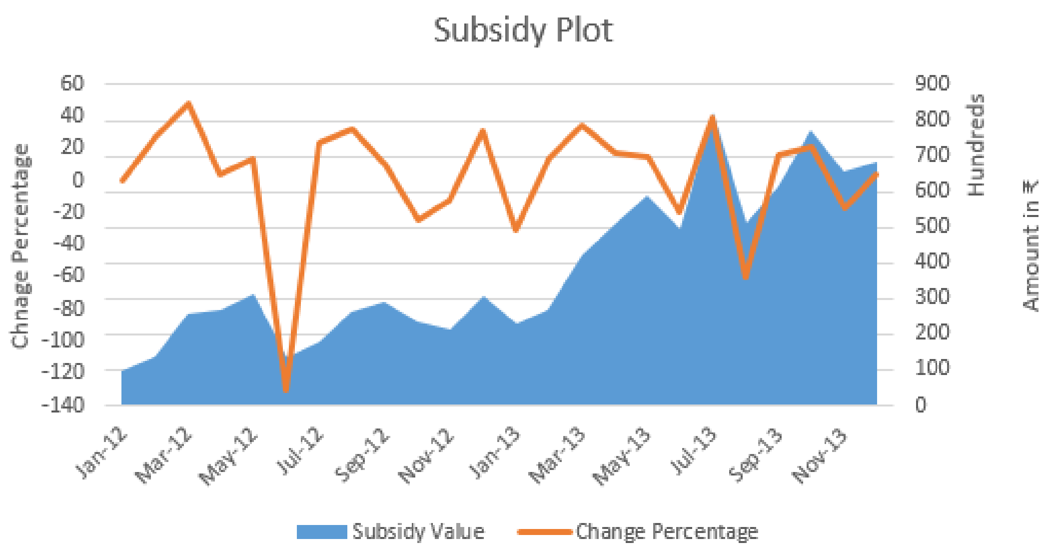
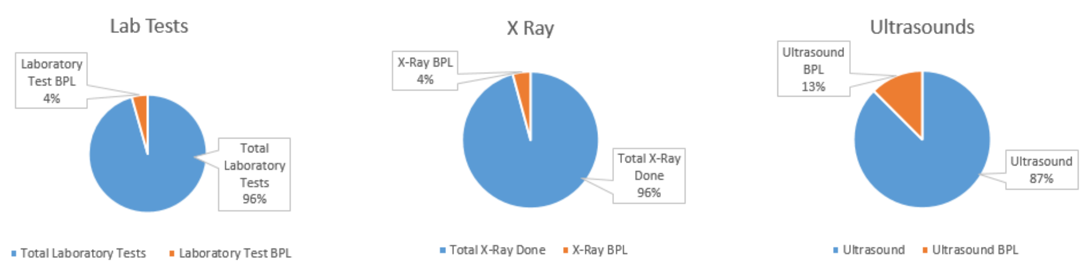


Figure 7.7: *The plot shows the variation in monthly subsidy provided to needy (BPL and RSBY) patients. On secondary axis we have the monthly change percentage. There has been a increase in subsidy provided over time.*

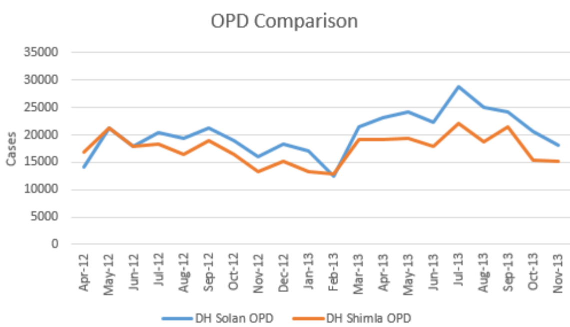


(a) Total Laboratory Test (b) Total X-Ray Done (c) Total Ultrasounds Done

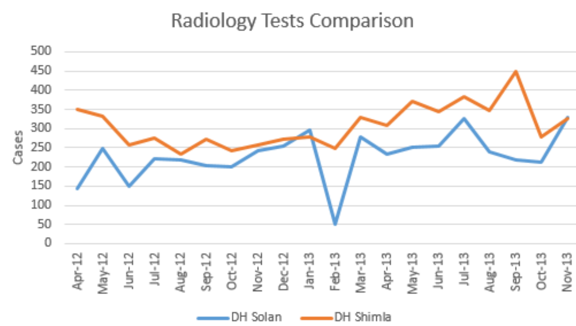
Figure 7.8: Charts showing resource utilization between patients. Compared to number of BPL or RSBY patients coming the number of tests that they undergo is atleast four times higher thereby resulting in additional subsidy.

7.2.3 Horizontal Analysis

This analysis is done across the hospitals. It is possible if we collect and compare hospitals on same indicators and in the same time period. Since a scalable deployment was planned in such a manner that exactly same indicators were used across all hospitals horizontal analysis becomes easy. With horizontal analysis we can find anomalies in data, we can study the trends in hospitals and compare different hospitals over their performance indicators. Figure 7.9 shows how number of patients attending OPD and number of Radiological Tests done follow very similar trends across both the hospitals. This is in contrast with Figure 7.10 where disease trends show very unusual trends in both the cases.

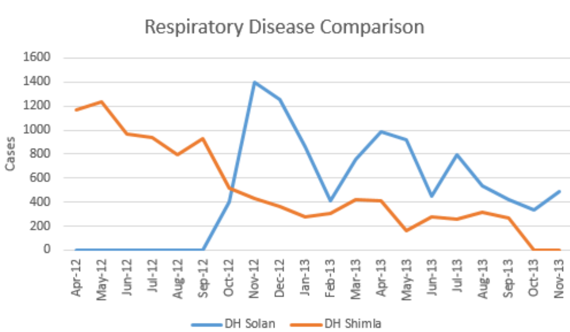


(a) OPD Attendance Monthly

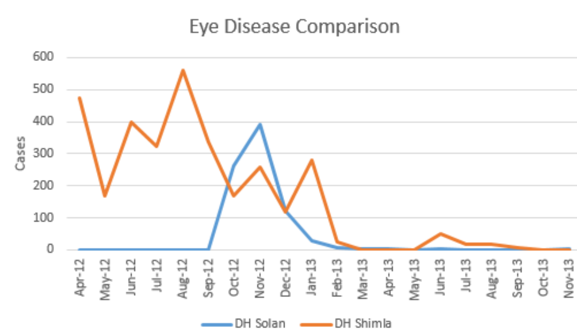


(b) Radiology Tests Done Monthly

Figure 7.9: These figure show that in cases where data is available both the hospitals show comparable trends.These can also be analysed for sudden change in values to gain useful insights.

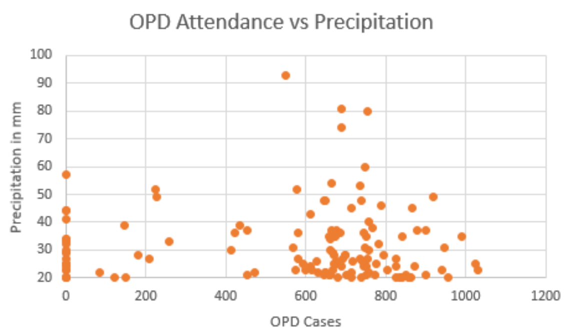


(a) Total Respiratory Disease Cases

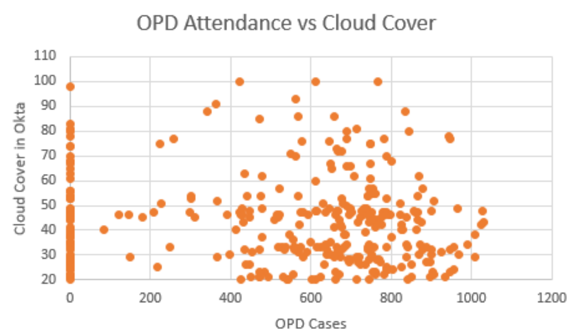


(b) Total Eye Disease Cases

Figure 7.10: These figure show variable trends in both the hospitals. This is because of lack of quality data in this case insufficient data points from the hospitals. Detailed inter-hospital analysis is not possible unless we have congruent data from both hospitals.



(a) Daily OPD attendance and precipitation.



(b) Daily OPD attendance and cloud cover.

Figure 7.11: Relation between OPD attendance and weather parameters like precipitation and cloud cover. These factors do not affect the attendance of patients in OPD.

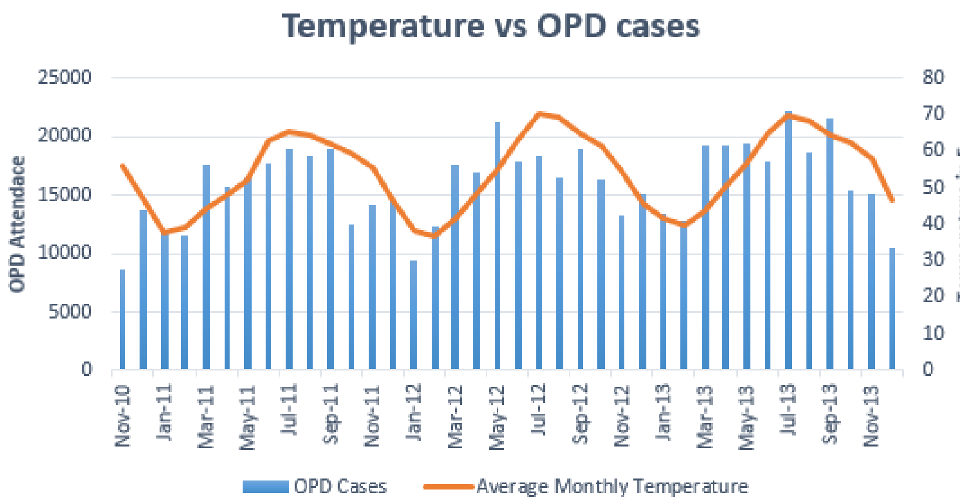
Possible Inferences

Geographically Shimla district is located adjacent to Solan district. As observed through the horizontal analysis above we find that both district hospitals have a very similar patient registration trends. Therefore not surprisingly they also share very similar trends for Radiological Tests. These two indicators are one of the few which have consistent data available for the entire period of study. On the contrary the other set of figures show diagnosis data from the OPD's. Solan has respiratory and eye disease data missing till September 2012 indicating that opd dashboard was not deployed but after this eye disease data is again missing from February 2013 onwards but respiratory disease data is present indicating that some issue with use in Opthalmology department.

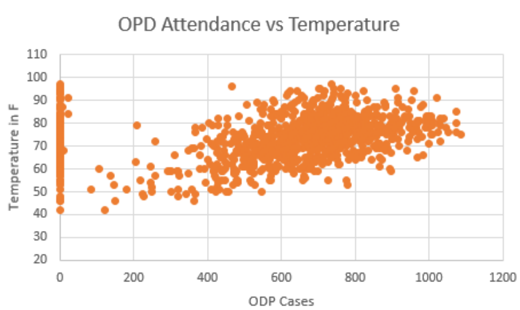
7.2.4 Analysis With Weather Data

After doing some basic analysis with the health indicators we tried to find clues into how external parameters affect patient and hospital activity. The first of the factors that we considered are weather values. We tried to find correlation between external temperature and number of indicators shown above. We found that there is moderate correlation between the average daily temperature and the number of cases reported in OPD. Almost all OPD departments show slight to moderate correlation between external temperature and number of patients attending the OPD. Total OPD cases had a correlation value $r = 0.24$. The value could have been higher but since there are no OPDs on Sundays it results in a slightly lower value. Orthopedics OPD cases had a correlation of $r = 0.237$ and Paediatrics OPD had a value of $r = 0.366$. If we do the same for monthly averages the values come out to be much higher. The value for monthly average temperature and monthly average OPD cases is 0.76. Figures 7.12 visualize these parameters.

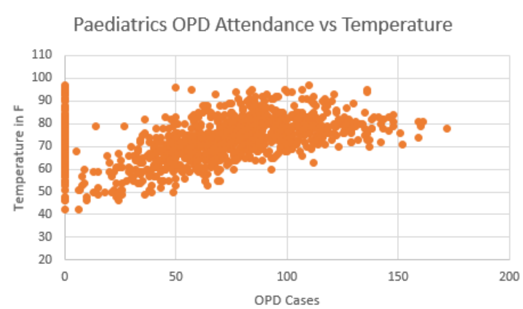
On calculating similar values for correlation between precipitation and cloud cover we find almost no correlation. The correlation value for total OPD cases with precipitation is $r = 0.074$.



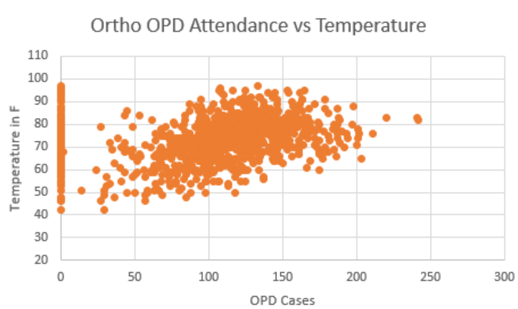
(a) Monthly variation of OPD cases with average monthly temperature



(b) Daily correlation of OPD cases and average daily temperature



(c) Daily correlation of Paediatrics OPD cases and temperature



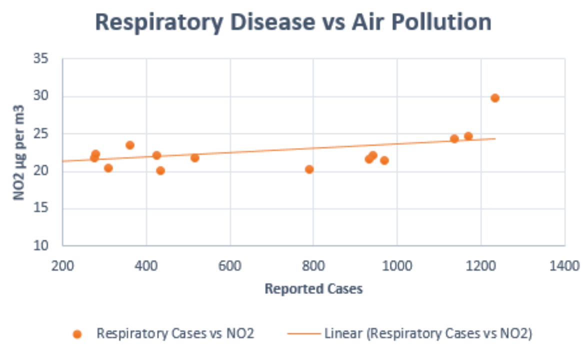
(d) Daily correlation of Orthopedic OPD cases and temperature

Figure 7.12: Relation between OPD attendance and temperature. Almost all OPD's showed a postive correlation between temperature and number of patients.

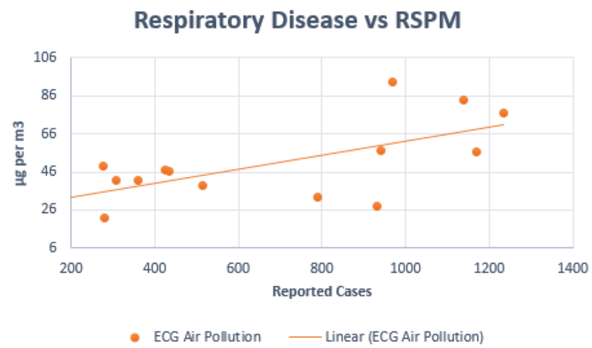
Similarly correlation value of total OPD cases with cloud cover is r = 0.0018

Possible Inferences

When we plot monthly temperature averages with monthly OPD case average we get a figure which nothing to imagination. We can clearly associate average patient activity with external temperature. Shimla being a hill station has cold climate. External temperature is found to vary between 30°F in winters to 75°F in summers. With extremely low temperatures in winters there is a significant decrease in OPD cases. To further consolidate our point we try to find correlations with individuals OPD departments. All departments show positive correlation with external temperature. After looking at temperature we looked into other weather factors like cloud cover and precipitation. Both these factors did not have any affect on OPD cases.



(a) Monthly Respiratory Cases and NO_2 pollution.

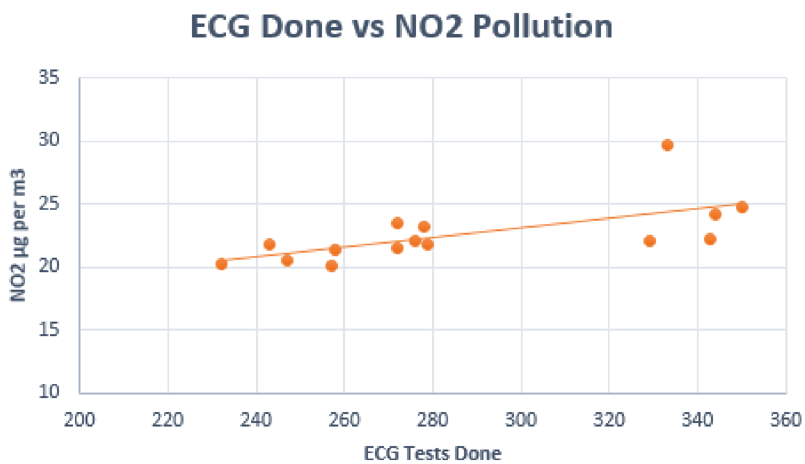


(b) Monthly Respiratory Cases and RSPM values.

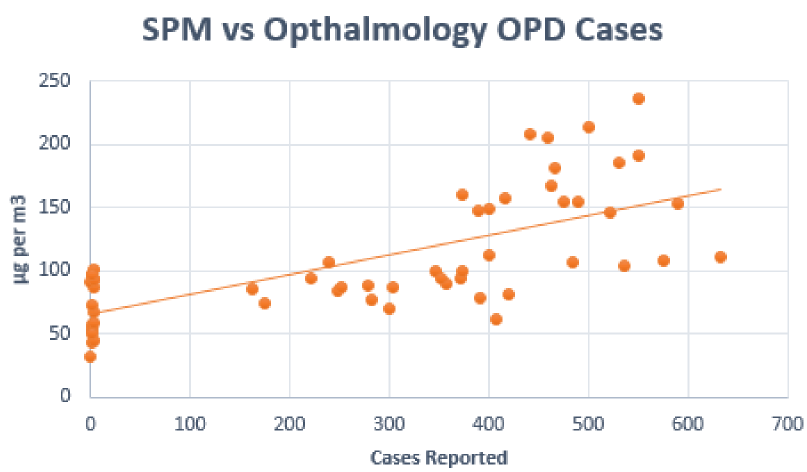
Figure 7.13: The figures show dependency between different Air Pollution values and Respiratory Disorders cases in the district.

7.2.5 Analysis With Air Pollution Data

The next part was to analyse air pollution data for any causal effects. The data that we collected from the Himachal Pradesh government site of pollution control board had around 2500 data points measuring value of NO_2 pollution, SO_2 pollution, Residual Suspended Particulate Matter(RSPM) and Suspended Particulate Matter(SPM). On an average 13 readings were taken every month. The reading for each of the parameters are recorded at every 6 hours intervals. We have continuous data from March 2012 to March 2013 followed by intermittent data available from March 2013 to December 2013. All the analysis is restricted to the above time period. We first started with trying to correlate different indicators with pollution values. The result was pretty much as expected. Almost none of the values show any correlation with any of the pollution parameters on a daily basis. If we do the same analysis on a weekly period some values show a moderately high correlation. Total number of Respiratory Diseases showed an increase in correlation with SPM and RSPM values. There was also correlation between total ECG done and air pollution parameters. Attendance in Ophthalmology OPD increased and showed a positive correlation with NO_2 , SPM and RSPM content in air with values 0.336, 0.58 and 0.67 respectively. If we do a monthly analysis we have an even higher correlation values. Correlation values for Total Respiratory Disorders with SO_2 and RSPM were 0.59 and 0.58 respectively. Similarly correlation of cases registered in ophthalmology OPD showed correlation of 0.66 with NO2 and 0.63 with RSPM. Total ECG Exams done also showed high correlation value of 0.72 with SO2 and 0.59 with RSPM.

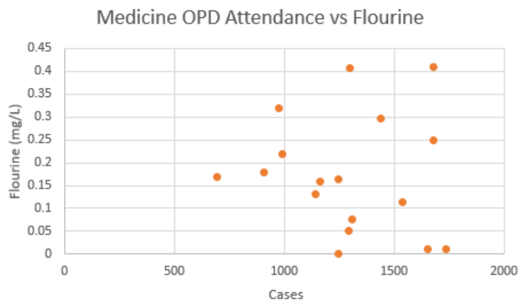


(a) *Monthly ECG's done versus NO_2 pollution.*

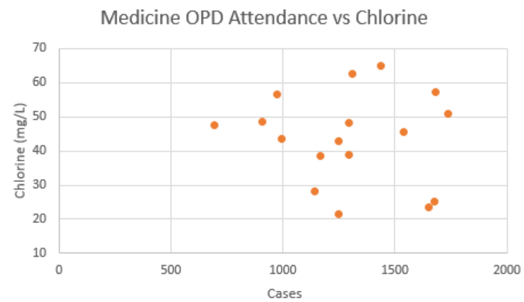


(b) *Weekly Opthalmology OPD cases versus SPM values.*

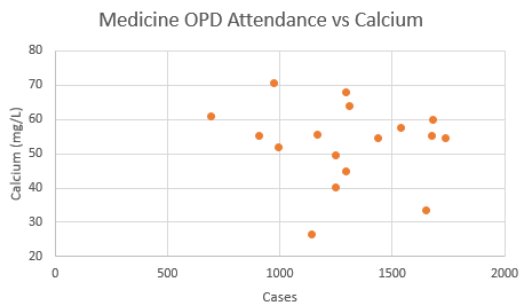
Figure 7.14: Relation between air pollution and different indicators showing positive correlation.



(a) Monthly values of OPD cases and flourine in water



(b) Monthly values of OPD cases and chlorine in water



(c) Monthly values of OPD cases and calcium in water

Figure 7.15: These figures show there is no dependency between different water pollution parameters and number of OPD cases. Almost all OPD's showed no correlation between temperature and number of patients.

7.2.6 Analysis With Water Pollution Data

Finally we also analyzed the data with surface water pollution. The data collected by the state pollution control departments has around 350 data points of different parameters relating to water pollution. The period of this survey was from March 2011 to December 2012. We tried to find if any of the hospital indicators that we have are related to the water pollution data. Most of the indicators showed no correlation with almost all of the correlation values lying between a range of $-0.2 < r < 0.2$. This may be attributed to the fact that Himalayan water in upper regions is typically clean from pollutants and water solutes that are added in the plains due to factory. Figures 7.15 show the how OPD cases fare with variation in different water pollution parameters. Also since the data points were very sparse we could not drill down to get weekly or daily correlation values as the results would not be very accurate with typically only about two or three data points available per week.

7.2.7 Possible Inferences from Air and Water Pollution Data

When we analyze the air pollution data with hospital indicators on weekly frequency we find some of them have low correlation. We then do the same analysis with monthly frequency, correlation with some indicators is clearly visible. There is an apparent positive correlation between NO_2 amount in air and number of respiratory disease cases turning up in OPD's. Similarly we clearly have more ECG's being done with increased NO_2 pollution and SPM in air. Also the number of OPD cases in Opthalmology department have increased with increase in SPM. To get more insight into whether air pollution directly causes these outcomes we require air pollution data for an extended period which is currently not available. In case of water pollution, almost all the parameters that are being monitored lie below the threshold for alarm. Water supply in Himachal is essentially clean and fresh due to numerous rivers flowing which carry fresh Himalayan water. This is also supported by the fact that we find no correlation between any of the water pollution parameters and hospital indicators.

Chapter 8

Benefits Challenges and Approach

An interoperable health information system yields numerous benefits. But these benefits come at the expense of a lot of challenges which need to be overcome before we can get a working system. In this section, we will briefly discuss some of the challenges that we faced from different aspects during the project lifecycle, some of these continue to be resolved.

8.1 Development Challenges

There was a lot of effort that went into getting a robust working module. The module went through a long development cycle which is briefly described here. The work on the module began in january 2012. A module had to be written in OpenMRS as well as changes were needed in DHIS2 to get a complete working system. The first version of the module which was developed was not used due to various logistical reasons. The first working version of the module was written by Bob Joliffe, a PhD student at the University of Oslo in December 2012. The module had the features which allowed creation of reports, import and export of reports as well as posting of reports to DHIS2. The module features were enhanced iteratively over the period of next year. First a feature of synchronization with DHIS2 was added. Then a feature of automated scheduler was added to ensure timely reporting.

8.2 Infrastructure Challenges

Infrastructure turned out to be a major bottleneck during deployments across the hospitals. Some of the major infrastructure issues which were faced during deployment are discussed below.

1. Governance: The process of getting the required infrastructure for the system might sound simple but is non-trivial and time consuming in the Indian scenario, especially in government institutions. For procurement of equipments in a large number, a tender is generated to ensure impartiality and cost effectiveness. Although necessary this process results in a delay since separate tenders are issued for purchase of equipments, installation etc.

2. Availability of Internet: Availability of the Internet was a primary need for a working module. Since all the other modules deployed in the hospital did not need internet access, the process of making the Internet available to the server room had been delayed. Hence even when the module was ready and deployed states had to wait for several weeks before the internet was made available to the server and report posting could begin. Another issue which cropped up later in one of the hospitals was that of scheduler not working correctly. On testing, we found that the Internet modem was shut down every evening after the OPDs since its power was drawn from the OPD ward which operates only till evening.

3. Maintenance of IT Infrastructure: Maintenance of the existing infrastructure is also important to assure long term smooth continuity of operations. Currently the server maintenance and backups are taken by system administrators from HISP-India but for a long term successful and scalable deployment this has to be taken care by local or government staff at regular intervals. Constraints on recruitments make it difficult for the hospital to hire the right technical staff required to make the system work routinely.

8.3 Architecture Challenges

In the approach to provide information exchange between healthcare systems, there has been a lot of debate over the model to be followed. Centralized and Decentralized models are the main contenders and in some cases we also have hybrid models which have features of both.

Table 8.1: Key benefits and challenges between two types of architectures

Parameter	Centralized	Decentralized
Information Exchange	Easy between modules on same system. Single system working as HMIS and HIS.	Difficult.Protocols to exchange data between different systems.
Robustness	Single Point of Failure	Robust Across different deployments
Scalability	Easy to scale for homogeneous systems	Tougher to scale than a centralized system.
Security	SSL/TLS, database encryption	To be incorporated in exchange protocols
Internet	Fast and 24x7	Moderate and only during information exchange

- Centralized: In case of centralized model, we will have all the information residing in a single system. In our case it essentially means a single system which will function as HMIS at the state level and HIS at the district level. All the patient information will lie in a single system which will have a common database. It can be very beneficial since it can minimize cost of maintenance and control of infrastructure becomes simple. In such a system, the problem of information exchange is minimized to a level of being able to access information from multiple modules on a single system, all the information residing in a single database. Complications arise due to the fact that with a large deployment the database will be very huge and effort will be directed to make system and information extraction efficient. Since we will have a single working system we also need to look into problems for tackling a single point of failure and reliable Internet connection across the state. Here we need to consider security of the system from two perspectives. Firstly, we have a data warehouse with entire state health information, including individual patient transactional information hence it needs to be secured and well protected. Secondly, we need to look into security of information transmitted over the Internet, mitigate possible eavesdropping cases by using SSL/TLS encryption mechanisms.

- Decentralized: A decentralized model for the state would imply that we have multiple standalone systems working without any external interference which share information through some mechanism. We have this model currently followed in the state. OpenMRS based HIS are deployed in all the district hospitals which work as standalone systems without speaking to each other. All the systems across the district are the same module and architecture wise, but each is a different implementation in term so of database, server etc... With a centralized system scalability can be easily achieved by making the required infrastructure available, including Internet. In case of decentralized model, we need the basic infrastructure (without internet) to start providing service and Internet is required only later on to support information exchange. Now since information is exchanged between similar but heterogeneous systems located in two separate places, the security of information transmitted also become important. Currently only aggregate-level information id transmitted from the hospitals to the state DHIS2 server but the information is not encrypted. This is one of the areas of improvement which must be looked into the future.

8.4 Technical Challenges

The main technical aspects in getting the project to work was defining the exchange process for two already working widely used open source systems. SDMX-HD was conceived as an open standard to allow information exchange between disparate health information systems. Its a XML based standard built which an implementation of ISO SDMX, an already widely used and mature standard for aggregate statistical information exchange. The reason why SDMX-HD to be written was that there was no agreed upon open standard till 2010 to facilitate information exchange in healthcare. Once the medium was clear the question was of implementation. Reports formats were defined and information was to be fetched through queries from the database. Since the modules were built before interoperability architecture was in place, they were in general made efficient from transaction processing. This resulted in some of the queries turning out to be very complex and time taking. Some queries take time in order if minutes to execute. Although use of OpenMRS concept dictionary allowed queries to be written in a flexible manner,

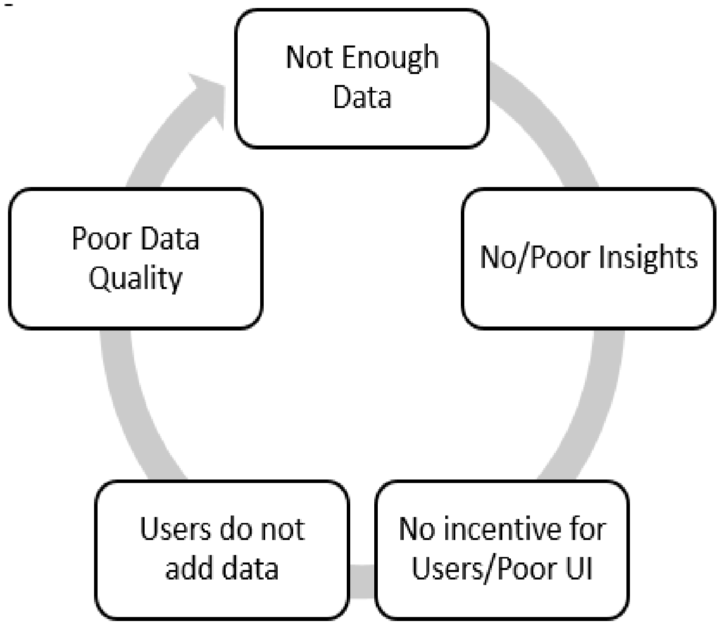


Figure 8.1: The viscous cycle of data demand and supply is shown. A system where users do not get any usage incentive tends be rapidly rejected and thereby prove ineffective over extended time period.

which could be used across hospital instances but minor changes in the information required or modules possibly required changes in query which then needed propagation across hospitals.

8.5 Usability Challenges

One of the major challenges which need to be addressed by information systems in the healthcare domain is that of usability. Many IT systems built in past and have failed because they lacked usability. In case of healthcare where the general users are doctors, nurses, pharmacists etc. are not very tech savvy usability issues tend to become a major deterrent to the adoption of the information systems. This issue becomes even more important in the context of developing countries which have over loaded facilities and over worked health staff. The ratio of patients per facility (beds, doctors, nurses etc.) is much higher in developing countries than in developed. The primary objective of the physicians here, especially in a government setting is to provide care to as many patients as possible. Enhancing usability of the system in a way that it helps users reduce load and improve efficiency can result in quick adoption of technology.

8.5.1 Resistance of doctors

In the context of developing countries, doctors as well as the nursing staff are heavily loaded and have a high number of cases per day. Till the time, doctors do not see any value coming from the system the extra minutes that they put in to punch in the information in the system results in no help to them. Information though important for administrative purposes does not help the doctors in achieving or aiding their short term goal of tackling the increasing patient load. From their point of view, its just adding to their work load since they get no information from the system but end up adding it to the system. Also doctors want to avoid committing to provisional diagnosis in their initial inspection without any tests. Due to the above facts there are no short term benefits seen by the doctors of using the system and the possible long term benefits are hampered due to the lack of information in the system.

Possible Solutions

Some possible solution to this issue can be found if doctors get to reap some short term visible benefits.

1. Providing Contextual Aid: One such possible short term benefit to doctors can be providing contextual aid speedy diagnosis. The contextual aid may include features like patient demographics, disease profile of the patients district, patients medical history and known allergies. These set of information can help the doctor to get a better understanding of the patients condition and make quicker and correct diagnosis. To provide patient history or clinical summary we can first look into the one available with the hospital or related group of hospitals. This will also provide patients with extra incentive to stick to a group of hospitals, a practice very commonly followed by private hospital chains but yet to be implemented in any government hospitals.

2. Better UI/ flexibility may be through touch screens and tablets: One of the ways to improve response towards the system can be by making it easier to use by using touch screens and more intuitive UI to enter the information. Another way to aid the doctors may be to make information available even away from their office desktops. The information can be made accessible on tablets which doctors may take and move freely through the hospital. They can also add more information, prescriptions on successive visits.

8.5.2 Resistance of clerks and other users

When a new system is deployed and the hospital staff need to shift to a HIS from paper records it results in changes in their functioning and routines. There can be multiple issues which might hinder adoption of the system. Some of the issues are linked to training while others are related to the system itself. The issues related to IT or the system need to be understood and should be taken care of as soon as possible. Some issues may be as a result of bugs in software which are easy to detect but there are other issues which result due to poor usability. These issues are more subtle and can be understood by doing a post deployment study including users. One such example comes from the registration department of the hospital. With time, the number of registered patients grew and search for already registered patients starting becoming slow. Given the long line, users at registration desk found it more convenient to mark the patient as a new visit rather than a revisit resulting in faulty clinical information of the patient. On looking into the problem, the cause was found in the fact that the search query which involved multiple joins was taking very long. Some simple changes in the table structure was all that was required to solve the problem. Here we see that the problem though trivial resulted in significant trouble to the users. Such usability issues should be resolved to help better and more efficient use of the system.

8.6 Analytics Module

The next step in this process is to build a module which will help in doing analysis on the collected information. Currently DHIS2 supports basic indicator visualization. The analytics module must include statistical functions such as variance, standard deviation, correlation etc. Some of the desired features from this module are discussed below.

1. Data Quality Monitoring and Improvement: There needs to be a lot of stress on getting quality data from the entire deployed health infrastructure. As we saw in figure 8.1 that quality data is important to build trust in the system and encourages users to adopt the system. This module can monitor data quality by comparing streams across the hospitals to look into sudden large variances across parallel streams from adjoining districts. It can look into ways to improve data quality by generating alerts in case of missing reports or data points.

2. Referral information analysis: One very significant information not being collected and analyzed is patient referral information. Patients at PHCs and CHCs are referred to district hospitals for particular kind of treatment. This information is not being currently captured in the system. Once this information is available, we can get a better picture about health status of state at PHC level. We can pin point lagging PHCs or CHCs and put special attention to improve their health profile by increasing the number of doctors, getting more equipment etc.

3. Epidemiology Information: The module will also clean data and replicate analysis shown in chapter 7, study of individual indicators for anomalies, looking for correlation between specific indicators and supplementing database with more information from sources such as weather, air pollution, water pollution and demographic data. We can look into how one or more of these factors affect our indicators through this module. Even after deployment, there is still a noticeable lag between data collection taking place in hospitals, report generation and any action being taken on the reports. This is explained by the fact the most of the reports are monthly or weekly. To improve health services in the state the turnaround time for any action should be minimized. This can be achieved by generating more daily reports. All the daily reports could be analyzed by the analytics module which can raise alarms based on user settings. For example, if authorities may set an alarm if the number or malaria cases cross a specific threshold value on a day or consecutive days. This will help build an early disease warning system for the state.

4. Scalability: The module should be scalable in sense that the analysis should be repeatable for new time periods. Adding new reports for study should be as simple as possible. In the current model it is not very simple since any new information which needs to be pulled from the hospitals requires new queries to be written. This problem can be tackled by use of a visual query builder to support simple queries. It should also be easy to accommodate for data coming from new hospitals.

8.7 Advocating Actions

The information makes no sense if no action is taken based on available information. Some of possible actions which might be taken based on insights from the current data are discussed below.

8.7.1 Through Financial Indicators

As we saw in the financial indicators, there is increase in the subsidy amount over the years. The possible reasons we discussed above was increase in number of patients coming under subsidized category. We also noticed a sharp difference in the ratio of number of tests undergone by BPL and general category patients. We also looked into average income per general patient and average subsidy per BPL patient. Such results can help in better management of hospital finances. Hospital should strive to reduce wastage and losses through unreasonable tests and procedures and improve overall efficiency. For starters, the lab can be made more accessible to the general public so that there are increases revenue from labs.

8.7.2 OPD indicators

When looking into disease profiles we found a high correlation between acute respiratory disorder and air pollution parameters. So monitoring pollution values can help authorities to be prepared for sudden high influx in the number of patients in a particular OPD.

8.7.3 Revisit Information

We studied revisit information to find that most revisits are on Tuesday but most new visits are on Monday. Consequently, most lab tests are performed on Tuesday. This overcrowding and spilling of tests to next days can be better managed if staff are prepared to tackle the Monday rush and hence tests are done accordingly.

8.8 Approaches

We discussed some of the issues pertaining to the deployment of any health information system and possible solutions. Any government or hospitals needs to put in a heavy up-front investment to digitalize its records and get a hospital information system. The benefits of such a system are only visible if the users actually use the system and are an active part of the making it better. This section we will discuss the steps which can to be adopted to enhance enhance user involvement and action to reap benefits from a deployed HIS.

8.8.1 Training

A very important component in adoption of any new technology is training. The users need to be trained at the new systems. They should be familiar with all the functions and feature and know whom to contact in case of any doubts. Also hospitals should employ a small in house IT team to trouble shoot failures and maintain the infrastructure. Also special emphasis need to be laid on the fact that users not only need to know how to use the system but they must be motivated on why to use the system.

8.8.2 Advocating use through peer pressure

Another viable way to improve the data quality is by advocating use through peer pressure and setting up positive feedback loops. Hospital staff should be encouraged to utilize the system and incentivized on proper use. For example, if doctors in OPD are filling up information in the system we can get information about how many patients each doctor is seeing. The doctor with highest patients per department can be given incentives to stimulate use of the system.

8.8.3 Building Capacity to analyze

One of the main advantage of this tedious exercise of interoperability to ensure information exchange is that of providing better support for decision making at all levels. To provide insights and present a true health picture of the state. One major obstacle currently in this regard is deciding which information to pull and what inferences can be drawn from it. Since not everyone is equipped to answer the above questions, states even after deploying such systems need to work towards building a group of people in-charge of studying the current health status, making sense of the information, drawing actionable insights and looking into processes of making information extracted richer by supplementing it with information from other sources such as weather, pollution, demographics, etc. The group should also look into which new indicators can be extracted and how to present them visually.

Chapter 9

Conclusions and Future Work

9.1 Conclusions

This paper presented a successful deployment of Dhisreport module for integrating two health information systems, OpenMRS and DHIS using a widely accepted statistical data and meta-data exchange protocol in healthcare domain called SDMX-HD. Currently, the module requires queries to be written for every new data element. This is one of the major challenges for the scalability of the module. Around 600 queries had to be written to fetch values for the set of 11 initial reports. Since the modules were designed to address the set of requirements for integration rather than reporting, some of the queries have turned out to be really complex. Therefore, if any new data element is required by the state a corresponding query has to be written for it in OpenMRS. This puts a significant burden on the developers. The problem can be addressed to some extent by providing a UI based query builder allowing implementers to easily write simple queries, but this would require that they have some forehand knowledge of database schema. Another way is to enhance the module to use existing cohort builder in OpenMRS to get values for simple queries without worrying about the schema. Though this will solve the problem of implementers being aware of schema it cannot realize complex queries through interface. We plan to consider these issues while improving the usability of our module.

There was significant effort involved in training the users. Since in most of the cases users lack the technical background, they take some time to understand the workflow. Further, users sometime tend to not synchronize the reports before sending and miss sending some of reports. This issue can be tackled by use of automatic scheduler. Since the frequency of reports if fixed an automatic scheduler can be used to send the reports to DHIS after synchronization of the reports.

The analysis of data is mostly dependent on the quality of data coming from the hospitals. Sufficient information can only be extracted from the data if it is continuously available and properly validated. This paper showed some of the simple analysis which can be carried out on the limited data available. This will be significantly expanded with more data becoming available over the coming years.

The project was carried out jointly by HISP India and IIIT Delhi. HISP India, a NGO which started in late 2000 has been working in development and implementation of customized HIS in South Asia. Although many interesting application are developed in a research environment round the clock very few actually get deployed and used at wide scale. This is one of the reasons why such collaboration proved to be beneficial for both the organizations.

9.2 Future Work

An important part of the project is to support ongoing customizations with the 'Dhisreport' module. Small new modifications and bug fixes keep coming time to time which need to be supported. In future when we have quality data from all the 20 district and regional hospitals much better and useful horizontal analysis can be undertaken. The analysis was mostly done on a local computer with tools like MS Office, it will also be beneficial if a statistical and mining tool can be built around or inside of DHIS2 which can allow detailed analysis of reports. The tool for starters will include features simple features like correlation, linear regression and trend lines, clustering and other such techniques relevant to the healthcare domain. Healthcare analysts and epidemiologists can then use this tool to derive deeper insights into prevailing health situations and thereby take better and informed decisions.

Also a very important and often neglected component of any Information System project is the post deployment evaluation. Researches have found that many such systems fail because of the lack of proper evaluation framework. They also point out to a set of indicators which can be used to evaluate a Health Information System [13]. This forms an important component of future work that needs to be carried out to evaluate the deployment.

Bibliography

[1] Health Insurance For Poor. http://www.rsby.gov.in/about_rsby.aspx. Accessed: 2014-06-21.

[2] H.P. State Pollution Control Board . http://hppcb.nic.in/index.html. Accessed: 2014-04-21.

[3] Polio Eradication : India Celebrates Victory Over Polio. http://www.unicef.org/india/health_3729.htm. Accessed: 2014-07-01.

[4] Weather Online: Historical Weather Data . http://www.worldweatheronline.com/premium-weather.aspx?menu=historical. Accessed: 2014-04-21.

[5] Braa, J., Kanter, A. S., Lesh, N., Crichton, R., Jolliffe, B., Sæbø, J., Kossi, E., and Seebregts, C. J. Comprehensive yet scalable health information systems for low resource settings: a collaborative effort in sierra leone. In *AMIA Annual Symposium Proceedings* (2010), vol. 2010, American Medical Informatics Association, p. 372.

[6] Brailer, D. J. Interoperability: the key to the future health care system. *Health Affairs-Millwood VA Then Bethesda MA-24 24* (2005), W5.

[7] Diehr, P., Yanez, D., Ash, A., Hornbrook, M., and Lin, D. Methods for analyzing health care utilization and costs. *Annual review of public health 20*, 1 (1999), 125–144.

[8] Haux, R. Health information systems–past, present, future. *International journal of medical informatics 75*, 3 (2006), 268–281.

[9] Hillestad, R., Bigelow, J., Bower, A., Girosi, F., Meili, R., Scoville, R., and Taylor, R. Can electronic medical record systems transform health care? potential health benefits, savings, and costs. *Health Affairs 24*, 5 (2005), 1103–1117.

[10] Jan Walker, Eric Pan, D. J. J. A.-M. D. W. B., and Middleton, B. The value of health care information exchange and interoperability. *Health Affairs* (2005), 10–18.

[11] Latifov, M. A., Mukherjee, A., Chakravarthy, V., and Sahay, S. Practical approaches to designing standards: the case of a district hospital information system in northern india.

[12] Lippeveld, T., Sauerborn, R., and Bodart, C. Design and implementation of health.

[13] Littlejohns, P., Wyatt, J. C., and Garvican, L. Evaluating computerised health information systems: hard lessons still to be learnt. *Bmj 326*, 7394 (2003), 860–863.

[14] Mamlin, B. W., Biondich, P. G., Wolfe, B. A., Fraser, H., Jazayeri, D., Allen, C., Miranda, J., and Tierney, W. M. Cooking up an open source emr for developing countries: Openmrs–a recipe for successful collaboration. In *AMIA Annual Symposium Proceedings* (2006), vol. 2006, American Medical Informatics Association, p. 529.

[15] Musen, M. A., and Bemmel, J. H. *Handbook of medical informatics.* MIEUR, 1999.

[16] Mykkänen, J., Porrasmaa, J., Korpela, M., Häkkinen, H., Toivanen, M., Tuomainen, M., Häyrinen, K., and Rannanheimo, J. Integration models in health information systems: Experiences from the plugit project. *Medinfo 11*, Pt 2 (2004), 1219–1222.

[17] Mykkänen, J., Tikkanen, T., Rannanheimo, J., Eerola, A., and Korpela, M. Specification levels and collaborative definition for the integration of health information systems. *Studies in health technology and informatics 95* (2002), 304–309.

[18] Pardon, S., and Berry, D. Mirthconnect howto document.

[19] Purkayastha, S. Hixen: An integration engine for multi-vocabulary health information using rest & semantic metadata mapping. In *Information and Communication Technologies (WICT), 2012 World Congress on* (2012), IEEE, pp. 679–684.

[20] Srivastava, S., Vatsalya, V., Arora, A., Arora, K. L., and Karch, R. Utilizing healthcare developments, demographic data with statistical techniques to estimate the diarrhoea prevalence in india.

[21] Tierney, W. M., Achieng, M., Baker, E., Bell, A., Biondich, P., Braitstein, P., Kayiwa, D., Kimaiyo, S., Mamlin, B., McKown, B., et al. Experience implementing electronic health records in three east african countries. *Stud Health Technol Inform 160*, Pt 1 (2010), 371–375.