A Cloud Based Disease Surveillance and Health Information System

By:

Adnan Bashir (abashir.mscs15seecs@seecs.edu.pk)

Supervisor:

Dr. Sohail Iqbal (sohail.iqbal@seecs.edu.pk)

Assistant Professor

Department of Computing (DoC)

A thesis in partial fulfillment of the requirements of the degree of Masters of Science in Computer Science (MS-CS)

In

School of Electrical Engineering and Computer Science (SEECS)

National University of Sciences and Technology (NUST)

H-12, Islamabad (44000), Pakistan

Approval

It is certified that the contents and form of the thesis entitled "Cloud Based Disease Surveillance and Health Information System" submitted by Mr. Adnan Bashir, have been found satisfactory, for the requirement of the degree of MS (Computer Science).

Advisor: Dr. Sohail Iqbal

Signature:

Date:

Committee Member 1: Dr. Asad Waqar Malik

Signature:

Date:

Committee Member 2: Dr. Safdar Abbas Khan

Signature:

Date:

Committee Member 3: Dr. Omar Arif

Signature:

Date:

Abstract

Outbreaks can overwhelm weak or fragile health systems that lack the tools, infrastructure, policies, and systems to keep communities healthy and safe. Timely detection, preparedness, and appropriate response are essential for limiting both the loss of human life and crippling political and socio-economic impact of disease outbreaks. Countries must build effective and sustainable disease surveillance and reporting systems that mobilize all levels of the health system including, communities for crisis response.

In Pakistan, currently there is no health information and diseases surveillance system. National Institute of Health has been given the task to detect disease outbreak (both current and future) and prepare appropriate response against it. The current disease surveillance system is completely manual. Diseases are being reported by sending hard copies or using the email system. Most of the disease outbreaks are being reported by the print and electronic media while the government has no or very limited information for the reported outbreak. This ineffective reporting system not only creates problems in managing the counter measures against a disease outbreak but is likely to cause a mass hysteria amongst the people.

District Health Information Software 2 (DHIS2) is an open source, web based software developed by University of Oslo and can be deployed in Pakistan for Disease surveillance and Health Information Management. It is being deployed worldwide in many countries and some institutions like the European Union (EU), and World Health Organization. A total of around 47 countries, both developed and under developed are in the deployment phases of DHIS-2. DHIS2 is not only a disease surveillance tool but is also used to process statistical data, validate the data, analyze and present it. Its data analytics and management platform is completely web-based and boasts great visualization features. The software has ability to create analysis from live data in seconds.

Certificate of Originality

I, Adnan Bashir s/o Raja Bashir Ahmed, hereby declare that this submission is my own work to the best of my knowledge. It contains no materials previously published or written by any other person, nor material which to a substantial extent has been accepted for the award of any degree or diploma at SEECS-NUST or any other educational institute, except where the due acknowledgment is made in the thesis. Any contribution made to the research by others, with whom I have worked at SEECS-NUST or elsewhere, is explicitly acknowledged in the thesis.

I also declare that the intellectual content of this thesis is the product of my own work, except for the assistance from others in the project's design and conception or in style, presentation and linguistic is acknowledged. I also verified the originality of contents through plagiarism software.

Author Name: Adnan Bashir

Signature:

Date:

Acknowledgements

I would like to thank my teachers especially my supervisor Dr. Sohail Iqbal and Dr. Asad Waqar Malik for their continuous guidance and support at each step. Their guidance and counseling helped me a lot in my research and thesis writing. I would also like to thank my other committee members Dr. Safdar Abbas and Dr. Omar Arif for their insightful comments.

My sincere thanks to Prof. Dr. Aamer Ikram, Executive Director, National Institute of Health for allowing me to continue my education in parallel with my job. Without his support and buck up this research would not have been possible.

Dedication

I dedicate my thesis to my family and friends. Especially to my mother, whose un tiring support, encouragement and push of tenacity kept me on my toes. I would also dedicate my work to my close friends Waqas Ahmed and Ahsan Gul who never left my side and supported and guided me at every step.

Table of Contents

Table of Figures

CHAPTER 1: INTRODUCTION

The countries around the world are upgrading their public sector management systems using Information and Communication Technologies (ICTs) in order to streamline the workflows in their governance. The introduction of ICTs in the public sector also called as e-government is expected to not only enhance good management and response but also is considered to be very cost efficient, dynamic and user-friendly.

One of the most crucial public sector entity is health. Health is also one of the top challenges mentioned in Sustainable development goals (SDGs) [1]. The stakeholder of the health sector includes public and the government. Thus, any delay can directly hamper the public [2].

This study explains the use of information technology to manage the health sector of a country in regard with the implementation of the integrated disease surveillance and response (IDSR). Appropriate use of information technology, to manage and use data, is a key component of strengthening health systems. The effective control of communicable diseases requires, effective response systems. Effective response systems depend on effective disease surveillance. Surveillance can be defined as the ‘ongoing systematic collection, collation, analysis and interpretation of data and the dissemination of information to those who need to know, in order that action may be taken’, or in short “information for action” [3]

Effective surveillance is only possible through real time data management which is only possible with the help of ICTs. ICTs, can provide a hassle-free data transfer from the geographically distributed locations to the central government institute with in no time.

1.1. Current System

At present, in some parts of Pakistan the existing health information system is based on District Health Management System (DHIS) system. DHIS was introduced in Pakistan after the devolution of Ministry of Health in the year 2001, in order to fulfill the needs of health management. DHIS was deployed with the help of Japan International Cooperation Agency (JICA) [4].The DHIS system lack features such as data dissemination, and analytical features required to predict the outbreak. The DHIS is used at the town level, with no integration with other DHIS system installed at neighboring divisions.

The systems work in silos hence having no mechanism of surveillance and reporting. Most of the systems are fragmented meaning that either they are limited to some places or diseases. Some of the systems for example the Polio programme and TB programme are working vertically, which not only results in data duplication, but also duplication of efforts.

If we look into the impact of these systems on the country's health infrastructure, then we would know that a lot of efforts are required just to get data on a single disease and then these efforts are repeated by very few health programs to get data on their disease of interest as each system works in silos. Also there is no data sharing between these programs. All of these programs have built their own custom made systems hence there is no interoperability between them of any kind.

1.2. Problem Statement

Unfortunately, in most of the countries, there is no proper Health Information System. Most of the work is performed manually and the information is shared through emails or messengers. However, such systems fail to predict or identify the outbreaks. Most of the time outbreaks are identified after a few days or even weeks [5].

Similarly, in Pakistan, there is also no such system installed to support health-related issues. However, a number of health information systems are deployed under different public initiatives. These systems work in silos each focusing on monitoring the activity or disease of its own interest. These information systems are functioning with varying degrees of success as surveillance systems [6]. While the country frequently faces emergency situations with regards to communicable diseases. It does not have a well-established system to provide regular, timely, and valid information on disease outbreaks and other health-related issues.

In Pakistan the information is present at first level care facilities only. So a culture of systematic and continuous data reporting does not exist. Moreover, the systems vary from place to place and level to level in terms of function, data collection and recording methods [7]. Thus, the existing systems are highly fragmented and often vertical, leading to duplication of efforts and unbalanced resource distribution. The entire systems would get benefit from a more integrated vision of disease prevention at all levels as well as resource sharing.

1.3. Integrated Disease Surveillance and Response

Surveillance and response systems function better when they are integrated and well-coordinated; this also makes better use of limited resources. Integrated Disease Surveillance and Response (IDSR) is a strategy devised by the World Health Organization for strengthening surveillance, laboratory and response capacities at each level of the health system (community, district, province and federal) [3].

The goals of IDSR are to:

- Strengthen surveillance and response for priority diseases at each level of the system
- Integrate laboratories into surveillance
- Rationalise use of resources
- Translate surveillance and laboratory information into specific and timely public health action

Pakistan has several health information systems that function with varying degrees of success as surveillance systems. Information is present in most first level care facilities, so a culture of systematic and continuous data reporting does not exists [7].

However, the systems vary from place to place and level to level in form, function, data collection and recording methods. However, these system are not integrated at any level. At present these systems are highly fragmented and often vertical, leading to duplication of efforts in some cases and unbalanced resource distribution in others.

An effective IDSR system allows early detection of and response to disease outbreaks. Surveillance data is also useful for decision makers and can be used for [3]:

- Priority Setting
- Policy Decisions
- Resource Allocation
- Planning
- Monitoring and evaluation of disease control programmes

A successful IDSR strategy requires the following [3, 8]:

- **Commitment** to implementing IDSR from all those involved

- Agreed **common case definitions** and reporting mechanisms
- Effective **communication** between levels of the system
- Simple, **timely analysis** and reporting
- Laboratory support
- Good **feedback** and rapid **response**

Implementing IDSR may face a number of **challenges**, including [8]:

- **Fragmentation** of surveillance systems: surveillance data may be collected by different government ministries, by the public and private health sectors, by the military, by academic institutions, by vertical programmes and by non-governmental organisations.
- **Separation of responsibility** of surveillance from the responsibility of response
- **Poor communication**, with no feedback of information to the lower levels of the system which collect data
- **Inefficient use of resources**, with duplication of roles across systems
- **Inconsistencies** in case definitions and data collected
- Lack of appropriately **trained staff**
- Lack of suitable **information technology infrastructure**

1.4. Structure of the Thesis

This document is divided into seven (7) chapters. The next chapter, entitled as "Literature Review" compares and reviews different disease surveillance systems implemented in different countries. It then compares the advantages and disadvantages of each system and summarizes the characteristics of an ideal disease surveillance system for Pakistan. In Chapter 3, after analyzing the characteristics of an ideal disease surveillance system, we propose an open source software District Health Information System 2 (DHIS 2) for implementation of IDSR in Pakistan. The chapter discusses various aspects of the software deployment, its configurations, data linkages, security etc. In chapter 4, the document focuses on the priority diseases and agreed case definitions for Pakistan. The mode of data collection that is aggregate based is also discussed in this chapter. In Chapter 5, the data flow of the system is discussed in detailed. Roles and responsibilities of each user which include, users at district level, provincial level and federal level are defined. Chapter 6, of this document is dedicated to the rights and privileges of different user levels. In the last chapter, future prospects and conclusion is discussed.

CHAPTER 2: LITERATURE REVIEW

The implementation of health information management system is a norm now a days. Almost every developed country in the world has a health information management system which is monitored by the government and used for policy making, trend recording and detection of outbreaks. The developing countries have either implemented this system or are trying to do so. Health is a basic right of every individual and it is the job of the government to ensure this.

In this study few research articles have been shortlisted which explain the implementation of disease surveillance system. The articles were selected keeping in view the current status of Pakistan which is a “developing country”. Implementation of disease surveillance system in the countries similar to Pakistan was studied. In order to see what the developed nations have achieved some articles were also selected which showed the implementation of this system in the developed and technologically advanced countries.

Domeika, Kligys et al.[9] explains that Lithuania had a manual disease surveillance system. The authors discuss the implementation of an electronic surveillance and reporting system in Lithuania for the prioritized communicable diseases. The manual reporting system was paper based and very time-consuming. It was found that only aggregate data was reported at the government level. Data duplication was also a big issue. It took a month to report a notifiable disease to the national level. However there were some the exceptions in case of few diseases which were under the category of high priority. Even urgent disease reporting took a week to reach to national level. Therefore, with the help of Sweden Government, Lithuania introduced electronic reporting system.

Two softwares SmittAdm and SmiNet-1 were under consideration for national wide deployment. These both sofwares were built in lotus notes [5]. The mentioned softwares were both suitable as they allowed the users to work offline. However, SmittAdm had few extra advantages over SmiNet-1 as it allowed the patient record integration and also was helpful at the country level in detecting outbreaks [5]. Another reason for choosing SmittAdm was that it could be easily and quickly acquired. Lithuania could easily acquire the software from Stockholm county. As a result, a new program, was developed which was very similar to SmittAdm. It was named as System for Data Collection and Analysis of Communicable Diseases (ULISAS). ULISAS changed the system from a monthly paper based aggregated reporting to a timely reporting system which also included patient data [9].

Howver ULISAS had a number of drawbacks as well and it needs to be further improved by:

- Integrating laboratory and physician notifications from the whole country.
- Outbreak detection algorithms and validation checks should be improved
- In order to extend ULSAS further a web application should be developed for easy acces to the system.

Rolfhamre, Jansson et al.[10] explains another software built primarily for the needs of Sweden for disease surveillance in the country by the name of SMINET-2. It is basically an upgrade of SMINET-1 [11] but more secure. SmiNet-2 was developed using standard edition of Java 2 while the web module used enterprise edition of Java 2. Two different types of database servers were used. MySql database server was used by county medical offices for their local services while Microsoft Server 2000 was used by the central server databases. For web services, Apache Axis in conjunction with a Jakarta Tomcat as application serve. The data was collected at county medical offices (CMO) and stored in their databases.

In total twenty two servers were being used within the system, a central server and 21 local county servers. SMINET-2 provided a number of advantages which included timeliness in the data flow, laboratory and clinical notification integration and custom made notification forms to handle more than fifty diseases.

Weaknesses of SMINET-2 were

- the output functions for the non-privileged user were not available and the user could only see static graphs, there was no alert system integrated.
- Algorithms needed to be implemented to detect clusters of patients.
- An early warning system should have be integrated with the system [10].

Epidemiologists and public health professionals are increasingly relying on geographic information systems (GIS) to predict, assess and study outbreaks in real time and to report to decision makers for necessarily actions. With the help of GIS the health professionals can easily visualize complex spatio temporal events. This in returns help to analyze data. Unlike old traditional maps, GIS have a variety of advantages like they help to pin point prevention and intervention programmes and they are also easily updateable which is very useful especially in less developed countries. EpiScanGIS is a software with the capabilities to visualize clusters. Reinhardt et al [12] explains that EpiScanGIS has an in built automatic cluster detection system. The system extracts the required information to be used by the epidemiologists like as number of fine type, cases, county, diameter of the circle and p-value in which the outbreak has occurred. It is evident that in order to appropriately assign resources and calculate the disease burden for preventive measures disease surveillance is very important and a must. It is particularly useful in the instances of infectious diseases to see where they spread [13]. EpiScanGIS is also very useful for laboratory surveillance, it enables the sender of a lab sample to query the database and extract data geographically, this helps the sender better understand the role of laboratory surveillance [12]. EpiScanGIS was used to provide reports from the reference laboratory every week to the public health offices which provided information on newly detected disease clusters with the help of geographical maps. EpiScanGIS made all of this achievable hence the system had now become a very important part of the surveillance [12].

According to Ward, Brandsema et al. [14],the main purpose of surveillance for communicable diseases is to launch a timely and appropriate public health response so that the disease spread could be prevented. It must be kept in mind that for effective reporting, the reporting should be accurate and timely. In The Netherlands, it’s the job of Gemeentelijke Geneeskundige Dienst (GGD, municipal health services) to generate appropriate responses against any disease outbreak. The laboratories and medical staff are required by law to report any patient diagnosed with notifiable communicable disease to GGD. As a result reports are generated at the GGD and sent to the respected chief medical officers (CMO) at the Inspectorate of Healthcare (IGZ). Some voluntary reporting was also carried by National Institute for Public Health and Environment (RIVM). As the data had to be sent to these different organizations, there were different steps and processes involved for each organization. The internet-based reporting system OSIRIS allowed all this to happen in a single, merged process. Only the authorized users at the IGZ, RIVM and GGD were given access to the system. OSIRIS has the ability to make the initial reports also available at IGZ and RIVM for early warning of disease outbreak. In the meantime the system allowed GGD to update and finalize the report [14]. Ward et al. [14] did comparison between the diseases reported by old paper based system during the year 2001 with the diseases reported by OSIRIS in 2003. For this study diseases were chosen which had a minimum of 100 cases per year. In order to determine timeliness of the system 3 different time points were defined.

- T1 (Day 1 of the illness which was entered in both conventional and OSIRIS reporting system)
- T2 (Date at which illness was reported at GGD)
- T3 (Date at which the illness was reported to IGZ and RIVM)

Three different types of delays were studied in the comparison:

- Total delay was defined as the time taken for report to reach national level i.e. RIVM and IGZ (T3-T1).
- The delay between illness being reported at GGD and IGZ is named as central delay (T3-T2).
- Median delays were calculated with an interquartile range. Median delays between both systems were compared using the Wilcoxon Rank Sum-Test [14]

As explained above only those diseases which had 100 plus cases reported every year were selected. The shortlisted diseases which were a part of this study were as follows

1. Hepatitis A
2. Legionellosis,
3. Pertussis
4. Bacillary Dysentery and
5. Malaria.

Another factor for choosing these diseases was that they represented distinct categories of communicable diseases in the Netherlands. The different categories included:

1. Enteric infection
2. Laboratory- notified infection
3. Vaccine preventable diseases

4. Respiratory infection and
5. Travel-associated infection.

The better timeliness did not affect the quality of data because the electronic report consisted of more information than the old system. The reason behind the improved timeliness was the reduction of delay between GGD and national authorities. This improved timeliness was due to the newly introduced electronic system OSIRIS as there was no other major change in work practices at GGD other than this system which could have resulted in the reduced local reporting delay (T2-T1).

This system was also very effective in administration and reduced 50% administrative workload at GGD level. The noted improvement in data quality is also important as this availability of more complete information should enable national authorities to respond in a timelier and appropriate manner [14].

Another approach to collect data related to disease surveillance is to use data mining techniques and gather data through online news sources. Brownstein, Freifeld et al. [15] in their article explain a similar system developed by the name of HealthMap. HealthMap is a free to access online, information system which helps in organizing data on outbreaks with respect to time, infectious disease agent and geography. HealthMap (http://www.healthmap. org/) is a real time multi-stream disease surveillance web application that continuously adds new reports on new and current communicable disease outbreaks. The system extracts, categorizes, filters, and integrates the reports, which in return helps in early detection and knowledge management. HealthMap uses data mining to extract data from electronic media in real time. The system depends upon a number of electronic media sources, which include online news sources for example Google News, expert curated discussions like ProMED-mail, and validated official reports from organizations like WHO [15].

However as the software relies on the web based media sources it violates the core concept of implementing the disease surveillance system in a country. The reporting of any disease outbreak should be done by the central government or disease surveillance agencies and authorities. The data flow of HealthMap violates this thing and in fact does completely opposite of it i.e. the data in HealthMap flows from online news agencies towards the governing bodies which is not only unacceptable but also can result in creating a mass hysteria among the general public while the government remains unaware of the situation with no preparedness and contingency plans. As according to Ward, Brandsema et al. [14], the main purpose of surveillance for communicable diseases is to launch a timely and appropriate public health response so that the disease spread could be prevented.. This is impossible to achieve if the government start relying on online news websites. It is also very difficult to filter out the hoaxes and wrong reporting.

Vlieg, Fanoy et al. [16] compared two national level disease surveillance systems running in Netherlands and China in their research. In China if a disease that comes under the category of notifiable is diagnosed clinically or confirmed through laboratory, it is obligatory to report the case to Centre for Disease Control China which collects the data for analyses. The health professionals use a web based system called Notifiable Infectious Diseases Reporting Information System (NIDRIS) to enter information on a standard form. The system provides a platform to report the disease cases. In parallel to NIDRIS another system by the name of China Infectious Disease Automated-alert and Response System (CIDARS) is being used by the Chinese government to facilitate early warning at different China CDC levels. When an outbreak is detected CIDARS notifies the country level CDC team via SMS. After getting notified the county level specialists, verify the information and send a team for field investigations for that particular outbreak. The field investigation team then reports back and the reports are then entered into CIDARS. Although CIDARS has been running successfully in China for early warning but it has a number of limitation like how to cater the false positives?, and also the management of large number of SMS signals that are sent to different office. CIDARS generates automated alerts for early warning with the help of fixed thresholds. A diseases crossing its threshold is immediately reported. China has the largest population in the world, the use of CIDARS helps the government to analyse, such a large volume of reported data and at the same time generate alerts at the county level. The system is fully automated therefore there is no need for expert opinions. However, the system can be customized to add zoonotic surveillance.

In case of Netherlands Vlieg, Fanoy et al. [16] explains that as per law all the notifiable diseases are to be reported to public health services (PHS) by the attending physician or the laboratory which confirms the test. After the case information is collected it is entered by PHS into the OSIRIS, a web based database for disease surveillance. In Netherlands, the experts use ‘barometer’ algorithm that carries out comparison of the number of cases of communicable during the past four weeks with the expected value or threshold. The system unlike China doesnot use SMS to notify the concerned departments whenever a threshold is crossed by a specific disease as in doing so the system would generate many false positives hence effecting the overall quality of data and increasing the work load. After every Netherlands early warning committee (NEWC) meeting a report was formalized which was then distributed amongst 2400 health professionals which are directly involved in communicable diseases. Also the reports generated from the meeting of zoonosis, hospitals and antimicrobial resistance are distributed among all the involved physicians. The flow of data and reports show that this system emphasizes on expert opinions rather than automated data processing and data output. The experts at RIVM decide whether or not to contact PHS for further investigation. The NEWC eliminates the use of SMS, as the system itself can alert all the stakeholders in time. It also eliminates the chances of false positives. Netherlands have a well-established health sector due to which there are very limited numbers of outbreaks, in addition to this the system provides short communication lines between the health professionals and government resulting in easy surveillance. The Risk Assessment and Early Warning (RAEW) focus only on the communicable diseases.

Santillana, Nguyen et al. [17] in their paper quotes alarming facts regarding the outbreak of Influenza and then discusses a solution for timely based response. It has been reported that the Influenza among all the other infectious diseases is the leading cause of death in the United States (US). It has been confirmed that around 50,000 people die each year by influenza like illnesses (ILI). The US Centre for Disease Control and Prevention (CDC) is responsible for monitoring and preparing appropriate responses against disease outbreaks. CDC also monitors ILI in the US by gathering information from physician's reports about patients diagnosed with ILI. Although the data collected by CDC provides very useful insights but it uses a lot of time ranging from one to two weeks long. The main purpose of disease surveillance as discussed above is to generate timely response, hence this time lag creates problems as the decisions are made on the information that is at least one to two weeks old. CDC requires a system that would provide the policy makers with real time data to monitor and manage ILI. According to Santillana, Nguyen et al. [17] for analysis the Electronic Health Record (EHR) is combined with the historical patterns of the disease and then these are boosted by dynamic and robust machine learning algorithms which in-return gives effective surveillance system at the local and national level. The developed system boost timeliness of data flow and also provides data specificity of the reporting sites along with accuracy. This shows the effectiveness of a cloud based system for disease surveillance and health information management system. There are a few IT firms that provide cloud services and mobile applications for health systems.

Athenahealth is one of the leading firms that provide health related IT based solutions. The firm has a database which includes claims for over 64 million people and health records of nearly 23 million people. The authors collaborated with Athenahealth team to measure the timeliness of the electronic system. A weekly aggregated data was collected by the system which included ILI visit counts, flu visit counts, flu vaccine visit counts, unspecified viral visit counts and total visit counts. The system provided near real time data capture and aggregation. The authors were able to obtain the aggregated data of a given Saturday-Sunday week immediately on the subsequent Monday. This showed that the data was available at least one week before the publication of CDC's ILI reports hence increasing the timeliness of the overall system.

CHAPTER 3: METHODOLOGY

3.1. Proposed System

This section summarizes key information on version 2 of the District Health Information System (hereafter referred to as DHIS 2), in particular as it relates to integrated disease surveillance and response (IDSR), and provides pointers to further information.

DHIS-2 is a modern Web-based health information system that is now in use in numerous low and middle income countries worldwide for surveillance and other purposes. The development of DHIS 2 is led by the University of Oslo as an open source project, funded by a number of global donors. DHIS 2 software is therefore free to install and use. The software is regularly updated and improved [18].

DHIS 2 offers advantages over other alternatives as a software platform for IDSR and can be rapidly implemented for IDSR in Pakistan with appropriate planning, resource allocation and accessing of relevant expertise.

DHIS 2 is a completely different tool to the older DHIS system still in use in Pakistan, despite the similarity in names. The DHIS system currently in use in Pakistan was developed in the 1990s and lacks many of the features required for modern infectious disease surveillance.

Users at district (or health facility) level can enter weekly surveillance data via desktop computers, laptops, and/or mobile devices, and can immediately access a range of reports, charts, maps or dashboards. The interface can be configured to use languages including English and Urdu.

The system is easy to use for data entry, requiring minimal training. Users can be trained to develop bespoke outputs for their own use. DHIS 2 can cope robustly with interruptions to Internet access and analysis can be done offline if required. DHIS 2 also has a range of validation functionality, to ensure that data collected is of adequate completeness and quality.

Users can also use the system to send messages to other users, or to add epidemiological “interpretations” to their analyses. The system can also display a dashboard to the public via its Web Portal [19].

DHIS 2 is a very flexible and adaptable system. The central server can be set up as a data warehouse to integrate disparate data sets, but also works well with other information systems (interoperability), so if a province preferred to use a different information system, DHIS 2 could interface with this system. More complex deployments are feasible for when some areas do not have Internet access or mobile coverage. DHIS 2 is also extensible, in that developers can create apps with additional functionality if required [20].

DHIS 2 also scales well to handling large volumes of data at country level. It is usually configured to allow appropriate access to data for each level. For example, users at province level would have access to all data for their province, and users at federal level would have access to all data.

Key advantages of DHIS 2 for IDSR are:

- more timely, accurate and efficient capture and entry of data (via the Web or mobile devices), enabling a more rapid public health response
- capture of both structured and unstructured data, supporting both indicator- and event-based surveillance
- online and offline use
- secure flow of data from district to federal level automated analyses including Web-based tables, charts and dashboards [18, 19]

3.2. Pre requisite Hardware and Software Requirements

DHIS 2 is recommended to run on a Linux server and therefore familiarity with Linux operating systems is required for installation and updates.

The server should have adequate processor (CPU) power and memory (RAM; a minimum of 12 Gb RAM is recommended) to handle the expected volume of data, should be physically secure and if possible should have a fail-over solution. It should also have a reliable network connection with sufficient bandwidth and an uninterruptible power supply [20].

Data backups will be obtained nightly from the database. The database is accessible to authorised users for *ad hoc* SQL **queries**. It is suggested that Network Access Storage should be used to store the data securely and also to create the necessary backups. Instead of using the physical servers it is suggested to use Virtual Environments and create virtual machines. These virtual machines have added security and also help in restoring the servers in case of any system crashes or failures. We can create different check points and virtual environment images to make restoration process easy and with minimum data loss.

DHIS 2 is under active development and several new **major versions** are released each year. The version which will be used for IDSR is DHIS 2 version 2.30. At the time of writing, the latest available major version of DHIS 2 was version 2.31. Major versions often introduce new features and make major changes to existing functionality, which might necessitate retraining of users. Version 2.30 itself can be updated (to a new **minor version** of version 2.30), for example after particular bugs have been fixed [20].

A diagram of data centre to host DHIS 2 is appended below for more clarification.

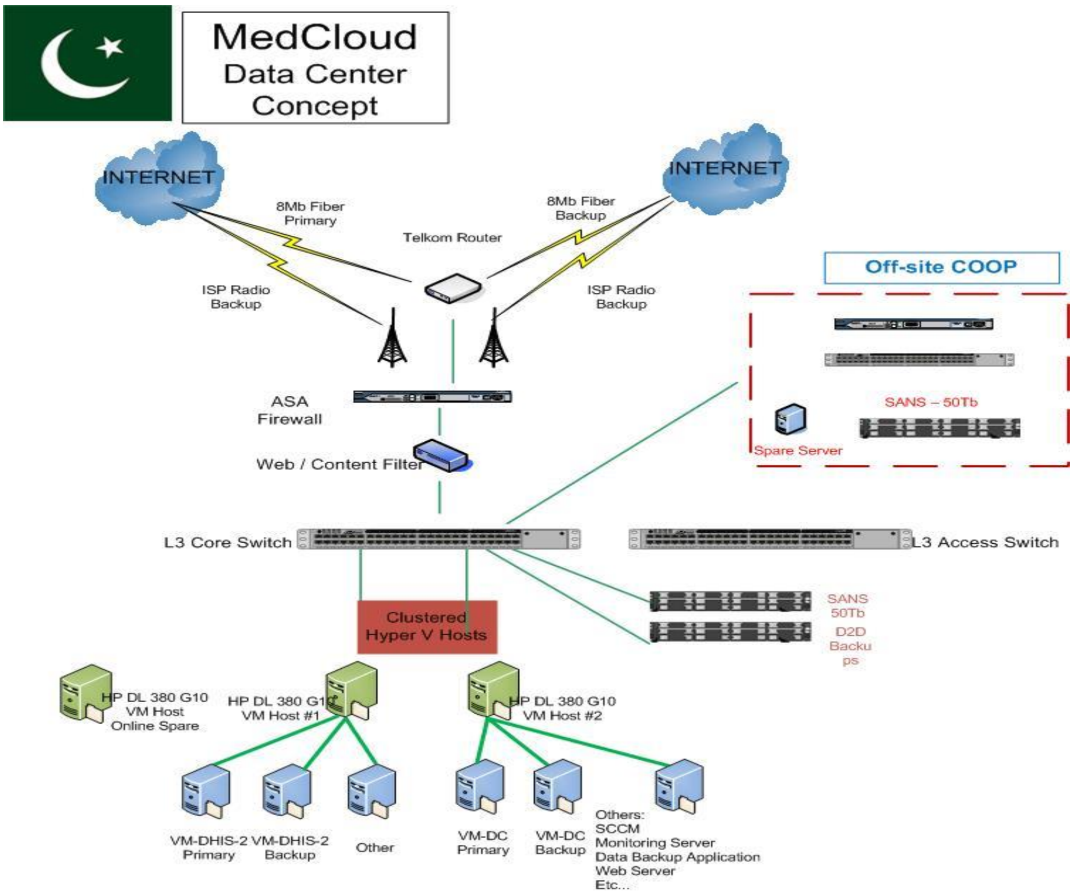

Figure 1 Data Centre Concept of DHIS-2

3.3. GIS Mapping

Another important feature of DHIS-2 is that is allows to upload the shape files into the system. This can help public health professionals in a number of ways. The data can be projected onto the map and clusters can be detected. GIS helps a lot in the early disease outbreak detection, hence making it easier for the decision makers to plan the response timely and appropriately.

GIS visualize complex spatio-temporal events and thus help to analyze data on geographic maps consisting of several layers of information. In contrast to traditional maps, GIS are updateable, and help to appropriately target intervention and prevention programmes, especially in less developed countries [18, 19]

3.4. Meta Data and Configurations

DHIS 2 framework is a very flexible and adaptable, which can be applied in a range of situations. It therefore needs to be set up (**configured**) with information specific to its intended application before use, known as **metadata**.

> ***"Metadata** is the data that is generated when you view data, or when you access data."[21]*

The first metadata usually required to configure DHIS 2 is for **organisation units**. For IDSR implementation in Pakistan, the organisation units involved will be health facilities, district offices, provincial surveillance offices and National Institute of Health (NIH). However it could be any other reporting unit such as a vertical programme office or a military base. For each organisation unit, the following metadata is required:

- Name and code of organisation unit (use of **standard codes** allows linkage of data to other information systems)
- Details of **contact person** including telephone number and mail and email addresses
- **Geographical coordinates** of the organisation unit (to allow it to be shown on maps) [22]

Each organisation unit is assigned certain data sets that it can work with. Organisation units can be assigned particular **levels** within the geographical hierarchy and/or organised into non-geographical groupings. The data flows upwards from basic health units and then up to the district, then to the province and finally at the federal level. The below diagrams explains the flow of the data with perspective of organization units.

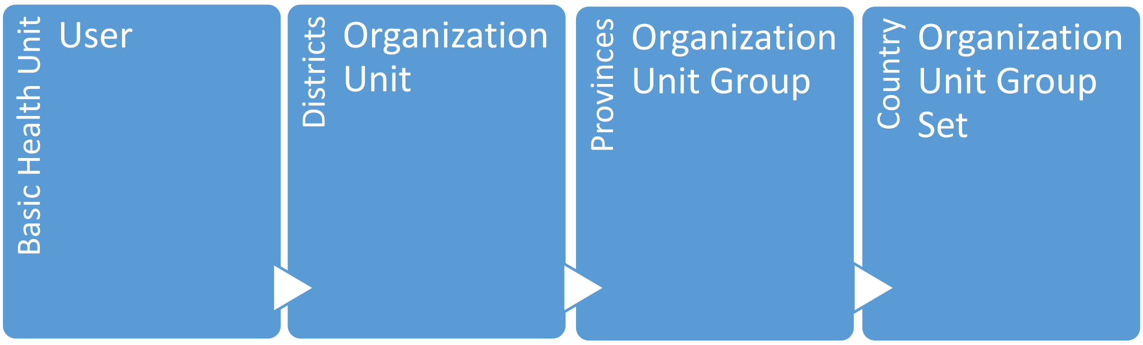

Figure 2 Hierarchy of Organization Units

The next metadata required to configure DHIS 2 is **data elements**. Data elements are single items of data that are to be captured by the system; for example, the number of cases of measles. Data elements can be text, any type of number, dates, times, email addresses or any other type of data which needs to be collected. For each data element, the following metadata is required:

- Name, code and description of data element
- Type of data
- Permitted values of the data element (e.g. gender could be either male, female or other)
- How the data is to be aggregated (e.g. if it is to be summed or averaged and at which levels it will be aggregated) [22]

Data elements may be used for the calculation of **indicators** in DHIS 2; for example, the data element for the number of measles in one year could be divided by the mid-year population to create an indicator for annual measles incidence. The metadata for an indicator specifies a formula to indicate how different data elements should be combined [18, 20].

Data elements may be grouped in a number of ways. For implementation of IDSR in Pakistan, the data elements for counts of new cases and deaths caused by priority diseases will be grouped into a single **data set**. A data set is simply a particular set of data elements. Defining data sets makes it easier for users to select data for analysis.

Other metadata within DHIS 2 allows for the definition of **case-based or event-based data sets**. The metadata required for **users** is listed earlier in this document. For DHIS 2 to be able to create maps, geographical information system (**GIS**) metadata on the locations of sites and the boundaries of areas is required.

Almost all metadata can be defined **manually** using Web apps within DHIS 2. Where this would be time-consuming, metadata can be prepared in a spreadsheet and **uploaded** all at once via the Web API into DHIS 2 as a CSV file or other data format. **Data integrity checks** can be run within DHIS 2 to identify problems with metadata.

It is important that **metadata is kept up to date** with changes such as closure, merging or renaming of organisation units, or changes of contact person details.

3.5. Data linkage

The central server can be set up as a data warehouse to integrate disparate data sets, but can also **import data** from other information systems. In case a province preferred to use a different information system, DHIS 2 may be able to interface with this system. DHIS 2 can import data in a number of data formats (CSV, XML or JSON). There are usually three steps to this process, which may be partly or fully **automated**.

- **Extract**: data is obtained from the source system as a file.
- **Transform**: the extracted data file is converted into a format that the system understands and any data errors are corrected.
- **Load**: the converted data file is uploaded into the system either manually or using the Web API.

This process may require specialist expertise. Where possible, data should be entered directly into the system, but where legacy systems exist to capture data it may be necessary to develop linkage solutions.

3.6. Data quality

> *“**Data quality** is defined as the completeness and validity of the data recorded in a public health surveillance system.” [23]*

Good data quality is when data is **up-to-date**, **accurate** (valid), **comprehensive** (complete; i.e. reported by all data sources) and **consistent** (for example, based on common case definitions). Poor surveillance data quality is an ever-present challenge which may lead to

inefficiency, **difficulties with data analysis** and **erroneous decision-making**. Potential problems include data duplicates, typos, data entered into the wrong fields, fakery, missing data, reporting delays or other issues. [23]

Maintaining good data quality requires a **commitment to data quality** among all those involved in providing and managing data. A clear **process** for data quality management with clearly assigned responsibilities is essential. **Training** of staff should give due attention to data quality.

It is important to **monitor** data quality at each stage of the data flow from the very beginning; errors are more difficult to correct the further down the line they are detected [24]. This is a key reason for providing **feedback and support** to surveillance data providers.

Data quality issues may arise when **excessive amounts of data are requested**, or when the **process of gathering or providing data is laborious**. It is important that **only essential data items** are requested. For many surveillance purposes, aggregated data is all that is required; case-based data should only be requested when this can be justified for public health purposes. Surveillance data collection **forms should be simple and easy to complete** [23].

In addition, periodic **data quality audits** should be undertaken to identify systemic issues [24] DHIS 2 has a number of features which aid data quality management. **Validation rules** can be set for data entered at district offices; pop-up messages appear to the user when the rules are violated. A **Data Approval** step can be introduced into the process; data is not included in analyses until signed off at a higher level [23]. The **Data Quality app** can run a number of procedures to identify data quality issues. DHIS 2 can produce a **Reporting Rate Summary** which summarises the reporting rates of data sets by different organisation units [20].

3.7. Ad hoc data queries or extracts

DHIS 2 can provide very flexible tabulation and visualisation of data within its analytical Web apps, which should be sufficient for most analytical purposes. CSV files can easily be downloaded from the system by the users for use in other software [20].

In some circumstances advanced users may wish to access data directly using **SQL database queries**. This can be done by the server administrator by directly logging into the database on the server and running SQL queries there but this would not ordinarily be necessary. SQL queries can be stored in DHIS 2 (**SQL Views**), with the results presented to authorised users through the Web API. The Web API also allows for data to be extracted directly for **analysis in statistical software** such as R [18, 20].

Standard reports can be designed in DHIS 2 which allow basic users to create formatted reports of specified tables or charts without requiring any analytical skills.

3.8. Basic principles of information security

> ***"Information security** is the practice of preventing unauthorised access, use, disclosure, disruption, modification, inspection, recording or destruction of information." [24]*

To protect **information security**, DHIS 2 will be configured to use an encrypted Web connection (known as HTTPS) to connect with users and the database will be encrypted with a strong password. DHIS 2 should also be hosted behind an appropriately configured firewall. Proper security protocols should be followed to make the system secure and protect the **personal identifiable data (PID)**.

Infectious disease surveillance data on individuals is often **sensitive** data (**personally identifiable data** or information; PID/PII). Patients who provide personal information to health facilities have a **right to privacy** and a reasonable expectation that their data will not be accessed or used inappropriately[25, 26]. Surveillance systems often depend on the **availability** of appropriate data and **the ability to share data** between different parts of a surveillance system. Storage and transfer of sensitive data involves numerous **risks** which must be managed appropriately; if not, serious **negative consequences** may ensue for an individual, for an organisation or for a surveillance system as a whole.

Risks in information security may arise from **people**, **processes** or **technology**. All three areas must be addressed; technology cannot fully compensate for risks from people or processes.

Technological solutions to protect information security include:

- Antivirus software
- Software updates
- Network firewalls
- Encryption of databases, communications and hard drives

- Authentication (confirming the identity of users at log in) and access control (limiting the data that users can access)
- Intrusion detection systems (systems for identifying hacking attempts on servers) [24, 26]

Information management processes should take account of the Caldicott principles [25]:

- Justify the purpose(s) of collecting PID.
- Do not use PID unless it is absolutely necessary.
- Use the minimum necessary PID.
- Access to PID should be on a strict need-to-know basis.
- Everyone should be aware of his/her responsibilities.
- Understand and comply with the law.[26]

Appropriate training, induction, policies, risk assessments and audits can help to ensure that the risks of data management processes are managed appropriately.

People are usually the weakest link for information security; **most information security incidents involve human error** [25, 26].

Common errors include:

- sharing of log in details
- weak passwords
- loss or theft of equipment
- use of insecure networks (e.g. personal email accounts such as Gmail)
- misdirected emails [24]

Simple measures people can take to protect information include:

- Log in details should never be shared.
- Passwords should be strong but not written down.
- Make sure software (particular antivirus software) is updated regularly.
- Ensure the physical security of data in paper form and computer equipment by locking it away when not in use.
- Be alert to risks, either from yourself or others.
- Do not use insecure email for transfer of sensitive data.[25]

CHAPTER 4: CASE DEFINITIONS OF PRIORITY DISEASES

4.1. Priority diseases

As resources are limited, countries usually develop a **list of priority diseases** for surveillance. Prioritization of communicable diseases for surveillance is usually based on the following considerations:

- **Burden** of disease and disease-associated mortality
- Potential for **outbreaks** or other serious consequences
- Relevance to national or international disease control **targets**
- Amenability to **intervention**
- Level of public, media or political **concern** [27]

Pakistan has selected the following diseases for the implementation of IDSR:

- acute haemorrhagic fever
- acute respiratory infection; severe acute respiratory infection
- **acute watery diarrhoea** (separately for ages < 5 and ≥ 5 years)
- diphtheria
- influenza-like illness
- measles
- Tuberculosis

4.2. Case definitions

The use of country-wide **standardised case definitions**, methods and tools facilitates the timely use of data at each level. As Pakistan has four provinces thus a consensus must be achieved on the already approved WHO case definitions [27, 28]. The government approved case definitions are shown in Table1.

Table 1 Government Approved Case Definitions

Indicator	Abbreviation	Case definition
Acute Haemorrhagic Fever	AHF	Acute onset of fever of less than three weeks duration in a severely ill patient with any two of the following: haemorrhagic or purpuric rash, nose bleed, blood in vomit/sputum/stool, other haemorrhagic symptoms; and no known predisposing factors for haemorrhagic manifestations
Acute Respiratory Infection	ARI	Any person presenting with severe pneumonia characterised by fever (temperature >= 38 degrees C) and one or more of the following: cough, sore throat or shortness of breath
Severe Acute Respiratory Infection	SARI	An acute respiratory infection with history of fever or measured fever of >=38 degrees C + cough + onset within the last 10 days, which requires hospitalisation
Acute Watery Diarrhoea	AWD	An illness characterised by three or more watery (non-bloody) stools within a 24 hour period and severe dehydration, in a person aged under five years
Diphtheria	Diphtheria	An upper respiratory tract illness with an adherent membrane of

		the nose, pharynx, tonsils or larynx
Influenza-Like Illness	ILI	An acute respiratory infection with measured fever of >=38°C + cough + onset within the last 10 days
Measles	Measles	Any person in whom a clinician suspects measles infection, or any person with fever and maculopapular (i.e. non-vesicular) rash and cough, coryza or conjunctivitis

4.3. Aggregated data

The System as mentioned and explained in the preceding chapters will only be collecting aggregate data. This aggregated data will help in assessing the disease burden of the country as well as it will help monitor the trends which in return will help the health professionals to predict and identify the disease outbreaks on time. This system will in return then help the response teams to generate appropriate responses against these outbreaks [29].

Below is a data collection form for the above mentioned priority diseases. The counts and deaths from each of these conditions will be routinely collected weekly by each health facility and reported to district health offices. This is referred to as **aggregated data** and is the simplest form of surveillance indicator data to collect [30].

Aggregated data is adequate for:

- Assessing the burden of disease
- Monitoring **trends**
- Identifying geographical variation
- Identifying **outbreaks** [29]

An example on-line form for collecting aggregated data is shown below.

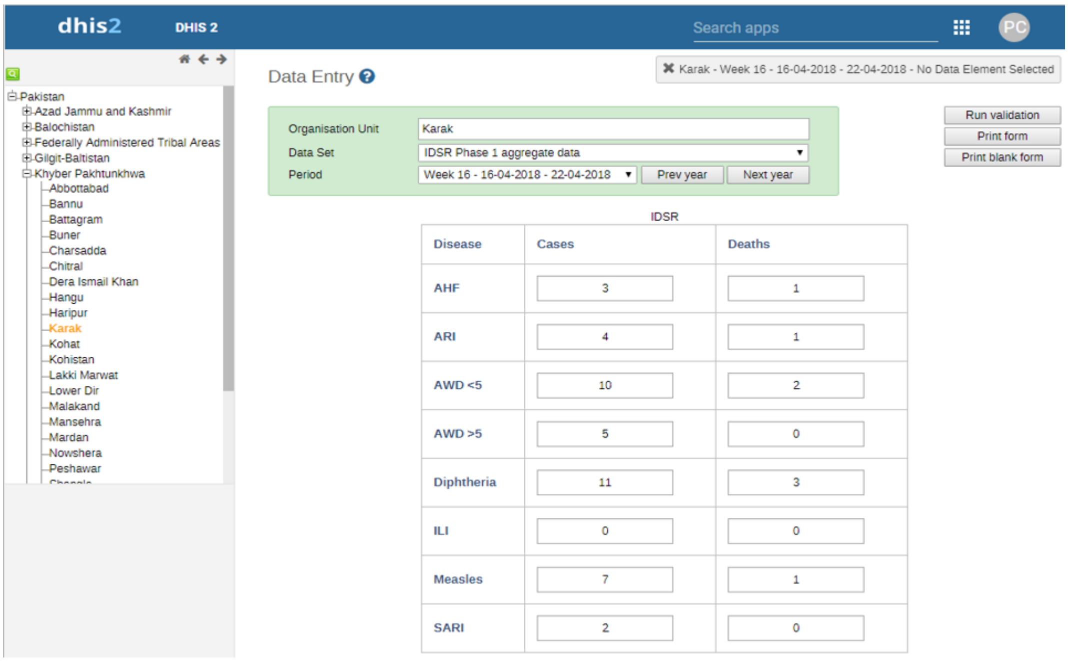

Figure 3 Simple example on-line data entry form for aggregated data

CHAPTER 5: PROPOSED SURVEILLANCE DATA FLOW

A hierarchically structured data flow linking the community, health facilities and public health practitioners at district, province and federal levels, with appropriate analysis at each level, allows the **public health response to be delivered at the lowest appropriate level**, while enabling **system-wide situational awareness** [27].

5.1. District health office role and responsibilities

District health offices have key responsibilities as part of maintaining the surveillance process:

- At the beginning of each week, **surveillance data collection forms** will be completed at all reporting health facilities, each containing aggregated numbers of cases and deaths of each priority disease for the preceding week, in accordance with the agreed case definitions (see list in previous chapter).
- Completed forms are then returned from each reporting health facility to district offices, where the data is checked for **data quality** and **entered into the online data entry system** (Proposed System) [23].
- Any health facilities which have not reported should be identified for appropriate action.
- The number of cases for any health facility should be examined for **unusually high numbers** with the help of data visualization tool of DHIS 2.
- After data entry, any **alerts** issued by DHIS 2 should be noted and appropriate actions taken.
- Further investigation and other actions (including collection of case-based data, collection of laboratory specimens, etc.) should be conducted as appropriate and advised. DHIS 2 has the capacity to collect case based data which is discussed in the future prospects.
- Appropriate **feedback** should be provided to health facilities providing data [27].
- District health offices may also need to coordinate with the **community** or other institutions or agencies at district level to raise awareness of priority diseases and

corresponding case definitions, to share information and to coordinate response activities.

Diseases requiring immediate reporting will be notified to the district by telephone initially [9].

Standard paper data collection forms, adapted from existing forms used by WHO, the DHIS/DHIS2 health informatics platform and/or legacy surveillance systems in Pakistan, will be used to capture the minimum data set for IDSR. It is important that health facilities submit routine data collection forms even when there are no notifiable cases or deaths to report (**nil reporting**) [31], as it is otherwise difficult to distinguish absence of disease from a failure in the system.

The following forms will be used at district level and below:

- **Immediate case-based reporting form** (to capture and share information from initial telephone call) [27, 28]
- **Weekly summary reporting form** for submission of routine surveillance data from health facilities to districts via **DHIS 2**
- **Laboratory reporting form** (where a public health laboratory is available) [7, 9, 10]

This on-line system will allow data entry for up to 14 days for a particular week, to allow for late reporting by health facilities, after which it becomes closed to data entry. However late reporting may lead to late recognition of outbreaks and incidents, and so **the majority of data should be reported at the earliest opportunity**.

The following priority diseases require **immediate reporting** [27] and collection of case-based data by district health offices and may require further actions as advised:

- Acute watery diarrhoea in a person aged five years or more
- Acute haemorrhagic fever

Other diseases which may require immediate reporting and collection of case-based data include [27, 28]:

- Measles
- Pertussis
- Diphtheria

District offices should also take **immediate action** for other diseases (including collection of case-based data) where the number of cases for a particular health facility triggers an automated alert in the on-line system.

District office surveillance staff will each be given access to the on-line data collection and analysis system of DHIS 2. District office users should log in to the on-line system each day to check for **messages** (including alerts) and to examine the specific **dashboard** for their district to identify adverse trends or geographical clustering of disease.

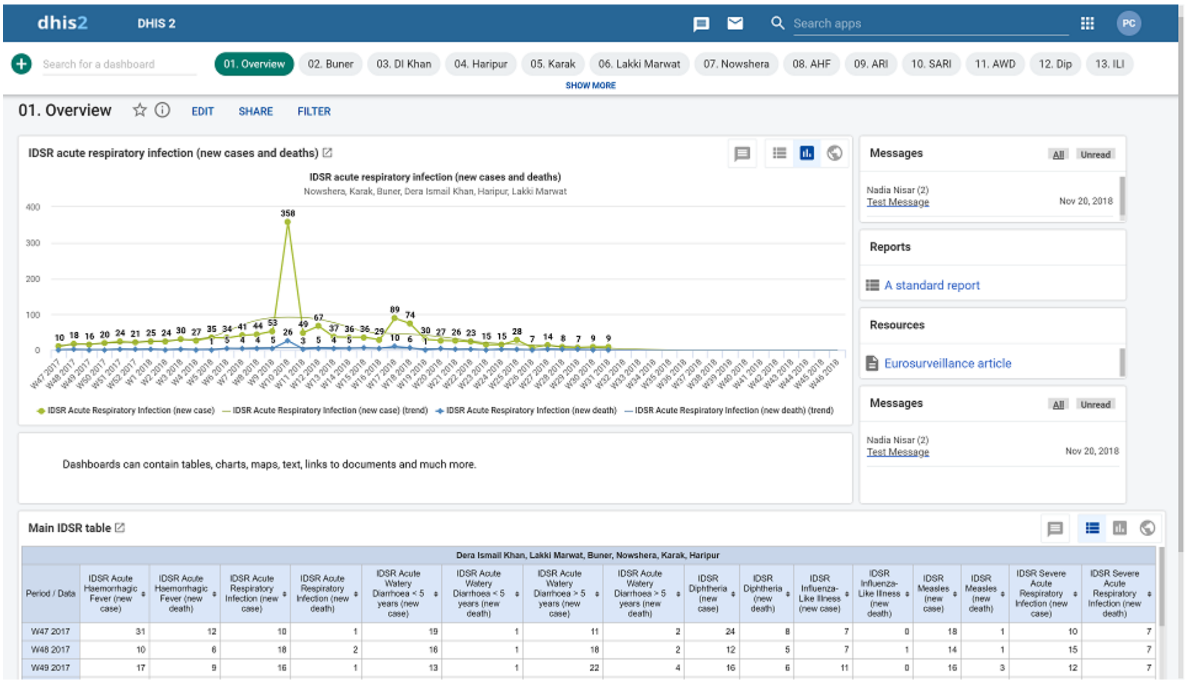

Figure 4 Example district dashboard

District offices which do not have a reliable Internet connection (but do have adequate mobile telephone coverage) are able to report data to the on-line system via an Android **mobile 'phone app**. The integration of Android Mobile app is also discussed in the future prospects of this system.

Information security of patient data is important at all levels of the system. District offices should store all patient-related identifiable data in paper form under lock and key, accessible only to surveillance staff. Computers used for data entry should require password login and should also be stored securely. **No identifiable patient data should be sent via insecure means** (e.g. via a personal email account) [24, 26].

Collection of laboratory specimens will follow standard protocols, including maintenance of the cold chain and correct labelling of specimens.

5.2. Provincial surveillance unit role and responsibilities

Provincial surveillance and response units have key responsibilities for overseeing and maintaining the surveillance and response process in their province:

- Monitoring **data quality** and taking appropriate actions to improve it
- Defining, monitoring for and responding to **signals in surveillance data** (e.g. alerts, adverse trends, clusters) at district level which could represent outbreaks or incidents
- Providing advice on response activities to districts
- Providing outbreak investigation capacity to districts
- Providing **feedback** to district offices in relation to data quality and surveillance analysis/interpretation, in the form of a routine **weekly surveillance bulletin** (and urgent outbreak alert communications where appropriate)
- Providing training, technical advice, supplies, surge capacity and other **support** to district offices
- **Liaison** with other province-level stakeholders

Appropriate **information security** measures will also be taken at province level [24, 26].

5.3. Federal surveillance unit role and responsibilities

The federal surveillance team at the Pakistan National Institute of Health (NIH) has key responsibilities at federal level for overseeing and maintaining the surveillance and response process [27]:

- **Oversight** of surveillance system functioning and data quality
- Defining, monitoring for and responding to **signals in surveillance data** (e.g. alerts, adverse trends, clusters) at province level which could represent outbreaks or incidents
- Providing advice on response activities to provinces
- Providing outbreak investigation capacity to provinces
- Providing **feedback** to provincial surveillance teams in relation to data quality and surveillance analysis/interpretation, in the form of a routine **weekly surveillance bulletin** (and urgent outbreak alert communications where appropriate)

- Providing training, technical advice, supplies, surge capacity and other **support** to provincial surveillance teams
- Defining and implementing changes to the surveillance system, in discussion with provincial stakeholders
- **Liaison** with other federal-level stakeholders and with reference laboratories
- IHR focal point arrangements [28]
- Server administration, DHIS 2 system maintenance and information security

Information is held in the on-line DHIS 2 surveillance system, which will be hosted at NIH. A range of robust information security measures will be in place.

5.4. Proposed timeline for reporting

The proposed timeline for routine data collection (which may be adjusted in some weeks to allow for public holidays) is:

- **Monday**: Health facilities complete routine **weekly data collection forms** (for data from the previous week up to midnight on Sunday) and return them to district offices
- **Tuesday**: Routine weekly data collection forms collated from health facilities; **data entered into the online system** at district offices; missing data followed up
- **Wednesday**: Standard tables, charts, maps and **dashboards** on on-line system examined at district, province and federal level for surveillance signals (alerts, adverse trends, geographical clusters) and appropriate **actions initiated**; weekly **surveillance teleconference** for each province; feedback to facilities as locally agreed
- **Thursday**: weekly provincial and federal **bulletins** prepared and circulated by provincial and federal surveillance teams; weekly **teleconference** of provincial teams with federal surveillance team

Note that diseases requiring immediate notification should be notified upwards to the next level on the **same working day** that the information is received.

Feedback to districts and provinces of results of routine surveillance analysis is available via dashboards at any time from the day following data entry. It is also feasible to provide access to dashboards for health facilities with Internet access.

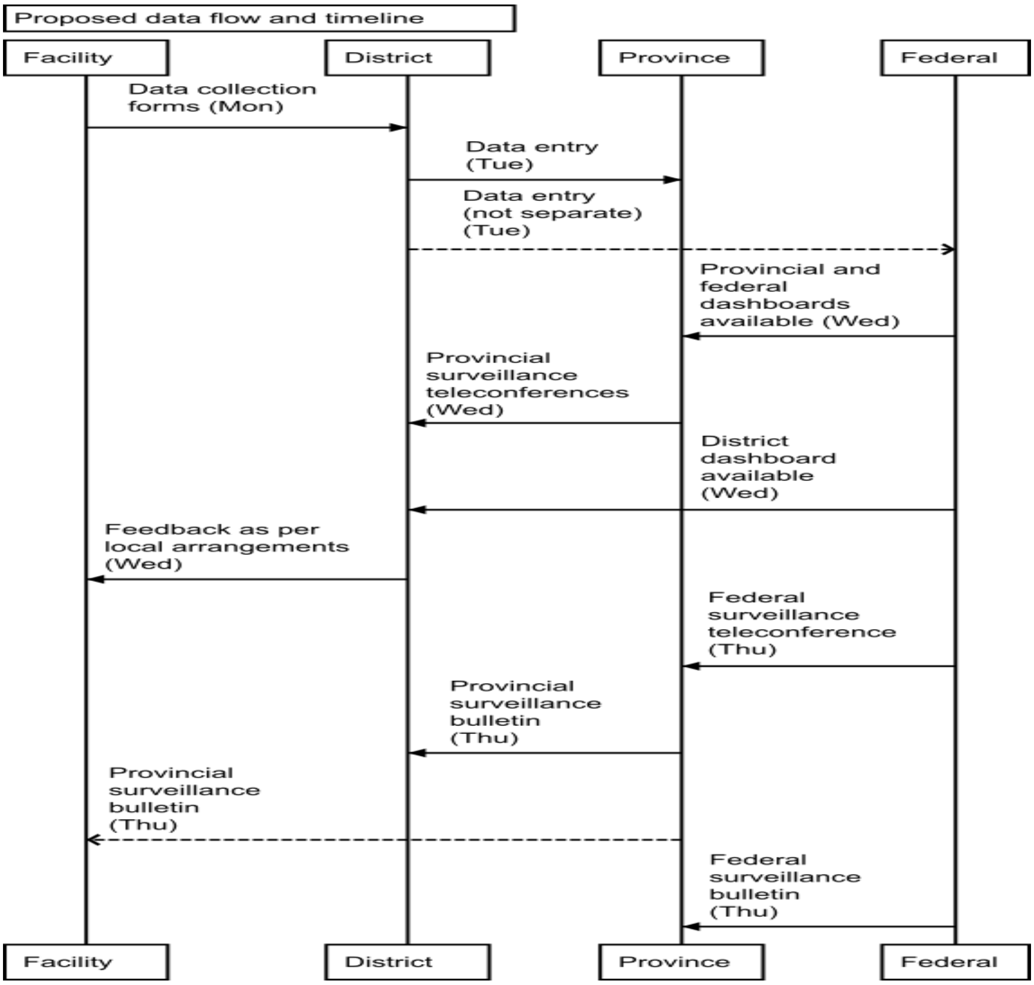

Figure 5 Proposed Data Flow and Timeline

5.5. Data management and the District Health Information System 2

Data management in an IDSR system is considerably facilitated by the appropriate use of **information technology**; analysis and detection of outbreaks can be considerably facilitated by the appropriate use of software.

Pakistan instituted Health Management Information System (HMIS) in the year 1992 in collaboration with USAID. After the devolution of Ministry of Health in the year 2001, in order to fulfill the needs of health management DHIS was introduced in Pakistan with the help of Japan International Cooperation Agency (JICA) [4]. Most provinces of Pakistan currently use the older version 1 of the District Health Information System (**DHIS**; sometimes more specific version numbers such as 1.3 or 1.4 are used) for collecting a range of different health data.

DHIS is based on Microsoft Access software and is now coming to the end of its lifespan. It lacks many of the features required for modern infectious disease surveillance.

A modern on-line health information system known as DHIS 2 (**District Health Information System version 2**), developed by University of Oslo, which has much greater functionality, is being deployed in different countries for disease surveillance and health information management. DHIS 2 has been customized to fit the needs of Pakistan [18, 19, 22]. DHIS 2 is a completely different tool to the older DHIS system, despite the similarity in names. DHIS 2 is an **open source** project, funded by a number of global donors. DHIS 2 software is therefore free to install and use. The software is **well supported** by developers and the community of users and is regularly **updated** and improved.

DHIS 2 is a rather **future-proof** solution for implementing IDSR in Pakistan and can be **scaled up to national level** with appropriate resource.

IDSR in Pakistan will use DHIS 2 for:

- **Data collection** (aggregated)
- Automated data **aggregation** (collating of data from a number of areas)
- Automated data **analysis** (production of tables, charts or maps; identification of possible outbreaks)
- Automated production of standard **reports** and presentation of Web-based **dashboards**

DHIS 2 will be hosted on a **central server** based at the Pakistan NIH and accessed by province and district level users via the Internet [20].

CHAPTER 6: USER RIGHTS AND PRIVILEGES

As we are aware that this medical data will be of immense important and we need to make all the necessary arrangements to make this data secure. We need to define the users from the highest level of hierarchy to the lowest level of hierarchy which means from Federal Level to Basic Health Unit Level. We can specify users at province or federal level that will have the permissions to **create new user accounts** for new users. This will usually be done **manually**. Users can only create new user accounts at **their own level and below**; for example, province level users can only create accounts for users within their own province. As described earlier, users can be assigned to particular roles or groups and given specific permissions for the data sets they can access and/or alter. Users can add further information about themselves (such as job title, qualifications, interests, languages spoken, birthday etc) within their **user profile**, which will be accessible to other users. The administrator of the IDSR system sitting at the federal level will have the rights to view and keep log of all the activities done by all the users. Users will be divided into a number of categories. Some are listed as under along with their roles and privileges

6.1. Users at District Level

Users at any level can be provided with log in details for DHIS 2 by providing the following information and appropriate authorisation:

- Contact details, including email address
- Organisation unit of user (district/province/federal)
- Data sets they are permitted to access
- Areas for which they can access data
- Whether they can enter, view, amend and/or analyse data
- Whether they should be given specific **user roles** or added to specific **user groups**

When a new user account is created on DHIS 2, the new user is given a “role” which describes what they can and cannot do. For example, a “district office” role will be created which allows a user to enter or amend data for their own district, but only view data for the rest of their province. System administrators responsible for maintenance of DHIS 2 would have a role which allowed them to change things across the whole system. Users may have more than one role. In general, those with a particular job title would receive a particular user role, which allowed them to do everything required for that job but no more.

By default, users of DHIS 2 are assigned to specific organisations (e.g. district offices) within the system. DHIS 2 “groups” allow users to be organised into non-geographical groupings. This means that in DHIS 2 messages can be sent to individual users, to specific organisations or to particular groups of users.

Users at district level will log in to the DHIS 2 Web app for weekly routine surveillance **data entry** or to access a range of reports, charts, maps and/or dashboards. The system is easy to use for data entry, requiring minimal training [19].

District level users can be trained to develop specific outputs for their own use or to share with others. District level users can also add comments (epidemiological **interpretations**) to specific charts, tables or maps and can also use the DHIS 2 system as a secure **messaging** system for reporting unstructured information [19].

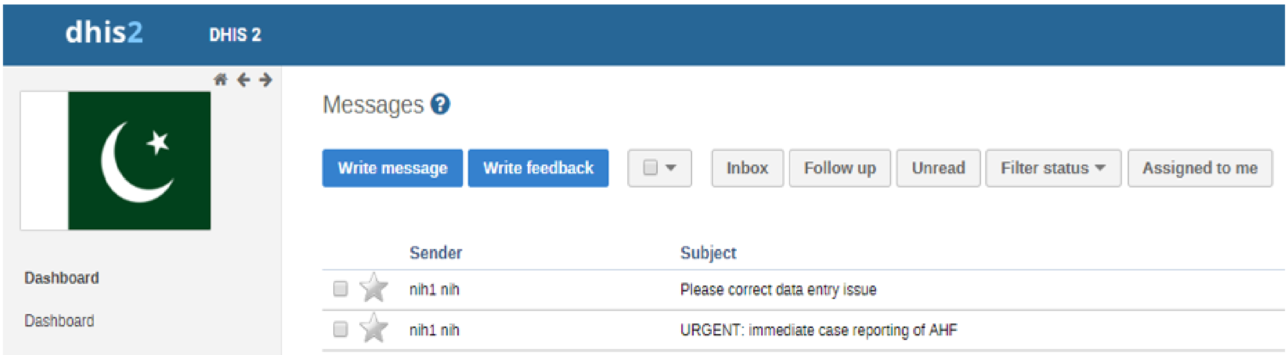

Figure 6 Messaging in DHIS2: receiving messages

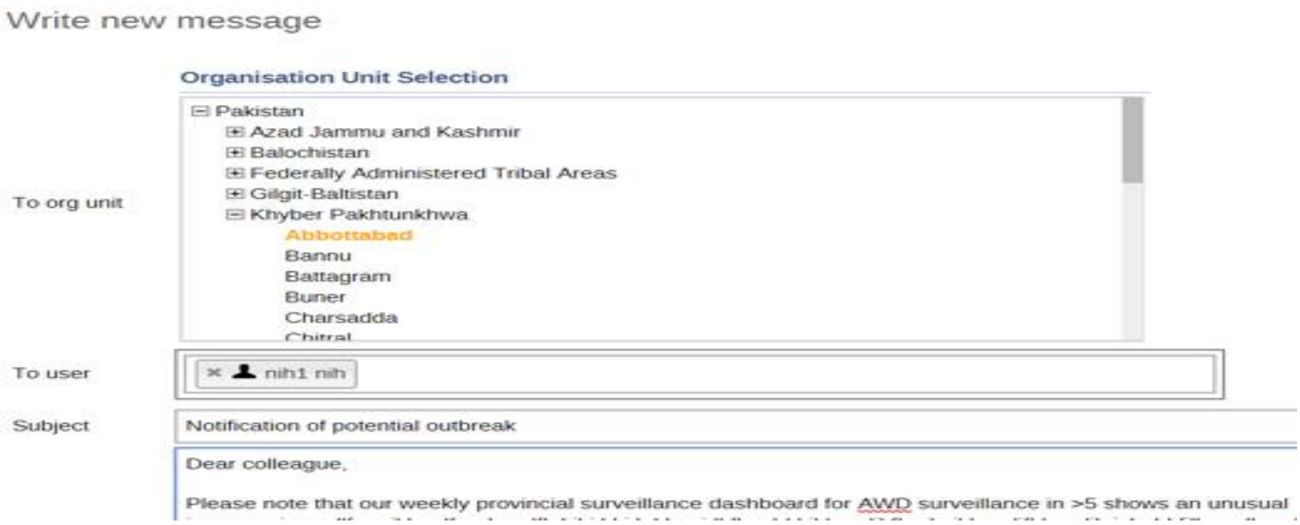

Figure 7 Messaging in DHIS-2: sending messages

6.2. Users at Provincial Level

Similar principles to those above for district users will apply to provincial users. However provincial users would not typically have responsibility for data entry and furthermore would typically have more extensive permissions to manage data for the whole province.

Provincial users may already have **historical data** (in an appropriate format) which they would wish to include in DHIS 2 analysis; this can be uploaded into the system by a user with system administrator permissions. Where provinces are collecting **additional indicator-based data** (such as priority diseases not included in the IDSR list), it would be feasible to incorporate additional data entry screens to enable this to be included in the system. Where provinces already have a satisfactory method of acquiring surveillance data [27], it may be possible to identify technical solutions to allow data to flow to DHIS 2.

By default, DHIS 2 routinely **updates** all tables, charts, maps and other outputs at midnight each day, incorporating any new data entered that day. The system can be configured to issue **messages** to specific users (within the system and/or via email) if certain conditions are met (such as a count of cases above a specified threshold in a particular area or a count of cases more than a given number of standard deviations above the mean) [19]; these are known as **surveillance rules** or trigger levels. Provincial users typically have the responsibility for defining specific surveillance rules in DHIS 2 [13, 27, 28].

Provincial users may have more advanced analytical skills than district level users. While the system provides a range of analytical capability, including production of dashboards with interactive charts, users with the appropriate skills can easily extract data for analysis in other spreadsheet or statistical software hence making the system **interoperable**. This can be done either by downloading **CSV** (comma-separated values) files from within the Web app or by connecting directly with DHIS 2 over the network (the **Web API**, or Advanced Programming Interface, which allows other software to directly extract data from DHIS 2). The system can produce a range of different user-specified tables and charts, which can easily be **exported** for inclusion in other documents.

6.3. Users at Federal Level

At NIH, surveillance and epidemiology users of DHIS 2 will participate in routine surveillance and epidemiology activities as described above. As DHIS 2 is hosted at NIH, therefore NIH also has the responsibility for server maintenance, system administration and security [20].

DHIS 2 is usually hosted on a networked server running the Linux operating system. This may be a physical server on your own premises or a virtual server in the “cloud”. Important responsibilities in server administration include: applying updates; monitoring usage of server resources; troubleshooting software or hardware problems; backing up important data; and identifying and addressing security issues.

The responsibilities for the implementation and maintenance of DHIS 2 at NIH are outlined in the next section, which is primarily aimed at advanced provincial and federal users.

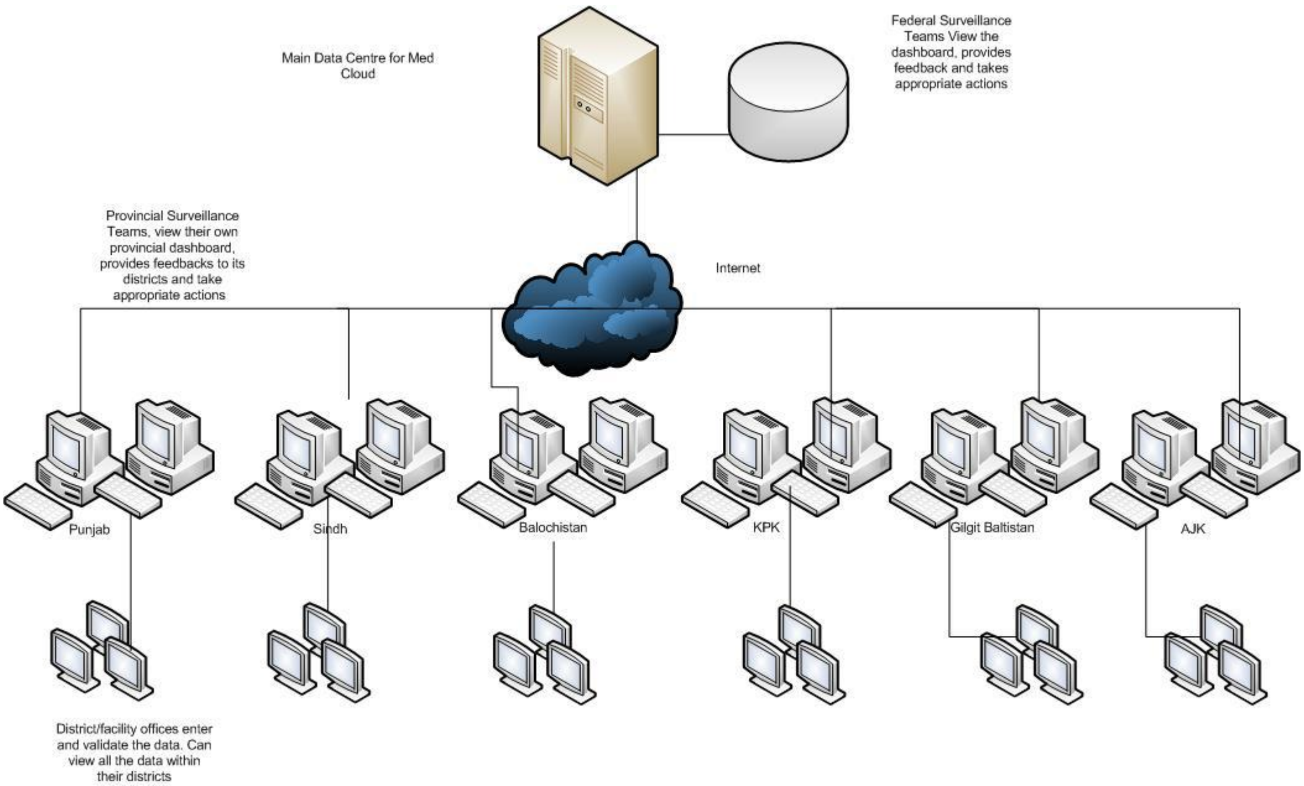

Figure 8 Hierarchy of Data Flow

6.4. Implementation and maintenance of DHIS 2

Implementation and maintenance of the DHIS 2 system includes several important tasks, including:

- **Installation** and updates (getting DHIS 2 up and running; updating to newer versions)
- **Metadata** configuration and maintenance [21, 32]
- **User** administration (e.g. creating user accounts and changing user passwords)
- **Acquisition of data** from other information systems

The implementation and maintenance of DHIS 2 at NIH requires both information technology and information management skills and experience; however, as **most maintenance of DHIS 2 is done through Web apps**, [20, 22] epidemiology and surveillance staff could also contribute to this after appropriate training.

CHAPTER 7: CONCLUSION AND FUTURE PROSPECTS

7.1. Conclusion

The implementation of health information system is a norm now a days. If compared with countries less developed than Pakistan like Tanzania, Kenya and other African countries [30, 33, 34], it is clear that we are much behind technologically and are still struggling to establish basic health information system. Currently in Pakistan most of the disease outbreaks are being reported by print or electronic media, while the government has no or very little information about the event. According to Ward, Brandsema et al. [14],the primary purpose of reporting specific infectious diseases is to trigger an appropriate public health response so that further illness can be prevented. This is impossible to achieve if the government relies on media. This can cause mass hysteria among the general public and also delays the response. It is also very difficult to filter out the hoaxes and wrong reporting.

This study explains a way to establish a disease surveillance system in Pakistan. The proposed system can be taken as the first step in implementing a much needed health information system in Pakistan. The system provides almost every feature of a modern day surveillance system and is easy to implement in Pakistan. As disease surveillance is part of National Institute of Health's objectives the idea of the proposed system was pitched in front of the management. The management showed keen interest in the proposed system and it is now being implemented in whole of the Pakistan for IDSR. This system can be scaled up to the country level. It is not only capable of running just aggregate data but also case based and event based surveillance can be done using this system. The next heading explains its further implementation and its expansion in future.

7.2. Future Prospects

The System as mentioned and explained in the preceding chapters will only be collecting aggregate data. This aggregated data will help in assessing the disease burden of the country as well as it will help monitor the trends which in return will help the health professionals to predict and identify the disease outbreaks on time. However for future we can shift the system from simple aggregate based data collection to the detailed Case based data collection.

Case-based data is the term often used to refer to data on individual cases, such as location, date of onset, age, sex and risk factors for infection. It can be recorded in a table called a **line listing**. It is often captured in outbreak investigations and may also be collected for surveillance purposes. It is more labour-intensive to capture case-based data, but it can be very useful for in-depth epidemiological analyses to answer specific questions for specific diseases or in specific situations [3, 7, 18, 22, 33]

Key case-based data to capture includes:

- Location (province, district, reporting health facility)
- Diagnosis
- Inpatient/outpatient
- Key dates: date seen, date of birth, date of onset of symptoms
- Patient identifiers
- Patient gender
- Patient address and contact details
- Patient vaccination history (where appropriate)
- Laboratory results
- Patient outcome (alive/dead/transferred out/lost to follow-up)
- Case classification (confirmed/probable/discarded)
- Date of notification
- Details of person providing information

The system also has capability to link to DHIS-2 Android Apps [33]. This can be done in future for case based or event based reporting. The field staff instead of carrying the laptops or collecting data on forms where network is not available can make use of the mobile app. The mobile devices can be provided by the federal government to the field staff and can be managed centrally by using Mobile Device Management (MDM) software like G Suite by Google. MDM will help secure the data and also monitor the field staff activities. MDMs can also help control the mismanagement of 4G mobile data and also can help with data security in case of lost or theft.

An example on-line form for collecting case-based data is shown below.

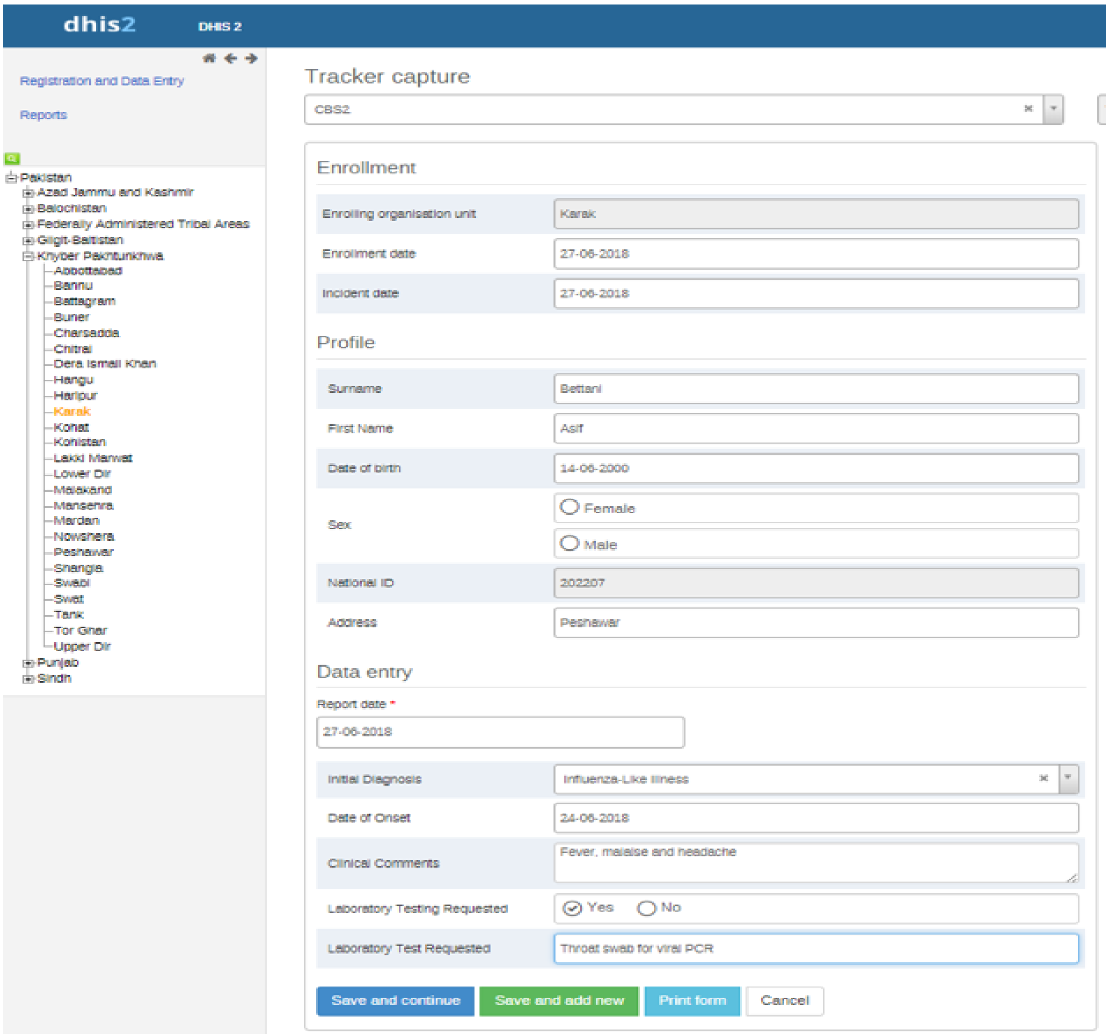

Figure 9 Simple example on-line data entry form for case-based data

REFERENCES

[1] D. Griggs, M. Stafford-Smith, O. Gaffney, J. Rockström, M. C. Öhman, P. Shyamsundar*, et al.*, "Sustainable development goals for people and planet," *Nature,* vol. 495, p. 305, 03/20/online 2013.

[2] H. Schmidt, L. O. Gostin, and E. J. Emanuel, "Public health, universal health coverage, and Sustainable Development Goals: can they coexist?," *The Lancet,* vol. 386, pp. 928-930, 2015.

[3] R. K. Phalkey, S. Yamamoto, P. Awate, and M. Marx, "Challenges with the implementation of an Integrated Disease Surveillance and Response (IDSR) system: systematic review of the lessons learned," *Health Policy and Planning,* vol. 30, pp. 131-143, 2013.

[4] J. I. C. Agency. (2001). *The District Health Information System (DHIS) Project for Evidence-Based Decision Making and Management*. Available: https://www.jica.go.jp/pakistan/english/activities/activity02_03.html

[5] K. Chandrasekar, *Use of Information Communication Technology (ICT) in communicable disease surveillance – a review of literature* vol. 2, 2012.

[6] S. Arshad, J. Iqbal, H. Waris, M. Ismail, and A. Naseer, *Health Care System in Pakistan: A review* vol. 3, 2016.

[7] K. Khowaja, "Healthcare Systems and Care Delivery in Pakistan," *JONA: The Journal of Nursing Administration,* vol. 39, pp. 263-265, 2009.

[8] L. Nakiire, B. Masiira, C. Kihembo, E. Katushabe, N. Natseri, I. Nabukenya*, et al.*, "Healthcare workers' experiences regarding scaling up of training on integrated disease surveillance and response (IDSR) in Uganda, 2016: cross sectional qualitative study," *BMC Health Services Research,* vol. 19, p. 117, 2019/02/13 2019.

[9] M. Domeika, G. Kligys, O. Ivanauskiene, J. Mereckiene, V. Bakasenas, B. Morkunas*, et al.*, "Implementation of a national electronic reporting system in Lithuania," *Eurosurveillance,* vol. 14, p. 19165, 2009.

[10] P. Rolfhamre, A. Jansson, M. Arneborn, and K. Ekdahl, "SmiNet-2: Description of an internet-based surveillance system for communicable diseases in Sweden," *Euro Surveill,* vol. 11, pp. 103-7, 2006.

[11] D. C. Miller and J. A. Thorpe, "SIMNET: the advent of simulator networking," *Proceedings of the IEEE,* vol. 83, pp. 1114-1123, 1995.

[12] M. Reinhardt, J. Elias, J. Albert, M. Frosch, D. Harmsen, and U. Vogel, "EpiScanGIS: an online geographic surveillance system for meningococcal disease," *International Journal of Health Geographics,* vol. 7, p. 33, July 01 2008.

[13] J. Cinnamon, C. Rinner, M. D. Cusimano, S. Marshall, T. Bekele, T. Hernandez*, et al.*, "Online Map Design for Public-Health Decision Makers," *Cartographica: The International Journal for Geographic Information and Geovisualization,* vol. 44, pp. 289-300, 2009.

[14] M. Ward, P. Brandsema, E. van Straten, and A. Bosman, "Electronic reporting improves timeliness and completeness of infectious disease notification, The Netherlands, 2003," *Euro Surveill,* vol. 10, pp. 27-30, Jan 2005.

[15] J. S. Brownstein, C. C. Freifeld, B. Y. Reis, and K. D. Mandl, "Surveillance Sans Frontieres: Internet-based emerging infectious disease intelligence and the HealthMap project," *PLoS Med,* vol. 5, p. e151, Jul 8 2008.

[16] W. L. Vlieg, E. B. Fanoy, L. van Asten, X. Liu, J. Yang, E. Pilot*, et al.*, "Comparing national infectious disease surveillance systems: China and the Netherlands," *BMC Public Health,* vol. 17, p. 415, May 8 2017.

[17] M. Santillana, A. T. Nguyen, T. Louie, A. Zink, J. Gray, I. Sung*, et al.*, "Cloud-based Electronic Health Records for Real-time, Region-specific Influenza Surveillance," *Scientific Reports,* vol. 6, p. 25732, 05/11/online 2016.

[18] F. K. Bjørn-Ingar Bergum, Magnus Øren, Vetle Falch, Tofik Sahraoui, " Analyzing Dhis2 as an Information Infrastructure," University of Oslo2015.

[19] U. o. O. UiO. (n.d.a, 8 November 2018). *DHIS2 Overview*. Available: https://www.dhis2.org/overview

[20] U. o. O. UiO. (n.d.a, April 2018). *DHIS 2 Implementer guide*. Available: https://docs.dhis2.org/master/en/implementer/html/dhis2_implementation_guide.html

[21] NISO-Press, *Understanding Metadata*, 2001.

[22] A. G. Renu Garg, "District Health Information System (DHIS2) Software in India," *Advances in Computer Science and Information Technology,* vol. Volume 2, pp. 39-42, April-June, 2015 2015.

[23] K. P. Fadahunsi, J. T. Akinlua, S. O'Connor, P. A. Wark, J. Gallagher, C. Carroll*, et al.*, "Protocol for a systematic review and qualitative synthesis of information quality frameworks in eHealth," *BMJ Open,* vol. 9, p. e024722, 2019.

[24] S. S. Fern Greenwell, *Health Information System Strengthening: Standards and Best Practices for Data Sources*: University of North Carolina at Chapel Hill, 2018.

[25] A. Greenough and H. Graham, "Protecting and using patient information: the role of the Caldicott Guardian," *Clinical Medicine,* vol. 4, pp. 246-249, May 1, 2004 2004.

[26] J. M. Appari A, "Information security and privacy in healthcare: current state of research.," *International journal of Internet and enterprise management,* vol. 6, pp. 279-314, 2010.

[27] WHO. (n.d.a, 2 July 2019). *WHO Recommended Surveillance Standards*. Available: https://www.who.int/csr/resources/publications/surveillance/whocdscsrisr992.pdf

[28] WHO, *International Health Regulations*, 3 ed.: World Health Organization, 2016.

[29] W. Raghupathi and V. Raghupathi, "Big data analytics in healthcare: promise and potential," *Health information science and systems,* vol. 2, pp. 3-3, 2014.

[30] J. M. Kariuki, E.-J. Manders, J. Richards, T. Oluoch, D. Kimanga, S. Wanyee*, et al.*, "Automating indicator data reporting from health facility EMR to a national aggregate data system in Kenya: An Interoperability field-test using OpenMRS and DHIS2," *Online journal of public health informatics,* vol. 8, pp. e188-e188, 2016.

[31] T. R. Talbot, D. W. Bratzler, R. M. Carrico, D. J. Diekema, M. K. Hayden, S. S. Huang*, et al.*, "Public Reporting of Health Care–Associated Surveillance Data: Recommendations From the Healthcare Infection Control Practices Advisory Committee," *Annals of Internal Medicine,* vol. 159, pp. 631-635, 2013.

[32] T. Guardian. (n.d.). *What is Meta data?* Available: https://www.theguardian.com/commentisfree/2014/aug/06/explainer-what-is-metadata-should-i-worry-about-mandatory-data-retention

[33] L. Pascoe, J. Lungo, J. Kaasbøll, and I. Koleleni, *Collecting Integrated Disease Surveillance and Response Data through Mobile Phones*, 2012.

[34] J. Karuri, P. Wagacha, D. Ochieng, and A. Manya, *DHIS2: The Tool to Improve Health Data Demand and Use in Kenya* vol. 8, 2014.

Appendix 1: Disease Reporting Form

IDSR Facility Weekly Report

Report to be submitted by each health facility/reporting site to District Surveillance Officer every Monday.

Indicator	New cases	New deaths	Case definition
AHF	[Acute Haem	[Acute Haem	Acute onset of fever of less than three weeks duration in a severely ill patient with any two of the following: haemorrhagic or purpuric rash, nose bleed, blood in vomit/sputum/stool, other haemorrhagic symptoms; and no known predisposing factors for haemorrhagic manifestations
ARI	[Acute Respi	[Acute Respi	Any person presenting with severe pneumonia characterised by fever (temperature >= 38 degrees C) and one or more of the following: cough, sore throat or shortness of breath
AWD < 5	[Acute Water	[Acute Water	An illness characterised by three or more watery (non-bloody) stools within a 24 hour period and severe dehydration, in a person aged under five years
AWD >=5	[Acute Water	[Acute Water	An illness characterised by three or more watery (non-bloody) stools within a 24 hour period and severe dehydration, in a person aged five years or older
Diphtheria	[Diphtheria (n	[Diphtheria (n	An upper respiratory tract illness with an adherent membrane of the nose, pharynx, tonsils or larynx
ILI	[Influenza-Lik	[Influenza-Lik	An acute respiratory infection with measured fever of >=38°C + cough + onset within the last 10 days
Measles	[Measles (nev	[Measles (nev	Any person in whom a clinician suspects measles infection, or any person with fever and maculopapular (i.e. non-vesicular) rash and cough, coryza or conjunctivitis
SARI	[Severe Acut	[Severe Acut	An acute respiratory infection with history of fever or measured fever of >=38 degrees C + cough + onset within the last 10 days, which requires hospitalisation

Appendix 2: Government Approved List of Priority Diseases

Government of Pakistan
Ministry of National Health Services, Regulations & Coordination
LG&RD Complex, Sector G-5/2, Islamabad
****

Islamabad, the 14th December 2017

NOTIFICATION

F. No. 1-31/Misc/FEDSD/2017 The competent authority, Ministry of National Health Services, Regulations and Coordination (NHS,R&C), Islamabad is pleased to notify the following Priority Diseases alongwith Zoonotic Priority Diseases for surveillance and Response in Pakistan:

Priority Diseases:

S. No	Diseases
01.	Tuberculosis
02.	Measles
03.	Hemorrhagic fevers (Including CCHF)
04.	Hepatitis B and C
05.	Malaria
06.	Polio and AFP
07.	Dengue/ Dengue Hemorrhagic Fevers/ Dengue Shock Syndrome
08.	Severe Acute Respiratory Infections (SARI)
09.	Cholera
10.	Diarrhea outbreaks
11.	HIV/AIDS
12.	Enteric Fever
13.	Neonatal Tetanus
14.	Rabies
15.	Meningococcal Meningitis
16.	Hepatitis A, E, and acute unspecified
17.	Anthrax
18.	Pertussis
19.	Diphtheria
20.	Bacterial Meningitis (unspecified, not meningococcal)
21.	Viral Meningitis
22.	Cutaneous Leishmaniasis
23.	Gonorrhoea
24.	Nosocomial Infections (Surgical site infection, neonatal sepsis)
25.	Encephalitis (Japanese, Unknown aetiology, arbovirus)
26.	Plague
27.	Mumps
28.	Botulism
29.	Leprosy
30.	Congenital Rubella Syndrome
31.	Syphilis
32.	Brucellosis
33.	Visceral Leishmaniasis

Cont'd------P/2

-2-

Zoonotic Priority Diseases:

Sr. No	Diseases
01.	Influenza
02.	Brucellosis
03.	Rabies
04.	Salmonellosis
05.	Crimean Congo Hemorrhagic Fever (CCHF)
06.	Anthrax

(Dr. Malik Muhammad Safi)
Director (Programmes)

Distribution:

1. Secretary, Primary & Secondary Healthcare Department, Government of the Punjab, Lahore
2. Secretary, Department of Health, Government of Sindh, Karachi
3. Secretary, Department of Health, Government of KPK, Peshawar
4. Secretary, Department of Health, Government of Balochistan, Quetta
5. Secretary, Department of Health, Government of AJK, Muzaffarabad
6. Secretary, Department of Health, Government of Gilgit-Baltistan, Gilgit
7. Secretary, Social Sector, FATA, Peshawar
8. Director General, Department of Health, Government of the Punjab, Lahore
9. Director General, Department of Health, Government of Sindh, Hyderabad
10. Director General, Department of Health, Government of KPK, Peshawar
11. Director General, Department of Health, Government of Balochistan, Quetta
12. Director General, Department of Health, Government of AJK, Muzaffarabad
13. Director General, Department of Health, Government of Gilgit-Baltistan, Gilgit
14. Director Health, FATA, Peshawar
15. NPM, Expanded Programme on Immunization (EPI), Prime Minister's National Health Complex, Islamabad
16. Director, Directorate of Malaria Control Project, NIH, Islamabad
17. NPM, National AIDS Control Programme, NIH, Islamabad
18. Programme manager, TB Control Programme, Prime Minister's National Health Complex, Islamabad

Copies:

1. Sr. PS to Secretary, M/o NHS,R&C, Islamabad
2. PS to Additional Secretary, M/o NHS,R&C, Islamabad
3. SPS to Executive Director, National Institute of Health, Islamabad
4. PS to Director General (Health), M/o NHSR&C