

Supporting evidence-based decision making at the district level in India:

A prototype for conducting spatial analysis in disease surveillance, based on the relationship between health events and facilities in the DHIS2 platform

Magnus W. Østeng



Master's thesis
Informatics: Programming and Networks
60 credits

Department of Informatics
UNIVERSITY OF OSLO

Spring 2018

Supporting evidence-based decision making
at the district level in India:

*A prototype for conducting spatial analysis in
disease surveillance,
based on the relationship between health events
and facilities in the DHIS2 platform*

Magnus W. Østeng
Spring 2018

© Magnus W. Østeng

2018

Supporting evidence-based decision making at the district level in India:
*A prototype for conducting spatial analysis in disease surveillance, based on the
relationship between health events and facilities in the DHIS2 platform*

Magnus W. Østeng

<http://www.duo.uio.no/>

Trykk: Reprosentralen, Universitetet i Oslo

Abstract

The District Health Information systems 2 (DHIS2) is a highly flexible and customizable platform for building health management information systems, for many developing countries. The DHIS2, by being a platform, supports innovation by third-party app developers, thus facilitating the targeting of more specific challenges upon an already robust platform. This thesis will cover the design, development, and testing of a web-based application called “Facility Burden Prototype” (FBP) as an extension onto the DHIS2 platform.

The research objective of the thesis is to determine to what extent and how the presentation of a visual relationship between cases of health events or diseases (e.g. reported patients with a specific disease) and the originating health facility, affects decision making, and how it should be applied at the district level in India. The decision problems covered in this thesis are *1) How to establish the existence of a possible disease outbreak? 2) Whether to monitor, investigate or control a possible outbreak?* The DHIS2, in combination with the FBP, will be assessed as a case- and syndromic surveillance system, considering strength and weaknesses in the context of the decision problems.

The analysis of requirements that goes into the design and development of the FBP have roots in the information cycle, and are split into five branches based a methodology within action research, called Multiview. The main requirements from these branches cover:

- Concepts and terminology must match the user level
- Pursue transparency in the process of generating information, through the user interface, for reducing any mistrust or concerns.
- Data should be organized and limited to “need only data”.
- Data should be comparable, thus have a context.
- Possibility for disaggregation of data.
- Ill-structured spatial decision problems should be supported by flexibility in the GIS.

As descriptive epidemiology, which is considered to be a working method of the end-users, concerns an area and not a set of health events derived from a facility, datasets in the FBP must be combined carefully to support outbreak detection in a geographical area. This makes the tracked entity-facility model a complementary model, focusing on an assessment of facility disease burdens, rather than the dominating model for establishing the existence of outbreaks.

The FBP arguably has the potential of facilitating a close feedback loop between the district and sub-district/facility level – A feedback loop that could provide district-level decision makers with complementary information that addresses the decision problems of concern. Also, the FBP, by displaying a visual relationship between health events and facilities, backed by a GIS with high spatial resolution, has the potential of targeting field investigations by having a reference to a residential location and enrolled facility. The result is a high level of spatial traceability that may lead to less resource-consuming investigations.

The modeling of the relationship between health events and the originating facilities also shows a potential of reducing the cognitive and psychological workload of the end-users, thus the strength of visualizing the relationship.

Preface and acknowledgements

The two last years at the University of Oslo (UO), ending with the submission of my Master's thesis, marks the end of a long journey of education. Despite the end of education, the learning will continue, as especially the field of informatics and software development are areas in rapid change and development. A challenge that the UO has made me able to meet.

The study of my Master's thesis has been very interesting but demanding. Working part-time as an IT consultant along with the Master's thesis implicates prioritizing, long nights and effective working. The thesis, in general, involves acquiring knowledge in multidisciplinary fields, covering informatics, public health, epidemiology, and psychology; some of which I had little or no prior knowledge or experience. From a technical perspective, the study of the DHIS2, a comprehensive health management information system, combined with new technologies within web development (Angular 2), contributed to the development of a totality of my skillset from a practical point of view.

At the end of 2nd semester 2017, me and my co-developer of the prototype, Nirujah Kirpanithy, conducted a field trip to visit our partner at the project, HISP India, in New Delhi India. The trip was a good contribution to both the development of the thesis, but also for getting insight into health issues of a developing country. We would like to thank the HISP India team, especially Harsh Atal and Arunima Mukherjee, for the hospitality and cooperation. I would also like to thank my supervisor Sundeep Sahay at the Institute of informatics at UO, for help with the conceptual aspects of my thesis, and also for the much-appreciated feedback and support. At last, I would like to thank my girlfriend Lene Aksnes for her understanding and support throughout the thesis.

All the errors in the thesis are the sole responsibility of the author.

Table of contents

1	Introduction.....	1
1.1	General introduction	1
1.2	The research question.....	3
1.2.1	Scoping of the research question.....	3
1.2.2	Scoping of the thesis.....	4
1.3	Theoretical starting point	4
1.3.1	Operational definitions	5
1.3.2	Health Information systems	6
1.3.3	Disease Surveillance Systems.....	7
1.3.4	Geographical Information Systems	7
1.3.5	The Information Cycle	8
1.4	The motivation for the research	9
1.5	Structure of the thesis.....	11
2	Context of research	13
2.1	Introduction.....	13
2.2	The DHIS2 platform.....	13
2.3	Disease surveillance systems	15
2.3.1	Case surveillance systems	15
2.3.2	Syndromic surveillance systems	16
2.3.3	Characterizing an outbreak in a disease surveillance system.....	17
2.4	Geographical information systems	18
2.4.1	GIS for spatial decision making.....	18
2.4.2	GIS Components	19
2.4.3	GIS functionality in public health	20
2.4.4	GIS in public health.....	21
2.5	Information use.....	24
2.5.1	Aspects of Decision making	24
2.6	Data interpretation	27
2.7	Data analysis and presentation	29
3	Methods.....	33
3.1	Introduction.....	33

3.2	Action research	33
3.2.1	Background.....	33
3.2.2	Characteristics of action research.....	34
3.2.3	Action research in IS	37
3.2.4	Adoption of method.....	40
3.3	Data sources	41
3.3.1	Field trip: 30. April – 16. May 2017	41
3.3.2	Field journal	42
3.3.3	E-mail.....	42
3.3.4	Meetings	43
3.3.5	Books and articles.....	43
3.4	The research processes – use of the method	43
3.4.1	Phase 1: Planning.....	44
3.4.2	Phase 2: Development.....	46
3.5	Validity	48
3.5.1	Internal validity	48
3.5.2	External validity	49
3.5.3	Reliability	49
3.6	Evaluation and reflection	50
4	Results	51
4.1	Requirement analysis	51
4.1.1	Introduction	51
4.1.2	Human activity analysis	51
4.1.3	Information analysis	54
4.1.4	Socio-technical analysis and design	56
4.1.5	Human-computer interface design	58
4.1.6	Technical design	59
4.1.7	Summarized requirements	63
4.2	The Facility Burden Prototype	66
4.2.1	Introduction	66
4.2.2	The development environment.....	66
4.2.3	Creating the prototype	68
4.2.4	Feedback and testing.....	77

5	Discussions	79
5.1	Introduction.....	79
5.2	Conceptual evaluation of the tracked entity – facility model	79
5.2.1	Decision problem 1: How to establish the existence of a possible disease outbreak?	79
5.2.2	Decision problem 2: Whether to investigate, control or monitor a possible outbreak?	81
5.3	Data quality	84
5.4	Presentation and Interpretability of information	85
6	Conclusion.....	89
6.1	Introduction.....	89
6.2	Findings.....	89
6.3	Contributions to research	90
6.4	Future directions.....	91
6.4.1	Domain specific application	91
6.4.2	Integrated part of the DHIS2 core	92
	Bibliography.....	93
	Appendix.....	97
	Appendix A - Feedback from HISP India	97
	Appendix B - User documentation	99

List of figures

Figure 1: Illustrates the dependencies within the information cycle, ranging from data collection to information use (Braa and Sahay, 2012; p.248)	9
Figure 2: Displays an example DHIS2 dashboard, combining several outputs like graphs, bar charts, tables, etc.	15
Figure 3: Heuristic Decision-making Procedure (Samitsch, 2015; p.17).....	24
Figure 4: Outlines the activities in an assessment process (Keppel and Freedman, 1995; p.6).	26
Figure 5: Factors that may influence decision-making efficiency (Samitsch, 2015; p.23)	27
Figure 6: Shows different organization of data presented to health care consumers in Hibbard and Peters experiments. The lower table summarized the data by displaying stars, which makes the consumers tend to choose Plan B (Hibbard and Peters, 2003;p.421)	28
Figure 7: Illustrating a cycle in action research. Originally from Gerald Susman, adapted by Rory O'Brien (O'Brien, 1998)	36
Figure 8: Shows a model of evolutionary prototyping originally from Connell and Shafer (1989) (Baskerville and Wood-Harper, 1998; p.99).....	40
Figure 9: Shows the adaption of IS action research methods, including Canonical Action Research, Multiview and IS prototyping. Multiview is used for initial planning, while a combination of IS prototyping and Canonical Action Research is used for the cyclic development of the FBP.	41
Figure 10: Description of the overall project process, from project start to final prototype.....	44
Figure 11: This final draft was used to model the functionality and graphical user interface for the development process.....	45
Figure 12: A small section of the use case diagram in this thesis, for illustrating the look of a use case diagram.....	47
Figure 13: Showing the excel sheet used for overall sprint planning and time management.....	47

Figure 14: Shows the dashboard on Trello.com, a project management tool for tracking tasks. The columns display tasks that are not started, in progress, finished and rejected tasks.	48
Figure 15: The figure shows relative emphasis of investigating and control efforts based on the causative agent and source/transmission mode (Goodman et al., 1990; p.11). The + symbols indicate the level of importance in the given situation.	53
Figure 16: Displays types of surveillance systems and associated analysis types and methods. (Lee, 2010; p.91).....	60
Figure 17: Shows the final use case diagram, summarizing all the functional requirements.....	64
Figure 18: Shows the final product as a screenshot from a web browser.....	69
Figure 19: Use case X1, select time interval.....	71
Figure 20: Use case X2, pick organization unit.....	71
Figure 21: Use case X4, pick program.	71
Figure 22: Use case Z1, view entity distribution.	72
Figure 23: Use case Z3, view organization units.	73
Figure 24: Use case X6, disaggregate data. Note that the age set in the filter determines what age range that will be included, not excluded.....	73
Figure 25: Use case Z4, View facility - tracked entity relationship.	74
Figure 26: Use case X7, display datasets. This example shows two selected datasets in the app. Each with different color code.	74
Figure 27: Use case Z2, show clustering.....	75
Figure 28: Use case Y1, view enrollment frequency. Shows a line graph of enrollments.	76
Figure 29: Use case Y2, view total enrollments.....	77

List of tables

Table 1: Suggested DHIS2 dashboard design guidelines by Jespersen (Jespersen, 2017; p.71).	31
Table 2: Table comparing characteristics of canonical action research with other forms of action research in IS (Baskerville and Wood-Harper, 1998; p.96).	38

1 Introduction

The research objective of the thesis is to determine to what extent and how, the presentation of a visual relationship between cases of health events or diseases (e.g. reported patients with a specific disease) and the originating health facility, affects decision making, and how it should be applied at the district level in India. The objective will be addressed through the design, development, and testing of a web-based application called “Facility Burden Prototype” (FBP). The FBP is thus instrumental to provide for the visualization to help assess how this visual relationship may contribute to supporting evidence-based decision making.

1.1 General introduction

At the end of the reign of Apartheid in the mid-nineties, the South African society experienced a high level of racial fragmentation, which was reflected in their national Health Systems and their supporting Health Information Systems (HIS). Lack of standardization of data collection tools, data definitions, etc. resulted in high integration barriers both horizontally and vertically, across organization units, health programmes, and administrative levels. Data was collected and pushed upstream for decision making at higher levels, resulting in a centralized design with little or no useful data left for action at the district level. South African public health activists and Scandinavian information systems developers, made a successful attempt to address the situation by a series of participatory design practices, organized within the Health Information Systems Programme (HISP), a network aiming to strengthen HIS building in developing countries coordinated by the University of Oslo (UIO). The HISP efforts led to the development of an open-source software, enabling information for local action. This software is known as the District Health Information systems (DHIS), developed with an emphasis on flexibility and decentralization (Braa et al., 2012).

The DHIS software is today still maintained and organized by the HISP network, and has since its initial approach evolved into a more comprehensive open-source platform enabling more powerful features in relation to collection, processing and presentation of aggregated and case-based data within multiple vertical programs

(University of Oslo, 2011). The current version, DHIS2, has become a highly flexible and customizable platform to target a range of developing countries and their needs for individual customization. One of the advantages of platformization involves the increased functionality and innovation by third-party app developers, which facilitates the targeting of more specific challenges upon an already robust platform. Third-party developed apps on the platform will be referred in this thesis to as “DHIS2 apps”, while the rest of the platform is discussed as “DHIS2 core”.

My Master’s thesis at UIO is rooted in the DHIS2 platform, as a third-party application that enhances the platform and thus increases the value for a targeted audience. The project is a participatory research-project with HISP India, a long-standing partner of UIO in the network. HISP India is a multidisciplinary team specialized in public health informatics, covering informatics, public health, and implementation (HISP India, 2018). Together with HISP India, which is the leading partner at the project, me and my co-developer of the FBP Nirujah Kirpanithy, aimed to research ways for improving evidence-based decision making in a disease surveillance context in India. The approach aimed to create a DHIS2 app (called FBP) for making a descriptive analysis of case-based data. Case data is traditionally collected on a census, counting health events or disease-specific cases. The level of granularity and number of attributes will vary between case definitions; however, my thesis depends on a descriptive analysis of cases in terms of place, time and person. This means that the system will be based upon case-data records rather than aggregated data. A primary focus will be on the “place” component, which will be conveyed using a Geographical Information System (GIS), displaying spatial distribution of cases on a map.

The flexibility of DHIS2 enables us to capture data representing a person, referred to as a tracked entity. The FBP derives the case data of interest from the tracked entities. The relationship between facilities (e.g. a hospital where the tracked entities are enrolled) and tracked entities is a key component in the research, assessing how and in what way the relationship may support evidence-based decision-making in both case-based surveillance and syndromic surveillance systems. Syndromic surveillance systems are most commonly designed to detect potential outbreaks faster, by uncovering clusters from a real-time data stream of syndromes, rather than confirmed cases of diseases.

Note that a visual demonstration of the relationship is not supported by the DHIS2 core at the time of writing, and will be a central component in the development of the FBP.

1.2 The research question

The research question for this thesis:

To what extent does a visual depiction of the relationship between cases of diseases or health events and originating facilities strengthen analytic capabilities in disease surveillance, for improved evidence-based decision making in the context of public health at district level India?

The “extent” will not be measured as a quantitative measure, but rather discussed and elaborated in the background of relevant research and literature of e.g. surveillance systems with transferable properties.

1.2.1 Scoping of the research question

As there was no consensus from the start regarding what kind of disease surveillance systems the FBP should support, the model will be incorporated and discussed in the context of case surveillance and syndromic surveillance systems. However, the application of the model is not necessarily limited to these systems. The model can be applied to all tracked entity configurations that hold a spatial property and their link to an enrolled facility. However, the utility is dependent on how the data structure is configured and for what purpose of the model is, in each situation.

The real-time capabilities will be originating from the facility-tracked entity relationship built on the DHIS2 platform. To research the utility of this tracked entity-facility model, two approaches has been derived for potential use of the model to support evidence-based decision making. More specifically, the kind of decision problems that health officials at the district level in India commonly encounter:

Decision problem 1: How to establish the existence of a possible disease outbreak?

Decision problem 2: Whether to monitor, investigate or control a possible outbreak?

The previous decision problems are compiled into a use case diagram presented in chapter 4, which describes the functionality of the prototype application, FBP. To be able to create a disease surveillance prototype, some properties rooting from the information cycle had to be considered, which also is incorporated as a part of the theoretical starting point.

1.2.2 Scoping of the thesis

- Political or economic considerations in the context of India are out of scope.
- Privacy concerns related to using of sensitive information (e.g. residential locations) of case data, will not be considered.
- Targets decision making at the district level in India, however the transferability to other developing countries are considerable, but not examined in this thesis.

GIS based on case data also has applicability to analyzing accessibility to health services, resource planning, etc. The utility within these disciplines are considerable, but is not a primary focus in this thesis.

1.3 Theoretical starting point

The theoretical starting point emerges from the thematic structure of the research question that can be roughly divided into four elements; HIS, Disease Surveillance Systems, Geographical Information Systems (GIS) and the Information Cycle. A basic understanding of HIS and disease surveillance systems are required to have a conceptualization of the context which the thesis is based upon. GIS will be a key analytic-component in terms of displaying the spatial dimension of the FBP, hence a visual relationship between cases of health events or diseases. The information cycle will provide the broader understanding of HIS, explaining the structure and interconnected parts of the systems that affect both evidence-based decision making as well as analytical elements, which is the focus in this thesis.

1.3.1 Operational definitions

Tracked entity

A tracked entity in DHIS2 is most commonly described as a person or patient that is registered in the system, but could also be used for describing other kinds of entities or objects. The DHIS2 platform has a built-in metadata model that enables the end-user to create entities with a user-defined set of attributes, and to link the tracked entity model to a program. Tracked entities can be tracked between different DHIS2 instances to keep a record of the movements and health data. A tracked entity needs a unique identifier in the system to successfully separate them.

Organization unit

In the context of DHIS2, organization units most often reflect the natural hierarchy of a health system, ranging from national to facility levels. However, the hierarchy is configurable and highly situational. Organization units may be configured with an ancestor and children or subunits.

Facility

Facilities in this thesis are referred to as organization units that most commonly do not have any subunits in an organization unit hierarchy. Typically, a health station or a hospital that has a fixed location can be considered a facility. A district is an organization unit with subunits and is thus not a facility. However, there can be health facilities existing at the district level.

Program

Tracked entities are enrolled into programs, e.g. a Malaria program, in a DHIS2 app called “Tracker Capture”. These enrollments represent events with a registration of a tracked entity.

Tracked entity-Facility model

This is referred to as a way of structuring or organizing data, both spatial and temporal for disease mapping. The model maps cases of diseases derived from tracked entities to the facilities they were enrolled, using visual relationship identifiers in the spatial dimension. Also, the model provides for a descriptive analysis of case data in terms of place, time and person, where “person” is a tracked entity. The “time” component will be the time of enrollment into a program at a facility. The “place” component is split between the residence location of the tracked entity and the

location of the facility. The tracked entity-facility model has much of the similar meaning as the “visual relationship between health events and originating facilities”, but the latter is used as a more general or generic term with no concrete relation to the DHIS2 platform.

1.3.2 Health Information systems

Public health refers to the strategies and effort to tackle health issues in a population. This means the allocation of resources to avoid and deal with illness, reducing epidemic outbreaks and addressing community health issues. The outcome of these efforts aims to improve life quality. As identified by Centers for Disease Control and Prevention (CDC) National Public Health Performance Standards Program (NPHPSP), the first two goals of public health services (Centers for Disease Control and Prevention, 2014), are:

- 1) Monitor health status to identify and solve community health problems
- 2) Diagnose and investigate health problems and health hazards in the community

These goals emphasize the collection of information, analyzing and acting upon it – a link that requires a suitable HIS. The purpose of a HIS is to deal with the aspect of the complexity of e.g. public health, due to large amounts of data originating from a sizeable population, often collected from many different data streams to provide health data in a useable format for decision makers at different organizational levels. To what extent the data collected in HIS will support evidence-based decision-making, is dependent on several factors as described in the information cycle (section 1.3.5), however, data quality is often a prominent challenge. Data quality is described with numerous properties like data accuracy, completeness, and timeliness, which are considered to have an impact on decision-making efficiency (Samitsch, 2015).

1.3.3 Disease Surveillance Systems

A HIS will serve as the backbone in a surveillance system, but may need to be tailored for supporting requirements, such as early detection and response to disease notification. A disease surveillance system is a specialized system that gathers data from one or more disease reporting sources and enables their monitoring over time. The strategy of disease surveillance systems may vary, but focusing on a narrow set of diseases is likely to enable higher data quality in the system. The very purpose of a disease surveillance system in the context of public health involves the strengthening of public health actions, making it critical that the system can support decision making and help to evaluate the effectiveness of public health actions (Lee, 2010).

1.3.4 Geographical Information Systems

“The literature of GIS contains many definitions. Often, the functions of capturing, storing, manipulating, analysing and displaying spatial data form the core of these definitions and the idea that GIS are designed to support spatial decision making is implicit [...]” (Densham, 1991; p.405).

In public health, GIS are typically used by epidemiologists, for example with the purpose of conducting disease surveillance, intervention monitoring and clustering (George et al., 2013). The strength of GIS is its ability to display a different kind of location-bound datasets at a certain point of time and to combine data sets and link them to a visual frame (Trodd, 2002).

Ever since John Snow, an English physician mapped cholera cases in London during the epidemic in the 18th century, the advantages of mapping diseases had proven its value in public health. A GIS provides the potential of supporting information needs necessary for public health, by mapping and structuring large amounts of data related to disease distribution and healthcare resources (Magnuson and Paul, 2013). GIS are often applied as a component in disease surveillance system to describe the “where” or location component of disease cases.

GIS-enabled decision-making systems is still largely underutilized relative to its potential, due to lack of awareness and expertise needed for configuring, as well as

the potential financial barriers of working with proprietary systems (Mak and Eisen, 2013, Lai et al., 2007). The FBP, as mounted on an open-source platform like DHIS2, lowers the cost of adopting a GIS. To overcome the technical barriers, it is important to lower the threshold for utilizing the GIS functionalities. Lai emphasizes the need to build user-friendliness and high accessibility to enhance GIS use (Lai et al., 2007). The ease of understanding data and its interpretability should meet the user level, and allow it to scale to a larger user base, as public healthcare professionals are often not educated adequately to efficiently use GIS as a tool (Magnuson and Paul, 2013).

1.3.5 The Information Cycle

The information cycle provides us with a model for understanding what affects information use in the context of a HIS. The steps or processes that data will encounter from it being collected to its effective use for e.g. decision-making purposes, make up the information cycle model (Figure 1). The processes are chained together, meaning that the outcome of each step is affected by the previous, which makes information use not only affected by data collection – but the entire circle (Braa and Sahay, 2012).

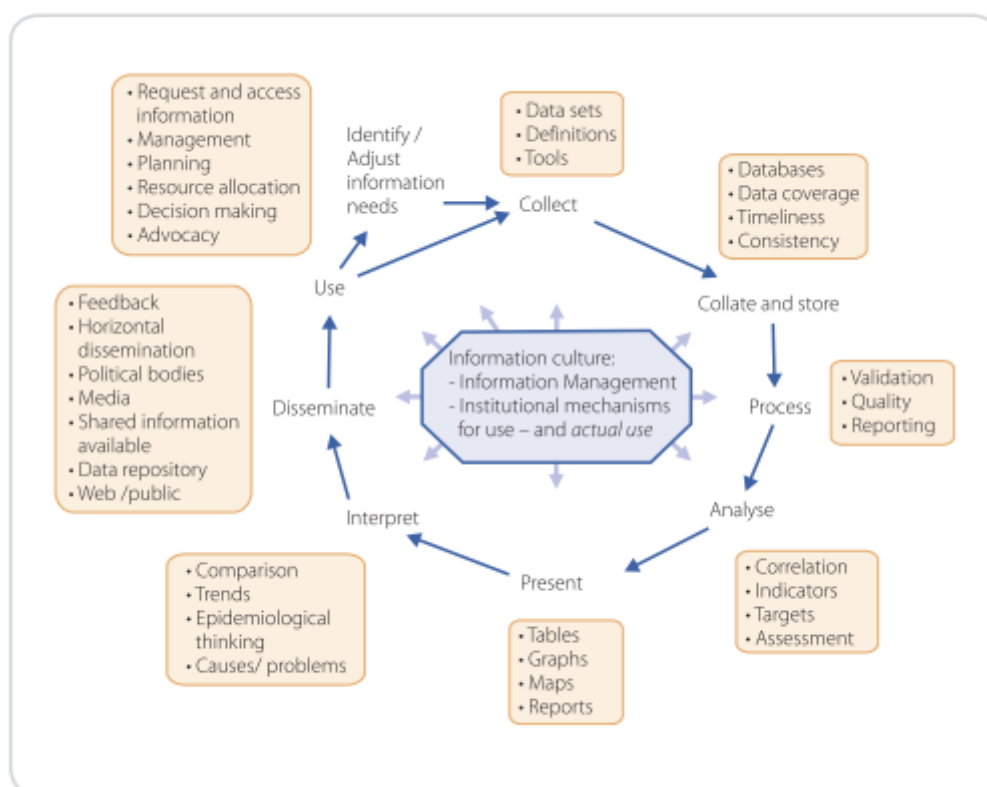


Figure 1: Illustrates the dependencies within the information cycle, ranging from data collection to information use (Braa and Sahay, 2012; p.248)

1.4 The motivation for the research

As in most developing countries, India faces major emerging public health challenges of both communicable and non-communicable diseases (Bhavan and Marg, 2011). While demographic changes and increased life expectancy contributes to spreading of chronic diseases, poor implementation of public health programs is one of the key factors that contribute to diseases epidemics and their poor management (Pokharel et al., 2007). For making most out of scarce resources, which is even more essential in developing countries like India, an effective HIS is required to effectively process huge amounts of healthcare data into more specialized management systems, that enables decision-makers to use the data for informed actions. Actions that save lives.

In public health, an effective disease surveillance system is critical for making decisions. Without such tools like DHIS2, the process of decision-making at all levels will face the risk of becoming subjective, representing a loose link between information and action. The worst-case-scenario is that decisions are not made at all

due to lack of data (Lee, 2010). An effective HIS can help to effectively utilize resources and help produce high quality health information, for supporting decisions that can be made to reach public health goals. The study by Thiagarajan and colleagues claims that the Indian HIS has a tendency of working like a reporting tool, rather than as management tools supporting decision-making (Thiagarajan et al., 2012). This results in a data led rather than an action led use of information, making the distance from decisions to actions longer, as the data is dispatched upwards in the hierarchy. Taking decision making further from the level where they should be undertaken, the district level - results in a heavy load of decision making at a higher level or having decisions not being made at all. Also, decision-makers at higher levels are often having a limited understanding of the local settings. Pushing data upwards also increases the time to discover a decision problem and to act upon it.

Timeliness of data affects the responsiveness of identifying outbreaks and building response. In terms of a syndromic surveillance system, where the responsiveness to deal with rapid clustering is essential, a week-old data may damage the decision-making process. Even if the decision is correct based on the current situation, the old data, may make potential outbreaks harder to delimit. The district level often has the right level of granularity and administrative resources, thus raising the need to provide them with effective information. Data incompleteness also has a substantial negative influence on decision processes, due to challenges in integrating information flows with all reporting units, especially those from the considerable private sector (Ministry of Health & Family Welfare Government of India, 2015).

In general, a successful surveillance system has many interdependent aspects, hence the information cycle, ultimately requires a synergy between humans and system-technical resources to create an information culture. Some Indian states and districts show signs of an increased use of information, helping to cultivate an information culture (Braa and Sahay, 2012). However, evaluations of the indicator based surveillance systems of the Integrated Disease Surveillance Programme (IDSP) in India, indicates insufficient data analysis and limited capacity to undertake analysis and response at the district and state levels (Deepak and Anil, 2014, Ministry of Health & Family Welfare Government of India, 2015). My thesis aims to strengthen the analytical capabilities of decision makers at the district level,

supporting them for effective use of information for action, timely response, and feedback to lower levels. I believe in the following quote:

“It is not because countries are poor that they cannot afford good health information; it is because they are poor that they cannot afford to be without it.” (AbouZahr and Boerma, 2005; p.582).

1.5 Structure of the thesis

The upcoming chapters in this thesis start by studying the literature to establish a solid foundation for the reader to become familiar with the disciplines drawn upon. Chapter 3 will cover the methodological framework for supporting and structuring the research of the thesis. Chapter 4 describes the analysis conducted in the development of the FBP, and the final implementation of the FBP. Chapter 5 will discuss my findings by assessing to what extent the relationship may influence evidence-based decision-making. Chapter 6 summarizes the thesis and specifies possible directions for further research and development of the FBP.

2 Context of research

2.1 Introduction

This chapter will present the background needed for putting the remaining chapters in the context of my research. The theory will mainly cover the fields of informatics but also touch upon theories and concepts related to decision-making, public health, and epidemiology. A considerable part of my theory is selected with the purpose of gaining a deeper understanding of the relations and dependencies within the different components of the information cycle, which feeds into the design of the FBP and is illustrated through some use cases of disease surveillance.

The chapter starts by explaining how data collection and visualization in the DHIS2 platform is conducted, by the use of integrated apps. The intention is to give the reader an insight into the DHIS2 platform which is the technical context of the FBP, together with the two disease surveillance systems of concern in this thesis; case and syndromic surveillance. The next subchapter entails the application of GIS, their implications for decision making in public health, GIS components, and functions. Three examples of surveillance systems with common properties with the FBP are then highlighted, for the purpose of transferring experiences from GIS supported disease surveillance systems. The last sections reflect the components of the information cycle, emphasizing information use, interpretation, analysis, and presentation. An evaluation of these components is central to establish an understanding of the key aspects that influence evidence-based decision making, which the FBP is intended to support.

2.2 The DHIS2 platform

In the context of the information cycle, the DHIS2 covers all phases from data collection to data dissemination. The platform has built-in a component for collecting or reporting data, called “Data Entry” - an electronic form for submitting aggregated data. The “Event Capture” and “Tracker Capture” modules are used for capturing case based data, where the main difference between them is tied to the tracker modules data structure, which enables follow up on patients over time. The event

capture module has basically the same functionality, apart from the tracking and registration functionality of patients in time and space.

An important strength of the DHIS2 is its high level of flexibility, enabling users to configure metadata structures, such as programs, tracked entities, organization units, etc., that suit the current situation and location. The flexibility enhances the platform into meeting the requirements for supporting local decision making in multiple contexts. The platform may, for instance, be customized to act as a disease surveillance system, by supporting data collection, processing, analysis, and interpretation, which is the backbone of a disease surveillance system. However, the flexibility comes with the cost of the training and education needed to configure the platform, like attending DHIS2 academies or workshops.

Visualization of data in DHIS2 is conducted by the use of multiple standard apps on the platform, like the “Data Visualizer” and the “GIS” app. The Data Visualizer supports dissemination of aggregated routine data and has several built-in chart types, e.g. line and bar charts. The GIS app displays a map with thematic layers and a range of overlays like facilities or boundaries between different organization units. The thematic layers enable the user to insert spatial data emerging from predefined metadata structures, given that the data content includes a spatial component. The GIS app is integrated with the metadata structures of the DHIS2 core, making it able to access much of the data content configured in the core.

Most of the apps that are related to data presentation also support exporting graphs, tables, etc., which then can be encapsulated into a dashboard. Dashboards may thus be created by combining different exports and customized to support situational analysis and decision-making needs. Figure 2 shows an example DHIS2 dashboard, screened on the start page of the platform portal.

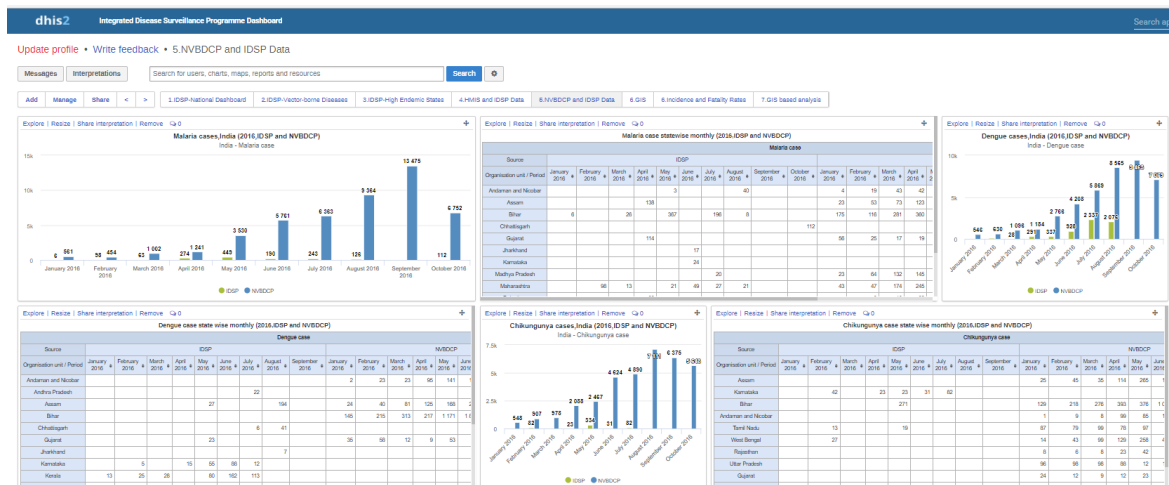


Figure 2: Displays an example DHIS2 dashboard, combining several outputs like graphs, bar charts, tables, etc.

The dashboards are an effective and flexible way of combining relevant data for evidence-based decision making. However, dashboards should be created with care, as challenges due to information overload, irrelevant data, lack of data context, etc., may undermine the decision-making process.

As pointed out, DHIS2 may be customized to act as a disease surveillance system. Next section covers two surveillance system types, which are common in public health, but also the context in which the FBP will be evaluated.

2.3 Disease surveillance systems

Disease surveillance data can be used for many purposes, e.g. detection of outbreak-prone diseases, building hypothesis and conducting research. The design of the system should be appropriate to help the data and systems to meet their intended use, analysis, and public health actions. Two of the most common types of disease surveillance systems are case and syndromic. These are now discussed.

2.3.1 Case surveillance systems

From a historical perspective, case surveillance systems are the most common type concerning individual cases or events of diseases. Data from case surveillance systems are often analyzed in terms of time, place and person, also referred to as descriptive analysis. Traditionally, this analysis type has involved limited statistical analysis. Analysis types common in case surveillance systems are descriptive (e.g.

case counts/rates), inferential (e.g. trend over time) and aberration detection (e.g. geographical clustering).

Inferential types seek to test hypothesis using statistical methods, derived from descriptive analysis. Statistical conclusions are often easier to communicate to policymakers for action (Lee, 2010).

2.3.2 Syndromic surveillance systems

Syndromic surveillance systems are constructed to detect fast emerging diseases, where decision-makers need to engage in very early phases of potential outbreaks and speedy response. By targeting syndromes, the system facilitates early detection, before lab confirmed samples return. Depending on the disease and available infrastructure, the lab confirmation of a disease takes some time. If earlier control measures were implemented, then fewer resources are likely to be needed for preventing epidemics. However, inaccurate syndromic surveillance systems may lead to less confidence in the system and wastage of resources during field investigations. Syndromic surveillance systems are facilitated by real-time data or close to real-time data. Syndromic analysis types are of aberrational character, including geographical clustering/aberrations and temporal aberrations. These are types that require efforts in real-time to detect clustering or abnormalities in space or time (Lee, 2010).

In spatial aberration detection, three general questions occur:

- *“Are there unusual aggregations (or deficits) of cases in particular areas? In other words, are there ‘clusters’ of disease?”*
- *Are there patterns in local rates observed in specific small administrative areas (e.g., counties or census tracts)?*
- *Are there links between the spatial pattern of a particular exposure (e.g., levels of air pollution, or more general notions of ‘exposure’ such as socioeconomic or demographic summaries from local or neighboring populations) and that of a particular health outcome (e.g., emergency room visits for asthma)?”*

(Lee, 2010; p.105)

A geocoded address is required to address the first question. However, to be able to determine if there are unusual aggregations of cases, there is a need for viewing the cluster in the context of the population at risk in a given area. GIS can be used for displaying spatial aberrations.

2.3.3 Characterizing an outbreak in a disease surveillance system

Statistical deviations from a historical baseline is a central component in outbreak detection, but must often be supported by complementary information to establish the existence of an outbreak. The measurement of the validity of disease surveillance systems for outbreak detection is often tied to several aspects. The list represents a selection of the factors that affect the detection and establishment of an outbreak in the context of disease surveillance (Buehler et al., 2004):

Case definitions: Defines the conditions of interests and response criteria's.

Baseline estimation: Estimations based on historical occurrences of disease cases, taking population size and geographical distribution into account.

Reporting delays: Untimely data has a negative impact on the performance of the surveillance system.

Outbreak characteristics: Characteristics of the causative agent and environmental factors.

Epidemiological analysis, interpretation, and investigation: Working methods of an epidemiologist, tools for analysis and response are important for factors contributing to the detection and response of outbreaks.

(Buehler et al., 2004).

The involvement of GIS in outbreak detection can be found at baseline estimation, outbreak characteristics and epidemiological analysis, interpretation and investigation.

2.4 Geographical information systems

2.4.1 GIS for spatial decision making

Despite its limitations, the map is the best way to structure spatial information (Carver, 2001) and for analyzing a spatial pattern representing a real-world problem. The GIS helps abstract a real-world problem into digital form, to make it manageable and analyzable. The usefulness of these abstractions is however tied to the validity of the assumptions made at each level of the abstraction process (Trodd, 2002). Errors in this process will affect the interpretation of the problem by giving false assumptions, creating a negative impact on the decision-making process.

The spatial decision supports systems (SDSS) differ from GIS by providing:

“[...] the user with a flexible decision-making environment that enables analysis of geographical information to be carried out in a flexible manner.” (Densham, 1991; p.405).

Some traits of the SDSS relates to the ability to help solving ill-structured problems, and helping users to explore decision options and in the use of analytical models (Densham, 1991). There exist multiple definitions of SDSS's, but largely they differ from GIS by how far in the decision-making process the system assists or interferes with, thus defining the level of complexity of the system. In general, the SDSS may vary, depending on the implementation support to the decision process. In contrast, the GIS supports discovering and analyzing aspects of the decision problem, like more traditional information systems.

A general limitation of GIS is the lack of temporal analysis, which is the “time” aspect of health issues reported. Comparisons between organization units or groups of people in an historical perspective are often required to view trends or patterns that may occur. Without a temporal dimension, it is also hard to have a reference to such measures as average or expected rates or values, as diseases vary over time. Thus, temporal analysis often needs to be a supplement to the spatial dimension.

Many GIS portfolios in the last decade have been stand-alone specialized tools with advanced statistical features, which are likely to offer a low level of flexibility in terms

of integrating data sources for public health purposes (Lai et al., 2007), This also represents a well-known lock-in strategy. Availability of geographical data for public health surveillance in general, is also a potential barrier for GIS which can be related to high costs of maps, layers and datasets - while integration of spatial and non-spatial data poses a challenge, due to the diversity of data sources needed for comprehensive public health surveillance (Lai et al., 2007).

GIS tends to disregard analytical modeling and decision-making processes. Not supporting a defined decision-making process may lead to different subjective interpretations of information disserted by the GIS applications (Densham, 1991). The analytic modeling is on the other hand largely used for predictive causes, and therefore is not the main focus of the FBP. A challenge with the process of decision making is often that the decision makers fail to define the decision problem or objectives (Densham, 1991). This is a well-known case, and it has been the motivation for embracing methods such as agile software development models, but can also to some extent be achieved by creating a flexible problem-solving environment. As Densham points out:

“To assist decision makers with complex spatial problems, geoprocessing systems must support a decision research process, rather than a more narrowly defined decision-making process, by providing the decision maker with a flexible, problem-solving environment” (Densham, 1991; p.403).

This indicates that Densham emphasizes flexibility to tackle complexity in a spatial decision environment.

2.4.2 GIS Components

In general, a GIS application consists of four components to abstract a real-world scenario, starting off with “Spatial-attribute relationship”. This represents the part of the system where a set of geographical coordinates in a topological context is related to attributes of datasets in the database. The relationship is based on a common identifier called a “geocode”, such as coordinates or a zip code. This is how the spatial is mapped with other kinds of relevant data, e.g. mortality rates of a district. A general challenge that threatens this relationship is interoperability, as database

attributes and geocodes may appear in different formats in the different systems (Magnuson and Paul, 2013).

Map projections and spatial reference systems refer to the component in which geographical data is referenced. This is most commonly a map, representing the earth by latitude and longitude. The map projection must be accurate to display a correct display of relevant data (Magnuson and Paul, 2013).

Map scale is the ratio of the distance on the map in relationship with the actual distance on the ground (Magnuson and Paul, 2013), e.g. for running in the forest a 1:50 000 map could be appropriate, while viewing the extent of Sahara a 1:500 000 000 might be more suitable.

Spatial data representation concerns how features or items are shown on the map, this is most commonly either through vector data or raster maps. Vector data are shown on the map which can be described as points, lines or polygons by coordinates. This could be a building, a road, a forest, etc. Raster data is linked to images such as aerial photos or satellites. Independent of which one is used, they should both be linked to a coordinate system. (Magnuson and Paul, 2013)

2.4.3 GIS functionality in public health

Common features applied for public health analysis includes:

Choropleth mapping is a mapping where a value important to public health analysis, is reflected on a map by color coding. This is used for showing the distribution of health issues on aggregated data. The failure to communicate spatial variations within the color-coded regions, is a disadvantage by choropleth mapping.

Distance measuring is a common feature of GIS software, which can be used for assessing the accessibility of healthcare facilities or modeling exposure to threats.

Spatial query involves making a query, targeting the database attributes and display the result spatially. The query may also be based on distance.

Buffer functions is a circular area around a center-point. Used as a graphical overlay to easily view e.g. spreading of a gas leak (Magnuson and Paul, 2013).

Spot maps are displaying single health event or cases of diseases on a map (Lee, 2010).

2.4.4 GIS in public health

GIS can be used to describe the spatial aspect of disease cases, “place”. The importance of information regarding the location of cases lies within the knowledge about the geographical extent of a health issue - where it occurs. Knowledge about the place is beneficial in terms of having the ability to provide information about disease-prone areas, giving decision makers the ability to distribute and prioritize resources (Mak and Eisen, 2013), which is even more important in low resource settings, like India. However, the required resolution of the geographical locations will vary according to what is the appropriate focus to describe the place of the health issue; it could be a district, a health facility, a street address, etc. (Centers for Disease Control and Prevention, 2012b). High resolution affects data accuracy, and with it thus improves decision making efficiency (Samitsch, 2015).

I will now provide some examples of GIS in public health, particularly in the context of disease surveillance.

GIS: Syndromic surveillance in public health

From November 2001 to November 2002, the newly established syndromic surveillance system of New York city health and mental hygiene was studied by Heffernan and colleagues (Heffernan et al., 2004). The system was processing 2.5 million visits of New York’s 39 participating emergency hospitals, and was mainly intended for detecting large-scale bioterrorism attacks, but proved to pick up correlating signals with influenza activity and suspected norovirus transmission. Data records from the hospitals included: date and time of visit, age in years, sex, home zip code, and free-text chief complain. Data was transmitted electronically daily, where half of the hospitals had a manual transmission. The data processioning and quality checking where received and processed into a new format. One analysis was carried out for each syndrome category for detecting citywide temporal abbreviation and clustering by either hospital location or patient home zip code. The intention was to detect moderate or large-scale events. In short, the city was divided into smaller

geographical areas, where the historical data from the last 14 days were used as a baseline for estimating the expected number of cases.

Investigation of spatial signals is based on a review of the emergency department, comparing expected with actual visits of each hospital. By focusing on the hospitals with the highest excess cases, a list of patients is compiled with their attributes and checked for errors that could be misinterpreted. Next, phone calls were made to alert hospitals of unusual activity. If necessary, a field investigation was conducted for interviewing patients or visiting a clinic at the hospitals. The syndromic surveillance systems managed to detect signs of communitywide influenza 2 weeks before an increase in positive influenza reports from laboratories. Also, norovirus and gastrointestinal illness in all ages were successfully detected and disseminated to the medical community in a timely fashion. However, the systems encountered many single-day clusters that were not identified as outbreaks (Heffernan et al., 2004).

Web-based GIS in Hong Kong

This study covers two models that have been successfully adopted by Hong Kong health authorities since 2004 for management of over 30 infectious diseases. The model is a web-based GIS that improves the dissemination efficiency and collaboration across institutional boundaries. The system enables system operators to combine disease-related data with spatial information to view disease incidents over the Hong Kong area. Data collection is carried out by practitioners and reported into the system as a suspected or confirmed case. This list is generated as an overview of the main functionality of the system, which is based on notifiable diseases reported by registered medical practitioners and hospitals:

- *“Function to permit query on disease case data based on date of notification, type, case district, etc;*
- *Ability to display cases based on geo-coded addresses and their corresponding disease codes on maps;*
- *Means to display disease cases based on specified disease groupings;*
- *Function to provide tabular summaries based on query criteria;*
- *Capability to view case attributes after performing spatial query (e.g. case district, home address, office name, office address, age, sex, hospital attended, etc.);*

- *Function to display aggregated number of disease cases in different districts with bar charts categorized by gender or age group;*
- *Capability to overlay locations of hospitals, schools and homes for the elderly on digital maps for visualization and action planning;*
- *Special symbols to differentiate outbreak and non-outbreak cases;*
- *Function to instantaneously compile summary reports in HTML format based on query criteria for follow up actions;*
- *Flexibility to process, integrate, display and query new parameters or data fields associated with a disease (i.e. data model needs to be flexible);*
- *Infectious disease data update at 15 minute or hourly intervals.”*

(Lai et al., 2007; pp.49-50)

The web-based GIS for surveillance of infectious diseases enables analysts to perform tempo-spatial queries, for detecting clustering in a timely manner, and provide a more coordinated response to outbreaks (Lai et al., 2007).

GIS: modeling accessibility to healthcare centers in Saudi Arabia

A study of applied GIS in health planning was conducted in Jeddah, Saudi Arabia.

The focus lies on the use of GIS to measure the accessibility of health centers.

Vector data such as lines and buffer rings are examples of means to measure accessibility in terms of distance. The accessibility is measured by combining three methods: 1) Proximity to a health center, defined as 2kms (which is the radius within which citizens of Jeddah should have to a health center), representing the catchment area between patients and health centers. This measurement can be used for planning locations of new health centers 2) Health-center-to-population-ratio is a way of measuring how many people a health center will serve. This analysis method needs a reference to the underlying demographic data/population density, for conducting this assessment. 3) The third model was carried out using sophisticated ArcGIS analysis of network formulas. The study provided health planners of Jeddah with three models for assessing the accessibility to health care services (Lai et al., 2007).

The above examples provide concrete cases of how spatial data was used for disease management. However, GIS modeling of accessibility is not the main use of the tracked entity-facility model, but it illustrates the range of applications. In the next

section, some general concepts around information use, and its different components will be discussed.

2.5 Information use

“Information use” refers to the use of information to support a process that produces an action or outcome, in this case, a public health action. The types of processes concerned in this thesis are decision making in the context of disease surveillance.

2.5.1 Aspects of Decision making

Decision problems can be solved by a range of approaches, e.g. by intuition, by experience, by advice from an expert, or randomly. These approaches are called descriptive models (Grünig et al., 2013). However, in prescriptive decision theory, Günig outlines two systematic-rational decision-making procedures that are available for practical purposes; Heuristic and Analytic procedures (Grünig et al., 2013). In short, the heuristic model is a simplistic and informal way of making a decision with a low application cost, but with no guarantee regarding the outcome. The heuristic model of decision-making in Figure 3 represents an understanding of how humans naturally make decisions. The input to the decision process is information and marks the involvement of the FBP in the decision-making process. The information should enable the decision maker to discover the decision problem, and to some extent help planning and to analyze the decision problem.

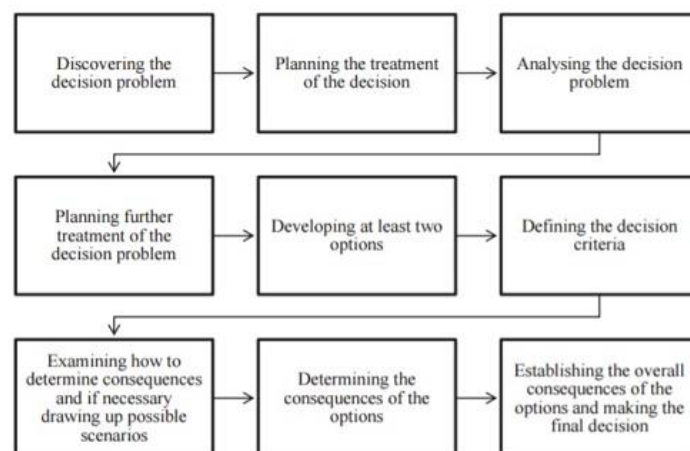


Figure 3: Heuristic Decision-making Procedure (Samitsch, 2015; p.17)

The analytic procedure is characterized by having guaranteed a near optimal solution, however, it will require a well-structured problem of application. A well-structured problem is defined by having quantitative aspects, and rules regarding what's an acceptable outcome. Also, the decision maker must be able to reach a solution within the use of reasonable time and resources (Grünig et al., 2013). This indicates that a well-structured problem is one of the multiple prerequisites for reaching an optimal solution.

The discovery of the decision problem is the starting point of any decision-making process. To be able to determine if there is a decision problem, a good understanding of the ideal and current situation is needed. The gap between the two different states will determine if an action is required to reduce the gaps between them (Grünig et al., 2013). To support decision makers in detecting decision problems, information systems help in providing information on the existing and desired situations. Sometimes, information provided to decision makers would come too late to correct a situation (Grünig et al., 2013), thus the need to build warning systems, which is often referred to as syndromic surveillance systems in a public health setting. A syndromic surveillance system seeks to detect geographical or temporal aberrations (Lee, 2010), and often offering support in defining the decision problem as the system logic e.g. algorithms automatically detects the gap between current and the ideal situation. However, such systems may often produce false alarms negatively impacting the decision-making process (Grünig et al., 2013). In a public health setting, the decision-making processes are often made by interpretation of information made available by data sources in information systems through an assessment, and then presented to decision makers.

An assessment: the decision-making process in a public health setting

An assessment is a key public health function, as pointed out by The US Institute of Medicine's Committee. Any public health decision-making processes will be based on an assessment to help identify public health challenges and to allocate resources to deal with the threats. An assessment is characterized as an ongoing and systematic process that is conducted on a regular basis or as a response to an event, including sub-processes such as collecting and analyzing available data, making the basis from which the assessment is made upon. The assessment

process will normally be conducted by a multidisciplinary team, including different roles within fields like medicine, epidemiology, administration, etc. (Keppel and Freedman, 1995).

Keppel and Freedman describe the assessment process in three steps (Figure 4): 1) Identify health needs by monitoring health status, 2) Identify available health resources and evaluate their effectiveness, 3) Present the public health challenges to health managers, policymakers and the public in a way that they support evidence-based decision making (Keppel and Freedman, 1995). The monitoring of health status and risk factors will normally involve the use of multiple data sources as one data source is unlikely to cover all the information needs of the assessment. A data source could be a routine HIS, disease surveillance systems or surveys.

Figure 4 shows the internals of an assessment process, as described by Keppel and Freedman:

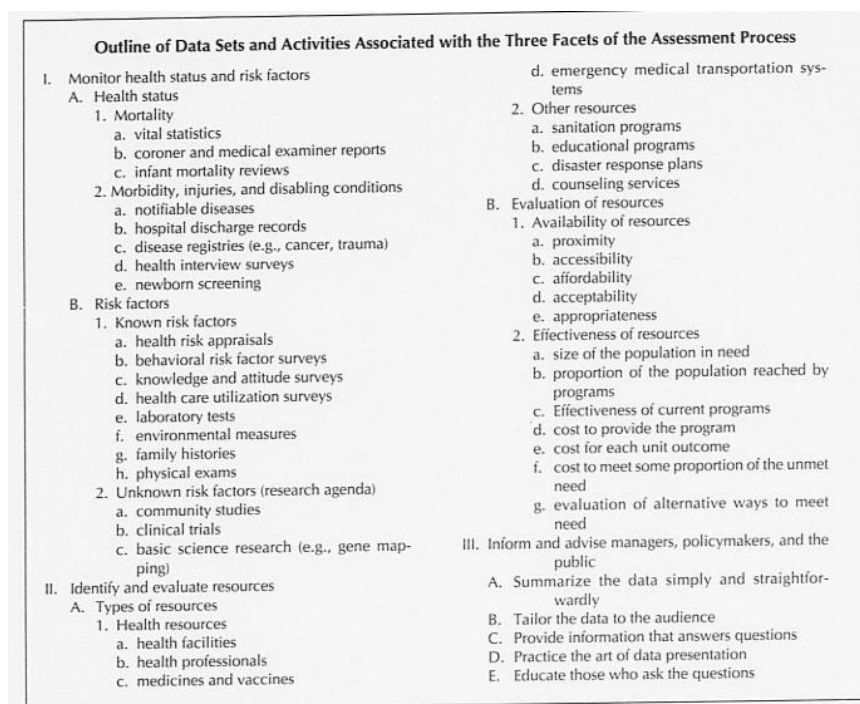


Figure 4: Outlines the activities in an assessment process (Keppel and Freedman, 1995; p.6).

Whether these exact assessment activities are being used by DSU's (District Surveillance Units) in India is less likely, but these activities are often used in some configuration for conducting evidence-based decision making in public health. The thoroughness and extent of activities used may differ between hierarchical levels,

e.g. district level with limited human resources are less likely to conduct a comprehensive assessment than at the state level. An assessment will thus differ in each local setting. The activities outlined in Figure 4 are information led activities that are common internals of a decision-making process in public health.

Decision-making efficiency and data quality

DECISION-MAKING EFFICIENCY			
Data objectivity	Amount of data	Data accuracy	Experience
Data believability	Data access	Ease of understanding of data	Knowledge about the domain
Reputation of data	Conciseness of representation of data	Interpretability of data	Restrictiveness and design of DSS
Value-added data	Manipulability of data	Representational consistency of data	Decision-making skills
Data timeliness	Data security	Self-efficacy	Data completeness

Figure 5: Factors that may influence decision-making efficiency (Samitsch, 2015; p.23)

Several factors influence decision-making efficiency, where most research has been regarding data accuracy, timeliness, and completeness, there are the data quality attributes of concern in this thesis. Data accuracy relates to whether the data is correct and reflects a real-world object or an event described. Timeliness refers to whether the data is old (Samitsch, 2015), reflecting the time interval between the capture of the data and the data being available for use. However, data timeliness must be put in the context of the use, as some data can be valid even though it is old. Data completeness refers to whether the data meets the expectations of the user, e.g. if all mandatory data is collected as expected, the data may be considered complete (crm4sure, 2018).

2.6 Data interpretation

A study of health care consumers by Hibbard and Peters (2003) examined how information is used to make informed choices when presented with comparative information. When decision makers are confronted with information, involving many variables in one single choice, their cognitive capacity tends to be challenged. The decision makers will then often make shortcuts by emphasizing one variable over the others, when comparing variables up against each other. Building such prioritizing is

challenging as interpreting comparative information is usually based on the assumption that the information seeker knows what is important to him (Hibbard and Peters, 2003), which is often not the case and can be understood as the goal or the desired situation in a decision problem. One way to lower the cognitive burden of the information presented to the consumer is to present them with real data from their own organizational context (Hibbard and Peters, 2003). This is can be achieved in multiple ways. 1) By implementing an information system that has the role of a decision tool, 2) Having a person in the middle (with technical expertise) to highlight important factors and trade-offs in the information, 3) Also, displaying information in a transformed way to meet the analytic capabilities of the users, e.g. by simplifying concepts (Hibbard and Peters, 2003).

Figure 6 shows how information is interpreted when the information is presented in different ways.

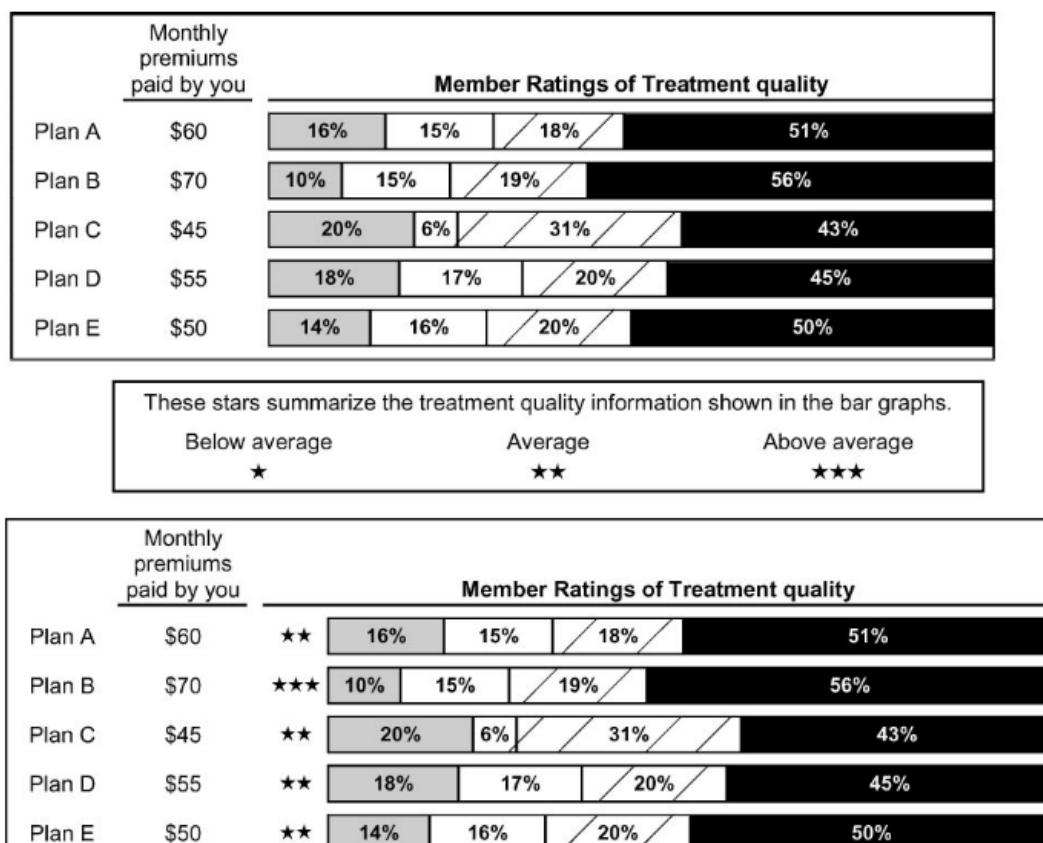


Figure 6: Shows different organization of data presented to health care consumers in Hibbard and Peters experiments. The lower table summarized the data by displaying stars, which makes the consumers tend to choose Plan B (Hibbard and Peters, 2003;p.421)

In this laboratory experiment by Hibbard and Peters, they showed two different information displays to the healthcare consumers, where the upper one (Figure 6) displayed treatment quality purely by use of percentage bars. The lower one displayed a rating of the treatment quality based on the percentage ratings. Based on the display in Figure 6, consumers were seen to be more likely to choose the plan with the highest quality. This indicates that the structure of the information presented or the way its communicated has an impact in terms of interpretability and readability of the information.

2.7 Data analysis and presentation

Lechner and Fruhling examined best practices for designing public health dashboards and made recommendations towards new guidelines (Lechner and Fruhling, 2014). These included:

- **Customizable, Actionable “Launch Pad”:** The dashboard should be customizable and functions that can be executed.
- **Supporting Correct Data Interpretation:** This can be conducted by data comparison of e.g. putting data into context from a historical perspective.
- **Information Aggregation:** Data should be summarized to show overall trends. This will reduce the cognitive burden and make the data become more readable.
- **Adherence to Convention:** A dashboard should have consistency regarding look and feel, and also has same the terminology throughout the dashboard.
- **Minimalist Aesthetics:** Only need to know information should be presented in a minimalistic way.
- **In-Line Guidance and Training:** The users should have visual feedback when making actions at the dashboard, e.g. making a query.
- **User Workload Reduction:** Reduce the cognitive and psychological workload of the users.
- **GIS Interface:** Use GIS whenever data has a spatial component to increase situational awareness.
- **Content:** content should meet the user requirements and information needs.

The new proposed guidelines cover:

- Use charts and graphs where sensible to show trends visually and for quick interpretation.
- Dashboards should have temporal capabilities whenever the data has a temporal dimension, emphasizing trends over time and relationship to historical data.

(Lechner and Fruhling, 2014).

Other studies on the design of dashboards have been conducted targeting more specifically the DHIS2 platform - presenting Jespersen's thesis on DHIS2 design guidelines (Jespersen, 2017), specifically for dashboards. "Developers" are defined as those people engaged in creating tools for the users to build their own dashboards. These guidelines were studied to help align the FBP approach of predefining a dashboard for decision makers.

User Created Dashboard - Design Guidelines		
Guideline	Developer Description	Dashboard Creator Description
Target Audience	Think about the dashboard creators and end users when creating the dashboard creation tools. Communicate with users for feedback and suggestions.	Keep audience in mind. If other people than yourself will use it, think what is best for them. Communicate with users.
Screen Boundaries	Create display media that scales well and make the dashboard responsive.	Do not overwhelm or confuse the user by filling the dashboard with items outside of the screen boundaries.
Display Essential Data	Do not have unnecessary precision in numbers. Create the display media so that the user can display data that is directly analysable without the user having to process the data.	Only display data that is essential for the evidence based decision making.
Data Needs Context	Provide display media that allows for more data context.	Give context to the data by displaying multiple targets for the measure.
Data Layout	Give the users flexibility to order the items on the dashboard as they want.	Group data that will be compared and keep the most important information at the top.
Choosing the Display Media	Provide good options for the users. Do not include graphs that is not fit for dashboards.	Use the display media that most effectively communicates the data.
Display Media Design	Design the display media so it effectively communicates the data that is shown. Reduce non-data pixels.	Does not apply to dashboard creators.
Highlight the Important Information	Provide users tools to highlight information manually or automatically (e.g. by validation rules).	Highlight important information if you want to capture the users attention.
Colour Palette	Provide users tools to choose colour palette.	Stick to a few colours and think about accessibility.
Make the Dashboard Attractive	Choose a design that is attractive, both for the display media and the dashboard itself.	Does not apply to dashboard creators.
Time and Updates	Provide tools for displaying when the data is from and allow halting of updates.	Display time from when the data is from. Allow for halting updates if necessary.

Table 1: Suggested DHIS2 dashboard design guidelines by Jespersen (Jespersen, 2017; p.71).

Jespersen's study was based on prototyping dashboards in cooperation with HISP Uganda, and includes the use of DHIS2 dashboards for data presentation, not third-party apps. While the transferability of this work is uncertain, they are useful as they have strong roots in theory related to visual and cognitive aspects of information dissemination.

This chapter has provided the reader with a contextual understanding upon the themes of concern in this thesis, covering the DHIS2 platform, disease surveillance systems, GIS, aspects of information use, data interpretation, and presentation. The next chapter will describe the methods used for the research.

3 Methods

3.1 Introduction

This chapter describes the methods adopted in the thesis. The first section describes the theoretical aspects of action research, which provides the basis for my methods. Action research will be elaborated in terms of definitions, principles, and paradigms. Next, the use of different data sources throughout the thesis will be explained, before focusing on how the method was applied and used in praxis. The chapter ends with a short summary, evaluating the use of method and data sources.

3.2 Action research

3.2.1 Background

Action research is a scientific method for researchers and scientists, originally rooted in the social sciences after World War II (Baskerville and Wood-Harper, 1998). Action research is most commonly found to be based on qualitative methods of data collection, but might also rely on quantitative measures whenever it is natural and practical. From a historical perspective action research has three underlying research paradigms:

Positivist paradigm

This has been and still is the dominant paradigm. Phenomena is discovered and verified by direct observations or measurement by independent researchers (O'Brien, 1998, Krauss, 2005). Positivist paradigm emphasizes an objective reality, while action researchers commonly will not try to remain objective, weakening the adoption of this paradigm. Action researchers also emphasize collaboration and participation above having independent observers to prove their results. By design, action researchers have pre-determined intentions of change, which conflict with the positivist researcher's assumptions of being objective and neutral.

Interpretive paradigm

The interpretive paradigm has a background in social sciences, emphasizing that the reality is founded on a mix of social and experimental understandings (Robert Wood Johnson Foundation, 2018). As characterized by O'Brien:

"[...] a belief in a socially constructed, subjectively-based reality, one that is influenced by culture and history." (O'Brien, 1998).

Action researchers acknowledge the strength of a social dimension, like a dialog between researchers and coworkers and co-learning. However, interpretive research methods rely on a more subjective role, whereas action researchers have a more active role (O'Brien, 1998). Ideally, in action research, the researchers aim to contribute to making a change to a system or organization, which is hard to achieve from a passive position.

Paradigm of Praxis

Some researchers claim that action research does not have the epistemological support necessary, introducing the paradigm of Praxis. "Praxis", a is a term used by Aristotle emphasizing that knowledge is derived from practice, and practice is informed by knowledge. (O'Brien, 1998). The ideals of this paradigm are considered the best match for action research, where theory and practice in a continuous and adaptive manner drive the research. Others have claimed that action research falls under the paradigm referred to as a non-positivist paradigm of reflective rationality, where validity is based upon the participant's experiences and perspectives rather than a generalized truth proven in a controlled environment (Zuber-Skerritt, 2001).

3.2.2 Characteristics of action research

There seems to be a consensus regarding the core definition of action research. However, different researchers emphasize various aspects of their definitions. To quickly demystify action research a simpler explanation, like of O'Brien's is presented:

"[...] a group of people identify a problem, do something to resolve it, see how successful their efforts were, and if not satisfied, try again." (O'Brien, 1998)

This narrow definition emphasizes problem-solving in an iterative way, which is similar to the agile methodologies adopted by software engineers. To separate action research from traditional problem solving and software development methodology, a wider definition may be considered:

“Action research...aims to contribute both to the practical concerns of people in an immediate problematic situation and to further the goals of social science simultaneously. Thus, there is a dual commitment in action research to study a system and concurrently to collaborate with members of the system in changing it in what is together regarded as a desirable direction. Accomplishing this twin goal requires the active collaboration of researcher and client, and thus it stresses the importance of co-learning as a primary aspect of the research process.” (Gilmore et al., 1986; p.160)

Gilmore points out that co-learning, systematically studying the problem at hand and making theoretical considerations, are considered important properties which makes action research different from “normal problem solving”. Action researchers more often target real world problems rather than experiments or fictional problems and are often invited into a domain for applying their methodological knowledge to solve a problem in an organizational setting. The action researcher will try to collaborate with personnel or co-workers and attempts to motivate them to apply their new knowledge themselves. That makes the social dimension a somewhat dominant property for defining action research (O’Brien, 1998), where human-processes are more dominant than generalizable truths (Baskerville and Wood-Harper, 2016).

As described, action research is an approach for making a change in a cyclical and participatory way, which also does not dictate the use of data sources, making it a flexible methodology. The flexibility makes it applicable as a base approach, to tackle a range of organizational problems. A unique position of action researchers is their rapid application of new learnings, unlike methods where the researcher has a passive observer role. On the grounds of being flexible, in a way that for some researchers seems more end-state oriented than process oriented, makes critics claim that action research may be described as a strategy rather than a research methodology (Denscombe, 2010). Other critics also have concerns about the scientific rigor of action research. Action researchers sometimes drift towards a

liberal action research type, where the practical concerns are more emphasized than theoretical and methodical aspects, often leading the researcher into becoming more like a consultant (Baskerville and Wood-Harper, 2016).

The process of canonical action research

The phases in Figure 7 will be explained stepwise:

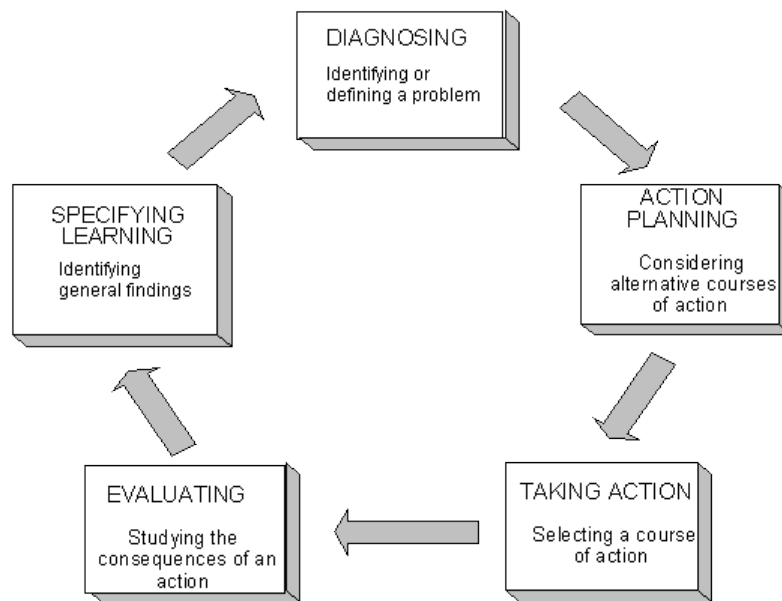


Figure 7: Illustrating a cycle in action research. Originally from Gerald Susman, adapted by Rory O'Brien (O'Brien, 1998)

Diagnosing: The first step will always be to collaboratively define a problem to solve a practical problem. A failure to correctly identify the problem will negatively influence the remaining steps of the iteration.

Action planning: Developing alternative approaches to problem-solving by researching the problem by use of available data sources. The course of action will be decided by the evidence-based decision making rooted from the data sources like literature, meetings, e-mails, etc.

Taking action: Implementing the changes decided upon in that cycle.

Evaluation: Considering the impact of the implementations, but also an assessment of the implementations that may include feedback. The goal is to measure if actions and interventions had the intended effects.

Specify learning: There may be several reasons for a positive or negative outcome of an iteration. This step involves analysis of the factors that contributed to a certain outcome, to stimulate learning for future iteration or problems.

3.2.3 Action research in IS

So far, the description of action research has been from a broad and general perspective, based on the original or “canonical” form of action research. Variations and adaptations of action research have been growing forth over the last few decades to becoming applied to an increasing number of domains like IS. The next section will look further into what Baskerville outlines an IS action research paradigm, and outlines four comparative characteristics that provide for an informed choice of method.

1) **Process model:**

- a. **Iterative:** A cyclical model, emphasizing problem diagnosis and action activities.
- b. **Reactive:** Like iterative, but focuses less on problem diagnosis than reflective analysis.
- c. **Linear:** A non-cyclical model with distinct phases.

2) **Structure**

- a. **Rigorous:** Strict and defined phases that are followed closely.
- b. **Fluid:** Loosely defined activities. Allows simultaneous activities.

3) **Typical involvement:**

- a. **Collaborative:** The research is integrating into in the organizational and socially, by becoming a co-worker of the subjects.
- b. **Facilitating:** The researcher is having an expert role but is still emphasizing cooperation with the subjects.

4) **Primary goals:**

- a. **Organizational development:** Improving the organization from a social point of view.
- b. **System design:** Create or modify systems in an organization.

- c. **Scientific knowledge:** Research that aims to contribute to greater understanding of a problem setting.
- d. **Training:** Increase the knowledge about a specific problem, from an educational point of view.

(Baskerville and Wood-Harper, 1998)

Table 2 summarizes types of action research in the IS domain, organized by the characteristics previously elaborated; process model, structure, typical involvement and primary goals.

	Process model			Structure		Typical involvement			Primary goals				Published IS examples
	Iterative	Reflective	Linear	Rigorous	Fluid	Collaborative	Facilitative	Experiment	Organizational development	System design	Scientific knowledge	Training	
Canonical action research	•			•		•			×		×		Baskerville (1993)
Information systems prototyping	•			•		+	+			•			Kyng (1991)
Soft systems	•				•		•		×	×			Checkland & Scholes (1990)
Action science		•			•		•		×		×		Reponen (1992)
Participant observation		•			•			•			•		Jepsen <i>et al</i> (1989)
Action learning		•			•			•				•	Naur (1983)
Multiview			•	•		+	+	+		•			Avison & Wood-Harper (1990)
ETHICS			•	•			•		×	×			Mumford (1983a)
Clinical field work			•		•		•		×		×		Hammer & Champy (1993)
Process consultation			•	•				•	•				Coad & Yourdon (1991)

Key: • signifies a dominant characteristic, + (or) signifies characteristics that will dominate in different studies, × (and) signifies characteristics that may occur together in the same study.

Table 2: Table comparing characteristics of canonical action research with other forms of action research in IS (Baskerville and Wood-Harper, 1998; p.96).

As a primary goal of this thesis is the creation of a system, rather than contributing to the social aspect of organization development, the following specific IS research methods will be examined: Information Systems Prototyping, Multiview, and Soft Systems.

Multiview

The Multiview framework provides methodical support for the thesis in terms of having logical stages for assisting in system creation: 1) human activity analysis, 2) information analysis, 3) socio-technical analysis and design, 4) human-computer interface design, and 5) technical design. However, these steps are considered rigid as the process model is linear, making the project less adaptive. If new information

influencing an early phase like *human activity analysis*, appears late in the development, it could have an adverse effect on the outcome. This increases the risk of failure, by not meeting user requirements. This model may however be useful as a linear model in the initial planning phase to create functional requirements.

Soft Systems

Soft Systems methodology combines a logical and cultural analysis into developing both system design and human activity systems. The cultural analysis looks into social and political aspects of a real-world problem. The logical analysis is based upon comparison with relevant systems, models, and the current situation. The logical analysis covering comparisons of relevant systems and situations is a meaningful approach, and is to some extent what has been conducted in this thesis by including literature of disease surveillance systems with comparable properties to the FBP. The method will however not be adopted because it has slightly more focus on the organizational development, at the expense of losing focus on system design of the FBP. Another contributing factor is the perceived high level of complexity of the Soft Systems model, which increases the threshold for understanding and adopting the method.

Information System prototyping

Baskerville justifies Information System prototyping as an IS action research method by pointing at five aspects:

“(1) it moves the design process into the user’s multivariate social setting, (2) it permits highly interpretive assumptions about observation allowing highly qualitative data, (3) it represents an intervention by the designer into the user work setting, (4) the designer is conducting participatory observation about the suitability of the design, and (5) the designer is studying the impact of design changes in the user’s social setting.” (Baskerville and Wood-Harper, 1998; p.98).

Points 3 and 5 describe an intervention into the users work or social setting. By not having access to the actual end-users of the prototype, the DSU’s in India, we cannot fully adopt this method, as user experimentation and testing are central elements (Figure 8). In our case, the HISP India team are intermediaries to the actual end-users. The idea of IS prototyping is linked to cooperative design, bridging the

competence of developers and end-users of interactive systems, into the participatory decision making of design (Kyng, 1991).

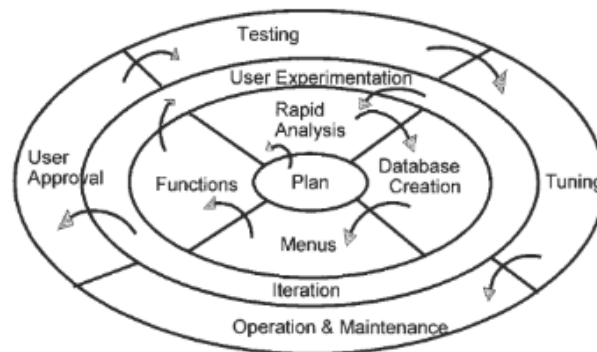


Figure 8: Shows a model of evolutionary prototyping originally from Connell and Shafer (1989) (Baskerville and Wood-Harper, 1998; p.99)

3.2.4 Adoption of method

The selected method for this thesis includes elements from Information System Prototyping, Multiview and canonical action research.

The Multiview method was adopted to support the initial analysis of the system design. These phases are essential to understanding the end-users, their work methods and their need for information, and for setting the requirements for the system design.

The information system prototyping has an advantageous structure in terms of being supportive of changes, as new information emerges from data sources, which may impose changes during the development. Also, the user-centric approach, involving testing and user approval, is fundamentally aligned with the aim of the research as the objective of the thesis is addressed through design, development, and testing of the FBP. This was the framework of the phases in the iterative development, but some of the cyclical elements were replaced by more flexible and generic phases of canonical action research. The elements of the IS prototyping are too rigid for supporting variations of objectives in each iteration. E.g. Some iterations will simply not concern creating of menus or databases, as described in Figure 8. The evolutionary process model of Connell and Shafer (Connell and Shafer, 1989) in Figure 8, serves as a starting point; however, the internals will be swapped in the

iterative development (database creation, menus, and functions), and replaced with a more generic approach drawn from canonical action research. These internals of each iteration were used: rapid analysis/action planning, acting, and evaluation.

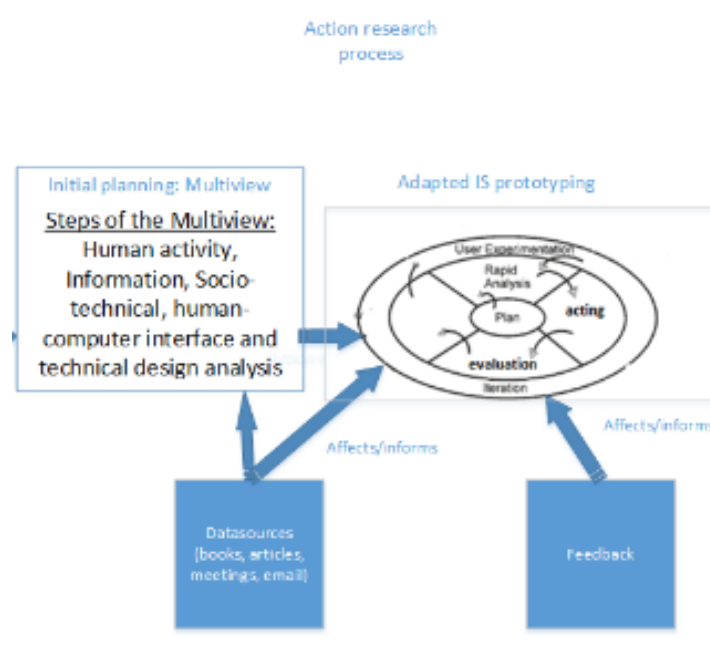


Figure 9: Shows the adaption of IS action research methods, including Canonical Action Research, Multiview and IS prototyping. Multiview is used for initial planning, while a combination of IS prototyping and Canonical Action Research is used for the cyclic development of the FBP.

3.3 Data sources

3.3.1 Field trip: 30. April – 16. May 2017

Already in December 2016, only a short time after the choice of Master's thesis, me and Nirujah Kirpanithy, who collaborated on the development of the FBP, but with different research goals, began initiating talks about a field trip to visit HISP India. As I had never visited a developing country before, we decided that this trip was important for putting both of us into the right mindset for understanding the context of this thesis. Also, the trip was important for really getting to know the HISP India team as well as the disease burden of India. The initial plan was to split the trip into two parts. In part one we were supposed to attend to the Disease surveillance level 2

conference in Goa, hosted by the HISP network. However, due to the overbooking of participants, we were unable to attend. As result, we traveled directly to the HISP India office located in New Delhi.

During our first week at the HISP office in New Delhi, several key persons were missing from the office as they were attending the Goa conference. So, we spent this time getting information about the disease burden of India and how the India health care system works. Dr. Gaurav Mishra of the HISP India team played a central role in this introduction.

In the second week, the focus shifted over to developing the research question and concepts. As we were having frequent meetings and idea discussions with the local team members, this became a productive week, resulting in us defining our respective research aims and building agreements upon initial mockups for the desired functionality.

3.3.2 Field journal

A lot of information was processed during the field trip, and events of relevance were documented in a field journal. The field journal was used for describing events, meetings or decisions made during the field trip, and later extended in terms of becoming a tool for documenting outcomes of important meetings or decisions along the project timeline.

3.3.3 E-mail

E-mail has been the primary tool for communication with HISP India, though there have been a few sporadic uses of video communication tools and chat-tools. E-mails were used for project coordination and feedback during the development process. At the end of each iteration along the timeline of the development process, a status report mail was sent to HISP India that covered descriptions of:

- Implemented features in the current iteration
- Issues or bugs
- Planned implementation in the next iteration
- Screenshots (if any) of implemented features

3.3.4 Meetings

Whenever physically collocated, we had meetings with HISP India representatives, mainly in Spring 2017 during the field trip or during the occasional visits of HISP India team members to Oslo. Meetings were the most productive and effective data source, covering a lot of information exchange in a short time.

3.3.5 Books and articles

Relevant literature is primarily found at UIO's literature search engine "Oria", providing access to a range of books, journals, thesis, etc. Secondary literature sources included; Google Scholar and UIO online sources. An overview of central literature sources:

- DHIS literature: from UIO intranet and UIO public web pages.
- Disease surveillance in India: Literature borrowed from supervisor and Oria search.
- General HIS/Disease surveillance: Articles UIO courses (INF5761), Oria search.
- Geographical information systems: Oria search.

3.4 The research processes – use of the method

When traditional software development is encapsulated into a research project, it spans various tools, practices, methods, and data sources, etc. For best organizing this complex landscape, a model was made of the research process from project start to final prototype (Figure 10). This model acted as a methodical roadmap, describing the overall research process adapted from Information System Prototyping, Multiview, and canonical action research.

The initial step in the process was to develop the concept and the research question for the thesis. Next step included the initial planning where all the phases of the Multiview method were used: human activity analysis, information analysis, socio-technical, human computer interface and technical design analysis. For the iterative development, a slightly adjusted version of information system prototyping method

was applied. Instead of only focusing on end-user experimentation as a feedback channel, also relevant literature was heavily emphasized in each iteration.

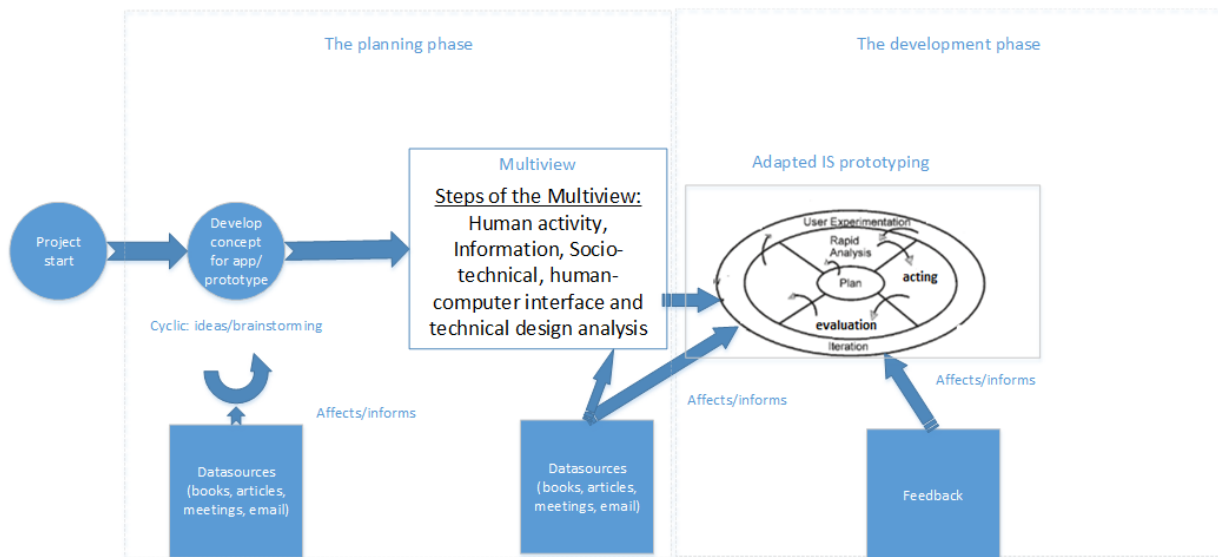


Figure 10: Description of the overall project process, from project start to final prototype.

3.4.1 Phase 1: Planning

Description

The planning process started out spring 2017. Together with the HISP India office, we researched different approaches on how to proceed with our project, and what kind of challenges we should try to address. Initially, we started to investigate how we could link climatic factors like temperature, humidity, and rainfall, into a DHIS2 driven application that could make a forecast of disease distribution in particular geographical areas. However, after discussions with HISP India and our contact person at the Centers for Disease Control and Prevention (CDC), USA, it appeared that the task was too comprehensive, and it would become hard to access the required climatic data sources.

Later, during our field trip to HISP India in May 2017, we began working on yet another approach. An application for modeling a tracked entity to facility relationship on the DHIS2 platform. In Figure 11 we show how this relationship may be utilized for disease surveillance purposes. The planning of the design consisted of three main sections: 1) a dataset selection part, much inspired by the inbuilt Data Visualizer app in the DHIS2 platform. This would have to at a minimum display data regarding organization units, programs, and a time interval, 2) A GIS application, representing

the spatial dimension modeling facilities and tracked entities, 3) A timeline or temporal dimension that could inform of facility burden over time.

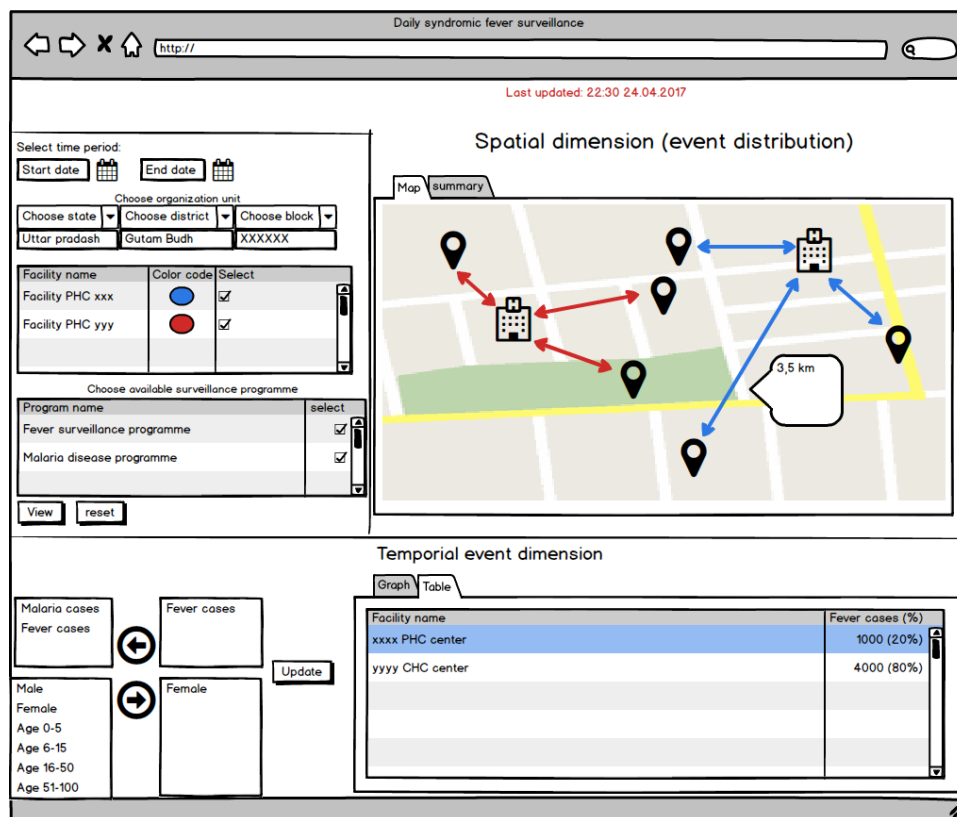


Figure 11: This final draft was used to model the functionality and graphical user interface for the development process.

Note that the planning process covered the main characteristics of the application, not how the application was to be developed after the lifetime of the project.

Methodological approach

From a methodological point of view, this phase was greatly influenced by brainstorming with HISP India, where several thesis ideas and relevant factors of feasibility, like scope, topicality, access to relevant data, etc., were discussed. Central data sources were mostly meetings and email with HISP India and the CDC contact point, however, relevant literature was included to improve knowledge of concepts prior to discussions.

During the field trip, we had several conceptual ideas that needed to be communicated successfully within the team, and between the team and central HISP India employees. Verbal communication became insufficient, so a mockup application called “Balsamiq” became a central tool in this phase. As in all software development – the customer often did not know their needs, before they actually saw

it. Balsamiq was used for describing functionality by the use of mockups. Several mockups were made during this phase, Figure 11 shows the final version, that represents the outcome of this phase, and input to the development process.

As soon as the concept was successfully agreed upon and communicated by the use of mockups, the planning phase addressed an analysis of the linear activities of the Multiview method: human analysis, information analysis, socio-technical analysis and design, human-computer interface design and technical design. This was necessary to get an understanding of the system and the end-users. Even though the conceptual phase involved several mockups of the user interface, these were mainly used for communicating the concepts, not for making a detailed analysis of the human-computer interface design. However, these mockups were also a useful starting point for conducting the human-computer interface design analysis derived from the Multiview method. More about the content of this phase in the chapter 4.

3.4.2 Phase 2: Development

Description

This section is a description of the development phase (Figure 10) which started in Autumn 2017 and ended late January 2018. The development was divided into five distinct iterations. Each iteration was manifested as a part of functionality, improvements or corrections.

Methodical approach

The steps used for the cyclic prototyping will be explained here:

- **Rapid analysis/Action planning:** Most often, one approach was developed, but different approaches were evaluated when it was natural. Data sources were used to influence the course of action, in terms of feedback by e-mail or literature. Note that sometimes the course of action was influenced by the available time left in the iteration or project.
- **Acting:** Implementing the changes decided upon in that iteration was manifested through use case diagrams.
- **Evaluation:** Sometimes it was hard to measure the effect directly, but user interface testing and feedback were actively used for measuring if the diagnosed problem was resolved, remained, or alternatively if the outcome

introduced had unintended bi-effects. The evaluation included the level of achievement of objectives, as well as the level of technical achievement. At the end of each iteration, designed screenshots and a summary were sent to HISP India team for feedback.

Tools

The tools described in this section were crucial for project feasibility, through improved management and communication of functional requirements, both internally and externally, for the length of the project.

Microsoft Visio and use case diagrams were used as a tool for communicating generic functionality throughout this phase (Figure 12). A use case diagram is a simple drawing that describes an actor and his or her intentions or required functionality in a system.

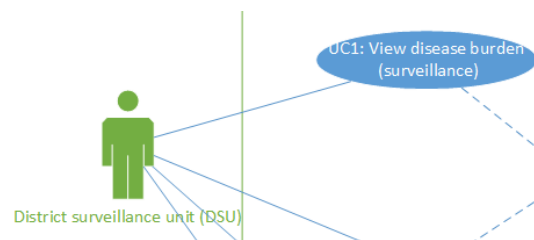


Figure 12: A small section of the use case diagram in this thesis, for illustrating the look of a use case diagram.

Next, Microsoft Excel sheet was used for overall project tracking (Figure 13). More concretely planning sprint durations and keeping track of external events like exams. It also describes the main features that are planned to be implemented in each iteration.

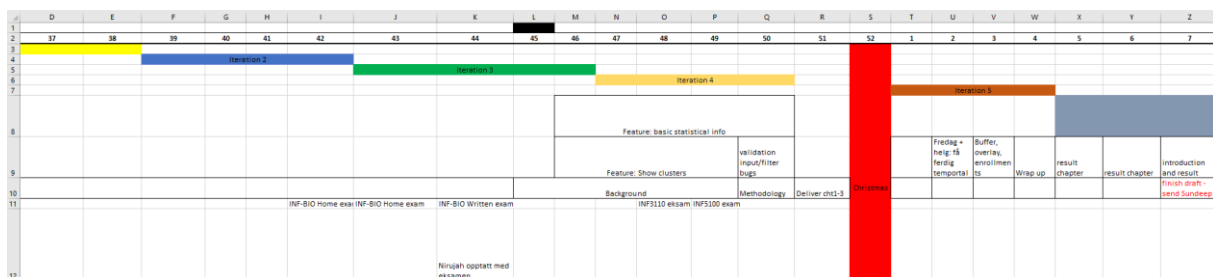


Figure 13: Showing the excel sheet used for overall sprint planning and time management.

However, the Excel sheet became too coarse-grained for tracking individual tasks and their progress. Trello, a simple project task tracking tool, was introduced to break down larger tasks into smaller ones, and to keep track of progress (Figure 14).

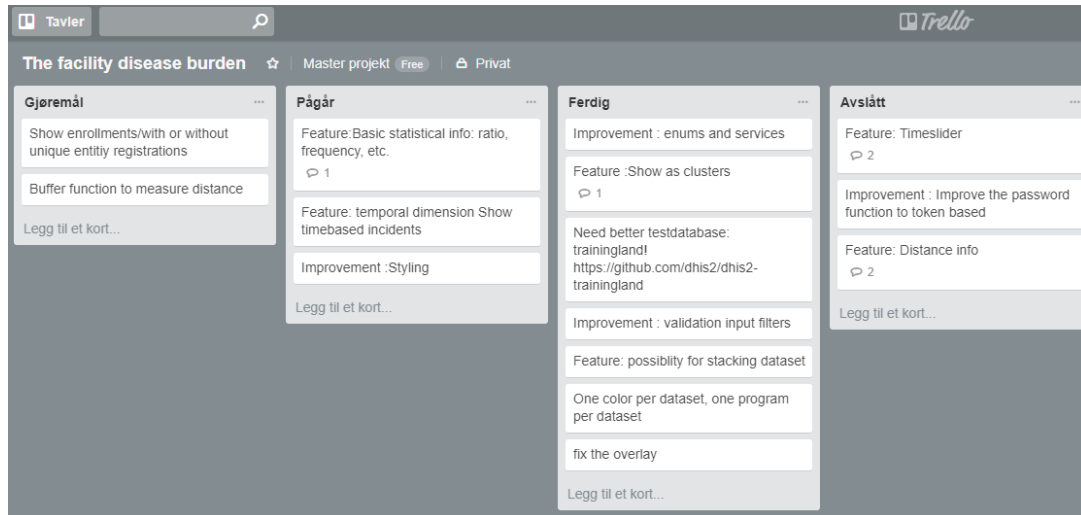


Figure 14: Shows the dashboard on Trello.com, a project management tool for tracking tasks. The columns display tasks that are not started, in progress, finished and rejected tasks.

3.5 Validity

3.5.1 Internal validity

Internal validity relates to what extent there was a strong link between central parts of the thesis and reality, and whether conclusions can be made from theory and method (Krumsvik, 2014). More concretely, this related to understanding whether the FBP will meet the requirements of DSU's in India to support evidence decision making. The greatest concern of internal validity was linked to feedback, more precisely having too sporadic feedback and not having feedback from the actual end-users.

The thesis has been a participatory project, making HISP India per definition the organization from which the research has emerged from. During the development phase of the FBP, there was sporadic feedback on the reports that were sent in each iteration. The two most dominant factors were lack of communication and difficulties with making a suitable test environment, so that our Indian coworkers could execute early prototypes. Another reason was linked to the content of iterations, where non-functional requirements were harder to test than functional. The bottom line is that earlier and more regular feedback could have influenced the design and choices.

The other influence on internal validity was linked to not running tests on actual end-users. The feedback from HISP India would not necessarily be sufficient to prove that

the system is designed in a way that really meets end-user's requirements. However, HISP India has a high level of knowledge of HIS in the setting of India, making their feedback based on local knowledge a valuable part of this thesis. The circle of participants was somewhat narrow, affecting the establishment of validity. As validity in action research is also based on the experience and perspectives of the participants, it is likely that including end-users like DSU and personnel with specific skill sets, could have been a positive influence. Committing acceptance testing with testing teams, consisting of relevant personnel from UIO within public health or epidemiology and testing with local DSU teams were also considered, but the latter would involve more field trips to India. Another idea was to bring the prototype for a DHIS2 academy for disease surveillance, as a lot of competence would be gathered in one place.

Another critical perspective was around the choice of literature. American and European literature is very prominent, increasing the distance the setting of a developing country, like India. However, at some point, there is a need to generalize the assumptions of background, education, and working methods of the end-users (DSU), as India is a large country of great diversity. To some extent, the variation of end-users was addressed by building flexibility, an otherwise dominant property of the DHIS2 platform.

3.5.2 External validity

External validity refers to what extent the conclusions of the thesis are transferable to the surroundings, as in other research projects (Krumsvik, 2014). Creswell claims that external validity is often not emphasized in qualitative research (Creswell, 2003), as the research is often very narrow, subjective and situational. However, comprehensive documentation of the research, as emphasized in the creation of the FBP, will make the individual researcher able to determine if the research is transferable.

3.5.3 Reliability

Reliability in research is linked to whether another researcher could replicate or reproduce the results of this thesis (Krumsvik, 2014). The process of making the

prototype is carefully documented, including analysis of requirements and choices, making the process of developing the prototype transparent. The HISP India has access to the final source code for deeper analysis of the technical implementation. The prototype has also been distributed to HISP India as a compiled application, for evaluation and testing.

Being developed with a research method that emphasizes subjectivity, increases the likeliness that projects with the same research question and methodology will produce another result unless the researcher is actively intending to copy the result. As in external validity, achieving a high degree of reliability in action research could be problematic, as the research is affected by situational aspects of a social and organizational character. The context makes it hard to replicate the research, thus affecting the reliability (Baskerville and Wood-Harper, 2016). This makes it important that the researcher documents choices, and the factors or context that leads to an outcome. However, transferability across qualitative research projects could become a challenge without any standardized patterns for conducting the research.

3.6 Evaluation and reflection

We had the potential at exploiting meetings as data sources, as meetings with local sources within the Global Information systems group of the Institute of informatics at UIO, would be helpful for e.g. acquiring technical DHIS2 related information.

We also learned that IS methods with the end-goal of creating or modifying systems, should include a plan for testing the prototype or system. The researchers should define a testing environment that enables the testers to test a working prototype of the product for each iteration. For each testing phase, enclosing a document to describe the structure of the requested feedback, could have been a necessary tool for making the testing more formal and comprehensive. Better and earlier testing would probably have increased the communication between the FBP developers and HISP India team while securing that all the expectations and requirements of the FBP are met. The feedback from the HISP India team on the final product can be found in Appendix A.

4 Results

4.1 Requirement analysis

4.1.1 Introduction

The purpose of this chapter is to make the process that led to the creation of the FBP transparent, more specifically review the process of making the requirements that have been implemented into the app. The requirement analysis was conducted as a part of the initial analysis of the FBP, but was adjusted throughout the iterative development process. This chapter will emphasize the functional requirements (use cases), but also include a portion of the non-functional requirement; what the system should do and how the system should behave. The only functional requirement set by HISP India, requires the FBP to support a spatial presentation of relationships between health events and the originating facility, making this the backbone of the FBP. The remaining requirements are addressed by conducting 5 separate analyses, derived from the Multiview method:

- 1) Human activity analysis, reviewing the work method of end-users to determine information needs required to enable evidence-based decision-making. This will to some extent overlap with the information analysis.
- 2) Information analysis, organizing data in the DHIS2 data structure, to highlight the options and limitations the system design of the FBP is bound to.
- 3) Socio-technical analysis and design will emphasis interpretability of data.
- 4) Human-computer interface design, focusing on dissemination of data.
- 5) Technical design, reviewing available surveillance system designs that support the research objective, as well as the previous steps.

4.1.2 Human activity analysis

One of the challenges of identifying requirements for the FBP was linked to the uncertainty of who the specific end-users of the tool were, and the skillsets they

possessed. Our initial understanding was that the tool was intended for being used by decision-makers at the district level in India, in the district surveillance units (DSU). The DSU's are elements within the Integrated Disease Surveillance Project (IDSP) – a government funded disease surveillance program in India. As of 2015, the positions available in a typical DSU in the IDSP were limited to 1 epidemiologist, 1 data manager, and 1 data entry operator. Most of the positions were contractual with a high level (41%) of unfilled posts of epidemiologists (Ministry of Health & Family Welfare Government of India, 2015). This indicates a problematic human resource situation in most of the districts.

The FBP will be designed to support the DSU teams in a public health assessment by discovering the decision problem, planning the treatment, and analyzing it. In general, decision making in public health is often based upon an assessment based upon the use of available data sources to systematically collect and monitor disease burdens and detect abnormalities. Next, an identification of available resources and their effectiveness, before the assessment is presented the decision maker (Keppel and Freedman, 1995). An assessment represents the set of activities conducted before the information is presented to-, and interpreted by decision makers. The public health assessments will support the decision makers in defining the quantitative aspects, defining an acceptable outcome, all which are components of a well-defined problem (Grünig et al., 2013).

Epidemiological approach

Epidemiologists, who are the data analysts, examine the surveillance data. The working method of the epidemiologist includes:

- 1) Collecting case-patients based on a case definition, which is analyzed by descriptive epidemiology, including the health event in time, place and person.
 - a. An epidemiology will start by analyzing the data. The data quality will affect this investigation.
 - b. Next, the user will seek to learn about patterns or trends of the health concern in time and place
 - c. Combine the data into presentation formats that can be communicated

- d. Identify areas or groups with heavy disease burdens and use that information to develop a testable hypothesis regarding the source or mode of transmission of the disease.

- 2) Use analytic epidemiology, field investigations, to test a hypothesis and find risk factors and causative agents

(Centers for Disease Control and Prevention, 2012b, Centers for Disease Control and Prevention, 2012c).

Decision making regarding a health issue from an epidemiological point of view, will determine the response options based on the knowledge about the source/transmission mode and causative agent (Figure 15).

		<u>Source/transmission mode</u>	
		Known	Unknown
<u>Etiology</u>	Known	Investigation + Control +++ Example: Hepatitis A in day care (6)	Investigation +++ Control + Example: Salmonellosis and marijuana (7)
	Unknown	Investigation +++ Control +++ Example: Parathion poisoning (8)	Investigation +++ Control + Example: Legionnaires' disease (9)

Figure 15: The figure shows relative emphasis of investigating and control efforts based on the causative agent and source/transmission mode (Goodman et al., 1990; p.11). The + symbols indicate the level of importance in the given situation.

The sources and modes of transmission are considered to most often be unknown, as most field investigations are a response to detecting the cause of an outbreak or possible outbreaks (Goodman et al., 1990). According to the investigation/control matrix of Goodman, both dimensions would be unknown in case of linking the FBP to a syndromic program (lower right corner of the matrix), as syndromic surveillance systems most often have no knowledge of the causative agent, nor sources or transmission mode. This would emphasize field investigations. Also, if linking to programs that are based on confirmed cases definitions, it would point to the need for field investigations (upper right corner in the matrix). Because field investigations can be a costly affair, both in terms of time and money, it is essential to make the

investigation as efficient as possible. By having information about whereabouts of infected people could greatly speed up the process (Goodman et al., 1990), thus affecting the following control phase. The tracked entity-facility model could support the planning of investigations, by providing the investigators/response team with a residential location of the reported cases, but also the location of the facility that had enrolled the patient. When having information about which facility the patient was registered, it also provides the investigators with traceability. E.g. If an ongoing investigation targets an area or a neighborhood with high case density, they need to identify and count cases or exposures (Goodman et al., 1990). If there is a high gap between reported cases and actual cases, one reason could be that they are enrolled as inpatients in the reporting facility. It is hard to track if a patient is an inpatient or outpatient, however knowing the two most likely whereabouts of a patient, provides information that contributes to decision making during a field investigation.

Human analysis requirements

Based on the analysis of the end-users, their work methods and information needs to address the predefined decision problems, a list of requirements are compiled:

- System designs must support the epidemiological approach, enabling the end-user to conduct a descriptive analysis of available case data.
- The system should support the epidemiologist by uncovering the etiology and source or transmission mode of the causative agent.

4.1.3 Information analysis

DHIS2

Practical considerations in terms of technical challenges and limits were made by analyzing the DHIS2 documentation. The DHIS2 core holds the raw data, which made a study of the platform's programming interface necessary. In the context of the information circle and the FBP, the DHIS2 platform will cover data collection, and to some degree data processing.

To access or query case based tracked entity data from the platform, three parameters must be set:

- 1) An Organizational unit, which is related to where the tracked entity has been registered.
- 2) Time interval, in which the tracked entity has been registered.
- 3) A tracked entity, which must be related to a program. E.g. a Malaria program.
- 4) Additional parameters may be set, to limit or specify the query.

The analysis of the data structure of both tracked entities and organization units show that they have properties that support map projections, in terms of latitude and longitude. However, both must be manually configured at some point.

To put the application in a practical setting, there is a need for addressing the platform specific metadata structures for supporting the FBP. The metadata structures of DHIS2 are highly configurable and flexible, which makes it able to support different contexts. However, two potential setups for integration with the FBP will be illustrated. Note that the application of the FBP is not limited to these examples:

One mixed syndromic/disease-specific program:

- 1) The tracked entity gets enrolled into the program with reported syndromes (e.g. Fever). The clerk puts a disease/syndrome attribute to “provisional” or “suspected”.
- 2) When the lab results arrive positive, the clerk set a tracked entity attribute to “confirmed”, e.g. a Malaria confirmed case. The tracked entity would also need an attribute saying what disease the patient has been infected with.
- 3) To have traceability of tracked entities across facilities, we need to consider this example: Two weeks after the tracked entity reports to another facility. The clerks search for the tracked entity in the tracker capture module, then finds him/her and enrolls him into the new facility.

The issue with this setup is traceability. The app could not differentiate if the tracked entity had a specific disease or just some syndromes in a historical perspective, as the attributes that identify a suspected or confirmed case is constantly overwritten. The benefit is a fast update and overview within one single program for both syndromic- and disease surveillance purposes.

One syndromic and one disease-specific program:

- 1) The tracked entity gets enrolled into the program with reported syndromes (e.g. Fever), and the clerk sets a disease/syndrome attribute to “provisional” or “suspected”.
- 2) When the lab results arrive positive of e.g. Malaria, the clerk enrolls the tracked entity in the Malaria-specific program.

This setup makes historical data available for temporal analysis, and makes it possible for traceability of tracked entities across programs, through the use of a unique identifier. This would be the preferred setup. The great flexibility in DHIS2 configuration enables the users to tailor their own program setups to fit situational needs, which makes it important for the FBP to extend the flexibility of the platform, by having loose dependencies.

Information analysis requirements

For the DHIS2 core to support the real-time analytic capabilities required in the FBP, it puts some constraints on the data structure of the platform:

- Programs should be separated by diseases.
- Both tracked entities and organization units must have a coordinate attribute, to enable spatial projections.
- Tracked entities must hold attributes that are possible to disaggregate data on, e.g. gender, age.

4.1.4 Socio-technical analysis and design

The relation between data interpretation and presentation of information was explored in a study by Hibbard and Peters (Hibbard and Peters, 2003). They examined how healthcare consumers made informed choices when presented with comparative information. Decision makers show a tendency of making shortcuts when being confronted with information involving many variables in a single choice. When comparing variables against each other to determine the importance, their cognitive capacity is challenged. The study concludes three means to address these cognitive challenges:

- 1) Implementing an information system as a decision tool.
- 2) Build human expertise.
- 3) Transforming information to meet analytic capabilities of the users. E.g. simplify concepts.

Human expertise cannot be an option, as the users are already the experts.

Transforming information or simplifying concepts is likely to be a better approach when dealing with healthcare consumers, as professionals often have an education or background that makes them comfortable with more advanced concepts and terms. Simplifying concepts for professionals could be confusing for the user, or even insulting. However, this study is based on health care consumers, not professional public health workers, but the cognitive aspects of comparative and organized information seems to be transferable.

Regardless of the user level, information must be organized in a way that communicates important factors and trade-offs. In a laboratory experiment by Hibbard and Peters, they showed that the way that the information is presented, e.g. to emphasize a certain property, may change the outcome of the decision (Hibbard and Peters, 2003, Rouse and Rouse, 1984). It is important to notice that the presentation may also be intentionally or unintentionally used to manipulate or mislead the user. However, the interpretation of comparative information is based upon the assumption that the user has a reference to an ideal situation or a goal, making this an important consideration before presenting information.

The end-user needs to identify a method to reach an ideal situation. This method will be the control measure for the given epidemic situation. The information system must help decision makers analyzing the decision problem and make progress towards identifying a causative agent and source/transmission mode of the disease, to improve evidence-based decision making. Decision makers tend to be reluctant to make decisions they may later regret (Tversky and Kahneman, 1974). However, if the decision maker has conducted a thorough analysis of the decision problem he/she may feel more comfortable (Wagner et al., 2006). Also, having users define their own datasets, may increase the confidence in the information, as often non-transparent information may not be considered reliable (Wagner et al., 2006). The

process will become more transparent if the end-user may generate his own information by combining datasets or data streams.

The literature also suggests that the transfer of terminology, name conventions, and concepts from the DHIS2 platform into the FBP is likely to lower the threshold for adopting the system (Lechner and Fruhling, 2014).

Socio-technical analysis and design requirements

- Concepts and terminology must match the user level
- Pursue neutrality in the presentation of information, to avoid misleading the decision makers to emphasize certain properties.
- Summarize information to avoid overloading the cognitive capabilities of the end-user.
- Pursue transparency in the process of generating information, through the user interface, for reducing any mistrust or concerns. E.g. by use of user-defined datasets.

4.1.5 Human-computer interface design

The actual representation will be of a visual character, which makes us consider some visual guidelines. Simon Jespersen (2017) has formulated a set of guidelines for making dashboards for a DHIS2 platform. His research is not specifically targeting Apps, but due to the similarities of DHIS2 dashboards and the dashboard setup in the FBP within the same platform or context, contributes to a high level of transferability. The guidelines that can be considered most useful are:

- 1) “Display essential data: data only needed for evidence-based decision making**
- 2) Data need context: provide a context**
- 3) Color palette: limit the number of colors”**

(Jespersen, 2017; p.69)

The end-user will need to have a reference to the spatial distribution in another point in time, to be able to put the data in context and determine if the gap between current and an ideal situation is large enough to act upon. As suggested by Lechner and Fruhling (2014), dashboards should have temporal capabilities, presented by charts and graphs, whenever the data has a temporal dimension.

Human-computer interface design requirements

- Data should be organized and limited to “need only data”.
- Data should be comparable, thus have a context.
- Possibility for disaggregation of data.
- Present temporal data, whenever the data has temporal aspects.

4.1.6 Technical design

As Rouse and Rouse (1984) state about information systems, the decision problems at hand have a strong influence on the system design:

“[...] if a problem can be pursued in a number of ways, perhaps some of them unanticipated, then there are likely to be many design alternatives and it may be quite difficult to choose the most flexible and effective alternative.” (Rouse and Rouse, 1984; p.133)

The thesis has thus been scoped intentionally to limit the types of surveillance systems into case- and syndromic surveillance systems – to investigate the effect of applying the tracked entity-facility model in that particular system designs. Even though the types of systems are predefined, a variety of analysis types must be selected (Figure 16).

Table 6–1 Types of Surveillance Systems and Associated Analysis Types and Methods

Data collection design	Analysis type	Analysis methods	Examples
Case surveillance	Case counts/rates (D)	Rate calculation, rate ratios, standardized rates	Number and rates of violent deaths in the US (5); age-standardized projected rates of arthritis (6)
	Trend over time (I)	Regression analysis to describe estimated annual percent change in case reports	Trends in HIV diagnoses in men who have sex with men (7); trends in pancreatic cancer (8)
	Geographic clustering (A)		Clustering of road injuries in Rome (9)
Syndromic surveillance	Geographic clustering/aberration (A)		Pesticide exposure (10)
	Temporal aberration (A)		Symptoms in companion animals after an industrial chemical release (11)
Supplemental surveillance systems			
Behavioral surveillance	Prevalence of behaviors (D)	Related to design (e.g., weighted analyses from probability surveys)	HIV testing from among men who have sex with men (12)
	Trends in prevalence of behaviors over time (I)	Age-adjusted rates, linear regression analysis for monotonic trends	Prevalence of smoking in the United States (13)
Clinical outcomes surveillance	Rates of clinical outcomes (D)	Incidence rates from probability sample	Prevalence of anemia among renal dialysis patients (14)
	Factors associated with clinical outcomes (I)		Factors associated with thrombosis in patients with HIV infection (15)

D: Descriptive; I: Inferential; A: Aberration detection.

Figure 16: Displays types of surveillance systems and associated analysis types and methods. (Lee, 2010; p.91).

Geographical abbreviation/clustering

There is a widely used “80% axiom” that 80% of healthcare and emergency management information has a geographical relevance:

- “Eighty percent (80%) of information needed for decision making has a location or spatial component (Folger, 2009; City of Boston, 2009); Yong et al., 2008).
- More than 80% of all healthcare transactions are believed to have significant geographical relevancy (Davenhall, 2003).
- As much as 80% of information used during emergencies is “spatial” information”

(Ric Skinner, 2010; pp.1-2)

The spatial component has a significant role in healthcare, the tracked entity-facility model being no exception. For this purpose, we use GIS as a facilitator for decision

making by modeling spatial distribution of cases. The analysis methods will be based upon manual abbreviation detection, by use of comparative time-bound datasets.

The two most common methods of presenting spatial information are Choropleth maps and Spot maps. Choropleth maps are dividing maps into regions based on aggregated, statistical data. The regions will be visually organized by color coding. However, because of aggregated data, local variation may not be communicated, which also is an unfavorable trait when demonstrating a visual tracked entity- facility relationship. Spot maps are displaying single health event or cases of diseases on a map, which also may convey spatial trends or patterns (Lee, 2010). Spot maps are the selected method for dissemination in the spatial dimension of the FBP.

A relevant analysis method that could have been applied to the tracked entity-facility model, is geographical clustering. This method is mostly relevant for syndromic surveillance, but has clear advantages in terms of algorithm-driven cluster detection, which lowers the manual labor linked to the discovery of clusters. The analysis method will not be used as an algorithm-driven geographical cluster technique because it would have turned the thesis into a study of algorithms. Also, it would imply a design that is geared toward syndromic programs, at the expense of potential integrational challenges with traditional case disease surveillance programs. A broader audience to demonstrate the application of the model is emphasized, thus discarding this method. However, future research may use the experience to develop more specialized systems.

Displaying the visual relationship between tracked entities and facilities will be presented with either polylines (lines on a map) or color coding. Because the FBP needs to support multiple datasets for comparative purposes that will put the data in context, the color coding is reserved for datasets, leaving the tracked entity-facility model to be described by polylines. Applying polylines on a map comes with a bi-effect of spatial noise when e.g. displaying 10 000 tracked entities, which also implicates 10 000 polylines on the map. The solution is to apply a hierarchical cluster method, effectively limiting the number of polylines to the number of clusters on the map, also increasing the performance of the GIS. This method will adjust the size of the clusters based on the scale level on the map and is found to be a suitable compromise. A hierarchical clustering method aggregates the information into a more

readable format, which complies well with the design guidelines of Lechner and Furling for public health dashboards (Lechner and Fruhling, 2014).

Temporal abbreviation

A temporal dimension is not strictly required to model the tracked entity – facility relationship, but abnormalities in disease cases cannot effectively be modeled in a spatial environment - the “time” aspect of health issues reported. Also, outbreak detection will be challenging without a historical reference of a “normal” or a comparative disease burden. Comparisons between organizational units or groups of people in a historical perspective are often required to view trends or patterns that may occur. For this purpose, we use charts and graphs. Analysis methods will be manual abbreviation detection of incidents (cases) over time. Without a temporal dimension, it is also hard to have a reference to such as average or expected rates or values, as diseases vary over time.

There exists a variety of analysis types and methods, however, most of them cover aspects like rates, ratios and regression analysis making them more suitable for aggregated data structures. These types are considered hard to apply a tracked entity-facility model on, due to the data structure of a DHIS2 tracked entity. A tracked entity will hold a more comprehensive data structure, like information about attributes. This makes queries more resource demanding than fetching aggregated numbers from e.g. a monthly report. The FBP will run in a browser environment at client-side, since in a developing country like India, there may be limitations in computer hardware performance, making the model basically unsuitable for aggregated analysis methods. However, if a system were made, based on aggregated data extracted from tracked entities, it would need to enforce specific data attributes upon the tracked entity structure, to collect required data. This approach would deviate from the intended use of tracked entities.

The information analysis of the DHIS2 structure indicates sufficient information to plot a graph of the aggregated number of tracked entity enrollments, over a time interval within the dataset. This will communicate the burden or load of enrollments at a certain point of time, emphasizing on patterns or trends. Line graphs are considered most suited for such displays (Kwapien, 2016). A trendline is added for each graph as an optional feature. A trendline is a suitable tool for displaying the general trends a

graph. However, the length in the time interval which the line graph spans, will decide the utility of the trendline.

A second output will be to show the total number of enrollments over the entire interval. A doughnut chart is used for that purpose. Pie or doughnut charts are not considered as the best display when having a great span in values, as the value range may result in losing the visibility of lower values (Kwapien, 2016). However, it's a very intuitive and easily readable chart, which also was complemented by displaying the raw value of a total number of enrollments. Another option is a bar chart, but again this would come with same weakness when comparing high and low values. When only displaying one single dataset, this chart has no real use, as the usefulness of this chart appears when comparing two or more datasets.

System design requirements

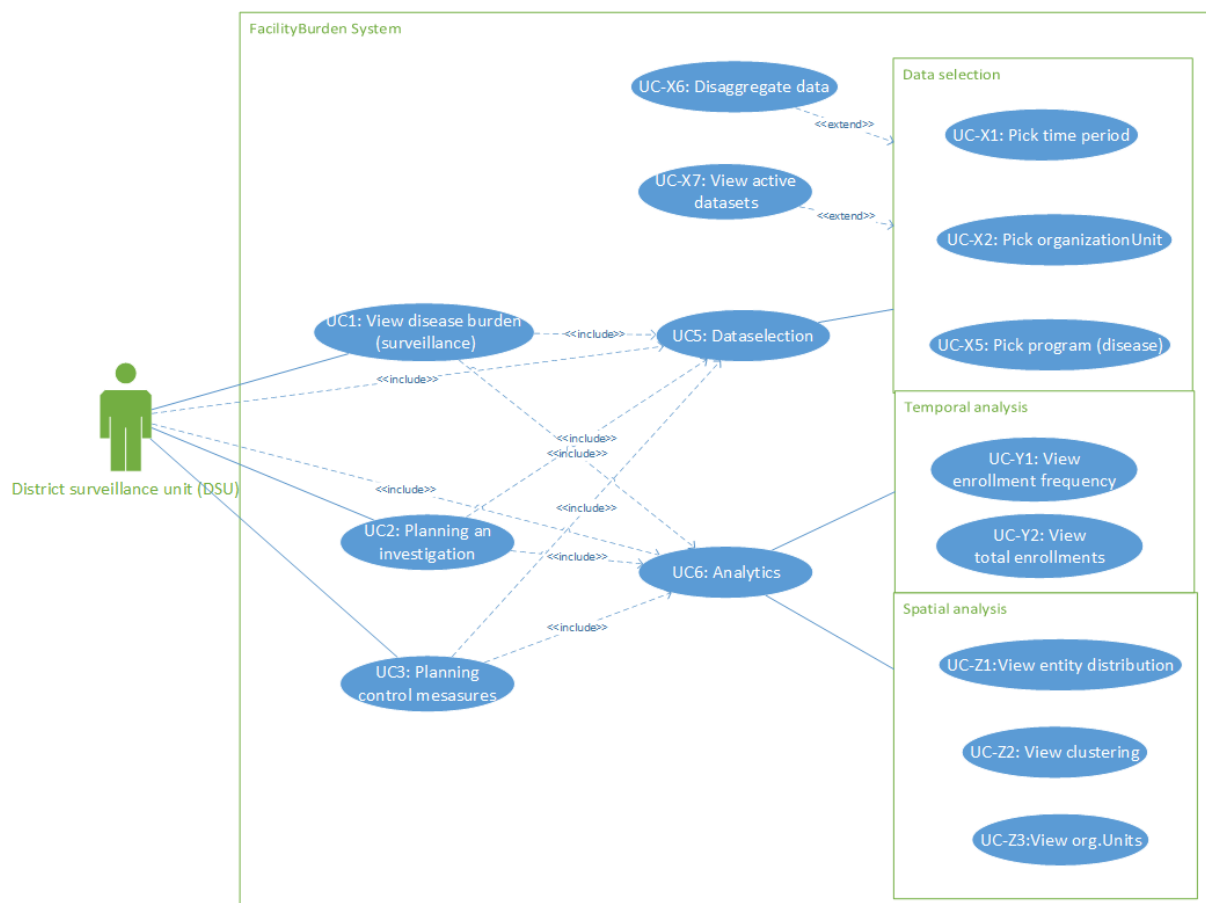
The system design has been influenced by the initial requirements of emphasizing spatial projections, while recognizing the temporal aspects used for supporting the epidemiological approach for establishing the existence of outbreaks, supporting field investigations, and implementing control measures. The requirements are:

- Case surveillance system and syndromic surveillance (predefined).
- Geographical abbreviation/clustering, facilitated by a GIS. Geographical clustering in this context means clustering for exploratory data analysis, not algorithms for defining clusters based on distance and frequency.
- Temporal abbreviation analysis for detecting time-bound patterns and deviations.
- If the FBP is to support decision making based on complex or ill-structured problems, flexibility in the spatial decision environment should be considered (Densham, 1991).

4.1.7 Summarized requirements

The initial system design resulted in three main sections: 1) a dataset selection part, inspired by the inbuilt Data Visualizer app in the DHIS2 platform. This would have to minimum display data regarding organization units, programs, and a time interval, 2) A GIS application, representing the spatial dimension modeling facilities and tracked

entities, 3) A timeline or temporal dimension that could inform of facility burden over time.



Use case 1 – View disease burden

and what. These keys limit the datasets, while the output is manifested in spatial and temporal information. As this is a real-time surveillance system, it could be used for determining if implemented control measures in case of an outbreak, are having any effect.

Use case 2 – Planning an investigation

The FBP can be utilized for planning investigations. An investigation may be planned or even carried out based on the “where”, “when” and “what” components of descriptive epidemiology.

Use case 3 – Planning control measures

In the case of having a known causative agent and/or source of transmission, the idea is to utilize the facility-tracked entity relationship to target the control measures in certain geographical areas.

Use case 5 – Data selection

- Select time interval: Limiting the enrollments of tracked entities to a selected time interval.
- Select organization unit: The user should be able to select an organization unit. Any selection of organization units that are higher in the hierarchy than a facility (e.g. a district), should recursively select all subunits.
- Select program: Selection of one program that can be linked to the organization units that were selected.
- Disaggregate data: Disaggregation of like age group or gender is important to support descriptive epidemiology, as hypotheses may be based on trends or patterns within a certain group in the population. In case of supporting a syndromic program, this may support in making a hypothesis about the causative agent.
- Show datasets: Adding a requirement for stacking user-defined datasets, which will provide a flexibility to combine user-defined datasets, adding transparency, as well as context to the data. This logic should assign one color for each dataset. A dataset should be based on the previously defined data selection requirements. The maximum number user-defined of datasets are limited to 6, to lower the complexity of analyzing data, and due to performance issues.

Use case 6 – Analytics

View tracked entity/facility distribution: We use GIS to abstract spatial data, representation by use of vector data on a map. In this dimension, we choose to display the concepts of a facility and tracked entity as polymarkers.

- View tracked entity – facility relationship: The requirement will be a visual relationship identifier. Two different options were faced when implementing visual relationship: 1) Using color coding, or 2) have a spatial vector link, like a polyline. The result was a combination of both. To organize the data, the spatial color coding should correspond to the given dataset.
- View clustering: A clustering approach would be a suitable balance between data abstraction and details. Also, the clustering implementation must be compatible with the selected map library. To limit the use of colors and make a consistent design, the colors of the clusters must be configurable by the software developer.
- View enrollment frequency: Use a line graph to support temporal trend analysis of enrollments. It should show the actual date the tracked entities where enrolled.
- View total enrollments: Showing a comparison of total enrollments between user-defined datasets.

4.2 The Facility Burden Prototype

4.2.1 Introduction

The FBP is an application for the dissemination of DHIS2 data derived from tracked entities. This chapter contains a review of the development process and the implementation of the use cases, environmental requirements, testing, and need for further development.

4.2.2 The development environment

Framework for web development

As of September 2017, a crossroad regarding the choice of main library or framework for the entire development process, was reached. The goal of the project, in a

practical sense, required uploading the product into a DHIS2 environment, as a third-party app. This explicitly involved making this a JavaScript-based application. To create a maintainable application, as well as getting some basic features out of the box, one of the most popular libraries/frameworks for web development; React or Angular2, seemed attractive. Neither of the developers in the project team had any prior experience with either of them, leading to a discussion of pros and cons from available sources. Regardless of the choice, it was implicit that it would require time and dedication to learn to utilize the technology. Angular 2 was chosen as a framework for developing the application, one of the main reasons being avoiding decision fatigue, since it provides a rich set of features out of the box and is a bit easier for beginners to grasp.

3rd part libraries

Leaflet is the selected library for spatially displayed data. It was selected because of being opensource, simple, lightweight and coming with a rich set of plug-ins. We also considered “Mapbox” another map library, but due to it being not free and uncertainty about to what extent this application will be utilized in further research, decided to drop it.

Styling

For saving time on styling the application, Bootstrap was used. This library, combining JavaScript and CSS to provide a lot of preset styling options.

DHIS2 environment

The DHIS2 instance was running locally with version 2.28, the latest version at the time of writing. This instance was used for the entire duration of the development process. This was connected to a PostgreSQL version 9.5, as recommended in the DHIS2 documentation for production environments.

Integration with a DHIS2 environment

There are some requirements needed for the application to load and utilized data from the DHIS2 core. The main data elements that concern the application is:

- 1) Tracked entity: This represents a certain person in the system with a unique id. The entity may include attributes according to needs. However, an absolute requirement to integrate with the application concerns having a single entity attribute with datatype: “COORDINATE”. The application will iterate the attribute

for a given entity and use the first match with a “COORDINATE” attribute to display the tracked entity on the spatial dimension of the app. This approach was chosen because of its simplicity, while the preferred way would be for the end-user to dynamically pick the relevant coordinate attribute, in case of more than one. Also, a unique identifier must be provided, or be auto-generated. We therefore assume that there will be only provided one single coordinate attribute. However, we acknowledge that this may not be flexible, and lead to less optimal scenarios in case of multiple coordinate attributes.

- 2) Program: the tracked entity must be attached to a given program. The only required attributes are a unique identifier and a display name.
- 3) Organization unit: The organization units need a unique id, a display name. Also, a coordinate field is required for facilities. The coordinate should represent a single point in latitude and longitude – not a polygon.

Test data

We used two different datasets which could be imported to populate a DHIS2 instance with metadata and data, for the DHIS2 instance:

- 1) IDSP2: We received a dataset called IDSP2 from the HISP India during our field trip. This provided us with some basic metadata structure, but most importantly a proper organization unit hierarchy of India, ranging from national- to facility level. Most importantly, the facilities had coordinates. Otherwise, the dataset provided no useful test data.
- 2) Trainingland: This is a dataset for conducting training sessions on the DHIS2 platform. The dataset was loaded to test whether it contained some useful test data, but it turned out it contained no more useful data than already included in the IDSP2 dataset, thus discarded.

4.2.3 Creating the prototype

This subchapter describes the development process, consisting of five iterations. An overview of the final product will be briefly introduced, to make it easier to follow along. Next, a walkthrough of each iteration and the included functionality will be described.

Final product

The graphical user interface has been organized into four compartments (Figure 18). At the left side the system, the user finds the data selection area, while the blue-labeled panels cover the spatial and temporal analytics sections. The “active datasets” table with green header displays the active data selection by the user. This list will expand as the user stacks datasets. Note that both the analytics panels may be dynamically positioned at the dashboard and dropped at a position that suits the end-user. This feature also enables the end-user to arrange the panels in a way that suits his or hers investigation.

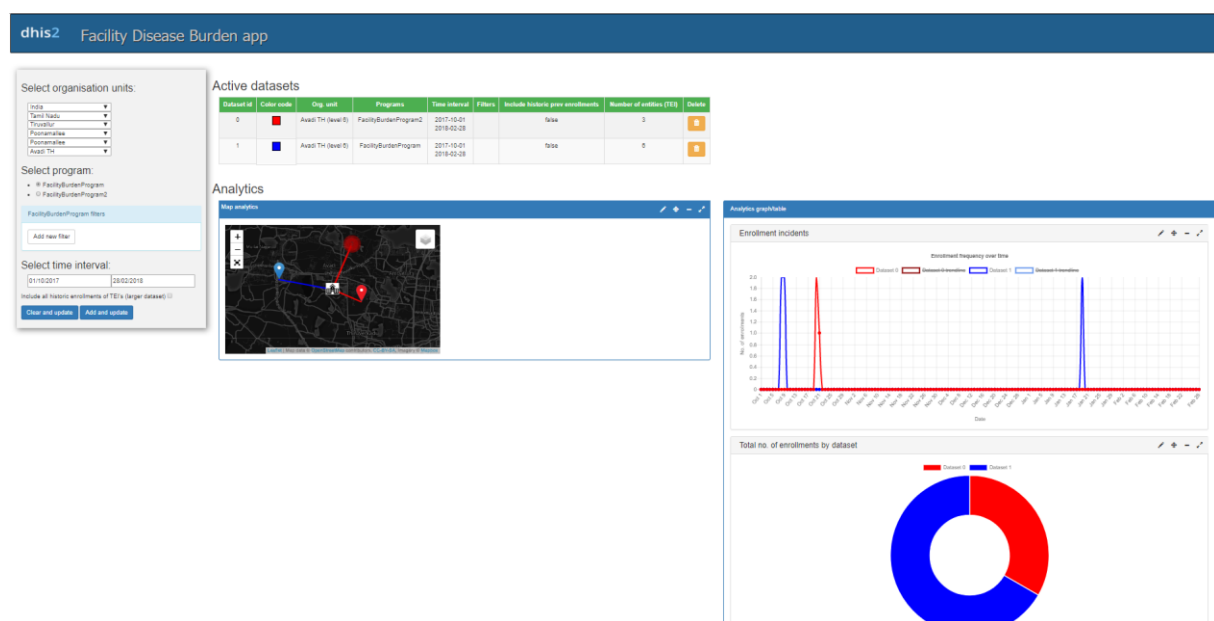


Figure 18: Shows the final product as a screenshot from a web browser.

Style

There were some minor irregularities in the styling, requiring the need to adjust the Bootstrap styling to fit small and medium screens. Also, the Leaflet map container requires an initial fixed height and width, but it should be possible to dynamically adjust it, according to the page size.

Security

The application is having a login page, using basic authentication. This means that the username and password are encoded but not encrypted. In its current state, the application is not secure enough for a production setting, without a SSL/TLS link (encrypted connection). To ensure proper security of authentication, the token-based OAuth protocol support on the DHIS2 should be used instead. In simple terms, it means that the user receives a token upon authentication, which is included in each

request to the DHIS2 instance. When uploaded to a DHIS2 instance the login page is not having any meaningful function, but it was used during the development phase of the prototype.

Architecture

The structure of the codebase of the application bears signs of being subject to an exploratory programming style, as the Angular2 framework was learned along with the implementing process of the application. However, this has a clean architecture between input-centric components carrying user input and the output components like spatial and temporal components on the dashboard. This architecture or pattern is based on the observer pattern, which means that every output component may subscribe to changes from the input components. So, any time the input form is submitted and the attached output components are notified, thus outputs are being generated.

The architecture has since the start of the project seen as a single page app, so there is no setup for multiple pages or any design for sending data for a given state from one page to another.

Iteration 1

The first iteration spanned from week 35-38, 2017, where the focus was on setting up the development environment, as well as implementing some basic features for selection of data from the DHIS2 core. This selection process represents the entire user input required for operating the application. The rest of the user interaction in the application is purely about manipulating already fetched data. This iteration can be divided into four major tasks/use cases:

Use case X1 - Pick time period

This use case provides the end-user with the date interval selection functionality (Figure 19). The end-user is intended to select “from” and “to” date, putting the boundaries on the size of the datasets. Note that this potentially generates very heavy datasets that are fetched from the DHIS2 core, depending on the number of enrollments in the given period.

Figure 19: Use case X1, select time interval.

Use case X2 – Pick organization unit

As displayed in Figure 20, the functionality of use case 5 involved implementing a multiple dropdown list that could list an entire organizational hierarchy in the DHIS2 API. Also, a feature that comes with this multiple dropdown lists, is the ability to select higher level units to include more facilities. E.g. the end-user may select a district level unit (level 3) to include data from a larger set of subunits. In the event of a level 3 selection, the program iterates down to the facility level and resolves the facilities. The application will only utilize organization units with a single point coordinate, as higher level units often return polygon coordinates because they represent an area, rather than a fixed facility.

Figure 20: Use case X2, pick organization unit.

Use case X4 – Pick program

This use case was about providing the end-user with the list of associated programs for the given organization unit (Figure 21). The programs are pre-defined in the DHIS2 core and put no constraints on the type of programs in the application. The end-user may select one program via the radio buttons.

Figure 21: Use case X4, pick program.

Iteration 2

Use case Z1 – View entity distribution

In this use case the spatial descriptions of tracked entities (Figure 22) were implemented. The tracked entities related to the selected organization unit and program is fetched from the DHIS2 core and will be displayed on a map, by use of the Leaflet library. The first occurrence of an attribute with a coordinate datatype is extracted and drawn on the map as a poly-marker in Leaflet. Tracked entities without any coordinate attributes are just skipped and not included at the map.

In iteration 2, we decided to make color coding on tracked entities based on the selected program. This was to enable spatial comparison between two or more programs. E.g. the end-user may compare the spatial distribution of programs like Dengue fever or Malaria, differentiated by a color coding. However, as describes in iteration 3, this setup was reversed.

The list of available colors is a fixed list, also having a corresponding list with predefined marker images to apply on the map. The list of colors is: red, blue, yellow, brown, purple and green.

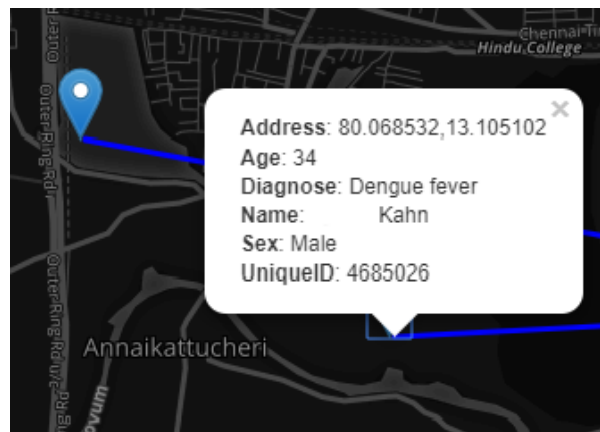


Figure 22: Use case Z1, view entity distribution.

Use case Z3 – View organization units

The program exploits the coordinate attribute of the facilities to display them on the map. The facility is displayed on the map with a customized poly marker, as displayed in Figure 23.



Figure 23: Use case Z3, view organization units.

Iteration 3

Use case X6 – disaggregate data

This functionality enables the end-user to disaggregate data based on the given attributes on the tracked entity in the selected program (Figure 24). All text and number based attribute types are configured to act as a key for disaggregation. However, not all attributes make sense to use as a key. Age and gender would generally be more usable than a last name.

Figure 24: Use case X6, disaggregate data. Note that the age set in the filter determines what age range that will be included, not excluded.

Use case Z4 – View facility- tracked entity relationship

The use of polylines was selected as the desired way to spatially demonstrate the relationship between a facility and all tracked entities that are fetched. The polylines extend from the coordinate of a facility till a related tracked entity (Figure 25). The color coding on the polylines will match the color coding on the given polymarker of a tracked entity.

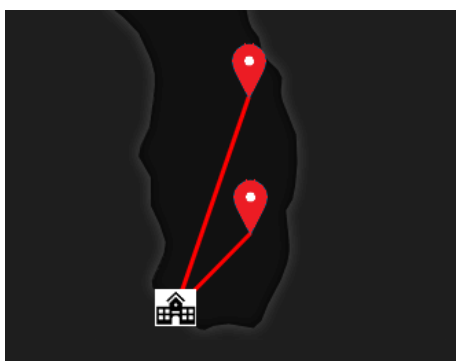


Figure 25: Use case Z4, View facility - tracked entity relationship.

Use case X7 – Show datasets

This use case is a result of an issue regarding data selection in the application. The FBP did not support comparison of facility burdens across organization units with no common ancestors. However, it appeared that user-defined datasets were a more flexible approach, enabling the end-users to customize their own data sets for comparison. When implementing the datasets, it was also noticed that it became easy to lose the overview of the selected datasets when having more than one active, introducing a table for summarizing active datasets with their input parameters (Figure 26).

Active datasets





Dataset id	Color code	Org. unit	Programs	Time interval	Filters	Include historic prev enrollments	Number of entities (TEI)	Delete
0		Avadi TH (level 6)	FacilityBurdenProgram2	2017-10-01 2018-02-28		false	3	
1		Avadi TH (level 6)	FacilityBurdenProgram	2017-10-01 2018-02-28		false	6	

Figure 26: Use case X7, display datasets. This example shows two selected datasets in the app. Each with different color code.

By having user-defined datasets, it could also support explorative analysis of spatial decision problems, by making the data selection more flexible, e.g. overlaying and comparing data.

Iteration 4

Use case Z2 – View clustering

The number of polylines and number of polymarkers, representing tracked entities, became a problem for readability of spatial data at the map. It was never an option to remove the polylines, as they were fundamental for visualizing the relationship. However, the spatial noise was greatly reduced by implementing a cluster plugin on top of Leaflet, as seen in Figure 27. This plugin collects polymarker within a given distance to each other into a cluster, graphically displayed as a circle with a number representing the number of hidden tracked entities/polymarkers. This plugin provided the necessary functionality, but two alterations were being made. First, the coloring on the clusters needed to match the polylines and polymarkers. Second, the polyline now representing the facility-cluster relationship was having to point to the center of the given cluster, while redrawing the polyline as a response to any map-scaling commands.

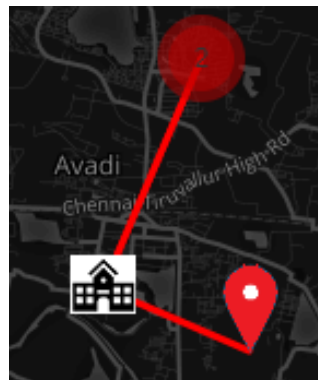


Figure 27: Use case Z2, show clustering.

Iteration 5

Use case Y1 – View enrollment frequency

To enable the end-user with a tool for assessing a load of enrollments within the selected dataset, a line chart has been added (Figure 28). The X-axis data is based on dates within the time interval, having “days” as the lowest resolution. In case of multiple datasets being active, the one with the largest interval will dictate the range of the X-axis values. The color coding in the chart is a reflection of the color displayed in the “Active datasets” table.

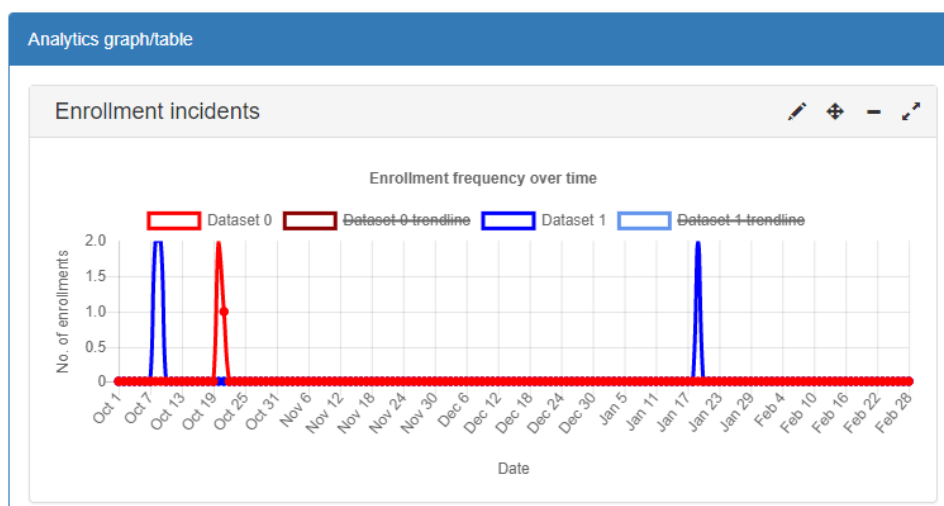


Figure 28: Use case Y1, view enrollment frequency. Shows a line graph of enrollments.

Y-axis data is the total number of tracked entity enrollments within the datasets. For each dataset, the application also generates a trendline of enrollments within the dataset. Trendlines are by default not displayed.

The metadata structure of DHIS2 is organized in such a way that it without further specification will only return the latest enrollment of a given tracked entity in time. By default, a tracked entity will be unique in a dataset, e.g. a patient gets enrolled multiple times with the selected time interval, only the latest enrollment will be included in both the spatial and temporal dimensions. The positive outcome of this is smaller datasets, thus increasing performance. In the context of getting a clear image of historical trends and developments, the app provides the possibility to include all historical enrollments that are critical to avoid misleading data. This will be enabled by checking the “Include all historic enrollments of tracked entities (larger datasets)” in the data selection section. As both modes may be desired depending on the setting, the choice is left to the end-user.

Use case Y2 – View total enrollments

In order to effortlessly compare total burden between multiple datasets, a doughnut chart was added (Figure 29). The chart adds up the total number of enrollments in the datasets and compares the totals.

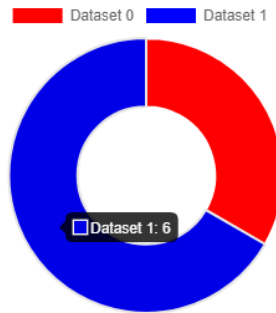


Figure 29: Use case Y2, view total enrollments.

4.2.4 Feedback and testing

No testing has been done other than basic interface testing during the development process. More testing during the development would be needed to ensure that functionality meets the requirements of personnel likely to operate this system. There have been no formal acceptance tests involving DSU operators testing the system, but rather depending on feedback from HISP India as they have experience on how DSU's operate and their requirements. Testing of the final product was conducted by the HISP India team and the full list of feedbacks can be found in Appendix A.

There has also been no performance testing. This is required to monitor the performance and behavior under stress, which can be related to how the application operates in a production setting, where large amounts of data are being processed. Only small amounts of test data have been used during the development process, as this has been manually inserted into the "Tracker capture" module in the DHIS2. However, the cluster plugin has been configured with so-called chunk loading, making sure that the map does not freeze when fetching large number of tracked entity data, during a data fetching operation.

Unit testing is common practice in software development and would have been beneficial for reducing potential uncovered bugs in the application. However, it would have been slowing down the development process, as that involves spending on getting familiar with unit testing in a JavaScript environment, and of course, writing the tests.

A user guide for the FBP will be provided in Appendix B.

5 Discussions

5.1 Introduction

Until now this thesis has covered the analysis and creation of a prototype for supporting evidence-based decision making within disease surveillance. This chapter will discuss the FBP, its potential and challenges for becoming applied for practical purposes in disease surveillance. The FBP will be subject for discussion within both syndromic- and traditional case disease surveillance. In the next section, other contributing factors to increased evidence-based decision making, like data dissemination, interpretation, and data quality, will be described.

5.2 Conceptual evaluation of the tracked entity – facility model

This section will start by discussing the application of the FBP, considering its strengths and weaknesses. As the FBP represents a very basic surveillance system it would be natural that it lacks all the essential analytic features for supporting decision makers. However, the core of the assessment will be linked to the tracked entity-facility model's influence on the predefined decision problems.

5.2.1 Decision problem 1: How to establish the existence of a possible disease outbreak?

The FBP is configured to enable connection to program specific data sources in the DHIS2 core. For decision makers, the ability to detect possible outbreaks in the FBP will mainly depend on the size of a geographical area and the rate over time, which an increase in disease cases occur (Lee, 2010). Two scenarios linked to this issue of detecting epidemics will be discussed:

The majority of patients are being enrolled in one or few facilities: An increase of cases in one or a few facilities are easy to notice when analyzing temporal enrollment trends. However, there are some challenges related to syndromic surveillance. E.g. Fever is a very common syndrome that can be linked to several

diseases, generating a lot of noise if the analyst is interested in linking syndromes to diseases. The information must be more specific. One way of solving this is to specify by combining common syndromes for e.g. Hepatitis A, like fever, vomit, dark urine, icterus (yellow skin). However, the FBP only support one filter, making it only possible to disaggregate data on one tracked entity property. This represents a weakness in terms of syndromic analysis capabilities.

The cases are near evenly distributed among the facilities in the district: This scenario makes it hard to detect temporal trends of enrollments on single facilities. However, by making queries that comprise all facilities in the district, it may indicate a temporal trend. Where it is hard to determine a temporal trend, the spatial dimension may, however, reveal a concentration of syndromic cases. The polylines will indicate the facility relationship so that the system operator may quickly make a phone call to the correct facilities for confirmation and to verify a possible trend and completeness of the data. Analyzing spatial concentrations of cases will give no meaningful information unless it is compared with an ideal situation. To perform such a comparison, the system operator would have to view the general baseline or average of e.g. Hepatitis A related syndromes of the entire district, since just one facility would not cover all the cases within a geographical area. However, a syndromic- or disease load may have local variations, making a district comparison too coarse-grained.

A second alternative to discuss is a user-selected geographical selection, for deeper location bound analysis. E.g. the system operator could select a suspected disease loaded area, get access to temporal trends within that subarea, and regardless of the facility relation, process data all the way up to district level. If a frequency enrollment graph for that given area was presented to an epidemiologist, he/she could be able to analyze the central location, shape, and spread of the curves, thus analyze if it is a possible cluster (Centers for Disease Control and Prevention, 2012a). This approach means more control at the cost of more manual labor to scan for clusters. Feasibility is higher at district- than the state level, due to the size of the area to scan, and clusters are often found in urban dense areas. In this specific scenario, this may offer the information required for making a judgment based on a syndromic area profile. This a step away from the facility-tracked entity model, as it pulls forward a facility independent analysis method.

Sometimes in descriptive epidemiology, there are no trends or pattern to detect, simply because the location of exposure differs from their residential location. This would spread out the tracked-entities, both over geographical areas and facilities, generating low or no signals. Also, traveling could contribute to lower the data completeness when they are not enrolled to a local facility before going back to their homeland, state or district. Signals closer to the source are often easier to detect, than far.

For syndromic surveillance purposes, the decision makers would be more dependent on an effective geographical clustering algorithm, to reduce the manual labor and shorten the time to detect clusters. However, syndromic clustering algorithms will differ from the hierarchical clustering algorithm applied in the FBP, which are designed for organization and explorative analysis of a cluster, rather than an analysis of clusters in time and space (Grubestic et al., 2014), thus making the FBP less optimized for detecting outbreaks in a syndromic surveillance setting.

5.2.2 Decision problem 2: Whether to investigate, control or monitor a possible outbreak?

Syndromic signals are often not enough for decision-makers to start a comprehensive field investigation, however, there may be an exception for customized bio-surveillance systems for responding to bioterrorism etc. However, the facility relationship provides the end-user with an overview of the involved facilities for faster verifying a trend or pattern. This is a way to improve evidence-based decision making at a syndromic stage by confirming hypothesizes, which again could narrow down the hypothesizes before launching a potential field investigation. The analysis functionality in the FBP also applies to disease programs of confirmed cases. A potential hypothesis the decision makers had about e.g. a Hepatitis A outbreak in a geographical area, would be confirmed as the lab results start ticking into the system. The possibility to effectively compare the syndromic and disease-related programs in both temporal and spatial dimensions may provide enough evidence to launch a field investigation earlier, rather than wait for the majority of a geographical cluster becoming confirmed cases, thus shortening the time before an investigation and following control phases.

To effectively be able to compare syndromic and disease confirmed cases may also help the decision makers in establishing confidence in the system, as well as acquire experience for research and adjusting hypothesis. In a given scenario, in which Hepatitis A confirmed cases enter the system, the epidemiologists know the incubation period, effectively narrowing down the time interval of interest. The FBP is also an arena with the potential of supporting the development of a hypothesis based on former outbreaks in historical data. As the system operators have access to information from earlier outbreaks, they may help analyze and compare for detecting common patterns. Even though the FBP is not customized for animation of disease patterns in a spatial/temporal environment, the flexibility for the end-user to define dataset may also be used for visualizing the spread or recoil of a disease, which is regarded as having a great potential in disease control and prevention (George et al., 2013).

In case the end-user having established the existence of a possible outbreak, a DSU is likely to send in a response team for field investigation. By using the FBP, the system operator may easily communicate location/addresses, size of the cluster, but also the major facilities that have registered cases in the possible outbreak. Spot maps, as used in the FBP, are useful in displaying geographical distributions of health event for targeting investigations, information that makes the response team able to organize and plan. However, spot maps are most useful when having a reference to the underlying demographical aspects, like the population at risk (Lee, 2010). For example, a cluster of 30 disease cases could be a trivial event in New Delhi, while being an indication of a severe condition in a rural village.

Lack of spatial information related to sources of infection, climatic and geological aspects, e.g. contaminated wells, seems to have negative implications on both generating hypotheses and implementing suitable control measures (Rai and Nathawat, 2017). Integration of these information sources could involve map overlaying datasets, a feature not implemented in the FBP. The information required in order to understand a health issue and discuss appropriate control measures will depend on the disease, as e.g. vector-borne may rely on information related to mosquito breeding grounds. Further development for strengthening the disease surveillance could include map overlays, for such as deforestation, mosquito, sewer

systems to model risk factors and modes of transmission. The investigators would be able to make faster testable hypotheses for field testing, thereby saving resources and time, and then again be able to start planning the implementation of control measure if needed. For the FBP to facilitate user-integrated map overlays will be a more attractive future solution, as it would require less maintenance by software developers, than built-in map-overlays.

Sometimes a field investigation will require interviewing the case-patients (Centers for Disease Control and Prevention, 2012c), but also for implementing specific control measures. This implies that knowledge of the location of the case-patients are of importance. Control measures can be implemented by interviewing case-patients, leading the investigators to the source of infection. E.g. a contaminated well. The FBP communicates the residence location and enrolled facility of the case-patients, thus supporting field investigations. Also, the outcome of a decision when deciding whether to investigate, control or monitor an outbreak could become different if the decision makers knew if the tracked entities were inpatients or outpatients, information that the FBP do not provide. This would require the FBP to connect to facility discharge records, which is also a method for measuring data completeness of a surveillance system (Heffernan et al., 2004).

From November 2001 to November 2002, the newly established syndromic surveillance system of New York city health and mental hygiene was studied (Heffernan et al., 2004). This system has many of the same properties as the FBP for spatial and temporal analysis. However, it lacks the support of the more advanced methods like statistical analysis and geographical clustering for syndromic surveillance. As in the New York example, the greatest strength of the FBP appears to be the combination of timeliness of data with a traceability of signals to facilities, which leads to reduced response time and more target use of resources. The healthcare professionals of the appropriate facilities may quickly be alerted on any trends, while the confirmation of trends from the facilities contributes to increasing evidence (Heffernan et al., 2004), thus more efficient decision-making of whether to continue monitoring or launch field investigations. Having datasets with a transparent tracked entity-facility relationship in a real-time surveillance system, is also facilitating a close feedback loop with facilities in the districts, establishing a basis for a

multidirectional information flow, and supporting both decision levels by having the right level of data granularity to support both district and sub-district level. The granularity of summarized case-based data is suitable for supporting information needs at the district level (Braa and Sahay, 2012). This close feedback loop cultivates an information culture between district and sub district level, but has little regard for state and national level. Utilization at higher levels would lead to performance issues when fetching datasets in time and space, demanding more efficient spatial data processing. Also, the presentation of data would need to be examined, to produce the appropriate granularity to support decision making at a given level.

5.3 Data quality

The steps from data collection to action in the information cycle are all dependent and affected by each other. Considerations conducted in the analysis of the FBP does not cover data quality, but this plays a major part in supporting evidence-based decision making. Most decision makers would refrain from making decisions with considerable consequences if they were making decisions based on inaccurate, outdated and incomplete data. We will consider the most prominent data quality properties:

Timeliness: At the moment an enrollment has been registered in the DHIS2 tracker capture, it would be available few seconds or minutes later in the FBP. From this perspective, the app is suitable for real-time surveillance, including syndromic analysis types.

Completeness: The data completeness is probably the greatest data quality issue. Records of all enrollments would need to be submitted into the DHIS2 platform, regardless of the size of the facility, e.g. a hospital in an urban area or a small health post in rural India. Any missing records would negatively influence the impression of disease burden. However, both the spatial and temporal dimension of the FBP will enable the end-users to quickly detect a suspicious low enrollment rate in given facilities, a reason for suspecting incomplete data. As discussed earlier the facility

discharge records may be used for measuring the completeness, or alternatively periodic aggregated reports or even yearly surveys.

Accuracy/resolution: The required resolution of the geographical locations will vary according to what is the appropriate focus to describe the place of the health issue; it could be a district, a health facility, a street address, etc. (Centers for Disease Control and Prevention, 2012b). In the FBP the required resolution is at the street level, as residential addresses are determined by a set of coordinates. Magnuson recognizes that general GIS limitations include resolution, data accuracy and data completeness (Magnuson and Paul, 2013). Sometimes, the resolution required is not supported by the GIS, leading to geographical inaccuracy and making it harder to determine the distribution of the health issue. In developing countries, there is a tendency of GIS's offering lower spatial resolution than appropriate for analytic purposes. For instance, because of questions regarding the reliability of locations such as villages (Mak and Eisen, 2013). This could be due to factors such as nomadic cultures without fixed infrastructures, but also because the GIS are not having the resolution to describe a rural village, making it very hard for a clerk to pinpoint a residential location. Reliability affects data believability, which can be linked to decision making efficiency (Samitsch, 2015).

The second issue of data accuracy is linked to unique identifiers. E.g., a patient may go to Facility A, get enrolled, however for some reason he travels to Facility B and gets enrolled there the next day. As India has no national unique identifiers, there is no guarantee that the clerk in Facility B will be able to find any records of this patient in the system. There is a chance that the outcome is double registration, which may affect the data accuracy. However, double registration may be more misleading in traditional disease surveillance than syndromic surveillance, as the signals from the last 1-3 days often are of most significance.

5.4 Presentation and Interpretability of information

Because of the local variations in training, GIS expertise, and practices between DSU in India, the data presentation and the usability of the FBP is based on simplicity, by showing essential data for evidence-based decision making. In the light of decision

problem 1 - detecting an outbreak, the presentational aspect of the information will influence the decision makers ability to detect the decision problem. The relationship between health events and originating facilities are on the basis of being visual, having the potential of reducing the cognitive and phycological workload of the end-users, which is a suggested public health dashboard guideline (Lechner and Fruhling, 2014).

The literature also indicates that the simplicity and consistency of the design and terminology will help make the leap from the DHIS2 core to the FBP shorter. The greatest barrier to utilizing the app lies in the dissemination of the concept itself, making a short introduction to the system necessary. The accessibility is high in terms of being web-based. However, the app is not tailored and tested on mobile devices, but has some adaptive features for supporting smaller screen sizes.

A web-enabled GIS platform used in over 30 government departments in Hong Kong in 2005, combined disease-related data with geospatial data for displaying spatial incidents of 31 diseases. The study indicates that web-enabled GIS platforms may enhance the data dissemination efficiency for monitoring and control of infectious diseases across organizational borders (Lai et al., 2007). This property increases collaboration by more easy information sharing, which could be especially critical for districts that do not have the human resources (e.g. due to epidemiologist vacancy) for analyzing data, relying on outsourcing of the analysis. This enables the decision makers to study the information that generates evidence for public health actions.

The interpretability of data should meet the user level, so lowering the technical threshold for operating the system will make it operable for a larger audience without a significant level of training. The public health care professionals are often not educated with the application of GIS as a tool (Magnuson and Paul, 2013), which may be a contributing reason for the perceived possibility of misuse and misinterpretation of GIS outputs (Joyce, 2009, Densham, 1991). Densham points out that subjective interpretations of GIS are caused by the lack of defined decision-making processes (Densham, 1991). This indicates that predefined decision-making processes, like a public health assessment, will produce a more similar decision outcome in different settings, than decision-making processes based on e.g. heuristic or experience based approaches. The FBP can not enforce a specific decision-

making process. However, an adaption of a system design of GIS, that supports working methods of end-users, ease of use, consistency in terminology, transparency in data processing, and organized data put in context, is likely to decrease the level of misinterpretation.

6 Conclusion

6.1 Introduction

The research question as described in the introduction of this thesis was:

“To what extent does a visual depiction of the relationship between cases of diseases or health events and originating facilities strengthen analytic capabilities in disease surveillance, for improved evidence-based decision making in the context of public health at district level India?”

The objective was addressed through the design, development, and testing of the FBP, which will be designed to support these decision problems:

- 1) **Decision problem 1:** How to establish the existence of a possible outbreak?
- 2) **Decision problem 2:** Whether to monitor, investigate or control a possible outbreak?

6.2 Findings

Considering predefined the decision problems, the FBP shows potential for supporting evidence-based decision making at the district level, in two main areas:

- 1) Facilitating a close feedback loop between the district and sub-district/facility level – A feedback loop that has the potential of providing district-level decision makers with complementary information that addresses the decision problems of concern, while sub-district/facility level decision makers are provided with timely warnings of unusual disease patterns.
- 2) Supporting evidence-based decision making for field investigations. The visual relationship between cases of diseases and facilities, backed by a GIS with high spatial resolution, has the potential of targeting field investigations by having a reference to a residential location and enrolled facility. The result is a high level of spatial traceability that may lead to less resource-consuming investigations.

The tracked entity – facility model has some limitations that makes it less suitable for being applied as the main model of analysis, for addressing the selected decision problems. As descriptive epidemiology concerns an area, not a set of tracked entities derived from a facility, datasets must be combined carefully to support outbreak detection in a geographical area. This makes the tracked entity-facility model a complementary model, focusing on an assessment of facility disease burdens, rather than the dominating model in decision problem 1. This makes the model not suitable, as the main approach, for establishing the existence of outbreaks.

The tracked entity-facility model is also likely to reduce the cognitive and psychological workload of the end-users, thus a strength of visualizing the relationship.

6.3 Contributions to research

This thesis has conducted a broad and exploratory elaboration of the utilization of visual relationships between tracked entities and facilities, thus expanding the insight of applying visual relationships in a GIS, on the DHIS2 platform. For the most part, the research has assessed the potentials and limitations of the model, with respect to decision problems that are common challenges in disease surveillance. The assessment of the model has a high level of transferability to non-DHIS2 environments with similar context. However, the practical implementations are likely to vary. The research does not cover comprehensive testing of the FBP, partly due to lack of access to actual end-users (DSU) and constraints of time and human resources invested in the project, which would have further refined and confirmed the results.

In the context of the information cycle, this thesis has sought to examine aspects of evidence based decision making, by including cognitive aspects of decision makers, decision making processes of the end-users, and their link to GIS and disease surveillance systems. The contributions may provide a suitable starting point, both theoretically and practically, for pursuing the application of the tracked entity-facility model in some of the further directions that will be outlined.

6.4 Future directions

The two main directions for further development, as suggested by HISP India (Appendix A), includes continuation of the FBP as a domain-specific application or as an integrated part of the DHIS2 core GIS app. In this section, I will suggest the content of these potential branches. HISP India feedback also indicates that further development of the FBP should emphasize the spatial dimension.

6.4.1 Domain specific application

By being applied as a stand-alone third-party app on the DHIS2 platform, there are potentials of addressing specific use cases in a disease surveillance setting, an approach that aligns with the development of the FBP until this point.

For further implementations based on the established use cases, there are two approaches that may be explored:

- 1) Differentiate in- and outpatients for further increasing the efficiency of field investigations, but also for establishing the existence of outbreaks, and measure the effect of control measures.
- 2) Map selection of specific geographical areas, for assessing the disease burden in an area with respect to the originating facilities. Useful in urban areas for inspecting complex relations between disease cases and facilities. This is a more efficient than sequentially inspect each facility disease load.

Optional use cases may include:

- 1) Assessment of resource situation at facilities. Outbreaks may lead to overburdened facilities, unable to effectively handle large quantities of disease cases. An effective monitoring of facility burdens is critical information for relocating and reprioritizing resources in the districts.
- 2) Another promising direction points towards health care availability analysis and resource planning. The outlining of the spatial distribution may give decision makers an estimate of facility catchment areas, which also by use of spatial distance measuring tools assisting in the development of new and improved health care services. Studies of routine surveillance systems from

developing countries like Eritrea and Morocco of comparable decision support systems with both temporal and spatial elements indicates a positive influence on decision-making at the district level (Lippeveld, 2001).

- 3) Data analysis emerged from the FBP shows the potential of being indicative of what types of health services being offered, at the facilities under surveillance.

6.4.2 Integrated part of the DHIS2 core

Integration of tracked entity–facility model as a feature in DHIS core GIS application may seem like a more sustainable long-term solution, as an integration with the main functionality of the DHIS2 core enables the model to be combined with the built-in analysis tools of the platform. However, this synergy may come with the cost of increased complexity of integration and maintenance.

Bibliography

- ABOUZAHR, C. & BOERMA, T. 2005. Health information systems: the foundations of public health.(Policy and Practice: Theme Papers). *Bulletin of the World Health Organization*, 83, 578.
- BASKERVILLE, R. & WOOD-HARPER, A. T. 1998. Diversity in information systems action research methods. *European Journal of Information Systems*, 7, 90.
- BASKERVILLE, R. L. & WOOD-HARPER, A. T. 2016. *A critical perspective on action research as a method for information systems research*.
- BHAVAN, Y. & MARG, S. 2011. Report of the Working Group on Disease Burden for the 12th Five Year Plan.
- BRAA, J., HEYWOODA, A. & SAHAYA, S. 2012. Improving quality and use of data through data-use workshops:
- Zanzibar, United Republic of Tanzania.
- BRAA, J. & SAHAY, S. 2012. *Integrated health information architecture : power to the users : design development and use*, New Dehli, Matrix Publ.
- BUEHLER, J. W., HOPKINS, R. S., OVERHAGE, J. M., SOSIN, D. M. & TONG, V. 2004. Framework for evaluating public health surveillance systems for early detection of outbreaks: recommendations from the CDC Working Group. *MMWR. Recommendations and reports : Morbidity and mortality weekly report. Recommendations and reports / Centers for Disease Control*, 53, 1-11.
- CARVER, S. 2001. The Future of Participatory Approaches Using Geographic Information: developing a research agenda for the 21st Century.
- CENTERS FOR DISEASE CONTROL AND PREVENTION, C. 2012a. *Section 4: Properties of Frequency Distributions* [Online]. Available: <https://www.cdc.gov/ophss/csels/dsepd/ss1978/lesson2/section4.html> [Accessed 15.03.2018 2018].
- CENTERS FOR DISEASE CONTROL AND PREVENTION, C. 2012b. *Section 6: Descriptive Epidemiology* [Online]. Available: <https://www.cdc.gov/ophss/csels/dsepd/ss1978/lesson1/section6.html> [Accessed 24.05.2017 2017].
- CENTERS FOR DISEASE CONTROL AND PREVENTION, C. 2012c. *Section 7: Analytic Epidemiology* [Online]. Available: https://www.cdc.gov/OPHSS/CSELS/DSEPD/SS1978/Lesson1/Section7.html#_ref38 [Accessed 14.03.2018 2018].
- CENTERS FOR DISEASE CONTROL AND PREVENTION, C. 2014. *The Public Health System and the 10 Essential Public Health Services* [Online]. Available: <https://www.cdc.gov/nphsp/essentialServices.html> [Accessed 20.02.2017 2017].
- CONNELL, J. L. & SHAFER, L. 1989. *Structured rapid prototyping : an evolutionary approach to software development*, Englewood Cliffs, N.J, Yourdon Press.
- CRESWELL, J. W. 2003. *Research design quantitative, qualitative and mixed methods approaches*.
- CRM4SURE. 2018. *Data quality overview* [Online]. Available: <https://www.crm4sure.com/data-quality/accuracy-consistency-audit> [Accessed 14.04.2018 2018].
- DEEPAK, K. R. & ANIL, K. B. 2014. Integrated Disease Surveillance in India: Way Forward. *Global Journal of Medicine and Public Health*, 3.

- DENSCOMBE, M. 2010. *The Good Research Guide For small-scale social research projects*, Open University press.
- DENSHAM, P. J. 1991. Spatial decision support systems. *Geographical information systems: Principles and applications*, 1, 403-412.
- GEORGE, J. M., PO-HUANG, C., TYLER, S., RACHEL, B., WILLIAM, K., BEREKETAB, L., HUI-CHEN, T. & CHRISTINA, W. H. 2013. Use of GIS Mapping as a Public Health Tool-From Cholera to Cancer. *Health Services Insights*, 111-116.
- GILMORE, T., KRANTZ, J. & RAMIREZ, R. 1986. Action based modes of inquiry and the host-researcher relationship. *Consultation: An International Journal*.
- GOODMAN, R. A., BUEHLER, J. W. & KOPLAN, J. P. 1990. The epidemiologic field investigation: science and judgment in public health practice. *American journal of epidemiology*, 132, 9.
- GRUBESIC, T. H., WEI, R. & MURRAY, A. T. 2014. Spatial Clustering Overview and Comparison: Accuracy, Sensitivity, and Computational Expense. *Annals of the Association of American Geographers*, 104, 1134-1156.
- GRÜNIG, R., KÜHN, R., CLARK, A., O'DEA, C., MONTANI, M. & SPRINGERLINK 2013. *Successful Decision-making : A Systematic Approach to Complex Problems*, Springer Berlin Heidelberg : Imprint: Springer.
- HEFFERNAN, R., MOSTASHARI, F., DAS, D., KARPATI, A., KULLDORFF, M. & WEISS, D. 2004. Syndromic surveillance in public health practice, New York City.(Research). *Emerging Infectious Diseases*, 10, 858.
- HIBBARD, J. H. & PETERS, E. 2003. Supporting Informed Consumer Health Care Decisions: Data Presentation Approaches that Facilitate the Use of Information in Choice. *Annual Review of Public Health*, 24, 413-433.
- HISP INDIA, H. 2018. *HISP Team* [Online]. Available: <https://hispindia.org/index.php/about-us/hisp-team> [Accessed 31.03.2018 2018].
- JESPERSEN, S. 2017. *Dashboard Design Guidelines for improved evidence based decision making in public health in developing countries*. Master, University of Oslo.
- JOYCE, K. 2009. "To me it's just another tool to help understand the evidence": Public health decision-makers' perceptions of the value of geographical information systems (GIS). *Health & Place*, 15, 831-840.
- KEPPEL, K. G. & FREEDMAN, M. A. 1995. What is assessment? *Journal of public health management and practice : JPHMP*, 1, 1-7.
- KRAUSS, S. E. 2005. *Research Paradigms and Meaning Making: A Primer*.
- KRUMSVIK, R. J. 2014. *Forskningsdesign og kvalitativ metode*, Oslo, Fagbokforlaget.
- KWAPIEN, A. 2016. *10 Dashboard Design Principles & Best Practices To Enhance Your Data Analysis* [Online]. Available: <https://www.datapine.com/blog/dashboard-design-principles-and-best-practices/> [Accessed 11.03.2018 2018].
- KYNG, M. 1991. Designing for cooperation: cooperating in design. *Communications of the ACM*, 34, 65-73.
- LAI, P. C., GATRELL, A., SPRINGERLINK, INTERNATIONAL CONFERENCE IN, G. I. S. & HEALTH 2007. *GIS for Health and the Environment : Development in the Asia Pacific Region*, Springer Berlin Heidelberg.
- LECHNER, B. & FRUHLING, A. 2014. Towards public health dashboard design guidelines.
- LEE, L. M. 2010. Principles and practice of public health surveillance. *Principles & Practice of Public Health Surveillance*. Oxford University Press.
- LIPPEVELD, T. 2001. Routine Health Information Systems: The Glue of a Unified Health System
- MAGNUSON, J. A. & PAUL, C. F. J. 2013. *Public Health Informatics and Information Systems*, London, Springer.

- MAK, S. & EISEN, R. J. 2013. *Geospatial technologies and spatial data analysis: PART 2: Use of geographic information systems and spatial analysis in infectious disease surveillance in North America and East Africa*, Oxford, UK, Oxford, UK: John Wiley & Sons Ltd.
- MINISTRY OF HEALTH & FAMILY WELFARE GOVERNMENT OF INDIA, M. G. 2015. Joint Monitoring Mission Report Integrated Disease Surveillance Programme.
- O'BRIEN, R. 1998. *An Overview of the Methodological Approach of Action Research* [Online]. Available: http://www.web.ca/~robrien/papers/arfinal.html#_Toc26184650 [Accessed 28.08.2017 2017].
- POKHAREL, P. K., PANDE, J. & NATH, L. 2007. *Health Care in Developing Countries: Challenges and Opportunities* [Online]. Available: www.pitt.edu/~super7/10011-11001/10091.ppt [Accessed 02.04.2018 2018].
- RAI, P. K. & NATHAWAT, M. S. 2017. *Geoinformatics in Health Facility Analysis*. Springer International Publishing : Imprint: Springer.
- RIC SKINNER, G. 2010. *GIS in hospital and healthcare emergency management*, CRC Press.
- ROBERT WOOD JOHNSON FOUNDATION, R. 2018. *The Interpretivist Paradigm* [Online]. Available: <http://www.qualres.org/HomeInte-3516.html> [Accessed 28.02.2018 2018].
- ROUSE, W. B. & ROUSE, S. H. 1984. Human information seeking and design of information systems. *Information Processing & Management*, 20, 129-138.
- SAMITSCH, C. 2015. *Data Quality and its Impacts on Decision-Making : How Managers can benefit from Good Data*, Springer Fachmedien Wiesbaden : Imprint: Springer Gabler.
- THIAGARAJAN, S., GUPTA, P., MISHRA, A., VASISHT, I., KAUSER, A. & MAIREMBAM, D. S. 2012. Designing an information technology system in public health: observations from India. *BMC Proceedings*, 6.
- TRODD, D. N. 2002. *Data and Datamanagement* [Online]. Available: http://gisknowledge.net/topic/methods_in_gis/petch_data_and_data_management_02.pdf [Accessed 27.05.2017 20017].
- TVERSKY, A. & KAHNEMAN, D. 1974. Judgment under Uncertainty: Heuristics and Biases. *Science*, 185, 1124-1131.
- UNIVERSITY OF OSLO, U. 2011. *Health Information Systems Programme (HISP)* [Online]. Available: <http://www.mn.uio.no/ifi/english/research/networks/hisp/> [Accessed 19.09.2017 2017].
- WAGNER, M. M., WAGNER, M. M., MOORE, A. W. & ARYEL, R. M. 2006. *Handbook of biosurveillance*. Amsterdam ;; Academic Press.
- ZUBER-SKERRITT, O. 2001. *Action Learning and Action Research: Paradigm, Praxis and Programs Effective Change Management through Action Research and Action Learning: Concepts, Perspectives, Processes and Applications*. Australia: Southern Cross University Press, Lismore.

Appendix

Appendix A - Feedback from HISP India

Feedback of the FBP.

Author: Harsh Atal, at the HISP India team.

Overall

The application satisfies the basic premise of plotting lines on a map which show the geographic relationship between health facilities and the people who visit them. It supports plotting multiple facilities at the same time which is useful for comparative analysis. It also features color coding the map lines which can be used to visualize disease or symptom specific traffic to a facility.

Following are the specific feedbacks gathered from the testing:

Navigation

- Login to be removed as it is an unnecessary step to the application
- Need a button to go back to the DHIS2 dashboard screen. Currently there is no way to go back to the DHIS home screen

Content

- Facility name(Lowest level) of the case should be present on the map along with the name of the facility that is selected for analysis.
 - Currently the relationships between the cases and the health facility can be seen in the map but there are only icons and the labels cannot be seen.
- If labels are present for each of the points along with the icons, then it becomes useful to take a screenshot for sharing and also looks more informative at a single glance.
- Pre Configured info graphics for common use cases
 - Although it is possible to select multiple facilities and plot them on a map but it would be helpful to have a pre configured tab which could show program wise (or disease wise) map of the facility which has the highest cases coming to it. Because the aim of the application is not only to visualise but to find out what it is that should be visualized. Such

a pre configuration will be very helpful in the initial investigation of facility load and resource allocation.

- Toggle for displaying distance on the lines
 - This ,so that the user can access the distance of the person from facility by the length of the line.

DHIS Specific

- Support for Non-Tracker programs
 - The app will be able to support more instances if it could also support Single stage events without registration. The coordinates could be taken from facility and the events coordinate attribute.
- The color coding is based on only 'attribute' which fulfills the basic idea of color coding but actually it falls short in actual implementation as attributes are scarcely used to capture diagnosis or disease condition.
- Support for data elements within the program stage or "Program indicators" and not just attributes.
 - The analysis currently supports adding multiple datasets within the same program but this is only limited to attributes. As the data entry of diagnosis and diseases happens usually in data entry , it would be relevant to have that also as a selection. In the same light it would be easier to support program indicators as that would cover the analysis possibilities in a wholistic way.

Future Work

Now that the basic premise has been implemented , effort needs to be put into either

- 1) honing the app as a domain specific application which supports use cases related to disease burden or
- 2) extracting a library out of it and plugging it in the default maps app of DHIS2.

While 1 is easier to do there lie potential benefits of doing 2 in the form of ready to use analytics and integration with the main functionality.

Side Note: The development effort diverges from the app's geographical emphasis and tries to cover other analysis in the form of chart and graphs There is a definite scope for such analysis within the app but it would be desirable to re-focus the development effort to the geographical map based analysis solution that it aimed initially to be and which is still its chief feature.

Appendix B - User documentation

This document will give the reader an introduction for getting familiar with the FBP. The document covers intention with the application, operating and technical requirements, and a guided tour, through the features and graphical user interface of the app. It is important to emphasize that the app is still a prototype.

Definitions

- **Facility:** is e.g. a hospital, health post. Most commonly at the lowest level in an organizational hierarchy.
- **Tracked entity (TEI):** for all practical purposes in this guide, a tracked entity is equivalent to a patient that can be enrolled into a program, in a facility.

Intention

The FBP is a standard web-app running on top of a DHIS2 environment, utilizing the data structure by use of the REST API of the platform. The overall intention with the app is to successfully model a facility-to-tracked entity relationship, for supporting evidence based decision making, a feature that of time of writing is not supported in the DHIS2 core.

Requirements and prerequisites

Users

The system users/audience are epidemiologists and public health professionals at district level in India, though it not bound to a specific location. The application is designed to support epidemiologists working methodology, by supporting the 4 of 5 W's of descriptive epidemiology:

- What (Supported)
- Who (Supported)
- Where (Supported)
- When (Supported)
- Why/how (unsupported)

Technical

DHIS2 version: have been tested on version 2.28.

Browser: Runs successfully in Chrome and Firefox.

Installation

The application has few, but fixed requirements to operate properly.

1) Add a tracked entity:

Go to the “Maintainance” app in the DHIS2. Add a tracked entity.

2) Add tracked entity attributes:

Go to the “Maintainance” app in the DHIS2. As a minimum, add the following attributes:

- a. Add a residence or address location, that has value type: COORDINATE. **NB:** the app will be looking for a single field with value type ‘coordinate’, if more than one attribute with such a value type is linked to a tracked entity, the first occurrence will be used for displaying the tracked entity on the map.
- b. Add more attributes as needed, however remember that only attributes with ‘number’ or ‘text’ as a datatype may be used for filtering/disaggregating data later.

3) Create an organization hierarchy

The only requirement is that the facilities being used in this app needs a location coordinate set.

4) Add a program:

Go to “programs/attributes” in the DHIS2 and add a program.

- a. Fill in mandatory fields
- b. Set Type to “With registration”
- c. Set tracked entity to the one you added previously
- d. Add the available attributes of choice, according to step 2.

5) Assign program to organization units:

Go to “programs/attributes” in the DHIS2, click on the program add “Assign to organization units”. Choose the facilities of choice.

6) Add tracked entities:

Go to “Tracker capture” in the DHIS2.

- a. Select your organization unit and program, and then “Register”
- b. Add one or more tracked entities.

7) Install the FBP

Go to the “App management” in DHIS2.

- a. Choose standard app tab.
- b. Upload .zip.
- c. Click on the app to open.

GUIDED TOUR

Login

Login with the same user as you used for logging into the DHIS2 core.

Dashboard

The screenshot shows the basic setup of the dashboard:

- The area in **red** represents selections of what data that will be fetched from the DHIS2 core (data selection).
- The area in **green** represents the spatial analytics section, a map for displaying spatial distribution of tracked entities and facilities that is included in the data selection.

- The area in **yellow** represents active datasets. Every dataset that has been added to the app from the data selection area, will be added to this table, for the user to easily remember what kind of data that has been fetched.
- The area in **purple** represents the temporal dimension, in terms of charts and graphs, and some other analytic measures.

The dashboard panels have a plug-in for arranging the layout on the dashboard. The panel buttons may minimize, expand or detach the panel from the dashboard. In cases where the system users are spending long time on analysis, they might want to optimize the panels for better view.



Data selection

Explaining the data selection from top to bottom:

- 1) Navigate the organisation hierarchy to facility of choice is found. You may choose higher level unit, e.g you want to display all facilities in your district in one selection/query. Remember that picking e.g. a level 2 unit, may reduce performance due to large number of tracked entities. This field is mandatory.

The screenshot shows a web interface for selecting query parameters. It includes dropdown menus for 'Select organisation units' (Sierra Leone, Bo, Badjala, Choose level), 'Select program' (Contraceptives Voucher Program, Facility disease burden program1, Facility disease burden program2), a 'TB program filters' section with an 'Add new filter' button, and a 'Select time interval' section with date pickers for /01/2018 and /03/2018. There is a checkbox for 'Include all historic enrollments of TEI's (larger dataset)' and a red text warning 'Program not set'. At the bottom are two blue buttons: 'Clear and update' and 'Add and update'.

- 2) Selection a program of choice, one per dataset. Mandatory field.
- 3) Add filter. This is a way of disaggregating data. You may disaggregate data based on the text and number attributes of the tracked entity. Keep in mind that it makes no sense to disaggregate on all the attributes, like birth name. However, age or gender is more useful, but all comes down to what attributes that was originally included when configured in the DHIS2 core. What you define in the filter is what gets included in your DHIS2 query, not the other way around.
- 4) Select a time interval. All tracked entities that was enrolled into the program in this time interval will be fetched.
- 5) Check the “Include all historic enrollments of TEI’s (larger dataset)” if needed. Let consider this scenario: A patient get sick, go to facility A and gets enrolled, two month later he goes to facility B in another village and get enrolled there. By default, a query to the DHIS2 core would only return the latest enrollment. So, if the system user wants a correct display of historic enrollments in the

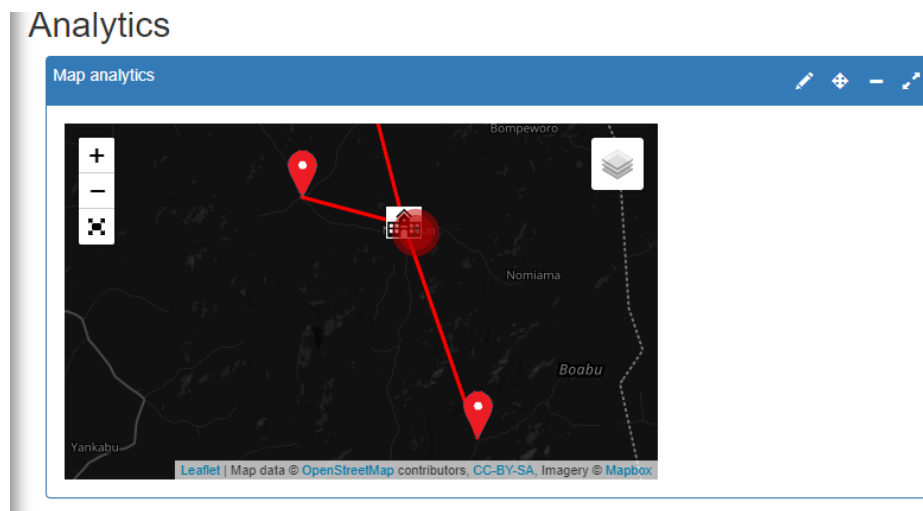
temporal dimension, this checkbox should be set. However, keep in mind that the application needs to process more data.

- 6) Push “Clear and update” to clear all previous data selections (if any) or “Add and update” to keep the previous data selections by stacking the datasets.

Unique color coding's will be issued to each dataset.

Spatial analytics

Shows spatial distributions of tracked entities. The polylines connects the tracked entities to the facility they where enrolled. The colors are unique to the dataset. The datasets are also made available for selection/deselection in the map overlay in the upper right corner of the map. The map controls consists of zooming and a fullscreen plugin.







Active datasets

Shows all datasets that are active. If they are stacked, more than one will appear.

The content reflects the parameters that was used during the previous data selection process. 'Number of TEI's' are the total number of tracked entities in that dataset.

Active datasets

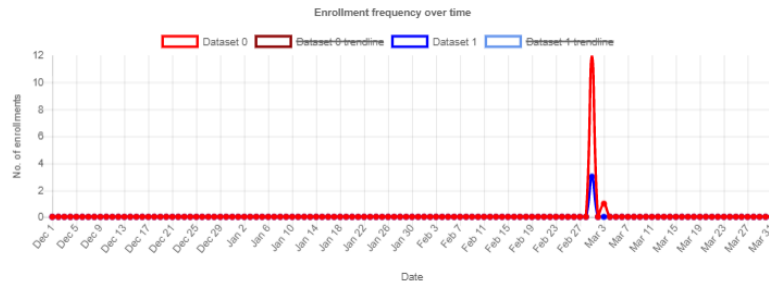
Dataset id	Color code	Org. unit	Programs	Time interval	Filters	Include historic prev enrollments	Number of entities (TEI)	Delete
0		Ngelehun CHC (level 4)	TB program	2017-12-01 2018-03-31		false	14	
1		Njandama MCHP (level 4)	TB program	2017-12-01 2018-03-31		false	3	

Temporal dimension

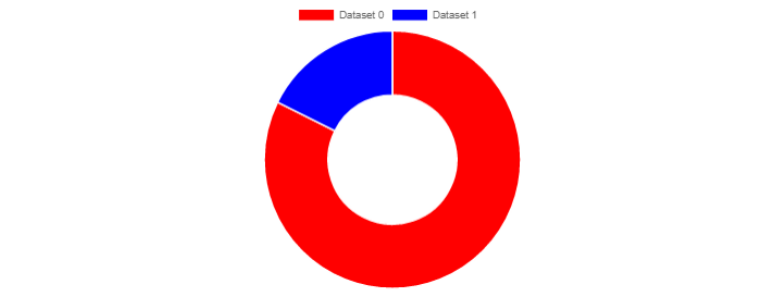
The temporal dimension consists of three different panels. From top to bottom:

- 1) Enrollment incidents is a line graph displaying the enrollment frequency over time in a dataset, hence the color coding. A trendline has also been added for displaying general trends of enrollments (burden) on the given facility. The app utilizes the enrollment date, that was generated in the 'tracker capture module' in the DHIS2 core.
- 2) Displaying a doughnut chart that used for comparing total number of tracked entities between across datasets. The chart is useless when only one dataset is active.
- 3) The last panel contains a highly experimental feature for testing on tracked entities. The module calculates prevalence for a certain attribute of a tracked entity, using the total amount of tracked entities in the dataset as a reference. E.g. if the system user wants to find the prevalence of a tracked entity being under 20 years old, this operation is possible to perform. However, the usefulness depends on how the tracked entity is configured and also on the reference group (the original dataset).

Enrollment incidents



Total no. of enrollments by dataset



Calculate prevalence of attribute

Select dataset by id:	0
Set prevalence group by filter:	<div>TB program filters</div> <div>Add new filter</div>
Prevalence:	0%