University of Oslo
Department of Informatics

Legacy systems and systems development in Mozambique

Bridging the gap between the old and the new, showing the need for change

Tore Christian Skobba

Cand Scient Thesis

5th May 2003

Acknowledgements

This thesis makes up the main part of my studies towards a Candidatus Scientiarum degree at the Department of Informatics, University of Oslo.

My initial thought when discovering the possibility to do a cand. Scient. thesis in a developing country was that this was something I wanted to do. As it would be a new and exciting setting for me to study in with the opportunity to meet and work with people with a different background than myself. Now, at the end of my research I have no regrets about the choice, as it has been a great way to end my formal education for now. During the research I had a good time and gained several new friends, and learned much. Not to forget all the fun of partying in Cape Town and Maputo, and seeing various beautiful scenery in South Africa and Mozambique.

Thanks go to Jørn Braa as my supervisor, for helping me understanding the data and assisting in the writing. I would also like to thank Sundeep Sahay and Jens Johan Kaasbøll for helping me with directions to stumble along during my writing, both during and after the research.

In South Africa thanks goes to Calle Hedberg, for providing a place to stay and giving advices on my program and general work. I also thank Greg Rudman and Pierre Janson in Softcraft for allowing me to work with them. Lastly I thank all the students I meet in South Africa for giving me such a nice time there.

In Mozambique, I would like to give a large thank to the HISP team for allowing me to work within and together with them, and for giving me a memorable time in Mozambique. Including the invaluable help you gave with fixing various troubles I got into. Thanks to you all, Jumo Lungo, Callu Mavimbe, Emilio Mosse, Leopoildo Nhasspossa, Bruno at the Ministy of health, Fumo T., Umberto, Baltazar Chilundu, Arthur Heywood, Mukama F. and Macueve G. and all the other I might have forgotten the name of. Thanks also go to the Ministry of Health in Mozambique and its employees, for giving me time to interview them, and for being so cooperative during the research.

Thanks also go to my parents for supporting me financially during my long study. Finally, I want to thank my fellow students and friends at *Parken* for support, encouragement and lots of fun. Parken forever!

Abstract

In this thesis I have participated in an action research project in South Africa and Mozambique within their respective public health care. The “replication” of a system from one country to another, presents a wealth of problems. In this thesis a research is done in the context of the transfer of a routine health information reporting system from South Africa to Mozambique. This process requires adaptation of the system to the new context. Part of this context is the already existing legacy systems as entangled parts of a larger information infrastructure. The existing legacy systems consist of institutionalized routines in existence, and an objective of the new system is to deinstitutionalize these, and institutionalize new organizational routines. This effort is in conflict with the installed base that includes the old legacy systems. To understand the tensions that arise as part of the effort necessary to align the new systems to the installed base, legacy theory will be used and evaluated.

Legacy systems theory has several methods for connecting the old to the new. One method is target database population, which consists of populating the new system with data from existing legacy systems by importing. As part of this thesis, a tool was developed to automate parts of the import process. The import process was highly complex and time consuming due to several legacy properties found in the existing systems; primarily lack of documentation, lack of interfaces and the general poor state of the existing systems. Data stored in the existing systems were found to be redundant, inconsistent, non-uniform and ambiguous. The tool is evaluated as a change tool in the context of legacy systems and information infrastructure theories. There are two steps in the use of the tool; import of data, enabling data previously hidden and unavailable to become visible, and analysis of the imported data to analysis and evaluate the existing systems. However, the tool was not found to act as a gateway, this due to lack of sustainability and support for continuous data import.

The existing systems and infrastructure is after completion of the import process evaluated by analyzing the data in the new system. The objective of the evaluation is to show the need for change of the health information systems within the organization. Several shortcomings of the existing systems emerged, including lack of flexibility, lack of integration, extensive fragmentation, problems in data collection, analysis and the use of data. There were also lack of conditions to enable new routines, sharing of information, and support for the heterogeneous user community.

TABLE OF CONTENTS

1 Introduction

In this thesis I will report from an action research project carried out within the Health Information Systems Program (HISP) in South Africa and Mozambique, and from my study of the current health information systems in Mozambique. In South Africa, HISP has taken part in the development of national standards for the collection and reporting of health data from the health facility /clinic /hospital level, via the health district and province levels all the way up to the national level. HISP has developed a database application called the District Health Information System (DHIS), which is supporting data collection and analysis at all levels. The standards and the application are based on data aggregated at the facility level, not on individual medical records. The DHIS is now a national standard in South Africa and is implemented and used in all districts and public hospitals in the country. Since 2000, DHIS has been translated, adapted and tested in Mozambique as well as in many other countries. During February-April 2002, I worked with the DHIS software development team in Cape Town, South Africa, and during April-June I worked on practical prototyping of the DHIS and some additional procedures I had developed in Mozambique. In Mozambique there are a number of different data sets that are being reported from the health clinics and hospitals and upwards, including hospital, vaccination, drug, mother and child health data. A major problem in Mozambique is that different datasets are supported by different, and often not appropriate, software and different reporting routines. The HISP program in Mozambique is currently trying to get several of the existing reporting systems replaced by the DHIS software. It is within this broad initiative this thesis has been carried out.

1.1 Research objective

The main problem addressed is as follows:

Investigate an approach to replace, develop and implement health information systems in developing countries, based on research in Mozambique.

I started out with an objective to develop and test a tool for importing and migrating historical data from old, outdated legacy systems (databases and spreadsheets). However, I soon realised that this task was not possible to achieve unless I included a broader study of the current information flows and information systems, or information infrastructure. Once I started my study of the information infrastructure, a third level, or area of focus, needed to be included to fully understand my primary task of importing data from the old to the new system; the use of health information and general health policy within the changing environment of the health organization. The changing of the information infrastructure was not only a technical change, e.g. to a "better" system, also the health sector and the intended use of information was changing. Integration of information from different sources and more local use of information for management was now the focus. This represented a break with the previous set of information systems, which are characterized with fragmentation, upward reporting of data and minimal feedback and local use. These factors led me to the following sub problem:

Analyse how to enable the implementation of a new infrastructure with the use of legacy theory to examine the interrelations between the old and new systems

This led to the following sub problems:

Study how data from legacy systems can be made available in a new system

Study the existing information infrastructure by importing data from the old to the new system. In this way shortcomings of the old systems and potential advantages of the new system, in terms of better data quality, may be studied.

1.2 Theoretical topics I draw upon

The problem identified above led to three main theoretical themes to be considered in my thesis:

1. Health information systems in developing countries
2. Legacy systems
3. Information infrastructure

1.2.1 Health Information Systems

The Alma-Ata declaration made by the World Health Organization (WHO) 25 years ago emphasises the role of the primary health care (PHC) in the provision of health services [WHO AA]. Today, an important part of the effort to strengthen the primary health care in developing countries is decentralization, including strengthening of health management at local level, enabling local data processing and use of information in local decision making. An important part in this change of focus is the development of routine health information systems [Lippeweld, 2001], which can give the local level the needed flexibility to collect, process and use information in local management and decision-making. In contrast to these proposed aims, the information systems of today are typically fragmented into several vertical programmes and only serve needs of the higher levels of the administration, and not enabling local use of data [Lippeveld, 2001]. The health sector is in a process of change and it is within this context this thesis is set.

1.2.2 Legacy Systems

Software systems are today used in almost all organizations worldwide; a majority of these systems were designed and implemented several years ago. Today, they are outdated and highly expensive to maintain and extremely difficult to extend to cater for new functional requirements [Sommerville, 1995]. These old systems are called legacy[1] systems; one property of legacy systems is according to [Bisbal, 1999] a system that resists modification and evolution. Thus the legacy system must be replaced with new systems or reengineered, or else they might hinder the overall business development. A large problem with decommissioning legacy systems, has its origin in that they are often critical and embedded into the operations of the organizations using them, and during their lifetime they have embedded business knowledge, procedures and data not stored or documented elsewhere [Robertson, 1997]. In this thesis I will look into the process of replacing legacy systems with a new system, in which an important part is the saving and transferring of knowledge stored in the legacy systems to the new system, i.e. data migration or target database population [Bisbal, 1999].

1.2.3 Information Infrastructure

"*Infrastructures are considered as always already existing; they are never developed from scratch. When designing a new infrastructure it will always be integrated into and thereby extending others or it will replace one part of another infrastructure*".

[Ole Hanseth, 1998, p. 13]

Public health care is a large and complex organization consisting of several levels with several different actors with different needs on each level, and different computer systems. Thus, they involve not only a few legacy systems, but several legacy systems being part of a larger information infrastructure, consisting of several elements such as work procedures, inscriptions and paper forms. Therefore, implementing the new system also consists of changing an information infrastructure and the installed base. The installed base is everything that already exists, consisting of a heterogeneous network of humans and technology [Hanseth, 1998]. The installed base is difficult to change and impossible to control due to its complexity and interconnectivity. This thesis will look into ways to align a new system to the installed base, i.e. how to adapt and implement a new system that will be compatible to the installed base.

[1] Defining legacy systems is difficult as it consist of several parts; old systems are not always legacy systems, for more see the theory chapter. Some theory uses the term legacy information systems. I have in this thesis consistently used legacy systems, however the two terms are denoting the same

1.3 Problem area

Mozambique, with a population of about 20 million, is a former Portuguese colony, and has since independence in 1975 been ravaged by war, floods and mismanagement. This makes Mozambique one of the poorest and least developed countries in the world, shown by its rank of 170 of 173 countries on the United Nations Development index for 2002 [UNDP HDR2002]. Today, Mozambique is looked upon as an African success story in the happening, with a large and stable economic growth, where the past is put behind, and the emphasis is on development and poverty reduction. The public heath care is struggling with limited financing and a large population to serve. Unfortunately, management is poorly supported by the information systems in use today, which are characterized as vertical programme specific information systems, only serving the central information needs. The local levels are typically collecting large amounts of poor quality data which are not used [Braa, 2001]. Within this context HISP is trying to decentralize and empower all levels of the health organization, in order to support management, especially local management. The new system, DHIS, is a system made to help institutionalize data use at all levels, building upon the experience from South Africa where the HISP initiative is ongoing since 1994.

Mozambique and the public healthcare are organized into 11 provinces. The provinces are seen as the information hub, and are also the lowest level with computers in the health care. The provinces are organized into districts of which there are a total of 131. Each district contains several health facilities and some larger health centres of, where one is the district centre. In addition, there are other units such as hospitals at the provincial and district levels.

1.4 Main problem addressed

The main problem addressed in this thesis is that of enabling change. The context is the “replication”, or replication of a routine health information system made for South Africa to Mozambique. In Mozambique, there are several existing systems which are interlinked, representing a larger information infrastructure. The deployment of the new system is viewed in the context of information infrastructure and facilitating change while trying to align with the installed base. Part of the aligning process is bridging the old systems to the new system. Legacy theory will be used for this; more specifically to understand how the new system will be populated by data from the old systems, by the development of a tool to automate the process.

1.5 Expected research contributions

The approach I have used may be labelled action research. The action part is the deployment of a new system and the analysis of the organizational change. Analysis of this process will be addressed using theories from the domain of information infrastructure, legacy systems and health information systems.

Expected practical contributions are the prototype development, implementation and testing of a tool to automate parts of the data import process. The actual use of the tool in Mozambique is separated into:

- Importing of data from existing systems into a new system
- Evaluation of the imported data in the new system

The evaluation of the imported data in the new system will allow a deeper investigation into the existing systems and the information infrastructure. This is important as the Mozambican Ministry of health is currently working on several vertical integration solutions. This research will show the technical difficulties and complexities involved with integration of the existing health information systems, and make the fragmentation visible to the policy makers.
The expected theoretical contributions are within:

- Health information systems theory in developing countries
- Legacy systems with an emphasis on decoding and saving of old data
- Information infrastructure and changing of those

Finding any research about legacy systems and information infrastructures in developing countries has been difficult. I believe that this thesis might help to increase the knowledge about legacy systems and information infrastructures in developing countries.

1.6 Limitations

The replication and implementation of an information system, or infrastructure in a new context is a highly complex and rich context to do a research. This has required me to have a focus. My focus is on the technical aspects of systems development, legacy systems, information infrastructure and health information systems. In analyzing the existing systems, the same technical focus is kept. While taking a broader view is necessary in order to identify sources of failure. Another limitation is the lack of existing theory. As [Bisbal, 1999] states the migration of legacy systems is an area requiring more research, as little is available. That combined with the setting of a developing country, means that available theory to study the problem is limited.

1.7 Structure of the thesis

In chapter 2 I will present the theoretical framework used in this thesis, which includes health care information systems theory in the context of developing countries, and implication for new systems. To assist with the understanding of the complexity of the current reality, and the changes needed, I draw upon legacy systems and information infrastructure theory.

Chapter 3 presents the background and context for the research, explaining the Health Information Systems Program (HISP) and District Health Information System (DHIS). In the end, I present general information and health related information about Mozambique, its history and current status to help provide the context of the research. Chapter 4 introduces the methods and action research approach used in conjunction with this thesis. Chapter 5 contains an outline of what I did and its implications.

Chapters 6, 7, 8 and 9 are related to my findings. Chapter 6 gives a snap shot of the existing infrastructure in the health care sector in Mozambique. Chapter 7 explains the decoding of one the existing legacy system in order to understand the data in the legacy system. Chapter 8 deals with the actual import process and related problems in adaptation of DHIS. In Chapter 9, the investigation of the existing systems and the larger information infrastructure is presented.

Finally, in chapter 10 my findings are discussed and conclusions are presented in chapter 11.

2 Theory and a literature review

In this chapter theory and literature used is presented using the following structure:

- *Theory about health information systems in general*
- *Health information system in developing countries, what is wrong?*
- *Success & failures of health information systems*
- *Complexity of the current reality and difficulties of change, looking at my case with legacy and information infrastructure theory*
- *Prototyping which is used in my research and by HISP*
- *Preservation of digital data vs. migration of data*
- *Healthcare information systems, concept for success and failure*

My main aim of the theory chapter is to show the existing situation from different perspectives, and the complexity of change.

2.1 Healthcare information systems

The organization of a typical public healthcare organization combined with the different information needs strongly influences the design of the information system, as [Lippeveld, 2001] states:

"*The healthcare information system structure should permit generation of the necessary information for rational decision making at each level of the health system, each of these levels has specific functions that require specific decisions to be made*".

[Lippeveld, 2001, p. 3]

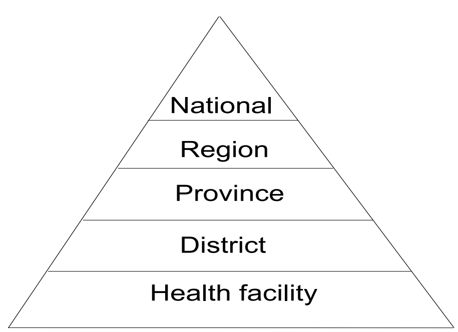


Figure 1: Typical information organization of public health care, the pyramid reflects the information amount needed at the level and number of units. Note amount of levels may vary.

Public health care is usually organized in strong pyramid fashion, where the information required at the level above is a subset of the information in the level beneath, often the information required at the levels above is a function of the information at the lower levels, such as aggregated data. The bottom levels have the largest information need and are the origin of the health care data and statistics being generated in the system. The different information needs among the levels is caused by the different responsibilities. The higher levels responsibility is to provide an enabling framework to work within and ensure the functioning of the levels below, the lower levels are responsible for health care delivery [Haga, 2001].

Of special importance are the lower levels, because evidence shows that most health programs are implemented on a district and health facility level by clinicians [Lippeveld, 2001], [Potomac, 2001]. Of special importance is the district, the district usually contains several health units of different type (stationary facilities, mobile units etc.), WHO (World Health Organization) have with the following definition explained how central the district is, or should be:

"A district health system based on primary health care is more or less self-contained segment of the national health system. It comprises first and foremost a well-defined population living within a clearly delineated administrative and geographical area. It includes all the relevant health care activities in the area, whether governmental or otherwise. It therefore consists of a large variety of interrelated elements that contribute to health in homes, schools, workplaces, communities, the health sector and related social economics sectors...."

[Lippeveld, 2001, p. 9]

2.1.1 Health information systems definitions

Health information systems are in "Last JM. A dictionary of epidemiology second edition, New York: Oxford University Press, 1988". Defined as:

"*a combination of vital and health statistical data from multiple sources, used to derive information about health needs, health resources, costs, use of health services and outcomes of use by the population a specified jurisdiction.*"

Not many systems are fulfilling this definition today, [Lippeveld 2001] states that in both industrialized and developing countries, the health care information systems are woefully inadequate in providing the necessary information support to individual care and public health activities.

Lately it has also become very popular to use the health information system in the management process. DHIS for instance is a system which primarily has as an objective to facilitate the management process. Thus the health information system is mixed with management systems. [Lippeveld, 2001] defines Management information systems as (based on another reference):

"*A system that provides specific information support to the decision-making process at each level of an organization*"

[Lippeveld, 2001, p. 3]

Combining the above definitions you have a system allowing for information collection from several different organizations/actors. The information collected should be available on all levels for all organizations/actors (finance, clinicians etc.), thus it is a complex systems[2], for a large organization with a large area of responsibilities from high tech operations to simple education about washing your hands.

2.1.2 Today healthcare systems status

Extensively research shows that several of today's health care systems (including information systems) are not fully adequate [Gravitz, 2000], [Lippeveld, 2001]. [Gravitz, 2000] states that: "Today health care systems are paper bound and fragmented industries, making them a primary target for IT". Problems and pressure on the health systems include:

- Strongly regulated, bureaucrat icy and lots of laws
- Resistant to change, huge complex organizations with a heterogeneous users group
- Largely for non profit and lack of financial funding
- Small investment in IT, 2-3 % of total budget, finance for example invest 5-10 %
- Pressure to contain costs and improves quality; need for better management on all levels (national to health facility)
- Establish evidence-based medicine. [Heeks,1999] States that 20-50 % of all major therapeutic intervention decisions involves little or no use of data; intuitive is used instead.
- Embrace consumerism, web is creating informed patients, increasing patient's expectations

Ref: [Gravitz, 2000] and [Holy Grail]

Note that although healthcare is one of the biggest industries in the world, healthcare is the largest national sector in the US for instance [Gravitz, 2000]; it is still having large problems with funding. In development countries this problem is even more acute.

There is interrelations between these elements, for instance the lack of evidence based medicine can lead to that proven useless therapies linger in practice long after its lack of effect is clear, leading to expensive and low quality health care. For developing countries not all these elements are important, for example embrace consumerism is not a big pressure on the public health care system in developing countries; few people have access to the web, and of the ones who have a large percentage probably have access to the private health care systems.

[2] Later in this chapter I will claim that it actually more resembles an information infrastructure instead of a system.

2.1.3 Properties of a healthcare information system

Extensive research [Heywood, 1994], [Lippveld, 2001] support the view that a healthcare information system should support the following processes:

- Collection of data: There are primarily 2 types of data being collected: Routine data and none routine data, I will primarily focus on systems working with routine data.
- Data transmission: Data must be transferred between the different levels and across different programs/organizations
- Process data: Process the data to ensure quality, validity, reliability, consistency and accuracy
- Analyze & present: Typically computing of indicators, coverage etc. This should enable data to be compared against each other across facilities and programs. The analyzing must be done on several levels (facility, district etc.). The presentation of the information should make it more accessible and easier to understand.
- Use information in planning and management: Compare facilities and trends etc. Evaluate how the progress toward the goal is? Action should be based on evidence and not solely on intuition.

Of special importance here is the collection of data and the analyzing of the data.

2.1.4 Data collection & data sets

As any other information systems, healthcare information systems are depended upon good data. For a routine healthcare information system, the selection of which data to collect is important due to the huge amount of data which it is possible to collect. This combined with the fact that it is usually clinicians, with several other tasks to do beside from data collection, doing the collecting of data, leads to a strong need to limit the amount of data collected to limit the workload. The actual selection of data to collect require a extensive study, requiring participation from all actors, clinicians, management, politicians etc. on all levels, health facility all the way up to national, important that all data collected have a purpose. All data elements decided to collect should be defined in a data dictionary, which basically is a list of data element names, with additional data, such as a short name, description of the data elements. The data elements are grouped into different data sets, a data set consists of several data elements grouped together because they are connected, for example DTP vaccination data could be grouped into a set of vaccinations data. Important that the sets are not closed, i.e. analyzing should be possible with data from different data sets. The size of the data sets are important, to large causes collecting information not being used, to small causes loss of important information [Heywood, 1994]. Ultimately the data sets should be clearly defined and consist of essential data elements [Heywood, 1994].

Data sets and elements may have different relevance due to different interests of the users. For example: The data element “fully immunized children less than 1 year” might be very important for a healthcare planer looking at the coverage rate of vaccinations, while for clinicians delivering DTP vaccinations it is of less importance, the elements “DTP dose 1”, “DTP dose 2” and “DTP dose 3” is of greater importance for them, as they are more interested in the drop out rate, drop out rate is the percentage of people not getting all the three doses, but only the first and/or the second. (Note “fully immunized children less than 1 year” typically will be a function of the DTP elements and several other vaccinations elements). However, the clinician can be interested in the “fully immunized children” in another setting. All actors have different information needs caused by their different responsibilities; essential information for some might be dangerous for other actors. As too much data leads attention away from the important data [Heywood, 1994]. Typically there is a trade off between what the national level want, the universal national standard of data collected, and the need for flexibility at local level, in information infrastructure terms the problem of homogeneous universal standards [Hanseth, 1998]. To allow all levels to have flexibility while at the same time supporting the national standards HISP uses a hierarchy of standards as showed in figure 4.

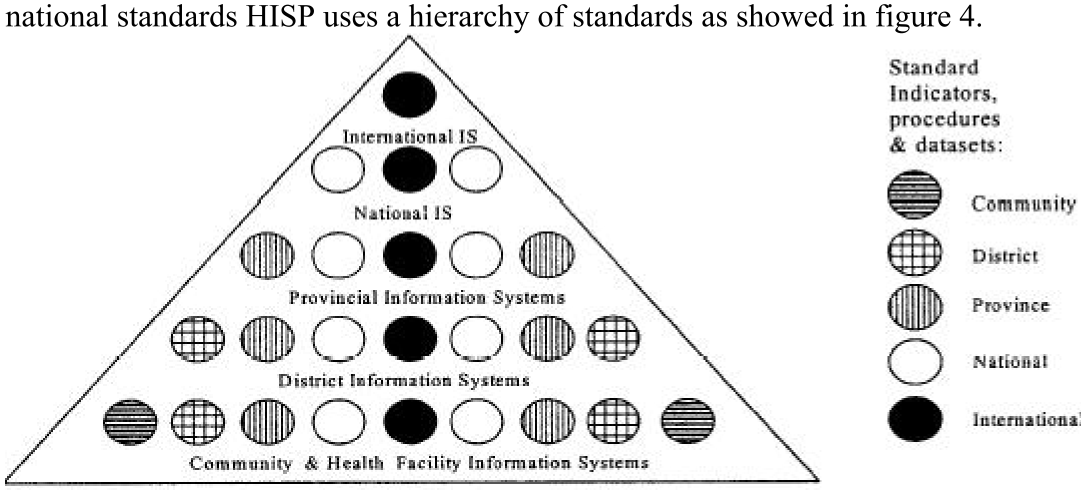


Figure 2: Hierarchy of standards, here each level has the freedom to define their own standards as long as they align with the standards at the level above. [Braa, 2000]. Note the primary source for data is the health facilities in this setting.

Ones the elements to be collected have been decided, it is necessary to ensure that the values collected are correct, important factors are the three C [Heywood, 1994].

- Correct: Is the data representing the reality? Check to see if data is within normal ranges.
- Complete: Submission by all (most) reporting facilities.
- Consistent: Are data in the same range as last year or other facilities?

Thus a healthcare information system should have functions and process to support/check the data values for these factors, basically ensuring reliability and validity.

In this section I have primary been speaking about the routine collected health data. Though, often information from other sectors is necessary, such as education, agriculture, and economics can be of importance [Lippveld, 2001]. Furthermore, the routine data is only from the population using the health facilities. In some poor countries a large part of the population done not have access to a health facility, thus to get information about them you need alternative ways, such as population based sample surveys or actively and routinely go out into the community and collect data [Lippweld, 2001]. The types of data being collected and/or stored in a health information system can be classified into the following groups:

Routine data: Activity data about patients seen and programs run, typical being collected as the clinicians are going about their daily business, examples are number of vaccinations given, number of deliveries. Typically data is being collected on custom made paper forms/sheets by clinicians.
Non-routine data: Special studies and other surveys collected by health workers and data collectors.
Semi-permanent data: Data seldom changing such as, population in the district, facility data (number of beds, number of staff etc.). These data typically does not belong to any program.
Permanent data: Data rarely changed, such as geographical data (roads, rivers etc.).

2.1.5 Analyzing and indicators

Raw data must be analyzed to bring more meaning and make it useful in planning and management, to do this DHIS uses indicators. The indicator approach is also what WHO recommends as shown by the following citation

"*Goals or objectives play an essential part in the formulation of rationales for implementing health policies, programmes and services. Indicators are the basic tools for monitoring progress towards these goals. They reflect the current understanding of achievements and the future directions programmes should take. Monitoring progress is essentially a process of comparison of indicators, over time and across populations.*"

[WHO INDICATORS]

An indicator is the difference between absolute and relative numbers. For instance: the data "1000 children fully immunized" carries no information for a manager evaluating the goal of having total immunization coverage, every child is immunized. This as the data does not indicate how many should have been immunized. Therefore an indicator is made by combining immunized children with how many infants should have been immunized. If the population of infants was 10 000 in the last example, then a typical indicator could be 1000/10000= 0, 1, or 10%, this is called a coverage factor [Haga, 2001]

Numerator
------------------- = Indicator such as coverage.
Denominator

Often the indicator is a result of routine information collected as the numerator, and then semi permanent data as the denominator, for example target population. As the indicators communicate how numbers compare it is possible to use them for data comparison between different sized units. The collected data and the semi permanent is often the responsibility of different staff, and for indicators to be useful both number must be correct. As was the case with the data elements the indicators must be carefully selected in order to be relevant and give meaning full information for the various users.

2.2 Healthcare information systems in developing countries

Health problems in developing countries are often different from the problems in developed countries; problems in developing countries are often caused by [Haga, 2001]

- Economical factors (example: malnutrition)
- Educational factors (example: lack of knowledge about prevention)
- Climatic factors (example: malaria)

While in industrialized countries other problems are of a bigger importance, such as the life style, and the fact that most of the health spending are on curative services in centralized large hospital. Thus the two health organizations need different information systems. Still there are similarities, for example the fragmentation and lack of information systems can be seen in both [Gravitz 2000], [Lippeveld 2001]. The amount of failure among the systems is large, [Heeks, 1999] is refereeing to research indicating that half of all healthcare computer-based information system fails.

2.2.1 Top down & centralized

Historically the emphasis for information required and analyzing has been on the top level, due to various reasons. Such as a political perceived need to have control, or centralized planning due to external subjective assistance (western experts with their own countries health information systems in mind). In addition donors for health projects are often only interested in programme specific numbers for the entire country [Haga, 2001]. These reasons and more have lead to many systems being designed and deployed from a top down perspective. The bottom level is looked upon as an information source without information needs. Data collection instruments and reporting forms have usually been designed by centrally located epidemiologists, statisticians and administrators (so called data people) [Lippeveld, 2001]. The emphasis for the top level information needs has also lead to extensively aggregation of data, this makes quality control and validation of data impossible, due to it being impossible to find for example holes in the data collected, such as none reporting facilities [Braa, 2001].

If analysis done at top level is sent down, it is often obsolete or irrelevant for decision making when received [Lippeveld, 2001], or the aggregation of the data makes local analyses impossible or irrelevant [Potomac, 2001]. In these systems aggregation is not a function but the way data is stored. Thus if there is a feedback mechanism it is usually poor, as stated in [Azelmat, 2001] regarding the health system in Morocco.

"The central level was not able to provide feedback to the periphery within acceptable timeframes. As a result, health service providers and managers limited data transmission to the central level, without much regard for quality"

[Azelmat, 2001, p. 3]

This also brings in another important factor, the quality of data.

2.2.2 Poor data

Due to the centralization of the information system the bottom levels are collecting data that is not relevant for them, thus they are not using the data they are collecting. This combined with poor training in collection of data and lack of feedback on collected data is leading to poor data quality [Lippeveld, 2001] and little data culture. Leading to the problem of: Data is not used because of poor quality, or data is of poor quality because it is not used. The poor data can be of several different types, such as:

- Missing: Lacking data, for instance a health facility not reporting
- Invalid: Impossible data, for example more vaccination given than total population
- Not updated: Old data, for example old population data making indicators wrong
- Transportation: Data typed wrong when transferred from paper to paper, or computer

Another important type of poor data is wrong collected data, the data might be correct but it is not needed. In addition the lack of a reference population might also lead to poor data as the data is not representative of the entire population [Lippeveld, 2001]. This is typically caused by portions of the population not visiting the health facilities due to various reasons, such as lack of money and isolated population groups. These errors are not caught as the clinicians are not aware of their catchments population, and the aggregation at higher levels lead to these data holes being hidden at national level. One solution to this has been to use survey based systems instead, as they can have built in routines for dealing with it. However, this has lead to fragmentation of the information systems.

2.2.3 Fragmentation

Different national health programs information needs varies, this had lead to several top down systems which are emphasizing the difference between the programs and meeting the specific needs of the specific programs[3]. The fact that 85 % of the information needed is common has not been emphasized [Holy grail]. These specific systems are sometimes based on survey data instead of routine data due to the advantages of survey based systems [Lippeweld, 2001]:

1. Enable more carefully data collection (allows for educated data collectors)
2. Less people involved and therefore less vulnerable to manipulation and less difficult to implement and deploy
3. Can easily be designed to meet the specific management and evaluation needs

A disadvantage of the survey systems is that data is not collected on routinely, thus it can not be used for day to day management. And often the way the systems are designed for a specific need means that the information can not be used by other. As the systems do not support sharing of information and are not possible to integrate with other systems.

In addition to these specific systems you usually have the general routine information systems existing, but not being used as it is not delivering the data needed for program management [Lippeveld, 2001], due to poor data collected or lack of collection of wanted data. Thus often the information systems are fragmented into vertical program oriented information systems. [Braa, 2002] is pointing out that these programs /survey systems tend not to be sustainable after external funding and support is withdrawn. A consequence of the fragmentation and the design of the systems is that data collection is redundant, i.e. several data elements are collected several times for different systems, leading to inconsistency and a higher than necessary workload for the data collectors, often the clinicians.

[3] Part of these top down programs are the WHO EPI programs to raise immunization coverage [WHO], which have proven to be highly successful in several countries.

2.2.4 Information for action, The New Routine Health information system

Looking at health information systems today you see several fragmented vertical top down systems. With a poorly functioning routine information system placed among them. Large amounts of data are being collected by the lower levels and reported upwards, without begin used at any levels, old data collection forms continue to be used despite having being scraped [Feldman, 1981]. Extensive research indicates that a comprehensive working routine information system would serve the organization and its users better [Lippeweld, 2001] [Haga 2001], but the problem is that the routine system require participation, commitment and support from all the actors. Thus it is difficult to deploy, and if the system fails, everything fails; thus there is a higher risk involved. All the investments in the custom made program specific systems are probably higher than the needed investment into one properly working routine system, but due to the fragmentation of the investors (donors) and the needs in the organization they are not able to get it rolling.

The need for data use by the clinicians to allow for more evidence based decisions and better management, combined with that fact that most health programs are implemented on a district and health facility level have been important parts in today's theory and policy of decentralization of the public health care [Lippeveld, 2001] [Azelmat, 2001] [Potomac, 2001] [WHO]. An important part in this process is empowering of the practitioners and clinicians at the facility and district level with data collection and use [Potomac, 2001], replacing today's frequently used approach of only intuition. [Feldman, 1981] is referring to case studies indicating a weak link between decisions and related information, lack of amount of data and information collected is not the problem, the problem seems to be lack of use of information in decisions making, or action. This as data collected is either not the right, or not available for processing in the organization. The new system must therefore emphasis collecting of data for actual use. A problem is that in most developing countries the information and data skills among the health workers are very poor [Heeks, 1999], and some workers are threatened by a system that is taking decisions on objective data and are afraid of automation [Lippeveld, 2001], as it can lead uniform reporting. Thereby making it difficult for skilled workers to shown them self [Heeks, 1999]. Thus there is a need for training and building a data culture in the organization to enable empowerment. If successfully done, it is believed that by enabling local use of data the data quality will improve, and better quantify of data and timeliness [Haga, 2001][Lippeveld, 2001]. Furthermore, cost will be easier to contain and the quality of the health care will improve, as stated in [McDonald 1993].

"Improving information systems and information flow can directly reduce health care costs by enhancing provider productivity or reducing the overhead associated with record keeping and payments. Bigger savings can be attained by providing health care

practitioners and managers with better information to manage the individual course of treatment and the overall process of healthcare."

[McDonald, 1993, p. 14]

A example is control of the drop out rate when giving successive vaccinations. For instance, DTP needs to be given in 3 doses to be effective, thus accurate count of each doses given is necessary in order to evaluate and take necessary action if the drop out is to large.

2.3 Complexity of the current reality and change

I will now present informatics theory I have been using in understanding the current reality and the necessary change process in the healthcare organization.

2.3.1 Legacy systems

Given that developing countries are poor, they are not computer free, many of them deployed computers into the health care several years ago [Heeks, 1999]. As such there is a large installed base of users and process in place. To help me analyze the existing systems and gain access to the data stored I will use theory about Legacy systems, or Legacy information systems to be more specific. However, in this thesis Legacy systems and Legacy information systems will be used to denote the same.

What is a legacy system? [Bisbal, 1999, p. 2] states that "*A Legacy information system can be defined as any information system that significantly resists modification and evolution*". [Robertson, 1997, p. 40] states that "*With legacy systems is that they are generally wired into the running of a business in a very substantial way*". As seen in [Heeks, 1999] and [Hanseth, 2002] the information systems are tightly integrated in the organizations using it. A recent example of the complexity and problems with legacy system is the year 2000 problem (which was very fashionable before the millennium change). Adjusting the old large systems to cater for four digit year's representation seemed to be an almost impossible task, the newspaper had articles elaborating on the consequences, loss of power, break down of financial systems etc. As not all the critical systems were guarantied to function afterwards.

[Bisbal, 1999], [Ulrich, 2002] also states the following properties about legacy systems:

- Legacy systems usually run on obsolete hardware that is slow and expensive to maintain.
- Software maintenance can be expensive: because documentation and understanding of system details is often lacking, tracing faults is costly and time-consuming. Old language used can make hiring skilled engineers difficult.

- Lack of clean interfaces makes integration of legacy systems with other systems difficult.
- Legacy systems are also difficult, in not impossible, to extend.
- Legacy systems are an important part of the business and the development of the business

Extensive research emphasis the fact that to be able to work with legacy systems you must fully understand them. Since legacy systems are so integrated into the entire organization they must be examined from all levels in order to be understood.

"*The legacy systems challenge must be tackled at an enterprise level because the installed base of systems and related data is too interdependent to tackle from a one-department perspective.*"

[Ulrich, 2002]

The implication of this perspective on my research is that it is needed to analyze the entire flow of health information from the national level to the bottom levels in order to fully understand the nature of the Legacy systems.

Legacy systems have shown themselves to have a good survivability rate, survived 2000, technical revolutions etc. [Ulrich, 2002], thus often legacy systems are old systems[4]. An important part of what legacy systems are storing is the so called business rules[5]. However, I will not work with movement of business rules, only stored data elements.

2.3.2 Legacy systems change theory

Replacing legacy system is difficult due to several reasons, important are the fact that they are critical and entangled into the organization. The replacement of the system can substitute a large change for the organization as the system are supporting routines which might have to be altered. In [Bisbal, 1999] several solutions are suggested:

- Redevelopment, rewrite existing systems. Described as redevelopment of the legacy system from scratch using a new hardware platform and modern architecture, tools and databases.
- Wrapping, to wraps an existing component in a new and more accessible software component. Surround existing data, individual

[4] Interesting if today's systems are tomorrow's legacy systems? It is difficult answer that question due to the fact that there seems to be no clear cut definition of a legacy system (no age limit for instance)? Only evolution can answer the question I believe, because most definitions are addressing the problem of living with the system several years after deployment and design.
[5] There are several definitions for what a business rule is, Object management group states that: "*Business rules* are rules that govern the way a business operates."

programs, applications systems and interfaces with new interfaces, in essence giving old components new operations or a "new and improved look". Typically the old systems serve as a server for the new "wrapped" system.

- Migration, move the legacy system to a more flexible environment. While retaining the original legacy system data and functionality. Legacy system migration moves an existing operational system to a new platform, while retaining the functionality and causing as little disruption to the existing organization as possible. A typical conflict is the needed added functionality to justify the cost of the project, this should be solved by first migration, and then add the functionality to the new operational system, thus making a small step in the beginning

Often these solutions can and are combined for one big environment; as they are applied at a component level, and a system consists of several components, some information is migrated while others are wrapped. Important is what to do with the business rules, collect and use them or throw them away.

[Bisbal, 1999] also presents several methods for the transition from the legacy systems to the new system:

- Cut and Run: Closing the legacy system and turning on the new system
- Parallel operations: The legacy system and the new system are used in parallel, once the new system is properly tested and trusted the legacy system is turned of.
- Phased interoperability: Replace a few of the existing components at a time; connect the new and old by gateways.

Each of these methods has disadvantage and advantages. The cut and run method is highly risky, due to the element of risk since the new system might not function properly. In addition the change which the new system is enforcing might be difficult to achieve in a flagging day, or a relative short transition period. The parallel operation in contrast, offers better testing of the new system. However, the cost and complexity of having two similar systems operational is high. The last strategy, phased interoperability seems to be relevant for my case. This as my import tool can be used as a partially gateway, allowing data from the already existing systems into the new system. Which can enable replacing of the systems one at a time, thereby enabling a slow and incremental change, i.e. the large change is divided into several few which may lessen the risk, as stated in [Heeks, 1999] and [Hanseth, 1998].

In the phased interoperability a few of the legacy systems components are replaced at a time. Gateways are used to allow the legacy system and the new system to interoperate. This means that the legacy system must be split into separate modules, or separate data portions that can be independently migrated.

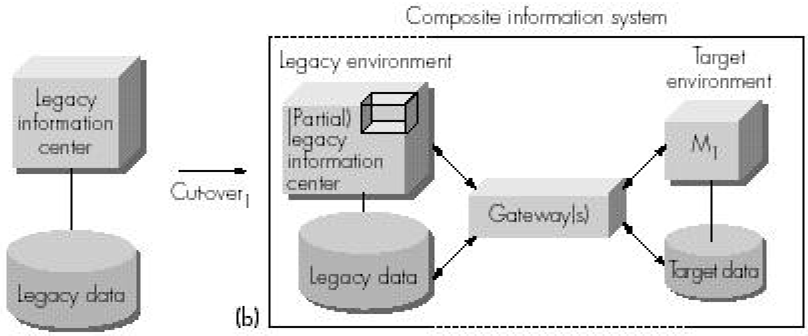


Figure 3: An ad-hoc phased interoperability strategy might be enabled by my tool. The difference being that my import tool is a one way tool, only from the old systems to the new systems. Figure taken from [Bisbal, 1999]

This is difficult due to the monolithic and unstructured nature of most legacy systems. And as such the phased interoperability is potentially highly complex and difficult, if not impossible [Bisbal, 1999]. However, earlier investigations have shown that the legacy systems in Mozambique is highly fragmented; this might be used as an advantage because it might mean that the legacy system are already separated into separate modules which can be migrated one at a time.

Important for the project is that [Bisbal, 1999] have found no successful projects using a comprehensive migration approach, the few successful migration like projects reported in the literature described ad hoc solutions to the problem. The reason for this is the complexity. Important to note that I am not doing a complete transition project as described above. I am only focusing on the data stored; in addition the new system is finished developed. This means that I am avoiding several of the complex problems, such as implementing and testing a system and partially running two systems at a time communicating with each other.

2.3.3 Information infrastructure

The description Health Information System might not be fully accurate. A more correct name might be Health Information Infrastructure as the health care information system described have several characteristics in common with characteristics of information infrastructures as described by [Hanseth, 1998], [Hanseth, 2002]. Investigation the theory I note that Health information systems must typically:

- Enabling for new processes and supporting a wide range of functions, not all decided when deployed
- Support information sharing among a large community of different users with different needs
- Be a part of a large heterogeneous socio-technical network encompassing humans, technological components and institutions

- New systems must respect the installed base, The already existing reporting processes, systems, computers, users etc. in place, a so called installed base
- Interconnectivity, everything is connected and interrelated.

This in contrast to the system characteristics of closeness, developed within a closed time frame, for a homogeneous group of users and tailor-made for a specific problem [Hanseth, 2002].

Looking at it from a higher perspective it also follows the classical economical view of an infrastructure, in that the more users uses the system the more rewards it gives to each user. This as it stores a more complete view of the health care situation, such as more reporting facilities, more reporting data elements. This is called network externalities [Hanseth, 1998]. The network externalities give rise to a classical lock situation. Ability to reward users and the organization is depended upon amount of users. That is, the systems requires change and commitment from several users before it gives rewards, but no one wants to change or give support before the effect can be proven. But the rewards are dependent upon support from all.

Not all information infrastructure characteristics are found in the health information system theory. None of the papers read mention anything about trying to build an underlying information infrastructure to build applications upon, which is something [Hanseth, 200 X] is referring to. In DHIS and others health information systems read about in the papers[6] there is no separation between transportation and for instance reporting. All elements, transportation, reporting and analyzing, are entangled in one system. This partly makes the infrastructure homogeneous; in contrast to the heterogeneous typical seem in infrastructures. For example, extending the infrastructure with new functionality, would consist of updating all applications wherever they are installed. This means that few users have the possibility to add functionality, and as such the systems are little flexible. In contrast to in a typical information infrastructure where you simply would add the functionality by a new application using the underlying transport layer, and leave the old applications in place. Another consequence of the system view is the lack of openness; the systems seem not able to communicate with any other systems, and as such are closed systems. There is no perceived need for an open standard for information exchange; with open I mean a clear and definable standard which other can use without scrutinizing the code. The lack of this can make usage of the data / information outside the systems, and organization difficult, and eventually requiring a gateway, as the only way to connect to the network is by using the

[6] Note I have not done an extensive study of health care information systems theory, only read a few from primarily one source (Rhino network) due to my focus. For instance several of the papers read about infrastructures by Hanseth is using health care as an example of the need for an infrastructure. But he is referring to different systems than the routine reporting system which I am working with.

system. This is in contrast to the use of standards in classical information infrastructure. The closed system implies that the developer is the only actor with power to add functionality. DHIS for instance is developed in Microsoft technology in such a way that all software is locked to Microsoft, such as Operating systems, Databases, Spreadsheets etc., Microsoft products is a standard for DHIS. Thus to be able to connect to the network you need Microsoft operating system, the office pack and the DHIS software package. These factors are reducing the flexibility in the information infrastructure, in contrast to the classical information infrastructure of Internet and the phone network. But it is worth remembering that Internet consists of several networks connected together by gateways and standards, and not one large homogeneous underlying network.

Which term to use?

The health information systems described have several important characteristics typical for infrastructures, at the same time they are lacking some. It is further made difficult by the different view points. For instance DHIS from a technical viewpoint is a closed system (not able to share information with other systems). While from a user viewpoint it can be seen as open due to the customizable data items to collect and customizable analyzing possible. Note also that [Hanseth, 1998] is stating that there is no clear cut definition for information infrastructures. And that there area different kinds of infrastructures (global such as internet, sector and corporate), and in that respect he is stating that the web of closely integrated systems in large corporations does not fit the system notion, and I would say the same is true for the health care information system.

In this thesis I will no longer discuss the differences between the infrastructure and systems, and I will use the following terms:

- Health care information system: Any health information system made for reporting of health related information, the systems may together form a network of systems, or infrastructure.
- Health care information infrastructure or information infrastructure: All the elements part of the information processes in a health care organization (such as paper forms, computer programs, processes).
- Infrastructure: Traditional infrastructure such as roads and the existing organizational facilities and units

I will use the legacy theory in order to analyze and understand the existing system, at the same time using information infrastructure theory to view the entire information infrastructure of several systems, paper forms, users etc.

Why use Information infrastructure theory?

The primary reason for using information infrastructure theory is to elaborate on the complexity of the reality the system is to be deployed into, and explain the difficulties of change. Important with information infrastructure development and evolution is the fact that you are never beginning on scratch, new information infrastructure it is always integrated and extending the installed base [Hanseth, 1998]. This means that the traditional design concept of making a system from scratch and exactly as you want it to be, does not hold true for infrastructures [Hanseth, 2002]. Infrastructures are evolving over time, where the existing infrastructure, installed base, strongly influences it [Hanseth, 2002].

An important element left out of this thesis is the problem of trying to predict eventually problems with DHIS in Mozambique due to its systems properties, because of now begin tried deployed as an infrastructure. For instance a claim could be that the lack of flexibility in the finished product is solved with prototyping in South Africa, while for Mozambique this might not be available. Thus the lack of flexibility might be a bigger problem in Mozambique; I will only briefly touch this area.

Actor Network theory (ANT)

Actor network theory (ANT) is a way of representing the reality in order to unpack the complexity of everyday life [Hanseth, 1998]. ANT is much used in Information infrastructure but it has its origin in social sciences. In ANT the selected area of interest is viewed as a heterogeneous network of human and technology. The term “Actor” denotes a node in the network, which can be anything such as human or technology. All the actors are connected together in a network by their reflexive influence upon each others. The technical artifacts (actors) all embody a pattern of use, this is called inscription. And this inscription can be used in order to influence the network, by translation wanted change into inscription in the technology.

As an example of ANT I will use the process of driving a car and the speed of the car. Typically the speed is determined by several factors, such as the engines capability and your own driving ability. If reduced speed is the wanted change the engines capability can for example be degraded, an inscription. However, degrading of the engine will also influence others areas, such as the maximum load of the car, this shows the interconnectivity. This interconnectivity means that it is important to set boundaries when using ANT, if not all problems turns out to be very large. Different actors are translation their interests into inscriptions of varying strength into the technology. With varying strength it is meant how flexible the inscriptions is, how easy to avoid. In the car example a speed sign on the road, would be a weaker inscription than an automatically photo box which takes pictures of car going faster than the wanted speed. From an ANT viewpoint, a computer program designer is inscription the technology with inscriptions of varying strength.

Installed base

An important part in Information Infrastructure is the concept installed base [Hanseth, 1998]; the installed base is the existing systems, existing work process, existing users etc., or a heterogeneous actor network (ANT) consisting of human and technology. The installed base is important due to its characteristics of being uncontrollable and influencing the evolution of the infrastructure. It is uncontrollable in the sense that no actors alone have total influence upon it, though several actors have limited influences upon it [Hanseth, 2002]. One of the reasons for its uncontrollable characteristics is because the elements are entangled in each other, high degree of interconnectivity. Therefore the indirectly consequences can be larger than the direct consequences[7]. The technology in an information infrastructure is looked upon as an independent uncontrollable actor, which can have the organizations processes and objectives translated into inscriptions.

What the installed base is varies depending upon what kind of infrastructure you are looking at, but an important element in information infrastructure is behaviour inscribed into already existing elements (such as applications, processes). A claim could be that in my case it looks like that the computer systems consist of several independent computer systems working independently, if one system is changed, no/little changes will spread to the other systems, and thus you could claim that the installed base is controllable since the amount of indirectly consequences are limited. But the installed base consist of more than technology, for instance the paper reporting forms (being used at lower levels) and processes are also part of the installed base. This as the forms and processes are connected to the computer systems as it is the information source for the systems, and a change in the systems can heavily influence the paper forms (for example frequency of data reporting).

Lock-in

The situation today regarding health care information system looks like what [Hanseth, 2002] is referring to as a lock-in, a lock-in is a situation where a technology has been adopted and gained an installed base. The size of the installed base makes the coordination effort and the switching cost huge, meaning that it is hard or impossible to develop competing technologies [Hanseth, 2002]. Lock-in are often a result of path dependency, which is that past selections have a large impact on future development [Hanseth, 2002], the classical case for a lock in and path dependency is the evolution of the QWERTY keyboard layout. Today keyboards have the layout they have due to decisions done almost a century ago regarding manual typewriters. And due to the size of the installed base today the coordination effort necessary to switch is huge [DAVID].

For the health care in developing countries today it seems like that the selections of computer systems and deployment of them in the past have lead the health care to a lock in situation, as the existing systems are tightly integrated into the large organization and have a large installed base. The need

[7]For example increased integration typical leads to increased side effects, which in turn can lead to less control.

to switch is documented [Lippeweld, 2001]. But switching is a huge coordination project and highly costly. The risks (and costs) involved with a change are large, and everyone is better off just sticking to the existing systems and processes. Alternatives for avoiding lock in are for example to define and adopt good universal standards [Hanseth, 2002], which means that no one is binding to a particular technology, standard can if correct used lead to flexibility. However, over time the standards will prove inappropriate to new circumstances. Then you will have a lock in, example IP ver4.0 not enough IP addresses for today, trying to release IP ver6.0 today. Ultimately the infrastructure must change, and to help with this it should be designed as flexible as possible [Hanseth, 2002], flexibility for the user's means that the installed base will less likely enter a lock in. While flexible technology makes it easier to leave a lock in once entered. And as such flexibility is important when dealing with information infrastructure and the change of those.

2.3.4 Information infrastructure strategies for change

[Hanseth, 1998] describes three generic strategies for changing an information infrastructure, these are:

1. Evolutionary: Slow, incremental process where each step is short and conservative
2. Daring: A faster process where each step is longer and more daring
3. Radical: fast changes, a radical break with the past

The radical and daring strategies are mostly difficult to implement due to the role and nature of the installed base and network externalises. The rewards of switching to the new system which no one is using are small; everyone is waiting for others to switch. However, radical changes have several times been a success, such as AOL e-mails switching to Internet in a day. But in general radical switches are difficult to implement, usually at least two preconditions needs to be in place for it to be a success [Hanseth, 1998]:

1. Easy change that everyone can do them at the same time (flagging day)
2. A central authority to organize and enforce the change

Looking at health information system theory I note that the complexity and size of the changes seems to large for doing in a flagging day. There also seems to be lack of a central authority with total control upon all actors, therefore the evolutionary approach seems to be the most appropriate.

Evolutionary strategy and gateways

[Hanseth, 2002] states that infrastructure grows as a self reinforcing process, the more users, the more valuable the infrastructure becomes, thus more user wants to use it. And once stared to grow the growth can be very fast. The growth, or development is uncontrollable by any actor alone; each actor has limited influence. Only the general direction of the growth can be influenced. Note that the self reinforcing process might be good or bad for the organization/installed base, thus being able to pull it in the right direction is important. Forcing it is impossible, only extraordinary events of large proportions can influence it alone [Hanseth, 1998].

The evolutionary strategy consists of changing a small part of the network, then making sure the newly added parts work properly. Then move on to change the next small part. This means that the change is spread out over time, each step being small. Thereby reducing the change of failure, as [Heeks, 1999] points out, higher amount of change, means a higher change of failing. The small changes help making the coordination projects smaller. In actor network terms the change is viewed as keeping the network aligned.

The fact that all steps are to be small, can bring about a problem with that the amount of reward for the first users will be small, this as rewards are dependent upon number of users. Therefore the first user might not switch due to the small reward in switching. One solution to this is to utilize the existing infrastructure by connecting the new part to the old infrastructure[8]. In order to connect two otherwise incompatibly infrastructures/networks to each other a gateway is needed, or converter [Hanseth, 1998]. The gateway thus allows two/or more networks otherwise incompatibly to co-exist and speak to each other while switching. The gateways are also important in the further development process where each new small step is connected to the existing infrastructure. In the case in this thesis, gateways could be used to allow for backward compatibility to primarily gain support from the first users, i.e. allow the first users the ability to utilize the existing infrastructure. Gateways can have other purposes, for example they can let several incompatible networks communicate with each other, thus the gateway becomes an important part of the network and a part of the installed base, and not something temporary as here [Hanseth, 200X]. Gateways can if probably used be an important tool in gaining support, extending and improving the installed base.

The evolutionary strategy has some similarities with the phased interoperability strategy from [Bisbal, 1999] regarding legacy systems. Important and relevant differences for this research are that the evolutionary approach of [Hanseth, 1998] has a higher emphasis on the installed base (cultivation), the need to grow slow and evolutionary and that the new systems should work like a part of the rest. The old part is not removed as in phased interoperability, whose aim is to replace the legacy system. Therefore in infrastructure theory the gateways become mandatory parts. While in legacy systems theory the gateways are looked upon as temporary solutions (Also a point made by [Hanseth, 2002X]).

[8] There are other solutions such as making the first version quick and tailored to a specific group of users [Hanseth, 2002], which will enable the system to quickly gain a installed base, but I believe this is not so relevant for me due to the already finished developed system

Cultivation
[Hanseth, 2002] is referring to other studies viewing IT development and implementation as a part of organizational transformation, this is much in line with what [Heeks, 1999] suggest might be successful strategies for development and deployment of IT system in the health care. In information infrastructure theory the strategy of change and development is called cultivation (of the installed base) [Hanseth, 1998].

Cultivation requires a close analysis of the way behaviour is inscribed in the already existing elements of an infrastructure, the installed base. This knowledge is used when designing the new system, or parts of it [Hanseth, 1998]. Cultivation is a conservative belief in the power of natural system to withstand our effort at design either by disarming them or by ruining them by breakdown. The designer can only influence the installed base, and not design something totally new, as the evolution is uncontrollable by any actor alone. Cultivation is a middle thing between design (with a high degree of freedom), and determinism (no human control) where the technology is following its own rules. Cultivation can also be used in the adoption process, in that analyzing of existing system can help with adoption of the new system to fit more with the installed base, thus be better aligned with the installed base. This requires that the new system is flexible. HISP refers to the continuous process of defining standards for the health care in South Africa as cultivation [Braa, 2000].

"By cultivation, we mean a slow incremental bottom-up process of aligning actors by enabling translation of their interests and resources already available from the base"

[Braa, 2000, p 4]

Thus cultivation is not only about developing new systems, but the slow change of an organization by negotiations and brokering between actors at multiple levels.

2.4 System development, prototyping

A prototype is an initial version of a software system that is used to demonstrate concepts, try out design options and generally, to find out more about the related problems and its possible solutions. Prototyping is an approach based on releasing prototype(s) and experimenting with them, the finished system is building upon experience gained from the prototype(s) experiments, and sometimes the prototype in itself [Budde, 1991]. Prototyping is a popular approach to use when the requirements are difficult to decide before the system is implemented.

The prototyping approach used in this thesis and within HISP is an incremental evolutionary participative one. The prototyping used is a continuous process with the objective to deliver a working system; new versions are continually released building upon requirements from the user, making the process incremental. This is in contrast to throw away prototyping, where the objective is to identify user requirements by evaluating use of the prototype, then make a working system [Sommerville, 1995]. The participative approach used is informal and to large degree based on improvisation, each interested user can have access to the development regardless of their place in the hierarchy. Typically advantages of user participation are that the knowledge upon which the systems is built is improved, the users more easily develop realistic expectations and increasing workplace democracy, giving the members an opportunity to participate in decisions that influence their work [Braa, 2002]. These elements can also be part of lessening the resistance to change and secure support from the installed base, DHIS for instance stood out in a positive way from other projects in South Africa by the use of prototyping [Braa, 2000]. Prototyping as used by HISP means that the translation of decisions into inscription in DHIS were not clear cut between the designers and users, it drifted as a cycle of translations in a process not controlled by anybody [Braa, 2000].

2.5 Digital data preservation

More and more data is stored digitally in today's societies; this has several advantages, such as higher availability of the data. Unfortunately digital storage has proven to be very bad longevity storage [Rothenberg, 1998], this combined with lack of maintenance have lead to data loss. The data loss can come about due to several reasons but most important are [Rothenberg, 1998]:

1. Decay and obsolescence of the media and or hardware used to store the data. Such as the standard music tapes used to store data on several old personal computers from 1980, today very few computers have hardware to read these tapes. Furthermore, the magnetic storage can decay leading to completely loss of data.
2. Software interpreting the data becomes obsolete. As a consequence data becomes inaccessible and or unreadable. Due to for instance coded data such as rating of an element with number from 1-5, is 1 best or worst? And, or the data might be stored in proprietary or undocumented formats, requiring original software which only run on old hardware.

These elements are often mixed; data is stored on old hardware, in old back ends and coded to save memory. Other problems regarding multimedia data is for example multiple formats, today we have several digital sound formats: *.waw, *.mp3, *.raw, etc. Decoding these formats in the future might be very difficult as you do not necessary now what the bits represent (standard ASCII chars, video, sound etc.). This is a problem likely to increase in the future; Due to the fast development it does not take many years for data to become lost.

"No one can access the reams of project information - equivalent to several sets of encyclopaedias - that were assembled about the state of the nation in 1986. ... It is ironic, but the 15-year-old version is unreadable[9], while the ancient one is still perfectly usable,' said computer expert Paul Wheatley. 'We're lucky Shakespeare didn't write on an old PC."

["The Observer" March 3 2002]

This shows that the problem of old data is very much alive, and will continue to exists in the future as well. [Rothenberg, 1999, p. 1] states a conclusion from a major Task Force on the Archiving of Digital Information in USA in 1996 as: *"there is, at present, no way to guarantee the preservation of digital information. "*, or in other words, maintenance of digital data is lacking. This shows that the problem of preservation of historical data is a huge problem.

Preserve the old data
The area of preserving digital data is large, with several different solutions, such as:

- Build systems with the functionality to store data in an archival format, i.e. store the data with documentation for interpretation, but this does not solve the hardware related problems
- Build an emulator of original software which will run on the computer of the future, for more se [Rothenberg, 1988].
- Allow data to migrate[10] to the new systems

Migrating of data is the solution my research is relevant for, as my work will consist of pulling data out from legacy systems, and then pushing them into a new system. Though, the focus during migration process are often only on the active data [Rothenberg, 1988], with the result that purely historical data is left to decay in the legacy system. Furthermore, data migrated to new systems can not said to be preserved as the fast development will probably led to the new systems becoming obsolete in a relative short time (1-10 years).

2.6 Healthcare success/failure concepts

As described earlier, health information systems in several developing countries are failing, or performing poorly. Why has this failure come about? [Heeks,

[9] Unreadable meaning don't have the software or hardware necessary to read the information
[10] "Migration of data", "target system population", or "import of data"; which term is most approximately to use in my case? I have decided on "import of data" to refer to the actual movement of data, which is part of the larger process of "target system population". The term "migration of data" will primarily be used in relevance to theory. This as I believe that "migration of data" also means that the data moved starts to "live" in the new system, i.e. being used. Which might very well take place in my case; it is my aim, but not my primary focus as other HISP team members are to work on related process (training etc.). The larger process which I am part of can as such be referred to as the migration of data.

1999] has a paper on success/failure[11] of health care information systems which provides me with important concepts to help in explaining the failure; the concepts will also be used to explain the shortcomings of the existing systems in Mozambique.

In [Heeks, 1999] a theory based on conflicts in reality gaps is used to explain the failure/success. "They argue that the greater the gaps between current realities and the design conceptions of a new health care information system, the greater the risk of failure", or the larger the amount of organizational change the systems requires the larger the change of failure (higher risk level). This is in conflict with the formal purpose of healthcare information system, with bringing about organizational change (more efficient organization for example). Thus a trade off is too made between the amount of change a system inflicts and the change of success, note there is not necessary a link between amount of change and size of benefits. Even though the gaps are necessary/ required, being aware of them and eventually eliminating unimportant gaps is important (for example a gap not leading to a wanted change). Basically you want to identify and control the gaps to allow for a higher change of success. The gaps are grouped into three types.

Rationality – reality gaps: A gap between what the organization and the users want, and what the systems has been designed to deliver (in scripted procedures). This gap can arise from several reasons, such as wrongly focus on the users needs, over or under representation of medical, managerial or technical users needs. And/or some user groups have been unable to express what they want from the system.

Country gaps: Gaps between health care systems from different countries caused by that all countries are different, thus they have their own reality, for instance different objectives and values. This gap can be reduced by adapting foreign software to the local reality. Due to highly different societies and values this gap is large between industrialized and developing countries [Heeks, 1999].

Private – public sector gaps: Gaps between health care systems in the public and private sectors caused by the different realities in the two sectors. Used to explain problems when transferring health care information system between private and public sectors,

Of importance to me are the country gap between the industrialized and developing countries and the rationality gap. Interrelations between the groups exist, for example most systems in developing countries are partly developed and financed by industrialized countries. This means that industrialized ideas and values influence the development, and as a consequence the specific custom made system can have a rationality gap caused by the country gap.

[11] The theory is mostly about systems that are failing in the deployment phase or soon after, while I am looking into an organization which has used several systems for a long time.

2.6.1 ITPOSMO model

[Heeks, 1999] bring forward a model of healthcare information systems conception called the ITPOSMO, called so because of its seven dimensions it looks at the systems (the gaps are grouped into the dimensions). **I**nformation, **T**echnology, **P**rocesses, **O**bjectives and values, **S**taffing and skills, **M**anagement and structures and **O**ther resources. All the gaps can be explained along these dimensions.

Information

Do the systems meet the real information needs of its users? The existing systems in developing countries seem to have large gap along information need among all the levels. The bottom levels seldom get any information from the systems, while the top levels are not getting all they want, and/or lots of it is poor quality, for example the lack of integration means that integrated planning is difficult. But do all the users, specifically lower levels user have an information need? Maybe not individually, but the organization have a need for information use and higher quality data collected among the lower levels as described in previous sections.

Technology

Does the system require large technology changes? Is the required hardware easily available? The technological infrastructure is limited in most developing countries, but they are not computer free areas as stated by [Heeks, 2001]. The existing systems which are top down require a few computers in central places (such as cities and other organizational centres), where the infrastructure probably is better than in rural areas. The new decentralized systems might have a large technology gap if they want to computerize the lower level(s). They need vast amounts of computers and various ad hoc solutions due to the poor infrastructure (lack of stable power supply, telephone etc). Important to note that a decentralized information system does not necessarily require more computers, analyzing can be done on paper, or responsible personnel could travel to a computer central to use computers.

Processes

Does the system support the processes the user wants to? Is it bringing about large changes in processes? The existing systems have brought about a large change of process, at the same time not supporting all required/wanted processes. However, the top levels are better supported than the bottom levels which has to collect vast amounts of data not serving any purpose for them. The new systems will require an even larger change in processes, as it will enforce a more systematic guideline or approach to clinician's daily work. This is a change that is seen as necessary and important. However, it might not be perceived by the entire organization, and therefore training and education is mandatory in order for the organization to change from within.

Objectives and values
Does the system support the objectives and values its user finds important? The existing systems have been designed to only support the top levels objectives and values, thus the large group of user at the lower levels sees no, or little value in the system. In addition the need to use the systems for management purposes is not being supported today; as there is little integration and the data is of poor quality.

Staffing and skills
Does the system require a higher degree of skills than is available? Both the new decentralized systems and the existing systems require more skills than is available today, one of the reasons for the failure of today's routine health information systems is the poor skills among the data collectors, the new systems will try to address this problem by education and local use of data. This point is important point because as [Heeks, 1999] is pointing out several skills, management, planning etc. among health workers in developing countries are poor and more limited than in industrialized countries. Eliminating, or minimizing this gap by training is seen as the most important part of the new systems, this will be more described later.

Management and structures
Does the system bring about large management and structures changes? The new systems want a shift in power downward, and will probably face conflict in doing so as stated about deployment of DHIS into South Africa in [Braa, 2000]. This conflict has probably come about due to that health care systems in developing countries are centralized and hierarchical [Heeks, 1999]. Thus the central and top levels will lose control due to the new system and might fight against it.

I will use the IPTOSMO in evaluating the existing healthcare information systems in Mozambique; by examining gaps along the seven conceptions.

The gap
The large gap between the new decentralized information systems and the organization is required as it is believed that today healthcare organizations are performing poorly. And the new health information systems will enforce better routines and work processes in the organization. However, the old existing systems also have several gaps today, but they seem to be smaller than the gaps in relation to the new systems. This is properly due to the old systems being part of the organization, and has influenced it over several years. The large gap of the new systems is pointing to that the systems will be a failure. Needed is therefore a strategy for deploying the new systems, the reality needs to be adjusted to fit the new concepts and systems through training and organizational change. Thus it is a process that contains more than just technological development and deployment, according to [Heeks, 1999] the

focus on solely technology is to narrow as it is a multi dimensional process. Therefore health care IT professionals must look upon themselves as change agents; they must implement and/or assist the change along the identified dimensions. This requires a broader set of skills such as communication, negotiation. A part of this is what is called Hybridisation agents, which are IT professionals with understanding about both the health care context and change management. I believe this notion of assisting change fits very well with [Hanseth, 1998] information infrastructure change theory, where the direction is influenced.

2.6.2 Rationality-imposing applications versus reality-supporting applications

[Heeks, 1999] is arguing that applications can be positioned along an axis consisting of rationality-imposing applications versus reality-supporting applications.

Rationality-imposing applications incorporate a whole series of assumptions about the presence of rational information processes, objectives and values, management structures etc. [Heeks, 1999] (or inscriptions). These rationalities must either be present in the organization, as a pre-condition for successful implementation, or they must be imposed. And as such the application has a large reality gap that needs to be controlled, or minimized. If the organization does not impose the rationalities inscribed into the application, then the application run a high risk of failure in that organization. Thus rationality-supporting applications usually need to be combined with a change in the organization. Example of a rationality-supporting application is DHIS.

Reality-supporting applications require fewer conceptions as pre-conditions or to be imposed to be met for the application to function successfully; thus they usually require a smaller organizational change [Heeks, 1999]. Typically examples of these systems are: Spreadsheets (such as Excel or Lotus), modern powerful word processing applications (Word and WordPerfect) email etc. These applications can, and are working successfully in a wider variety of organizational environments. Another benefit of them is that the cost and risk usually is lower due to them often being shelf ware[12].

[12] Inexpensive because they often are shelf ware (finished developed) and have a large user group to cover the development costs. This in contrast to customized software which typically must be developed for a often small user group and therefore have a higher price and risk, however the customization can enable much higher efficiency.

2.7 Summary

There is a need for a change in the routines and processes within public healthcare in developing countries today. Routine health information systems can be an important part of this change, if successfully deployed it can enable better management, better processes and better control, thus enabling better healthcare for the public. However, there are several old fragmented systems in place, storing large amount of data which can not be lost. This data must somehow be transferred to the new system in order to make the change easier. The size and complexity of the existing organization combined with the large needed change make the change difficult. Theory emphasis the need to change slowly and evolutionary, allowing the organization slowly to adapt and influence the change to the new information systems inscribed processes.

3 Background

The work described in this thesis is part of the Health Information System Program in primarily Mozambique. This chapter gives a description of HISP and the District Health Information System (DHIS) made by HISP South Africa. In the end the Mozambican context is presented.

3.1 HISP description

Health Information System Program (called HISP) vision is:

"To support the development of an excellent and sustainable health information system that enables all health care workers to use their own information to improve the coverage and quality of health services within our communities"

[Source: www.hisp.org (January 2003)]

HISP started in 1994 in South Africa as a local pilot project in three districts, emphasizing prototyping and local use and analyzing of data. It was initiated by researchers from Norway and the Universities of Western Cape and Cape Town (South Africa) [Braa, 2002]. Over the years HISP have developed the DHIS software which has become the national standard in South Africa today, 2002. HISP primary focus is on developing countries, implementation is primarily funded by donors, such as EQIUTY/USAID, whereas the research is funded through NUFU (Norwegian University Council/NORAD) [Braa, 2002].

HISP started in South Africa after the end of Apartheid[13]. After Apartheid South Africa started a process of reconstructing the entire society in order to provide its citizen's equal rights and opportunities, before different groups had different right and opportunities. An important part of this was the reorganization of the public health system to provide all citizens with basic health care. Important was the "right for health" as described in the African National Congress (ANC) (winner of first free elections) national health plan, a healthy population was deemed as a prerequisite for social and economic development [NHP]. A bottom up approach sensitive to local customs emphasizing prevention rather than cure was selected, typically as World Health Organization (WHO) has defined Primary health care in the Alma Ata declaration [WHO AA]. An important part of the reorganization was implementing a new national health information systems as the old were highly

[13] Apartheid comes from the words apart and heid (hood). In Webster defined as policy of segregation and political and economic discrimination against non-Europeans groups in the Republic of South Africa. An good example is the health care during apartheid which emphasized curative hospital services for primarily whites citizens, 60 % of the resources where (and are) used by the private sector serving 20 % of the population [Braa, 2002].

fragmented an adapted to the old unequal fragmented organization [Braa, 2000]. And it was from this setting that HISP grew.

3.1.1 HISP organization

HISP is, and has always been a collaborative research and development program between several universities and governments organizations, such as ministries and departments of health. This allows them to shape an ongoing process at the same time as being an outsider academic actor, having emphasis on elements such as innovation, research and development of new strategies. Several courses and students doing / and have done their degree (master & PhD) within the HISP team are part of the academic factor. HISP core actions and elements are [Braa, 2002]:

- Participatory prototyping (of DHIS and the adoption process), software should support process and not drive them.
- Bottom up, empowerment and training of local health workers to use data, establish routines and procedures for handling information at health facility and district levels.
- Support the process of defining good health data standards and indicators
- Scaling and sustainability. Important that the processes should be institutionalized and gain a critical installed base to support sustainability and scaling of action, i.e. changes are working and evolving after the researcher leaves

The emphasis on bottom up has not excluded work on the top levels, as HISP has always worked on networks and not singular units as that often results in non sustainability [Braa, 2002]. Important part has been the enrolment and alignment of key actors on multiple levels of the network.

Technical and social challenges have both been important, correct alliances and support from key actors are just as necessary as good software, a combination have been the key to success in South Africa. A success in that the data coverage in South Africa has never been so good, 95 % national coverage by the end of 2001, politicians have started asking for data reports. However, still vast amounts of work remains in building an information culture and a support network.

3.1.2 HISP internationally

HISP has over the years grown to several countries creating a network where Norway and South Africa are the central hubs. India and Mozambique being the primary nodes, followed by Malawi, Tanzania, Cuba and Mongolia, and with work starting in Angola and the Dominican Republic resulting from independent work of the nodes.

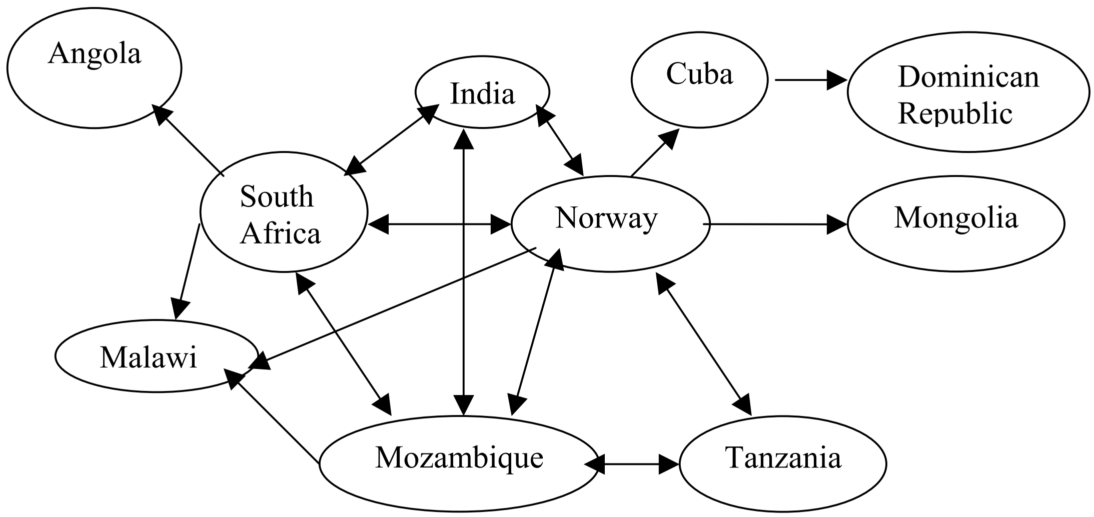


Figure 4: The HISP network, figure taken from [Braa, 2002]

In accordance to HISP ideas the adaptation process in most of the countries is driven by participatory prototyping/ adaptation of the DHIS and training of local people to enable an information culture to be institutionalized.

3.1.3 HISP Mozambique

Mozambique was the first country outside South Africa that HISP moved to; a HISP team was formed in January 1999 with an emphasis on three provinces. Two of these provinces were severely hit in the flooding in 2000; leaving only Niassa in the far north untouched. Support has been gained from the National level, but actual backing has proven to be less than expected [Braa, 2002]. The HISP team in Mozambique consists of 5 PhD students (3 medical and 2 informatics), 2 senior officials from the Ministry of health and a South African expatriate expert, the team is tightly integrated with the faculty of medicine at the Eduardo University in Maputo, Mozambique. All the students are registered at the University of Oslo (international cooperation is an important part of HISP).

Preliminary studies have found similarities in Mozambique to the situation in South Africa [Braa, 2000], such as excessive data collecting and little local data use, while there are also dissimilarities. Such as that Mozambique have national data sets and a unified structure consisting of several stable national vertical health programs, imposing standards and policies at the lower levels [Braa, 2000]. This was not the case in South Africa. Furthermore, in Mozambique the province level is looked upon as the information hub in contrast to the view in South Africa that the district is the information hub. Note that the districts in Mozambique are smaller and have fewer human and other resources than a typical South African district. As a consequence the provinces are viewed upon as an important actor in having DHIS rolled out to the districts and institutionalized [Braa, 2000].

HISP has started on a new participatory prototyping in Mozambique which includes adaptation of the DHIS software. Including translation of DHIS to Portuguese, and the installation of DHIS on several computers at district and province level. Important parts have been the making of a data dictionary, defining minimum data sets and set up of the infrastructure in DHIS. The installation has been combined with training and work shops. In which primarily the district staff have been trained to collect, analyze, and use data, and DHIS. However, there have not been built any routines for report of DHIS data, and all formal collecting and reporting of data is done without DHIS, DHIS is not institutionalized.

Due to the lack of resources and infrastructure in the districts in Mozambique, the idea of implementing DHIS in all districts is not feasible. Several different solutions are being evaluated, such as making some districts as “Information centres”. Which staffs in surrounding districts travel to regularly to report, analyze and use information.

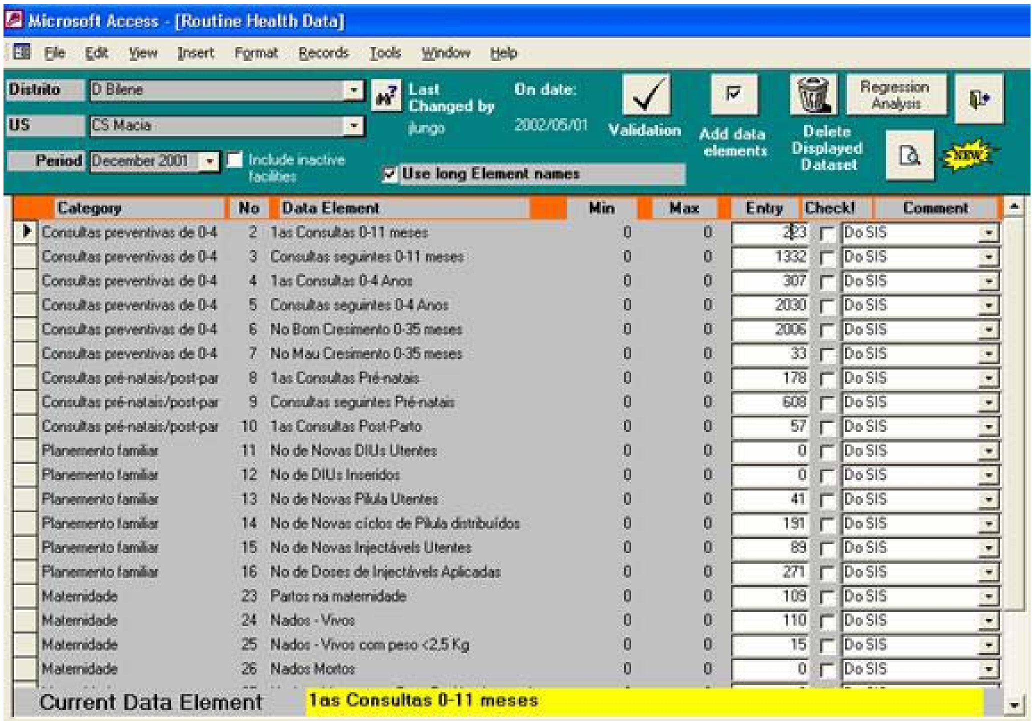
3.2 DHIS description

DHIS is a routine District Health Information System designed by using participatory prototyping in South Africa. It was initiated and is managed by HISP. DHIS is designed to support collection and analyzing of a relative small set of data items, its primary purpose is to enable for use of data at all levels, including the bottom levels. Thus being a bottom up empowerment tool. Development is done in South Africa by a local computer firm, Soft Craft. DHIS is seen as an enabling factor, and not the driving factor in accordance with HISP principles. Though, it is still an important factor as the information systems is an important factor in having changes institutionalized, and thereby heighten the change of sustainability.

The software is based on a Microsoft Access relational database as a back-end storing all data. With Microsoft Excel (for data analysis), Word (for making reports) and forms made in Visual Basic (for data entry etc). Visual Basic is the glue and the tool used to develop the graphical user interface (GUI), the code is open and free; but the modules from Microsoft being used are not free and open. A deployment of DHIS needs Microsoft licenses for at least 400-500 $, operation system and the Office pack. Excel and Word are closed systems making export of data outside a Microsoft environment difficult. Though, Access has several export tools. Currently DHIS is running on most of the available Microsoft operating systems and office pack combinations[14], due to

[14] As of today DHIS is being used on Win95B (second release), Win98, Win Me, Win 2000, WinNT and Win XP in combination with Office 97 sr-2b and all later Office packs. Further increasing the diversity is the fact that each of these systems again comes in different version (typically updated by the service packs from Microsoft).

requirements of the healthcare in South Africa, which does not have the money to update all software and hardware regularly.



Picture 1: A screen capture of DHIS with some of the data elements imported from SISC during the research. English is the language due to it being the version used by the researchers.

Which data elements to collect are fully configurable to allow for needed local flexibility, in order to enable local use of the data. In addition to this several customized validation rules can be added to the data elements to support better quality data, ranging from simple minimum and maximum values to more specific ones, such as no more vaccination given than stockpiled. DHIS is an important tool to achieve the HISP goals, but not an end in itself. Consequently, implementing DHIS consists of large amounts of training in data use and collecting. As DHIS is a tool that requires data understanding and skills to be properly used.

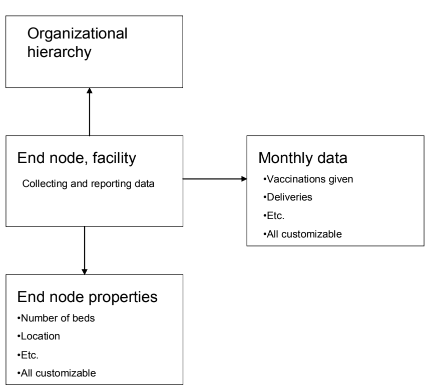


Figure 5: Information stored in DHIS and the connections

In addition to the routine reported data elements, DHIS is storing the organizational structure and the infrastructure, which are all the health facilities linked to the organizational hierarchy. For each facility several properties are stored, such as number of doctors working there, population.

This is not perm-anent data as facilities are upgraded, closed and opened. Therefore it is called Semi-perm-anent data. The semi perm-anent data is an important part in calculating indicators such as coverage factors. The organizational structure links all units to its parents (mother parent relationship). Typically all facilities have a mother district, the mother district again have a mother province, the province then have a region, and the region is connected to the national level.

The prototype approach objective of delivering a working system and the concluding of the prototyping process is difficult to define for DHIS. This as DHIS has been working for several years now, with new versions continuously released, fixing bugs and adding functionality. Some of the new versions are due to DHIS being implemented in other countries. Though, the evolution and new requirements from the South African public health care organization also influences this process.

The implementing in Mozambique and the other countries is viewed upon as a new pilot processes, instead of replications of the HISP experience and DHIS from South Africa. This means that today DHIS is being extended by prototyping in several countries, and it is necessary for DHIS to become international software. For example, the first Portuguese version of DHIS for Mozambique was translated by editing the code as the English language was hard coded into the code. The process of making software international is by [Yeo, 2000] described as separating the software into modules according to culture sensitive parts and internationals parts:

1. International core, culture independent: Typically the program logic
2. Localized part, culture dependent: Cultural sensitive parts such as text.

This process for DHIS has been started, as shown by the recently released multi-language version of DHIS, where all text strings have been exported from the code to a database. However, several more modifications are needed to increase the flexibility so that it can be adapted to different countries. The separation between international core part and localized parts is not a clear cut and easy one.

I sum up DHIS by stating the 5 principles for DHIS, taken from [Heywood, 1994]:

1. Support the district-based primary health care
2. Collect essential data used to calculate indicators
3. Encourages decentralized use of information by health workers.
4. Includes all service providers at all levels
5. Is integrated with and supports other information systems

3.3 Mozambique

Picture 2: Map of Mozambique with main cities

Mozambique, a Portuguese speaking republic lies in East Africa along the coast, sharing borders with Tanzania in North, with Malawi, Zambia and Zimbabwe to the west, and Swaziland and South Africa in the South. Mozambique is today (2002) one of the poorest and least developed countries in the world as shown by its ranking of 170 among the 173 countries ranked on the UNDP human development index for 2002 [UNDP HDR2002]. However, today the country is having a large economic growth and is among several looked upon as an African success story in the happening [USAID MZ], with the last ten year being exceptionable stable, before that a long war for independence, followed by an even longer civil war have shattered the country. The capital Maputo is located in the southern parts of the country along the coast. The climate is tropical to sub tropical.

3.3.1 Mozambique history

The land area now known as Mozambique have been populated by humans (or Homo sapiens) for approximately 100 000 years, approximately 2000 years ago humans with iron tools and weapons (group of human called Bantu) migrated into the area, setting up towns, which grew into trading posts with links to Africa, Middle East and India. Arab influence in these ports grew strong as it has along most of the east coast of Africa.

Around 1500 the Portuguese came to the country to establish supply points and trade. During the 5 centuries of Portuguese colonial rule there were made few social investments in the country, i.e. little educating of local people. And the Portuguese exploitation of the country led to harsh conditions for its inhabitants. After the Second World War the worldwide trend of independence among Western colonies also reached Mozambique, and in 1962 Frelimo (the Mozambique liberation front) was created, on 25 June 1975 after several Guerrilla campaigns the republic of Mozambique was pro-claimed. The Portuguese pulled out fast, leaving the country short of an educated workforce. As an example there were just 80 doctors in the country after independence [Roemer, 1992]. The government turned to the former east bloc for help and started on a socialistic path.

The Cold war was quite warm in these parts of Africa, with South Africa and Rhodesia (today Zimbabwe) being ruled by white minority government fearing Communist influence and independence movements[15]. Mozambique, together with other neighbouring countries provided safe bases for the Guerrilla forces fighting for independence/democracy in South Africa and Rhodesia (Zimbabwe). The response by South Africa and Rhodesia was to start a destabilizing campaign against Mozambique (and other countries); an important tool in this was Renamo. Renamo was a guerrilla force created and funded by external assistance which became active in the late 1970, thus leading Mozambique to civil war. South Africa and Rhodesia also followed up their destabilizing campaign with several cross border raids; as a consequence the Mozambican economy was shattered. When Rhodesia gained democracy and changed name to Zimbabwe the funding and training of Renamo was taken over by South Africa. In 1983 Mozambique was hit by a severe drought and opened more up to the west to gain required relief. In 1989 the ruling party, Frelimo abandoned Marxism, a new constitution was made and the following year a multiparty election where held. Frelimo won it with a strong Renamo on the second place. A UN negotiated peace agreement ended the civil war in 1992. Since this the economy has had a high growth rate and the country have been relative stabile. Though there have been several setbacks from natural disasters, cyclone and flooding, the latest flood in 2000 heavily influenced the HISP group as two of the test districts where hit hard. In addition use of land mines during the war is a legacy affecting people every day, mines combined with flooding have lead to it being even more difficult to neutralize the mines.

[15] Independence/Democracy movements were often seen as Communistic movements in both South Africa and Rhodesia, even if they did not adhere to Communist ideas. In South Africa selected members of ANC were put on trial for being communists, and thereby traitors. This as a consequence of the cold war.

3.3.2 Mozambique organization

Picture 3: Map of Mozambique with all provinces except the Maputo City province.

Mozambique is organized into 11 provinces (including the Maputo city province), as shown on the map to the left. Each province has a capital and is divided into districts.

In general the level of development is best along the coast, and closest to Maputo. The further away from Maputo the poorer the infrastructure becomes. The interior consist primarily of undeveloped land with a poor infrastructure. Most of the population is living in rural areas along the coast, 71 % rural and 29 % urban. However, there is a migration to the urban areas today.

3.3.3 A quick glance at Mozambique today

The last 10 years have been exceptional stable for Mozambique, with the inflation under some control, now running at 10 % a year, and a high economic growth, in 2001 at 9.2 % [CIA] and [WHO MZ]. As such Mozambique is today by several looked upon as an African success story in the happening. However, the last elections were not unproblematic, corruption is still at large [NORAD MZ], and the countries health statistics are not good [WHO MZ].

The population of Mozambique is today estimated at approximately 19.000 000 [INEMZ] with a growth rate of 1.13 %. The illiteracy is estimated at 57 % (% of population above 15 who can't read and write). With 70 % of the population living below the poverty line, a national plan have been made to fight poverty, its goal to reduce the amount of people below the poverty line to 50 % in 2010 [NORAD MZ]. Unemployment is running at 20 % with most of the population employed, or working with agriculture (81 %). Though, most of the country is undeveloped, arable land is 32.000, 4 % of a total 800.000 sq km total land area, in which again irrigated land is 1.000 sq km [CIA]. As of today Mozambique is an accepter of aid from several countries (totally 632.8 Million USD [CIA]), and parts of its foreign debt have been reduced through forgiveness and rescheduling which have brought it to manageable levels.

Mozambique economy is very typical for developing countries, having large exports of raw goods, such as food and cotton. And import of pre processed food, machinery, equipment and textiles. Its foreign trade balance is negative [CIA]. The main trading partner is South Africa, which also have done large investment in the country, such as the construction of a new modern highway from the South African border to Maputo. There are claims that the new economical factors have mostly benefited the southern part of the country, Maputo and the surrounding areas. The last turmoil in Zimbabwe have provided Mozambique with an opportunity to further increase its economic growth, and develop its land by leasing out land to farmers from Zimbabwe who have lost their land (currently it is illegal to own land).

The latest years have seen relative large investment in modern communication in Mozambique, and as a consequence, modern services such as ATM, Internet and mobile phones are available in several parts of the country. The mobile industry is growing fast, as of 2001 there were 100.000 mobile phones in the country, outnumbering the old cable phones of which there were 90.000. Observation done was that in most places the most modern and clean buildings were the petrol station (foreign money) and the shop selling pre paid phone cards. In Maputo Internet is easily available, but having quite slow connections. The Internet industry is not very mature, or advanced, shown by the fact that all traffic was routed to USA, even connections between local Internet Service Providers (ISP). However, during my stay the local ISP were about to connect themselves together to build an internal Internet. The modern services available are mostly out of reach for average Mozambique, due to the high prices compared to the wages. Annual gross domestic product pr. citizen was in 2001 according to [CIA] 900 international dollars (int $), while [WHO MZ] puts it at 697 international dollars (int $) in 2000[16].

3.3.4 Mozambique healthcare history

The leadership of the independence movement Frelimo, consisted of several former health care workers [Roemer, 1992]; this combined with the political

[16] The difference in these estimates are difficult to explain for me, the one year difference (i.e. WHO is having 2000, CIA 2001) can probably not explain it, the difference is too large. Checking both sites I found out that both were using international dollars (Intl $)to calculate the amount, international dollars are dollars which takes into account the different purchasing power between currencies (i.e. price level between countries, more Big Mac's for a USD in South Africa than Norway). Calculating the international dollar rate for developed countries is usually easy, while for developing countries it is harder. So this in combination with the year difference might explain the difference, or they might use two different sources. The lesson learned should be that the numbers which are presented here in this chapter should not be viewed upon as exactly, because of the difficulties in finding good statistics for developing countries. The problems with different static's on different international organizations where also found in relation to other sources, such as between WHO and UN. Also note that since the National statistics bureau in Mozambique is in Portuguese I was unable to use it.

domination played by Frelimo in the post colonial period in Mozambique have lead to that public health care have been deemed as important. And within this prevention had a high priority, emphasizing immunization, latrine construction and general health education [Roemer, 1992]. As a consequence the National immunization programs were started in 1976, succeeded in 1982 by the WHO EPI[17].

During the liberation war public health care was free in the northern regions liberated by Frelimo. And in the new socialistic government the Ministry of health got and held a strong position, almost having total authority of all health care of which almost all were nationalized, including all the medical missions. No other agency of government had any significant health functions, even the medical school at the university were controlled by the Ministry of health, the only exception being traditional healing and the armed forces [Roemer, 1992]. After independence the number of health facilities and their size were greatly expanded, but the civil war had a large impact on the health care sector as health stations and personnel were looted and/or killed/destroyed, at the same time as transportation and communication was disrupted by the guerrilla. At the time of independence there were 426 health facilities, in 1986 this had increased to 1.326, at the same time 595 had been destroyed or looted (thus actual number of facilities established had been 1.921) [Roemer, 1992].

3.3.5 Healthcare status in Mozambique today

Mozambique is a very poor country and has limited money and resources to spend on healthcare, as a consequence a large part of the health care budget are made up of foreign aid. Total health expenditure is 4.3 % of the GDP, which translates to 30 Intl $ pr. citizen[18] [WHO MZ]. This is above the 10 US$ which WHO believes is necessary to deliver a basic primary health care. Life expectancy is 44.8 years [WHO MZ], and a large part of the population is of young age (i.e. 42.5 % is between 0-14 years, 54.7 % between 15-64 years and only 2.8 % is above 65 years). The statistics for Mozambique are in direct contrast to for example Norway. As is showed by the following example looking at death among the population group of females aged less than one year in 2000:

- In Mozambique 16 % died (55.798 deaths out of a population of 347.381 [WHO MZ])
- In Norway 0.3 % died (92 deaths out of a population of 27.915 [WHO])

[17] WHO initiated the Expanded Program on Immunization (EPI) in 1974, its purpose is to improve immunization coverage and decrease morbidity and mortality of vaccine preventable diseases. A variety of approaches are used and it has proven to be highly successful in several countries.
[18] Using the average exchange rate it is only 9 US$ pr. citizen [WHO MZ]. And US $ is needed to buy medicine, gear etc. as Mozambique does not have any medical industry. Both numbers are low compared to Western countries and other African countries.

The infant mortality rate is running at 139 deaths/1.000 live births [CIA], with the maternal mortality rate running at 1100 pr. 100.000 live births [UNDP]. An important cause of this is probably the fact that only 27 % of the deliveries are in a health facility, and in total only 30 % of the deliveries have a skilled attendance [WHO MZ]. This also points to that the public health care has difficulties in reaching out to the entire public, which again can be related to the large problem of HIV/AIDS. It is estimated that 12.6 % to 16.4 % of the adult population in Mozambique (age 15-49) have HIV/AIDS, and 1.5 million people is living with HIV/AIDS in Mozambique in 2001 [CIA] (the HIV/AIDS epidemic also influence the age structure and every part of society, such as large amount of children without parents). An important part in coping with this epidemic is education of the population and supplying of preventions (specifically the risk groups). But this is difficult if the public health efforts are not covering the entire population. Here a proper working health information system might assist with assessing how large parts of the population are covered. Some parts of the health care systems are performing well, such as the immunization programs (EPI). They have proven highly successful in raising the immunization coverage from 5 % to 80 % for several diseases [WHO MZ], thereby preventing around 3 millions deaths annually. However, these (EPI) programs are designed and financed by external donors and are not sustainable, thus the programs will probably function poorly when external funding is withdrawn. Thus it is deemed as important to try and make these successful EPI programs sustainable. It is believed that a proper working sustainable health care information system might allow the programs to become sustainable. Or as [Lippeveld, 2001] is stating that a proper working healthcare information system have the possibility to integrate both the individual and public health interventions together, thereby working a as a glue.

3.4 Summary

I have in this chapter presented HISP, which is a collaborative, cross faculty action research project which is aiming at improving health services to the poor and marginalized. HISP have developed the DHIS software package which aims at supporting local management and health care delivery by enabling flexibility at all levels. The DHIS is an important part in having the HISP aims institutionalised. Currently HISP is working on implementing DHIS in Mozambique and a HISP team is active there, building upon Ministry of health officers, medical and informatics students and experts from South Africa. Civil war, flooding and mismanagement have left Mozambique and the public health care in shatter. However, today the past is put behind and the focus is on rebuilding the country and reducing poverty.

4 Methodology

The settings of this study required me to use several different methods in order to gain the necessary knowledge to conduct my action part. This chapter explains my action research approach selected and the methods used. And why they all were selected.

4.1 Research settings

Collecting of data for this thesis was done abroad for a 3 months period, this meant that I was unable to go back and investigate problems discovered afterwards. And my empirical research was primarily done during the time I was on site. However, afterwards I had means to be in email contact and talk to former research colleagues, but less extensive than on the site study.

4.1.1 Background & motivation

My previous knowledge about health care information systems in developing countries before embarking on this thesis was nil. I had been in Africa twice as a tourist, but never in Mozambique and South Africa. Doing this research in Africa represented for me a unique change to work in a new environment which was a valuable experience, and fun. The main motivation was:

Gain a better understanding of the complexity of change by investigating problems associated with deployment of IT systems and the required change of the organization.

The background for the thesis is the replication of the DHIS system from South Africa and into Mozambique and other countries. My research was conducted within a HISP project in Mozambique; therefore my research is mostly about the Mozambique context. To assist with deployment of DHIS, the new system I will try to make a tool allowing for DHIS database populating, an important part of the migration process requiring more research as stated by [Bisbal, 1999]. My research might also increase the understanding of problems associated with change in large organizations.

4.1.2 Location and period of study

Collecting of data for this master was done in South Africa and Mozambique in the first half of 2002. I have divided the study in three parts. One part consists of preparatory studies in Norway and South Africa, one part consist of the main research being done in Mozambique, and the last part consist of writing up my experience and trying to draw conclusions.

- Pre study: November (2001), January and February 2002 in Oslo
- Pre study: 20 February - 14 April 2002 in Cape Town, South Africa
- Research: 16 April – 22 May 2002 in Maputo, Mozambique
- Post study: Back home in Norway working from 23 May 2002.

In Mozambique my research was conducted in Maputo (Capital) and on two fields' trips to the provinces Gaza and Inhambana lasting a week each.

4.1.3 Aim of study

I had several aims during the study, but the main aim of the research project was to:

Gain a better understanding of the current installed health information systems in Mozambique, and develop a tool to allow the DHIS to be connected to the installed base.

DHIS needs to be connected, aligned to the installed base in Mozambique to ease deployment. I will develop a tool which will, if successful, connect DHIS to data stored in the existing systems, thereby making the deployment easier and increasing the change of success. An important factor with the tool is that it should be general enough to be used in other countries.

An important part of the process is the investigation of the existing systems in Mozambique, in order to understand the data stored and understand parts of the installed base. This research is also important to show a "customer need", by evaluating existing systems and eventually show the weakness of those.

4.1.4 Teamwork

In Mozambique I worked within a HISP team which had similar objectives as I; my practical aim in Mozambique was a sub objective for HISP. The HISP objective was the deployment of DHIS and building a data culture within the health care in Mozambique. The fact that I have been part of a team has influenced my research in several ways which will be dealt with in the respective sections.

4.2 Research approach

Critical elements for my selection of the research approach were:

- Short time for the actual investigation (16 April – 22 may 2002).
- Limited availability of documentation of the existing systems due to factors outside my control

- Limited availability of general information before, during and after the study. Since most of the documents was in Portuguese, and I lacked resources to get them translated.

This meant that I had limited information available before, during and after the study, and limited time on site. In addition, I had no clear problem statement and no hypothesis to test before the research. These factors meant that I needed a flexible approach that would allow me to change focus and research area as I gained a better insight of the problem. My motivation of gaining a better understanding of the complexity of change meant that I should be able to understand factors influencing change. Thus understanding change was important, I selected Action Research. As [O'Brien, 1998] states that action research is an flexible approach in which you learn while doing, thus doing action research trying to change the organization would allow me to gain a better understanding of the factors influencing the change process.

This has influenced my research in several aspects, such as when investigating the existing systems I was looking for errors, and not so much positive elements, I have of course tried to be objective. Another way to do an action research could have been to look at ways to improve the already existing healthcare information systems, and recommend improvements. This would have required a bigger focus on the working parts of today systems.

4.3 Action research

As the name suggests you have two outcomes in action research [Dick, 1993]:

1. Action, bringing about a change
2. Research, increased understanding of the case

My main aim was increased understanding of the complexity of change and writing a thesis, thus the action part was a bi product for me, and wherever a change actually happened is not that important. As a documented, evaluated and analyzed failure is also a contribution. Action research is by most writers viewed as a cyclic process containing at least the elements showed in figure 6 [Dick, 1993]. Other important elements are for example action planning, considering alternative courses of action. The cyclic process enables iteration, accordingly making action research an exploratory approach allowing adjusting of the research as you go along

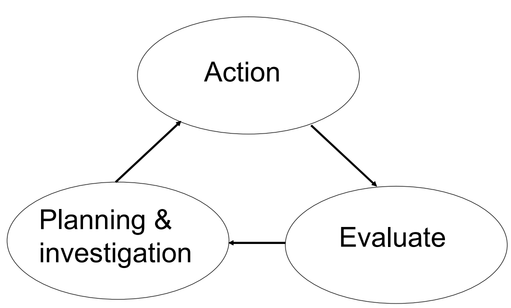


Figure 6: The basic cyclic process of Action research

[Smith, 1993], or as [Dick, 1993] states that in action research you let the data decide the next step. This in contrast to standard research where typically the theory is investigated, data collected and in the end analyzed, or a hypothesis is made, then tested out and evaluated. Action research enables a more flexible approach where you can let data collected (and analyzed) decide which theory is appropriate. The emphasis on change and action implies that responsiveness is an important element, as [Dick, 1993] states that action research without modifying probably is not action research; this is in contrast to the need for a research to be replication able, this and other trade offs will be elaborated on in the last section of this chapter.

4.4 Research cycle

Due to lack of information before the study and the complexity of the reality investigated I had to use a cyclic process in my investigating (see figure 7). In the beginning I had a fuzzy idea of the infrastructure; and I sought to clarify my understanding by interviews or other ways of data gathering, such as analysis of the systems. Data gathered was often analyzed straight away; because that enabled me to gain a better understanding after each cycle. Former findings were continually questioned in the process, some findings were strengthened, other falsified, and the falsification of former findings meant that I often had to renew my thinking, both conclusions and what to do next.

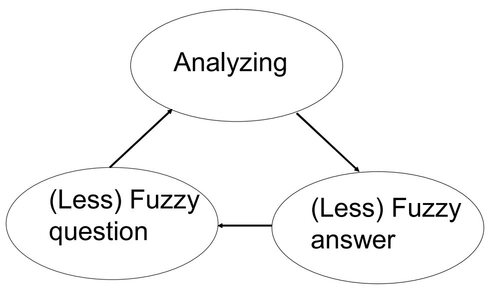


Figure 7: Showing my investigating process of the existing reality.

When I believed I understood the existing systems well I started to implement my change, use my tool. I had to go several cycles while using my tool as I discovered new problems as I went along, thus my research and action part were not strongly separated. During the implementing of the action I also continued my investigation (i.e. interviews etc.). Once home notes, interviews, findings etc. where set up against more theory, which increased my understanding of the reality and helped me to generalize.

The limited time on site also had several consequences for my research. Firstly I only had the change to try and implement a change once, thus if my tool were not working. I would not have time to develop another tool, note here that the tool might function perfectly, but not working in that it is not supporting importing of data. To compensate for this, I will use experience gained by other HISP team members working with the same aim as myself. Secondly evaluating eventually change was difficult Due to the fact that the changes I was looking for takes a long time, organizational change. To compensate for this I am using communication with former HISP colleagues while writing the thesis. And thirdly, having a tool ready for the case in Mozambique, which was to be general enough to be used in other countries, required that I had to almost finish the development before Mozambique. This means that the tool was developed according to my ideas of a general situation, and might not fit the specific Mozambique situation. I might also be biased towards using my tool in the action phase, i.e. don't want to throw away all invested time and resources.

4.5 Methods

[O'Brien, 1998] states that action research is a holistic approach allowing several different tools and methods to be used. Therefore my selection of methods was dependent upon the settings, I choose primarily qualitative methods as they usually are more flexible than quantitative methods [Dick, 1993]. Secondly, my research consisted of trying to understand a naturally occurring phenomenon in the real world, and for this qualitative method is better than quantitative methods [Smith, 1991], as the richness of the data is greater.

However, quantitative methods would in several situations have been appreciated, for example when investigating how the different users used the existing systems. I might have developed questionnaires, or conducted systematic observations, this would have enabled me to compare the different users and parts of the organization against each other, and the data gained might be less biased. But due to the limited time I was unable to do this (for example questionnaires would have been made, translated, dispersed and then collected). However, my analyzing of the health data collected in the database might be clarified as a quantitative method.

My methods of information gathering were:

1. Open interviews
2. Random and planned observations
3. Study of systems and information stored
4. Study of documentation

4.5.1 Open group interviews

I conducted group interviews because there were always several other HISP team members present and part of the interview, all on equal terms. Furthermore, I often needed help with translating and did not have a translator, so the students in the HISP acted as translators.

The objective and aim in most interviews were to find out more about the infrastructure, with an emphasis on the computer systems for me, and the use of it. Follow up questions, and follow leads gained in the interviews were important during the interviews, due to the fact that follow up interviews with the interviewee was sometimes impossible due to several reasons. This made it necessary to get the most out of each interview. As a consequence the interviews were flexible instead of rigid, no pre determined rigid interview pattern. Most often we agreed only on the directions and general elements to investigate before starting the interviews. Other reasons for not having rigid interviews pattern were the other researcher's participation. A rigid interview pattern would most likely not fit well with their interests, and we did not have the time available to sit down in cooperation and make rigid interviews pattern. I believe the following citation from [Smith, 1991, page 74] shows my reasons for selecting open interviews: "*One aim of the interview is to develop an understanding of the respondent's world so that the researcher might influence it, either independently or collaboratively as might be the case with action research*"

Problems with open interviews are that they are very subjective, for both the interviewer and interviewed. It is difficult not to be coloured by ones own views, assumptions and prejudices. One advantage is that you get lot of relevant information, but it is qualitative, thus it is difficult to compare the different interviews to each other, as the interviews are very different. There is also the danger of misunderstanding and misinterpretations. Particularly I felt that was the case when the interview was conducted in English, as it is a secondary language for all of us. For more errors, see the section in this chapter specifically dealing with errors sources.

A typically interview would have the following setting:

- Interviewee: 1 user of the system of interest, note interviewed both entry clerks, administrators and decisions makers.
- Interviewers: 3-5 students including me, sometimes also one Professor
- Location: Typically the office where the interviewed worked
- Duration: 20 – 90 minutes
- Language: Portuguese and/or English

The numbers of interview conducted were 11, including follow up interviews.

4.5.2 Observations

My observations were divided in two groups, planned and not planned. Not planned primarily came about due to waiting in different offices, thus I was able to see how people actually worked and noted interesting observations, but the observations was not systematic. I did not look for anything particular. For instance once I waited I saw how the entry clerk entered data from the paper forms into the computer. This allowed me to later understand why some of the data stored in the computer systems looked so strange, error in entering process. Though, I only saw this once, but that observation combined with analyzing stored data allowed me to generalize. The fact that I was waiting might have led to that the person felt safe to do as he/she regularly did, thus I was not perceived as a threat of any kind, in contrast to the informed observation where the observed can be heavily influenced by the fact that they are being observed. This rises the ethically question of observation without informing the subject. In my case I do not see any ethically problems, as most of the observations were random and quite impossible not to observe. Further, it was part of the research and all observed people knew why I was there. Furthermore, I never went anywhere to solely observe and camouflage my presence.

Planned observations were such as looking at walls and taking pictures of analyzed diagrams and graphs, asking for permission; this gave a clue to how information was used in the organization at the different levels we visited. Note that I am not using systematically observation; therefore none of the observations are quantitative comparable.

4.5.3 Study of systems and information stored

My primary aim was to see how the information flowed trough the different systems. One important aspect was to check data, information, quality and how the information was used, then compare this knowledge with information gained in interviews. Included in here is

- Technical/statistical analysis of data stored to check for reliability and validity, one process typically was searching for obviously wrong data such as more children signed out than born. DHIS, SQL and Excel were used as my analyzing tools
- How and what information was used in yearly publications. I typically looked at the printouts from ministry of health and checked to see if I could trace the numbers back to the computer systems
- How data collection forms were used. I checked the data forms entered at the data production unit

Basically I was analyzing the information flow from paper forms in the health facilities, up to the computerization in the provinces and the national level, and then the paper reports from the ministry of health. I was never creating or modifying any data my-self, simply checking already existing data. In order to verify and understand my findings I was comparing the results from the open group interviews and observation with the quantitative data found during the information analyzed.

4.5.4 Study of documentation

I hardly used any documentation in Mozambique, as most was paper based and in Portuguese. However, the fellow Mozambican researchers explained the necessary Portuguese documentation.

4.6 Tool development process

As part of the learning experience of the HISP project and my thesis was the development of a tool to use for DHIS database population process. The tool would if successful, facilitate the change process (as described in the theory section and other). In addition to that tool, one other HISP team member made his own solution, or tool.

4.6.1 Database population

Database population is one part of the legacy replacement process as described by [Bisbal, 1999]. Database population, usually presents a wealth of problems, especially if the database is complex with triggers referential integrity etc. My project involved taking data out from a dbase ver. III, spreadsheets and text files which have no support for referential integrity and other advanced functionality, into a more modern database environment, MS Access. Important was the mapping of the data elements into the corresponding names in DHIS, and converting the data format so that DHIS could import it with its import functionality, basically making a text file, described later on. Another important part was selection of which data to import; both which source and which data in that source. This meant that I had to try to find an independent dataset that could be imported. In addition there might be a need to do some data cleaning if the data are of poor quality, as stated by [Bisbal, 1999]. My tool might be used as a gateway, though it is lacking several important functionalities that should be in place to allow it to be used as a gateway, such as transaction management and consistency check.

4.6.2 DHIS import tool

The tool should connect to a data source consisting of tables and columns (can be database, spreadsheet, text file etc.). After that the user selects which data he wants by selecting columns of data in the data source. The primarily task of the tool is to get monthly reported data, this data is identified by:

- WHAT: A data element being reported (example: Incidents of Malaria)
- WHERE: The unit reporting, (example: Eduardo health centre)
- WHEN: When the reporting is for, typically a month (example 2001 January)

This is the minimum of data that must be available for each element for it to be imported into DHIS.

DHIS store more data for each element than this, such as last user editing, date and time for last change. This information might not be available in the old data source; therefore the user has the ability to insert a default values for any values in the DHIS import file. In addition data selected for import must be mapped into the DHIS data dictionary, mapping consists of renaming the values and data elements from the old data source to fit the data in DHIS. Example: in the old system Malaria was coded as MAL, while in DHIS Malaria is stored as Malaria, then the mapping would be MAL → Malaria. In addition to the data dictionary mapping, other values might also need to be mapped, for example different representation of date. After the mapping a DHIS import file can be made. Then DHIS can import this file and give eventually error reports to the users. Then the user must manually correct the errors.

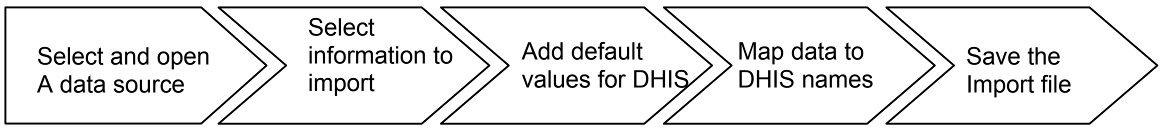


Figure 8: The flow in the DHIS import tool

For more about the tool see later chapters.

4.7 Evaluating the change

Alternative actions will be investigated by using experience from the other approaches used by other HISP team members. In addition to my research there were two projects with the same aim as my action part; one during my stay which consisted of another student making a specific ad hoc tool to solve the same problem in Mozambique. This tool was developed during the fieldwork. And a second project started and completed afterwards, consisting of a HISP full time member, i.e. not a student doing a master, who imported data from the existing systems. He did not make any tools, but simply used his inside knowledge about DHIS and the source systems databases to manually import the data into DHIS, i.e. used SQL quires and other finished non specific tools.

4.8 Trade offs

In this research specific local knowledge have been emphasized in order to understand the specific reality and bring about change, thereby sacrificing global knowledge and generalization, as stated by [Dick, 1993] is a standard criticism of action research. To compensate for this I will try to generalize and point out eventual discrepancies between my findings and the theory. My need for responsiveness means that replication is difficult to achieve [Dick, 1993], even if DHIS is not deployed the organization might have taken steps to correct the errors found, as the HISP team delivered a rapport on the errors we found in the existing systems. I have no control over the eventual change following my intention. However, most findings should be replicable, such as the analyzing of the existing systems, as long as no one alter the historically data stored.

My participation in the HISP team implies that I have been biased towards DHIS and the HISP team's ideas (the HISP team ideas is also biased). Furthermore, it has also leaded me to be biased in my theory selection, some of it provided by them, and research. The fact that the HISP team and I are participation strongly in the project my also influence the research, such as problems with informed observation [Smith, 1993].

4.9 Error sources

I will now explain eventually error sources that might have influenced the research. First of all in some of the interviews we, the researchers were a large group, 3-5 students, coming from the university. As a consequence the interviewee might have felt a bit frightened by us. Thus not always telling the truth, but what he/she believed we wanted to hear.

I specifically remember one time when we, the interviewers used group pressure on the interviewed to say something:

- Thesis: Information was transferred quarterly (4 times a year) on floppies from the provinces to the national level (Ministry). This was based on interview of staff at the Ministry of health, behaviour inscribed into the software and what I had been told before I went to Mozambique.

- During an interview at province level with the responsible person for information, we were told that print outs were sent quarterly, but floppies where sent at random and only when someone called and asked for it. We did not believe this, so we asked several following up questions, leading to that she changed her statement to what we originally wanted to hear (our thesis). But analyzing of the information stored at province level vs. national showed that she had been right and the Ministry wrong, follow-up questions showed that most people at the Ministry was aware of this fact!

This was a case that was easy to detect. But it shows how easy it is to influence the interviewee; it also shows how I used the combination of interviews and system analyzing in reaching and verifying conclusions.

I was not always able to meet the people I needed/wanted to meet. Also sometimes the time for the interview was a bit short, and thirdly some follow-ups were not possible due to short time and person not available. As a consequence of this I might have lost some critical information. In addition to this, there was also some language problems as I sometimes did not get information directly from the source, but as a summary from a translator, i.e. not directly translated. Note the translator was also a researcher. This combined with that lots of data was coded, and in addition there were obviously some special and undocumented rules that were being used. Therefore my view of the data might have been damaged by some undocumented rule I was not aware of.

Information gained from observation was not planned, and could not be compared quantifiably. Lots of my observations are therefore specific for the places I have been, but I generalize based on findings in analyzing data from other provinces.

Getting the old systems to run on another computer by simple copying the installed files was troublesome due to several conflicts between the copied installed files and the new computer, such as paths. Note this is also a problem with new software as well; it is difficult to copy finished installed software to new computers, and then run it due to missing systems files etc. on the new system. This combined with the language problem might have influenced my findings and the research into the functionality of the existing systems. However, this process did not alter actual data in the systems, as data was viewed directly in the back end.

To compensate and eventually eliminate these errors sources I narrowed my focus during the research. In the beginning I was looking at three different information systems. In the end I only did a meticulousness investigation and imported data from one system. The rest of the systems were not investigated with similar meticulousness. In addition, I used and compared data from different levels and systems with my findings, this to evaluate and verify findings and conclusions. I believe that these actions have to a high degree removed the changes of serious errors.

5 What I did and my contributions

In this chapter I will firstly give a deeper look into some problems addressed during this research. Then I will describe my research approach in concrete terms: where I went, when, what I did and whom I meet. In the end I will describe the expected contributions of this thesis.

5.1 Central problems addressed during the research

The "replication" of a health information system; or any other system from one country to another is a highly complex procedure consisting of several problems. Within this broader context my focus is on the transfer of data from existing systems to the new, or target system database population, which [Bisbal, 1999] states are an area that has not yet to been extensively researched. In my case, and probably more generally, inclusion or compatibility with historical data is a requirement when shifting to a new database system. This leads to a number of problems related to linking or mapping the old data to the new system. The interconnectivity of various elements, such as the systems, the data stored and the organization require that I must fully understand the existing systems and their use, understanding both from a technical viewpoint and a social; or application area viewpoint. Full examination of the existing systems would have included general business knowledge and procedures. However, I did not try to move this information, because that would have required a full migration of all business knowledge and as such required a different approach, such as reverse engineering [Sommerville, 1995]. This again would have required a substantial deeper investigation of the existing systems, including the design and code, and the development of an almost totally new system. Organizational change is an objective for the larger project of which the database population is part. Thus changing much of the old business procedures is one important part of the entire project. It is therefore not required to transfer and translate the old and outdated business procedures. My focus has therefore been on populating a finished system, DHIS and on the required sub problems, such as the necessary adaptation of DHIS to the Mozambican context in order to allow for data to be imported.

My task of importing the data requires that I first get the data to import. I can-not expect that the existing systems posse's data export functionality, thus actually finding, reading and understanding the data stored is the first step.

1. Finding the data: where is the system storing its data? This can further be broken down into several layers. Where in the country? Is the data replicated or distributed? Where on the computer and on what computer?

2. Reading the data: Need to be able to read the data once found, what is the back end, or data file? Is it proprietary and closed or open? Custom made or based on third party storage systems, such as database vendors?
3. Understanding the data: What does the data/values measure? What does a coded column name such as “D1” mean?

These sub-problems must be tackled for all the existing systems, which I chose to import data from. The next step is to decide which data to import, which again is dependent upon several sub-problems.

1. Actual selection of data: Import what data from which system at which level in and in which order? This will require full understanding of the existing systems and the requirements from the new users of DHIS. For instance, should only active data be imported? Order of import introduces the problem of prerequisites before data can be imported? Or upholding eventual referential integrity within the data.
2. Data cleaning: Should poor and obvious wrong information be imported as it is or corrected if possible? How to deal with eventually inconsistency due to redundancy data storage?

Finding and implementing the prerequisites before data can be imported is a part of the adaptation phase for DHIS. The total adaptation phase of DHIS is a resource and time consuming task, which is too large and complex for this thesis. I will therefore only work on the necessary adaptation for population to take place, focusing on technical difficulties. Other and more important parts such as training of the organization will not be extensively looked into. An important issue in the adaptation phase is the fragmented nature of today’s systems, as described in the theory and the next chapter “Case description”. This indicates that the organizational procedures might be very varied and fragmented. Therefore the flexibility of DHIS will be an important element in the adaptation. This as the introduction of DHIS as the single system to replace all the systems, and alter the organizational procedures is a move towards greater standardization, and possible less flexibility for the different programs as they will become more entangled into each other. However, flexibility can be viewed from several aspects. And it is believed that DHIS with trained staff and a support organization constitutes a more flexible system than the old ones.

5.2 What I did and where I went

I will now briefly give a time line of my study, describing what I did, where I went and whom I met. Note, all dates are 2002 unless otherwise specified.

5.2.1 Norway

In Norway I started to prepare for the fieldwork by reading relevant literature and learn more about the necessary technical parts. To learn more about DHIS I participated in a Norwegian team consisting of students working on a Norwegian translation of DHIS. We worked on the multi language edition so we only had to translate the Strings, i.e. we did not alter any code.

In Norway the work on the DHIS import tool was started, including learning more about Java and the graphical user interface functionality (Swing). A first basic version was made which allowed for import from structured text files. Other requirements were difficult to decide due to the lack of information, as only some data files from Mozambique were available. And these files required additional information to be fully understood. For instance, some of the files were parts of a back-end database, and I did not now which type of database (i.e. IBM DB, Oracle, SYBASE etc.). In addition the data was coded and I did not have the necessary information to decode it.

5.2.2 South Africa

On the 20 February 2002 I left for Cape Town, South Africa. In South Africa the HISP software coordinator, Mr. Calle Hedberg, provided me with a place to stay. In addition he gave me several advices regarding the needed flexibility of the tool. As an observer on several meetings I was able to learn about the HISP and DHIS status in South Africa today. However, my work in South Africa was mostly technical and related to the development of the tool.

For a week I was working within Soft craft, the South African company that is developing DHIS. There I looked at what they had made in order to decide if I could use any of it. Some tools were found which had some of the functionality, which I needed. But I was unable to use it because it was coded in Visual Basic and my knowledge about Visual Basic was poor. I decided that it was to risky and time consuming to learn Visual Basic and then translate the necessary code to Java. Instead I worked with an engineer in Soft Craft for some days that introduced me to ODBC drivers, Open DataBase Connectivity. This application-programming interface, API, for database access provides a standard interface to several databases. This combined with the text parser made in Norway meant that I had the general technical knowledge to start work on a finished version. I could use JDBC.ODBC (JDBC.ODBC is the Java version of ODBC) to connect to the different databases, and my text version made in Norway to connect to the necessary text/Excel files.

5.2.3 Mozambique

On the 14 April 2002 I entered the train in Cape Town to travel to Maputo, Mozambique, where I arrived on 16 April. In Mozambique most of my work was done as a member of the HISP team. I was situated in Maputo were most of my work was done. However, I had two fieldtrips of one week each to the provinces which proved invaluable in understanding the existing installed base, both technically and socially. My fieldtrips had an emphasis on the provincial levels, but I also visited districts and health facilities to get an overview and confirm findings at the higher levels, and to satisfy my curiosity for how it was at the bottom levels.

In Mozambique a taskforce was formed to partially work on the population of DHIS with data from the existing systems, consisting of the original HISP Mozambican team and several master students in informatics. In the beginning two professors from the University of Oslo were present to guide and support the research. Added to this were several medical students on the last field trip who looked at the current usage and quality of the health data in the district and facilities. My technical focus meant that I primarily met with key people who dealt with the information systems, i.e. few clinicians. I primarily worked with the informatics students, which were present all along the research.

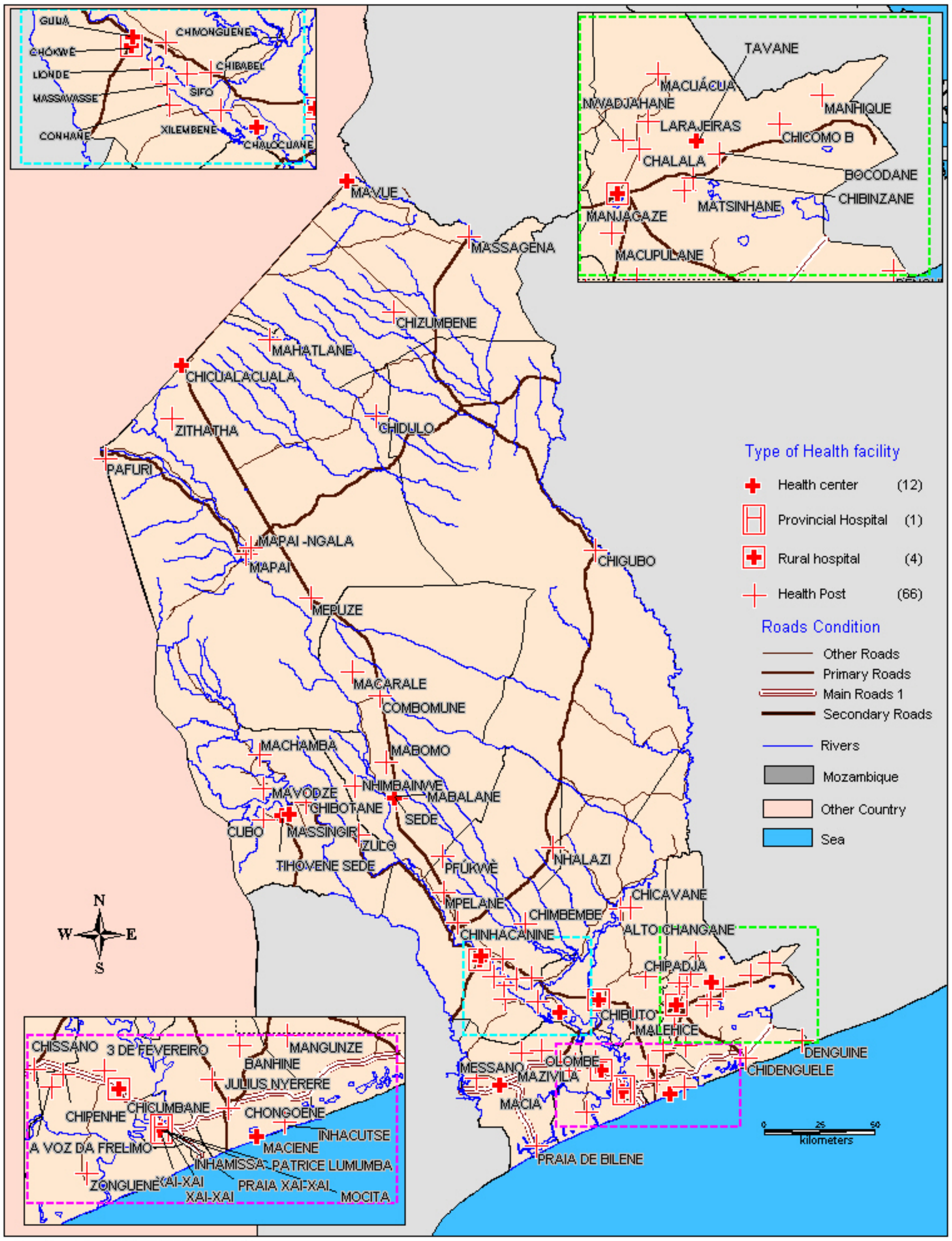
Maputo

In Maputo I was working at the national level, and not the Maputo province level. In Maputo most of the technical decoding of the existing systems was done, as most the technical knowledge was located there. Essential was the proximity to the Ministry of health and access to administrators and general information managers, information key people. The Ministry was very helpful as they were interested in the research as well, we were therefore allowed to investigate theirs computer systems and interview several key actors. The national SISC program, including its database was transferred to my computer; the SISC is the routine health information system being used today in Mozambique. The investigation of SISC started with an emphasis on decoding the back end; the following pattern of investigation was followed:

1. Observations and investigations into the back end database
2. Based on intuition and knowledge make hypothesis about the system
3. Used open interviews with SISC related staff to determine the validity of the hypothesis.

This procedure was repeated several times as more information was obtained in each round. First, I had just received some random tables from the national level; this was used to decode some of the tables in the database. Obtaining the entire national database enabled me to discover even more. Later, I also went to a province to get their provincial database, but to really manage to decode the data I needed the original forms from the districts, thus I had to leave the computerized parts to be able to fully understand the backend. This was the strategy chosen because obtaining the documentation and technical knowledge about the SISC software proved difficult due to several factors.

I also completed my tool, and added new functionally based on actual import of SISC data. In addition I participated in HISP lectures and HISP team meetings as a member. At the same time as another HISP student started on a specific solution for automatically data import from SISC to DHIS.



Produced by EHP/USAID (Aug 2000)

Picture 4: Map of Gaza with the health units. Note the density of units along the coast as the highest population density is there.

Gaza, Xai-Xai district 21-26 April 2002

On 22 April I visited the Gaza province as part of the HISP taskforce. The Gaza province has approximately 1 000.000 inhabitants and contains 12 districts. Our base was set up in the Xai-Xai district where the Gaza provincial headquarters is situated (15 km away). The district Xai-Xai have a population of approximately 200 000 and contains 12 health units, note the City Xai-Xai is not included in these figures. The reasons for selecting Xai-Xai was its proximity to Maputo, 220 km, and its reasonable infrastructure concerning telephone and computers, though the power cuts were far more frequent than in Maputo. In addition Xai-Xai is situated along the main road of the country and little off road driving was needed in order to get to the other districts capitals. These factors meant that it would be easy to go back to do follow-ups. Xai-Xai was severely hit during the flooding in 2000, but looked to be back to normal now. However, concrete skeletons of old hotels and promenades along the beach reminded me about the country's turbulent past, and Xai-Xai history as a tourist destination.

My main emphasis were on the provincial headquarter in Gaza. Which I visited several times conducting interviews with information workers and transferring most of their data systems, including the provincial version of SISC, Spreadsheets for health facilities and hospital data, and the provincial BES system, BES is a system for reporting of diseases. A lot of work went into finding a proper list of health facilities as each system had their own list. An HISP team member and the SISC responsible person concluded this after two days work. As was the case at national level, the people at lower levels proved very helpful, though, one was too busy to be fully interviewed.

In addition to the provincial headquarters at Xai-Xai I also visited the Delana-Misina and Julius Nyherere facility and some other facilities. The health post I visited did not have any electrical power and were using gas coolers for storing of vaccination etc. Interesting was the huge amount of mothers with their babies waiting for immunization. With this workload it is easy to see why they do not use much time on filling out different forms for the top levels. Though, even at the facility they were using paper and pencil to make graphs of various types. On the 26 April we went back to Maputo.

Gaza and Inhambane 5-11 May 2002

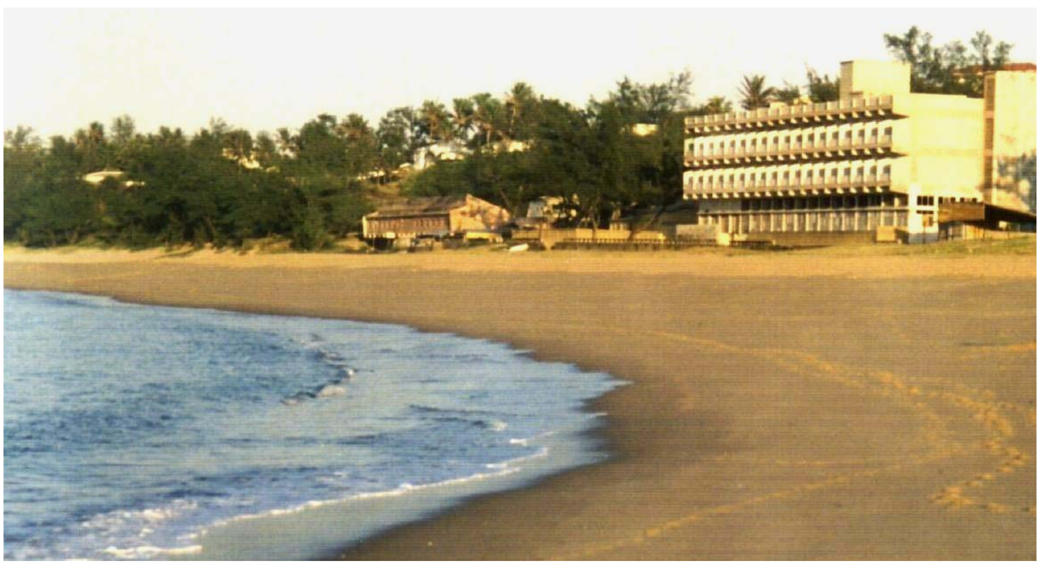
On 5 May I went back to Gaza province staying in Xai-Xai again, this time I was part of a larger group that consisted of the HISP taskforce and several other medical students. The plan this time was to finalize the import of data and manage to use DHIS with imported SISC data. This would enable the medical students to use DHIS with real data from Mozambique. Furthermore, it would enable better training of the local staff with DHIS as they would work on their own data. This time the HISP taskforce was split up to allow the medical and informatics students to cooperate. I was sent to Macia districts headquarter without a computer, which limited my work. I was allowed to investigate and copy the original paper forms being used at the district and sent to the province, this allowed me to compare data stored at district level on paper vs. data stored at provincial level on the computers. Which gave a better understanding of the data stored, as I later in my findings conclude with that some data element are ambiguous and that there are some discrepancies between the district and province level. On 9 May import and configuration of DHIS with SISC data for the Gaza province for 2000, 2001 and parts of 2002 was completed.

Picture 5: Entrance to the Gaza provincial administrative centre, photographer: Jumo Lungo

Picture 6: Parts of Chicumbane health centre, photographer author

Picture 7: Juluis Nyhere health facility, photographer author

Picture 8: Chicumbane Rural Hospital, photographer: Jumo Lungo



Picture 9: Xai-Xai beach, photographer author

Picture 10: Maputo city seen from HISP apartment, photographer author

The next day, 10 May, I and two other informatics students travelled to the Inhambane province farther north visiting the provincial headquarters in Inhambane City. Our objective was to get and import their provincial SISC database into DHIS. But once we arrived we learnt that the computer running SISC had been sent back to the capital (Maputo) for repair and the time of return was unknown. We then started to work on the health facility list for Inhambane and imported some data from the national SISC database to enable us to show DHIS to the key information people. On 11 May we went back to Maputo.

Maputo

The finishing work in Maputo consisted of writing a report for HISP and the Ministry of Health about our findings, on both the computer and paper based system. And visits to the Ministry of health to show the results of our work, DHIS populated with SISC data. There was also discussion on the next step for HISP. At the same time my research was continued by investigating of other existing information systems at national level, though, I did not manage to import data from other systems due to the required pre processing combined with my limited time left. On the 22 May I left for Norway.

5.2.4 Norway

In Norway notes were reviewed and writing on the thesis was begun. Contact with the HISP network was upheld (both internationally and HISP Mozambique). I continued the DHIS population process by transferring some SISC data for the Niassa province in Mozambique, which meant that in total SISC data for Gaza, Inhambane and Niassa had been imported into DHIS. In addition to this my tool was extended to cater for another file format (EPI file format, which will be explained later). And a user manual was made and posted on the web. The tool and user manual was on request sent to HISP team members in India and Cuba.

5.3 My contributions

My contributions have been in two main areas. For HISP I have done action research into problems associated with implementing DHIS in a new country context. While at the same time my experience can enable better understanding of processes of change and possible approaches.

5.3.1 HISP contributions

For HISP I participated in research and made a tool to import data, which might be used both in Mozambique and elsewhere. Thus I am separating my HISP contributions in two parts

HISP Mozambique

My contribution for HISP Mozambique has been along the following:

- Enable DHIS and the old systems to be evaluated
- Allow for stepwise and easier implementing
- Migrate and preserving of historical data

Enable DHIS and the old systems to be evaluated
DHIS had been trying to gain a foothold in Mozambique for several years. But the Ministry of health has said: "You need to show us what you can do!" The Ministry is not able to evaluate DHIS. The problem with DHIS is that in order to show the system you need it populated with data, and as such used for a prolonged time, thereby having it implemented. Alternatively, and as have been done, is to implement it in a province combined with training of the staff there. Problems with this approach are that it requires two systems to be supported in parallel, with the extra amount of work this requires. Furthermore, as data reporting is looked upon as something done to only support the top level, it is difficult to get the lower levels to use a system the top level does not have any interest in. In addition, the training of the staff for collecting and use of data for their own purposes is a huge and time-consuming task. Thus, the interconnectivity of the installed base means that isolated implementing of DHIS is very difficult [Braa, 2002]. Populating of DHIS might enable HISP to quickly show what DHIS can do without investing large amounts of time and resources, thus maybe breaking the deadlock by connecting the new to the old. And more importantly, allow before obscure data in the old systems to be analyzed, and as such show how today systems are performing and possible show the need to change, or a "customer need" for DHIS. In addition this research might also give the Ministry of health an indication of how difficult it is to make vertical integration solution on top of today's systems.

Allow for stepwise and easier implementing
When implementing and starting to use DHIS it would have an empty database in the beginning, as a consequence searching for trends and other functionality would not be available from day one. This might imply that the old systems should be used in parallel in a transformation phase. Import of data can remove this gap and allow DHIS to be used with its full potential from the start.

Secondly, my tool can allow for an evolutionary (stepwise) deployment of DHIS. For examples in Mozambique there are several systems to replace, replacing all of them, a revolutionary approach, would most likely fail due to complexity and the large amount of changes needed. Large amounts of training would be needed and a strong central authority to enforce the change, which the Ministry of health might not have. By connecting DHIS to the existing systems an evolutionary approach might be enabled, replacing one system with DHIS at the same time as data is being imported from the other systems that are still operational, thus enabling integration of all information in DHIS with small disruption of the installed base. As a system is successfully replaced HISP can move on to replace the next system with DHIS, thus spreading the change over a prolonged time thereby increasing the chance of success [Hanseth, 1998] [Heeks, 1999].

These factors imply that a phase when DHIS and the other systems are used is going to be highly complex and probably will lead to inconsistency between DHIS and old systems. This consistency problem is also referred to in [Bisbal, 1999] as one of the reasons making a migration approach difficult. So in the end the change has to happen; my tool is a temporary tool to assist in the transformation phase by enabling some backward compatibility but not full. It can not be used for ever due to its shortcomings, in contrast to gateways which typically becomes a part of the infrastructure [Hanseth, 1998].

Migrate and preserve historical data
The installed base in Mozambique has invested lots of resources and time in the information stored today, and this information would probably become inaccessible when the old systems are decommissioned and the knowledge about them forgotten. By migration the information to DHIS, some of it will at least be preserved, or saved a little longer. And it might also lessen the resistance toward the change, as they do not feel like losing so much invested work.

HISP Internationally
Mozambique was the first country HISP moved to, as a consequence document fieldwork and tools developed to assist in Mozambique will most likely be important in other countries with similar HISP projects. The international contributions for HISP are along the following elements:

- Development of a general DHIS population tool by doing a case experiment in Mozambique and South Africa
- Adoption of DHIS to new country contexts

The problem with replacing several existing information systems tackled in this thesis is likely to be common within most of HISP countries. [Heeks, 1999] is stating that some developing countries deployed computers in the management of health care in the mid 50, Mozambique was late with 1992. The international aspect influenced my development in that I could not make many specific solutions for Mozambique; this combined with the short development time limited the tool functionality. In order to evaluate the difference between the general and the specific approach, my findings will be compared against findings from a fellow student who made a solution specific to the Mozambican context.

In addition I have been working on the adaptation of DHIS to the Mozambique context, an important element in this was the flexibility of DHIS, and could DHIS function with today’s business rules in Mozambique? This has been briefly evaluated, and I have pointed out conflicts found and the solutions we used.

5.3.2 Systems development contributions

My main systems development contribution has been along replacing legacy systems within an information infrastructure, and problem associated with the change of the larger information infrastructure. Within this context, I have selected to do a case experiment on transferring data, or migrating data from the legacy systems to the new system as part of an aligning / adaptation phase to the installed base. And problems associated; such as the necessary decoding of legacy systems back ends. The fragmented nature of the existing installed base implies that I have also been able to look at several different systems to see how much data can be transferred (vs. necessary work). The migration and decoding of data is also relevant for data preservation. I have also been able to look at problems associated with legacy system in developing countries and consequences of lack of sustainability. In addition I have looked into other rewards from such a project, primarily showing the capabilities of the new system vs. the old.

5.4 Summary

Decoding and importing data from old legacy systems require that the old systems must be fully understood. In order to do this I travelled to national level, provincial level and the facility level in Mozambique. In addition I stayed a while in South Africa in order to learn from the South African HISP experience. Expected contributions from the research are more experience and tools for HISP, both Mozambique and International, in addition my work might facilitate the implementing of DHIS in Mozambique, showing a customer need and making the actual organizational change easier. Findings in relation to legacy systems theory within an information infrastructure are also expected.

6 Situational analysis, a case study

In this chapter I present general findings from selected theory, and HISP research about the existing information systems and information flow. It purposes is to show the reader a simple picture of the information infrastructure and systems in Mozambique today.

6.1 Organization of the healthcare

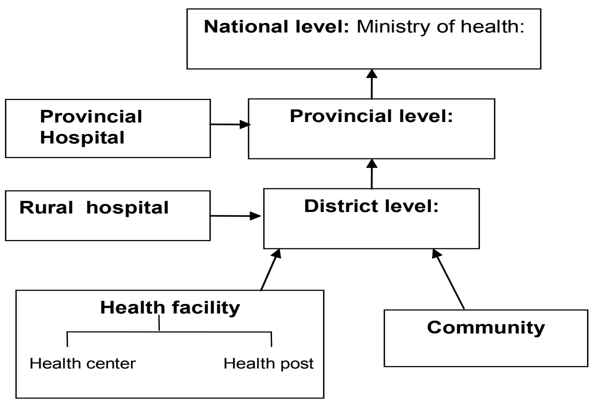


Figure 9: Showing the organization of health care in Mozambique.

The healthcare in Mozambique is organized in a typical pyramid manner. Patients are to enter the public health care network at the most peripheral point, i.e. the facilities, and then be referred to the more technically developed units as necessary, such as hospitals and health centres [Romer, 1992]. As shown in figure 10 units are reporting data up to the next level in the hierarchy. The 11 provinces directly underneath the Ministry of Health each have a centre that contains an administration & management department, together with other departments such as, Pharmacy and Community Health [2Braa, 2001]. In addition, the provincial centre has a hospital and other miscellaneous medical units attached, such as a medicine depot. The province is viewed as the information hub in Mozambique; as such the administration and management of the different programs and the healthcare are mostly conducted in the provincial headquarters [Braa, 2002]. The provinces are divided into districts of which there are a total 131 [2Braa, 2001]. Each district has an information centre consisting of one to three staffs responsible for data collection, analysis and transmission to the higher levels. Some districts also have a rural hospital. Below the districts are the health facilities, each district usually has 5-15 health facilities [2Braa, 2001]. However, actually finding the total number of health facilities in Mozambique is difficult due to the poor state of the information systems, as my analyzing will show. In the health facilities most of the actual health work is conducted. The health facilities are separated into health posts and health centres, in which the latest are larger and have more staff and service available to the public. One of the health centres in each district serves as the district centre. Often the health facilities, particularly the posts, are staffed by poorly trained personnel and have a shortage of drugs, this due to the economical situation and lack of training. As a consequence people sometimes bypass the peripheral point and go directly to the centres, which then becomes congested [Romer, 1992].

The facility staffs are collecting data on a day-to-day basis on several program specific forms, and then on regular intervals, but different depending upon the program, are sending the data to the district centre. The district then collects and reorganizes the data and sends it further on to the province, which again reorganizes data and sends it to national level [2Braa, 2001]. As the figure show both the district and the provincial level gets data directly from attached units, typically hospitals.

6.2 Health information systems and computer use

The current health information system called SIS, dates back to 1982, in 1989 it was revised due to constrains related to lack of defined objectives, complexity of the forms, duplication of data and data collected with no function. The revision simplified the system and made a minimum data set to avoid duplication. These changes lead to that the number of forms where reduced from 60 to12, in addition several indicators for use at district and facility levels were added to the collection forms [2Braa, 2001]. This show that local use of information has been an intention in the design of the system, but this has not ensued [2Braa, 2001].

During the years several program specific systems have been added, as a result today's information infrastructure is fragmented, consisting of several vertical health programs emphasizing the information needs of the central / high levels, and not enabling local analysis of data [2Braa, 2001]. Some consequences of the focus on solely the central level information needs are:

- Data is extensively aggregated. [Braa, 2001] is describing how institutional deliveries reported are a function of reporting health facilities i.e. January 10 deliveries reported and 1 reporting health facility, while April had 40 deliveries and 5 reporting health facilities. The aggregation means that actual reporting facilities are lost.
- There is a large gap in terms of needed information locally and actual information collected, solved some places by locally designed forms [2Braa, 2001].

Figure 10 shows most of the programs currently active, note that the figure shows the programs and not necessarily the computer systems. All most all the programs are part of the SIS, or the National health information system, except for Tuberculosis (TB) and finance.

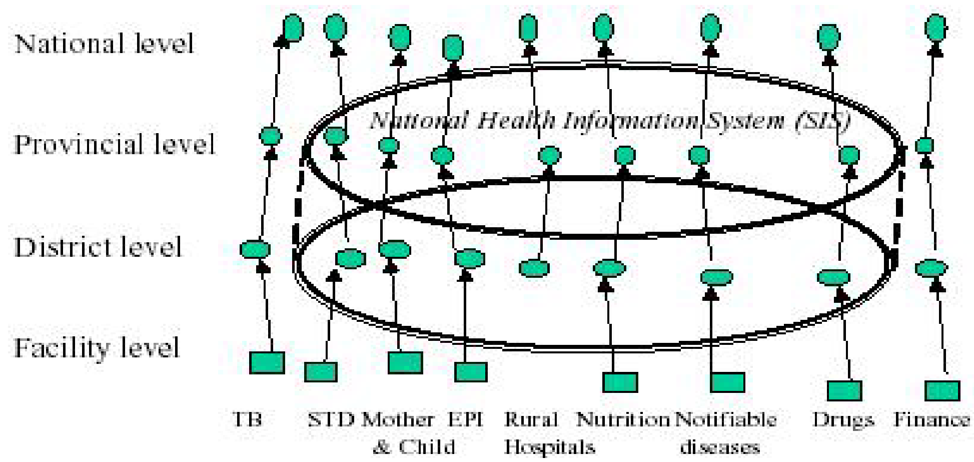


Figure 10: Flow of data on the vertical health programs, note there is no vertical communication between the programmes. Figure from [2Braa, 2001].

6.2.1 Existing computer information systems

The national and provincial level is currently the only levels fully computerized in Mozambique as of today, 2002. The amounts of computerization in the district vary as a case study done by HISP in Gaza, Cuamba and Niassa province showed. The study, [2Braa, 2001], found that 5 out of 15 districts had a computer, but the computers were used for general work as the current health information systems only are for the province and the national level. Word processing was found to be the major task among the computer users in both the district and provincial level. The knowledge about computers was found to be poor among the users and most were dependent upon external assistance from the capital or other sources. The quality of computers varied on all levels, this was also observed by myself as I found a large variety of computers from Pentium II to old 486, all running a variety of Microsoft operating systems from Win95 to Win XP. A problem with the computers is the lack of sustainability, I myself often saw computers not being used due to lack of knowledge about installing and / or transferring the software / data, or defunct hardware such as a monitor.

Today all actual reporting and transportation of data to the province level is on paper forms, the forms are entered into eventually different computer systems at the province or the national level. The first computer based health information system was introduced to the provinces and the ministry of health in 1992, the system was specifically designed for Mozambique and was made for reporting on routine data such as immunization, family planning, drug distribution and consultancies. Data was entered in the province from several forms from the districts, and then electronically on a floppy disk transferred quarterly to the national level. This system is called SISC and is still in use today as the routine health information system. After SISC, several more vertical computer based systems have been added to the province and national level, often supported and still supported by foreign aid. For example, lots of the computers observed had small stickers saying USAID or Italian Aid. To get an overview of computer use I asked the Ministry of health about a list of the computer programs currently deployed, and I got a non-comprehensive list:

1. BES(weekly) spreadsheet & Monthly Epidiomogly (BES)
2. Spreadsheet for Tuberculosis data
3. Pilot AIDS,HIV system(SIDA)
4. Community health
5. SISC
6. Mental health care is experimenting with a database.
7. Pharmacy (Stock etc.)
8. SIMP (integration and analyzing tool)
9. DAG experimental pilot system for finance
10. Database program management by maintenance units
11. Database of human resources
12. Other miscellaneous systems, such as a CD containing all health units and related information such as location and size made by US AID some years ago.

None of these systems integrates with each other, except for SIMP, which has been made in house as a analyzing and integration tool. It pulls existing data out from the systems at provincial and national level. This means that SIMP probably does not address the problem of the poor quality of data collected as found by [2Braa, 2001]. SIMP was started as a response to the Ministry of health request for more integrated planning; other responses have been the design of several spreadsheets to store organizational data, such as data about the health units. This shows that the Ministry of Health is aware that today's system is not function as they want.

6.3 Information flow in the facilities and the districts

I will now briefly describe the information flow in the healthcare system, with an emphasis on three specific data sets in the facilities, district and provincial level:

- SMI = Deliveries, mother and child data
- PAV = Vaccinations/immunization data
- SIS = The rest of the medical data such as hospital data and notifiable diseases

The emphasis on the two specific data sets is due to:

- The fragmented environment, looking at the entire information flow would be a large project

- Due to my focus on specific computer systems it is necessary to understand the corresponding information flows in facility and district level, SIS and particularly SMI and PAV are essential parts in SISC, which was my main focus.
- The importance and use of these data sets in the healthcare

6.3.1 Health facilities level

Health workers at the health facilities as well as in the community collect data on immunization, mother and child health, family planning, drugs, outpatient's statistics, hospital etc using the SIS forms. However, there are problems with the SIS forms such as duplication, and gaps in needed information. Some of the gaps are solved by the use of locally designed forms, which sometimes leads to further duplication of data. For example, preventive consultations 0-4 years were found to be collected on 3 forms in a district (Chibuto, March 2002). Currently there are approximately 270 elements being collected, and the concept of essential data set seems to be absent. In addition to the SIS data there are a number of other vertical programs collecting data, such as DTS/SIDA, which are not part of the SIS.

Looking at SMI and SIS it is being aggregated into a monthly report and then sent to the district, while the data on PAV are being sent to the district on daily basis[19] (without being aggregated). The data is collected on one form, and then the report to district is produced on a different reporting form. E.g. for SMI there are three different forms for data collection, which are then combined into a single reporting form which is sent to the district.

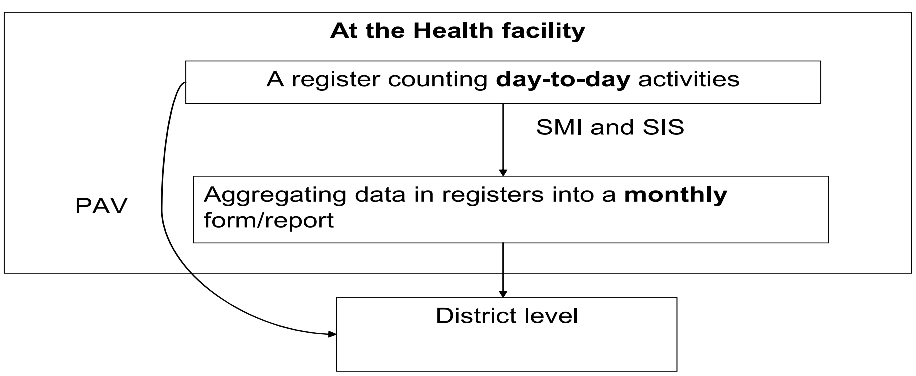


Figure 11: The flow of SIS, SMI and PAV in the facility, figure taken from an HISP report.

[19] It is not necessarily transferred to the district each day, but when eventually transferred it is one form for each day.

There were observed local use of data in the facilities, such as graphs. However, when asked for meaning of these graphs the staff had problem analyzing them.

6.3.2 District level

At the district the person responsible for SIS receives all the reports for all the programs from the health facilities, and then he distributes them to each individual responsible person for SIM and PAV. Each one prepares the report on the particular program and then they sit together to produce a district report, which has to be checked by the district director before being sent to the province. After checking the report the district director sends it back to the person responsible for SIS, who then sends the report to the province.

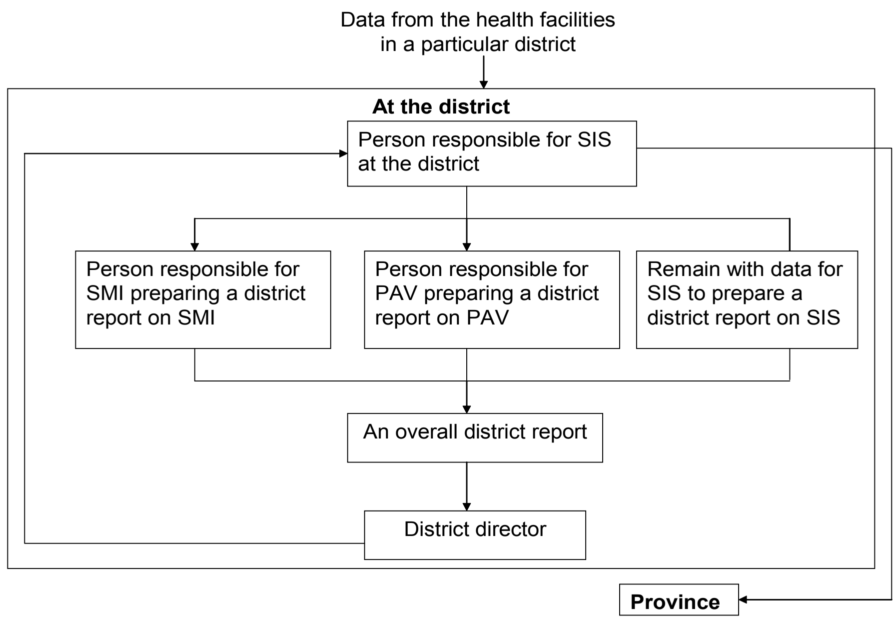


Figure 12: Figure showing the information flow for SMI, PAV and SIS, figure taken from an HISP report.

At the district data from the health facilities are being aggregated into single reporting forms and sometimes separated into different reporting forms. E.g. A single reporting form received from the health facilities on SMI is being reported into three different reporting forms from the district to the province (the new form contains all the facilities with sub set of the total SMI data set, i.e. different grouping of the data). The reporting on SMI data from the district to the province is done per health facility aggregated monthly, while PAV data is aggregated monthly pr. district. All forms used are stored in well organized folders going back several years at the district.

As in the facilities, there were also lots of graphs and data in the district offices, again the lack of data understanding was evident as shown in the table below (taken from wall at Medico chefe office, Mandlakaza).

Vaccine	Target population	Vaccinations	Coverage
BCG	1424	4805	337%
DPT1	1388	4835	348%

Problems with this table are as follows [Source: A. Heywood]:

- The target group for BCG and DPT1 is different, whereas it should be the same Due to the fact that both are vaccinations for infants or nearly the same age (4% total population)
- Coverage larger than 100 % is impossible, either the number of vaccinations is wrong (to large) or the target population is wrong (to small)

After investigation HISP discovered that population data was not based on actual population or estimates, but calculated by the province annually and sent down. However, no one trusted these figures. The HISP team attended two different data analysis meetings, and the conclusion was that in general the staff making analysis and writing reports appeared not to be trained in data analysis, in addition there were a lack of adequate tools for analysis of data. Data is entered on paper, despite the fact that several of the districts have computers with software installed (DHIS, Excel, Word etc.), and staff have been given some training. There seemed to be no feeling of ownership of the data and no pride in the reports made. Feedback was virtually non-existent, unless there were glaring gaps or mistakes in data entry. There was no evidence of active data use in planning of districts activities, supervision or in systematic monitoring and evaluation of program activities. However, most were aware of the poor quality of data as the following comment shows

DDS Chibuto "I know it [data quality] is not good, but I can do nothing"

6.3.3 Provincial level

At the province reports from the districts are being aggregated in each three months, whereby the three-month report is cumulative i.e. three month, six months, nine months and annually. The person responsible for SIS receives all the reports for all the programs from the districts, and enters the data into the SISC system. The reports are then distributed to each individual responsible person for the SIM and PAV programs. Each one prepares the report on the particular program, while the SIS person also prepares an overall report for all programs (SIS, PAV and SIM). The overall report is compared with the reports from the individual responsible persons by the SIS person, and then the overall province report is sent to the National level.

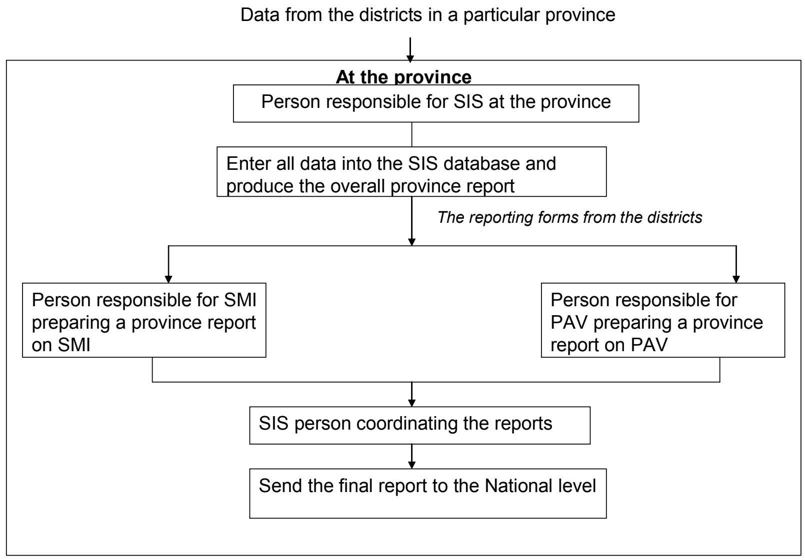


Figure 13: Showing the information flow at the province level for SMI, PAV and SIS, figure taken from an HISP report.

6.4 Summary

The findings at the facility and district level have been very much the same as [2Braa, 2001] found. Data is collected methodically and reported to the fragmented information systems in the levels above, there is little data use and data skills at the lower levels. However, there are several procedures in place which tries to increase data use, such as use of graphs, information meetings and indicators on forms. The quality of the data collected is poor due to several reasons such as large amounts collected, lack of data use, and lack of feedback. Collected data which should cause alarm are not triggering any action and there is little evidence of use of information in management and deciding action

7 Empirical work, decoding existing system

In this chapter the findings from the investigation into the existing routine health information system, SISC, is presented. The primary goal was to enable stored data to be found, read and understood, this in order to enable the import process of data into DHIS.

7.1 About SISC

SISC is part of the computerized routine health information system in Mozambique; it comes in two versions which are very similar, one for the national and one for the province. I will in this chapter refer to SISC as both systems, when I am specifically writing about one of the versions I will write SISC Provincial or SISC national. SISC main data is:

- Reporting and analyzing of monthly routine health related information.
- Storing of the infrastructure including properties for the facilities and districts

SISC was developed and deployed in 1992 - 1994 in Mozambique as the first computerized data reporting information system; it was developed by a foreign engineer who is gone today. The system is based on Microsoft DOS, and has no support for mouse. However, it can run in a DOS window in win95, but not win98. Regarding hardware it runs primarily on old computers, 386 and 486. However, with some difficulties we managed to get it to run on our computer, win2000 and Pentium II, but it was not stable in this environment.

7.2 Lack of documentation

Documentation and know how about the system was almost totally absent, or unavailable due to several reasons such as:

- Source code were stored on floppies, no one know there whereabouts of these floppies
- Source code and documentation where supposedly stored on an old 10+ year old Toshiba laptop. The battery was defunct or missing, and as a consequence the computer was unusable. The laptop could have been opened and the hard disk connected to a desktop, this would have required extra computer gear, such as an adaptor, in order to connect the hard disk to the desktop computer. However, there were no guarantee for that the hard disk stored the data wanted

Source code and documentation might have been attainable given enough investments of resources and time. However, it's eventually condition was unknown, and given that SISC was written 10 years ago combined with young students might have led to that the source code and eventually documentation might not be valuable.

7.3 SISC information flow

Important part of SISC are the paper forms, the structure of the database is build up based on the paper forms, each form entered has it corresponding relational table. All information in SISC is coming from paper forms reported and aggregated down in the health organization. The figure below shows the flow for the most important reported data sets of SISC.

PAV=Vaccination data set
SMI= Prenatal, maternity (including deliveries) and mother and child data set

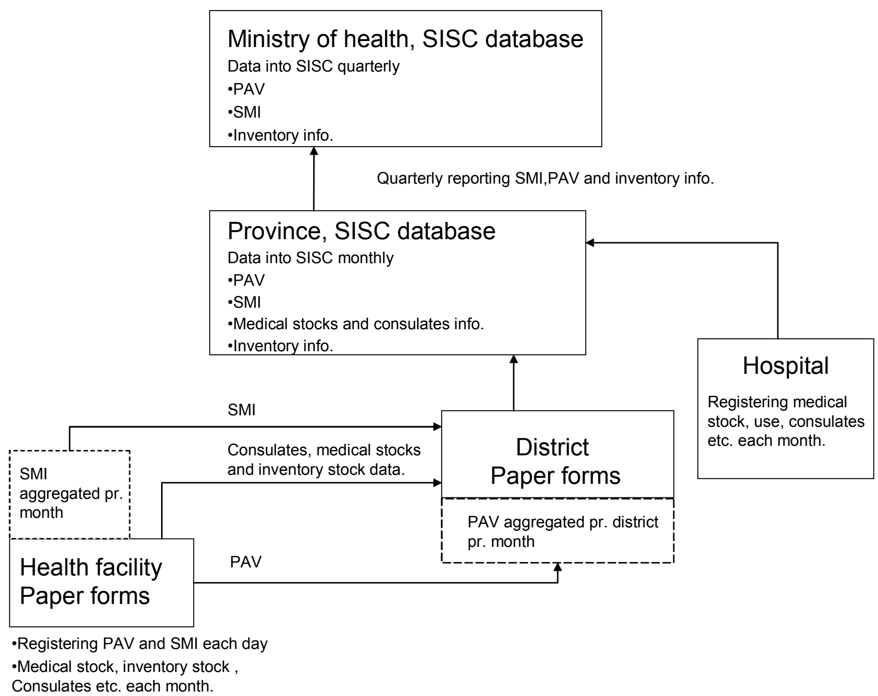


Figure 14: Showing the information flow for the SISC elements, from health facility up to national level. For a more complete picture of the SISC flows see Appendix A.

As the figure shows most of the information is aggregated after first being reported.

- PAV reported daily at health facility, ending up in SISC aggregated to month and district.
- SMI reported daily at health facility, ending up in SISC aggregated to month.

- Consultancies (consultas) pr. month.
- Medical stocks, inventory, bed usage etc. is not aggregated based on how it is being reported (monthly).

Note that SISC is a closed systems without functionality to export and import data to and from other systems.

7.4 SISC user interface and functionality

SISC is an old program and its functionality and user interface is poor compared to today's standards. SISC resembles old MS-DOS based software, menu driven where the users selects a standard selection from a menu by moving the cursor with the arrows on the keyboard. Advanced features are selected from top menus which are accessed by holding down ALT or other control keys combined with a letter key. Given its age, it does not have a poor user interface; often used selections are easily available and I had few problems understanding the standard flow of the program, even though it is in Portuguese. In addition to the SISC executable file there are several other tools which can be used, such as a graphical analyzing tool etc. These tools have also been investigated.

SISC functionality can be divided in three parts, data entry, control & validation and analyzing. The two versions have some varying functionality.

7.4.1 Data entry

SISC national gets data from the province on a floppy made by the SISC province program; i.e. no data is entered at national level. The staffs simply insert the floppy from a province and select an import function in SISC national. SISC national then automatically recognizes which province and time (year, month) the data is for and transfer it to the national database.

The SISC provincial have manual data entry from the forms, typically done by a dedicated staff, the procedure is as follows:

1: Selects which form to enter

```
22/04/02      Ministério da Saúde - Sistema de Informação      GAZA 2002

        +-------------[ Introdução de Dados ]--------------+
        ¦                                                  ¦
        ¦ A04  : PAV Resumo Mensal para Distritos          ¦
        ¦                                                  ¦
        ¦ B05  : Stocks/Métodos Anticonceptivos            ¦
        ¦ B06  : SMI Cons 0-4 Anos/Vig.Nutric.             ¦
```

```
¦ B07  : SMI Maternidades e Parts. Trad          ¦
¦ B08  : SMI Cons Pre Nat/Post Parto/PF          ¦
¦                                                ¦
¦ C04  : Consultas Externas e Stocks de kit A - CS ¦
¦ C04H : Consultas Externas dos Hospitais        ¦
¦ C05  : Consultas Externas e Stocks de kit B - PS ¦
¦                                                ¦
¦ J01  : Recursos - Rede Sanitaria/Camas         ¦
¦ K01  : Recursos - Pessoal Niveis 1/2 (K02)     ¦
¦                                                ¦
¦      Retorno ao Menu Principal                 ¦
¦                                                ¦
+-------------------[ Ajuda-F1 ]-------------------+
```

Above is a screen capture of the "Select form to enter" in SISC provincial. The user moves between the selections of forms by using the arrow keys. The leftmost column indicates the form number, while rightmost column is the name of the data set. First selection is form A04 which contains "Vaccinations data (PAV) Monthly summary for Districts"

.

2: Then chose year, month and place (district or health facility)
3: A new screen is shown resembling the paper form with boxes for each data element, all elements are as default set to 0, the user moves around the boxes with the arrow keys and enter data by using the keyboard. When finished a special key is pressed and the user is returned to the select form to enter screen.

7.4.2 Control and validation

SISC have a tool called VERI.EXE which automatically checks the structure of the database and reports on the status of data in the tables, both monthly reported data and infrastructure data, including duplicate records as the database have no functionality for keys. The results are stored in a text file when finished. In addition to this the SISC provincial can also check the data and report to the user about missing reports from districts. The SISC national have a function to report on empty databases and check for faulty tables/databases. Note there is a difference between an empty database and no database at all, it does not check for no database. In addition it checks for missing reports from districts, but this does not seem to be perfect as it only report up to which month data is reported, it does not tell the user if it is 1 or 10 units who have reported.

Examples:

- SISC will report missing record if 10 out of 10 districts have reported January through November, complains that data is only up to date up to November, but only 8 % of the data is missing.
- SISC will not report missing record if 1 out of 10 districts have reported January through December. No months are missing, but 90 % of the data is missing.

SISC contains few validity rules, the only one I heard about was that the system checks to see if a unit reporting data has the ability to report the kind of data it is reporting. There is no validation of the data entered, and as my research will show the program accepts for instance a data element to increase 600 % for a month.

7.4.3 Analyzing

SISC includes a tool called GRAFICOS.EXE which is used to analyze data and make graphs; in addition indicators such as coverage can be made to assist the health staff in province and national level in managing. This is unfortunately where SISC most show its age, the user interface is very poor and it is difficult /impossible to export the graph made to other tools for making reports. An interesting difference in the SISC national and provincial version is that the SISC national only have the possibility to make reports and analyzes on a quarterly basis, while the SISC provincial has the ability to make monthly reports and analyzing. This was done due to a paper shortage at the time of implementing and deployment, and today, when the paper shortage is not as important they are still affected by the inscribed procedure. This shows the lack of flexibility in SISC will I will return to later in the thesis.

7.4.4 Updating of infrastructure data

When the program was implemented in 1992 it was initialized with empty databases for each province for 1992 to 1999, including the infrastructure data and empty tables for reported data. Some claimed that the program was never intended to be used after 1999, but this was unconfirmed. After 1999, new databases for each year have been made by running a commando in SISC which initializes a new database for a year. However, there have been problems with the year 2000, for instance some files made by SISC were found to have 2095 as their creation data, this can be due to the operating system and/or system clock not handling the year 2000, or simply just wrong entered system time. And it does as such not prove that SISC was unable in coping with 2000 and the preceding years.

Adding of a new health facility can only be done at the national level, while deleting and renaming of facilities can be done at both the province and the national level. I was told that adding of a health facility was a large and complex process, and the person did not mention eventually how the changes done at national level should be migrated to the province, as there were found few indications of data going from national to the provincial level. There are no ways to alter the districts.

7.5 *Database analyzing, what does SISC store?*

The database used as back end is Dbase version III. The data in SISC is organized into one database for each year for each province, containing the infrastructure and the reported data, all having similar Meta data. The database is located in a folder with the year as the name (for example “1998”). In this folder each table is stored as two files, one .CDX file containing system information for that table and one .DBX file containing the column names and the actual data (rows)[20]. I was able to open the tables, the .DBX files in MS Excel and also to import the entire database to MS Access. Both these tools were used extensively to analyze the database. The difference between the SISC versions is that the SISC province version has one database for each year for the province. The SISC national have one database for each year for each province.

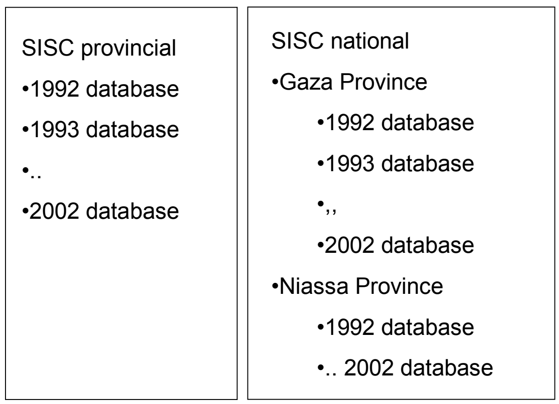


Figure 15: Showing how the database in the two SISC versions is organized, the provincial database is replicated at national level.

The actual databases in SISC provincial contain 21 tables, while the SISC national databases contain 14 tables, replicated from the provincial. Thus the national database is a proper subset of all the provincial databases, all tables found in the national are found in the provincial. However, the provincial are most often better updated as data is entered monthly, while the national is updated quarterly.

7.5.1 Information flow and decoding the data

As previously stated the tables correspond to the data collection forms. To explain this I will use the vaccination data set (PAV) flow as a case example, see figure 17 on next page for an illustration. At the health facilities the workers are using tick sheets to enter the data on form A01 and A02, these tick sheets are then sent to the district centre where they are aggregated to district (sum all health facilities PAV data in each district) and month on form A03. Form A03 is then sent to the province, where its content is moved to form A04 and entered into table A04 in SISC. The table A04 is then quarterly exported together with

[20] I have not gone in-depth into how Dbase is build up, and the DBX files contain more than the column names and rows. However, I have been able to gain access to the column names and rows by solely using those files, i.e. ignoring the CDX files. Furthermore using database driver to gain access to the database also returned little information, no keys etc. Therefore I believe that the CDX files only contain system information.

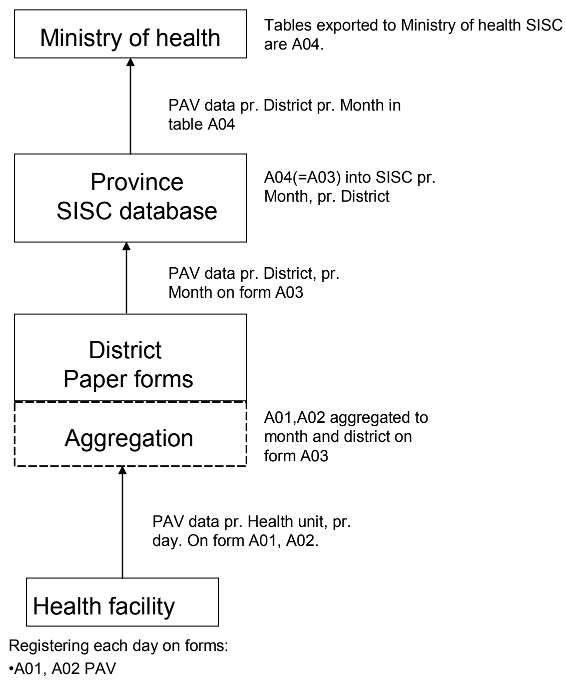


Figure 16: Figure showing the flow for vaccinations data

most of the provincial database from SISC provincial to the corresponding SISC national database. When the form A04 enters the SISC database it entry boxes are mapped to columns in table A04, this is of course transparent for the user who has an interface where data is entered. For example, on form A04 you have the following elements, see table 1 and appendix B and C for an example of form A04.

VACINAS / GROUPO ETARIO	0-11 MESES	12 MESES DU MAIS	TOTAL APPLICADAS
BCG.	876	26	902
POLIO APLICACAO PRIMARIA	770		770
POLIO 1A DOSE	827	27	854

Table 1: Some of data elements on form A04; values are examples.

The table A04 in SISC contains several columns with names such as BCG1, BCG2, POLD1_2, and VATMA_1. These columns names are representing data elements on the forms and the rows actual data. Thus there is a mapping from the A04 form data elements to the columns in A04, but it is not one to one as A04 contains more columns than A04 have data elements.

ANO	MES	DCOD	SNUM	BCG1	BCG2	BCG_DESP
0	01	04	1	258	58	0
0	01	11	2	355	32	53
0	01	03	3	195	32	0
0	01	07	4	109	0	38
0	01	02	5	68	2	14
0	01	01	6	878	46	586
0	01	08	7	82	2	110
0	01	05	8	93	0	37
0	01	10	9	652	50	0
0	01	06	10	193	12	95

Table 2: Table A04 with some columns and rows.

As shown above the A04 columns names in the SISC database gives no or little information, and the forms are essential in order to decode the database and understand the information stored.

A04 Form entry boxes	A04 Table columns	Value
BCG 0-11 MESES	BCG1	876
BCG 12 MESES DU MAIS	BCG2	26
POLIO APLICACAO PRIMARIA	PAP	770
POLIO 1A DOSE 0-11 MESES	POLD1_1	827
POLIO 1ª DOSE MESES DU MAIS	POLD1_2	27

Table 3: Example of mapping from A04 Form to SISC database columns; to the left are the names on the form. In the middle the corresponding column in the SISC database, to the right the actual value as found on the form and in the SISC database.

Note that aggregated values on the forms are not stored in SISC, for example "Total Applicadas". The example table A04 was the easiest to decode as for instance column name BCG1 was fairly easy to map to the BCG elements on the A04 form. However, other tables were more difficult, for instance table B07 with column names D1 to D8, for a comprehensive mapping of SISC elements see the end of appendix H.

Further research enabled me to decode the entire A04 table and reach the following conclusions. ANO and MES is year and month, DCOD is a district code. ANO MES and DCOD together make the primary key for table A04; that is a district can report once each month, each year. Unfortunately, the database system did not have support for keys so this constraint was upheld in the software. SNUM is a column being used by the program and does not contain any health related information. The rest of the columns are for data from the form A04. All tables are most likely on Boyce Cod normal form. However, due to that all columns were not fully decoded it is impossible to verify. Also note that year is redundant as the name of the database indicates which year the data is for, this is also stored in each data report table within the database.

This case of A04 is representative for the way information moves and is stored in the SISC system, the difference between the other information flows is primarily how it is aggregated. Everything is aggregated monthly, while some data is aggregated to district while other is pr. health facility. By analyzing the tables and interviews the following rules which could be applied to the tables storing monthly reported data were found:

1. Tables with leading columns equals: ANO, MES, DCOD, UCOD (primary key) is monthly data pr. Health facility, UCOD is a code for a unit, DCOD is a code for district, UCOD and DCOD uniquely identifies a health facility in a table UNIDS containing all facilities for the province, typically a foreign key. DCOD also uniquely identifies a district in table DISTS containing all districts.
2. Tables with leading columns equals: ANO, MES, DCOD (primary key) is monthly data pr district, where DCOD uniquely identifies a district in DISTS.

3. The rest of the columns in a table are data elements for a form or system information.

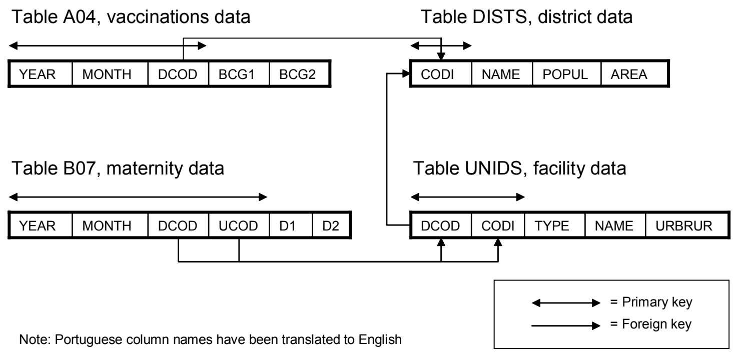


Figure 17: Showing some tables and relations in the SISC database, note several columns not part of a primary or a foreign key is left out.

By applying these rules and using the forms I reached these conclusions about the tables in the database:

- Table A04 is PAV data pr district pr. Month
- Table B06 is SMI data, children consultancies pr. Health facility, pr. month
- Table B07 is SMI data, Maternity data pr. Health facility, pr. Month
- Table B08 is SMI data, pregnant consultancies pr. Health facility pr. Month.
- Table B07_REG is SMI data, Deliveries pr. District pr. Month.
- Table C04_REG is Total consultancies pr. health centre (a large facility)
- Table C04_HOSP is Total consultancies pr. Hospital
- Table C05_REG is total consultancies pr. Health facility

For the infrastructure I found the following tables:

- Table DISTS is district data (primary key would be district id); in addition the population and other elements are stored.
- Table UNIDS is health facility data, each health facility is identified by a facility code and a district code (primary key), district code is a foreign key to DISTS. In addition several properties about the facility are stored. Such as wherever it is located in an urban or a rural setting, type of facility (Post or center), services available and type of data it is supposed to report, divided in:
    - PAV: Vaccinations data
    - SMI: Prenatal and child and mother data
    - MAT: Maternity data, which is a subgroup of SMI data.

There are more columns in several of the tables storing information which I did not manage to decode, or were deemed as unnecessary to decode as they were not used, for instance table CO4 contains columns for medical stock and use, but these were not in use. Although there were no rigid rules for what was used and not used between the provinces, for instance some provinces seemed not to report consultancies; Niassa used it very little, while Gaza used it, similar observations were done regarding use of table J01 in Maputo province vs. other provinces in 2001. This was not investigated anymore as the focus was on vaccinations data, maternity, prenatal and child and deliveries (PAV, MAT and SMI) data. As these seemed to be the ones which all provinces were reporting. Tables found deemed as not, or little used were:

- The tables J01, J0_REG, K01, K0_REG, C05, and C04 are not or little in use today. However, some of them have been used before.
- The tables KIT and CAT RESP are look up tables (medical supplies); they are not storing any useful information for DHIS.
- The tables B08_REG and B06_REG are system tables; they store which user did what and other misc. system information.

I spent little time analyzing these tables. Primarily the tables and columns not being used was supposed to store medical stocks, use, inventory, number of beds etc. While I found that the SISC database was not being used for this information I saw that the paper forms are used and the health facilities are reporting the data, but it is not entered into SISC.

As previously stated the corresponding province databases in the national database contain fewer tables than the actual provincial database. The difference is that the national version is not storing any C0X tables, KIT and CAT_RESP. Consequently the national version contains primarily information about PAV and SMI. The extra data in the provincial version consists primarily of tables for hospital and health centres with information about consultation.

Note: I looked at Niassa province (up to Dec 2001), Gaza province (up to April 2002) and the National database (up to April 2002). And as such these are not absolute findings as I have been unable in obtaining all provincial databases.

7.6 Summary

The legacy systems characteristics of lack of documentation, source code, know how of the system and dependency upon old hardware was the framework for the investigation of SISC. This implied that research into the backend combined with organizational routines at all levels and data collection forms were necessary in order to decode data stored. SISC was found to be in two versions, one national and one provincial; it is MS-DOS based and shows its age of 10+ years. Data was found to be partially replicated to the national level; both SISC versions use Dbase ver. III as the back end.

8 Empirical work, Import process

In this chapter my findings into the import process is presented and explained. I begin with explaining the DHIS import functionality and the database design, as it is an important element in the import process. I then explain my work on the actual development of the tool and related problems. In the end I explain the use of the tool in Mozambique in importing data and related problems such, as the necessary adaptation process of DHIS.

I this chapter little actual code from the tool will be shown; to view the code and documentation see the CD included as appendix J, it contains the source code, an executable jar file and the documentation in html code, see the readme.txt on the CD for more.

8.1 DHIS import functionality

The DHIS import functionality influenced my work and must be explained in order to understand the tool developed. DHIS have the functionality to import both:

- Monthly data: Monthly data reported by health facilities
- Organization Unit data (OrgUnit data): Data about the infrastructure and the organization of the healthcare.

This is the data DHIS mostly consists of, and each of these has their own file format. A problem was that I did not manage to find documentation for the OrgUnit file format, only the monthly data format. However, the two file formats are technically very similar, but I have been unable in understanding how DHIS uses the OrgUnit data at import. Therefore my tool does not support making import files for infrastructure data, OrgUnit. When I am referring to DHIS import file I am referring to the import file for monthly data.

I opted to use the import functionality in DHIS due to the functionality it offers. Such as check for duplicates in imported data and between imported data and existing DHIS data in the database, then prompt the user for appropriate action at import, and control that referential integrity is upheld. An alternative to use the DHIS import functionality could have been to use SQL quires to insert data directly into the DHIS back end database. Given some advantages, I found it overall to be too risky and a few reasons are showed below:

- Little control on duplicates in the imported dataset and data already existing in DHIS, which is a problem as the quality of the source data is unknown.

- A powerful tool, being fast and efficient, but which could have severe consequences if wrongly used (i.e. replace correct existing data with wrong data)
- Error message from SQL could make the tool unusable for the user
- Changes in the DHIS database might make the tool useless until modified

These risks were further increased if the users were to be none experts.

8.1.1 DHIS import file and database design

The DHIS import files is in plain ASCII text format, with column names in the first line, then the values in the following lines until the end of the file (i.e. there are no line or file terminating except the standard ones (line change[21] and end of file). The file structure is organized by the use of the following rules:

- "," is the separator between the elements
- String is identified by ""
- Float, or double is identified by X.XX, where X= Integer
- Integer is identified by any number X
- Date is identified by "yyyy/MM/dd", sometimes the date also includes time, such as HH:MM:SS

Note that there is no way to separate string from a date. But the column headers and the data format is a good way to recognize what type of data it is. See below for a small example of some columns from a DHIS monthly import file (For more see the Appendix with the user manual for the tool).

"strDataField","strOrgUnit","Period","dblEntry","strComment","strUser"
"Adrenaline 1/1000 (1ml) ","kz Vryheid Mobile 3","2001/08/01",10.00,,"Jabo"
"Amoxicillin 125mg/5ml ","kz Hlobane Clinic", "2001/08/01",4.00,,"Jabo"

To explain the file structure I will now describe the first line in the example above ("" denotes data from the file):

Health facility "kz Vryheid Mobile 3" have in "August 2001" reported "10.00" "Adrenalin 1/1000(1ml)" and it have been entered into DHIS by "Jabo" who has given no comments on the entry, shown by the ",," (an empty data entry).

Note: In this example period is represented with the starting day of the month, therefore "2001/08/01" translates to "August 2001". However, DHIS have the ability to interpret other date representations, such as "August 2001".

[21] Due to the fact that DHIS is developed in Microsoft tools they use the Microsoft standard for line change, line change then carriage return, while Linux for instance uses only one for both line change and carriage return

The data element name (strDataField), name of heath facility/unit (strOrgUnit) and month (Period) makes out the primary key in the import file and in the table tblMonthlyData, the table containing all the reported data in DHIS. The column dblEntry contains the reported value. These are the minimum set of elements which must exist for import of date. The rest of the columns in the import file are information about the data, such as yes/no check status, last update time, last user editing. Important referential rules in the table tblMonthlyData in DHIS and as a consequence are rules for the import file (Due to the fact that the content in the import file goes into table tblMonthlyData) are shown in the figure below:

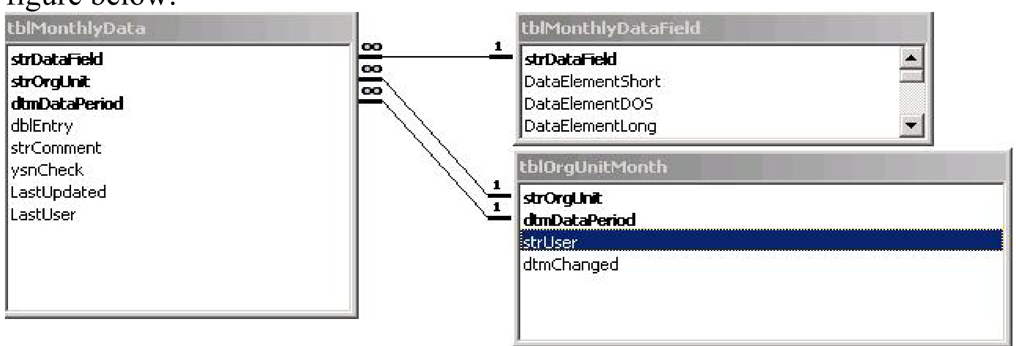


Figure 18: Screen shot of the relationship from table monthly data.

strDataField is a foreign key to tblMonthlyDataField which consists of a list of all the allowed data elements names and information about them, such as short name, long name etc.

strOrgUnit and **dtmDatePeriod** is a foreign key to tblOrgUnitMonth, which consist of the health facilities (strOrgUnit) and a reported month. From tblOrgUnitMonth column strOrgUnit is a foreign key to table tblOrgUnit consisting of all OrgUnits and their properties, such as number of beds etc.

The solution of storing the data elements (strDataField) and facilities (OrgUnit) as rows allows for adding of data elements, facilities and units without altering the Meta data. In contrasts to SISC were it was necessary to edit the database design in order to add a data element.

The referential rules described above means that the DHIS database must have been correctly set up with the infrastructure and data elements before importing of monthly data can commence. Importing of monthly data with strDataField and strOrgUnit not existing in the corresponding tables are refused by DHIS. In addition, the data might have to be mapped as the name of the health facilities (strOrgUnit) and the data elements (strDataField) is not necessarily the same in the data source as in DHIS, for example data in SISC is coded, while DHIS stores data un-coded, or the coding is hidden from the user. Also note that only monthly data is allowed in DHIS and it is only facilities which are allowed to report data in DHIS; i.e. the end nodes, leaf, in the organizational tree.

8.2 About the DHIS import tool

The objective of the tool is to make DHIS database population easier when importing from data sources. With a data source I am referring to a clearly defined and structured list of data in a file or a database. Having the column names in the first row, then rows with data elements separated by a separator. To make the tool possible to program in 2-3 months development time I had to define a small focus, in the beginning I wanted the program also to do consistency check on the data to import, and assist the user with quality control, but due to the limited time this was dropped. The important factors left were:

- Be used in several different countries with several different data sources
- Not all the data in the data sources were to be imported, thus needed was functionality to select specific content in the data source. Note specific content means entire columns of data, and not specific rows.
- The data needed to be customized mapped to fit the DHIS data dictionary
- Add default values for particular DHIS columns not having any corresponding values in the data source, the data source does not necessarily store as much additional data about the data entries as DHIS does. For instance if a source data element did not have any maximum value (intMax), then my tool should add a user customized value as the maximum value.

I choose to divide the solution in three modules:

1. Open the data source: The user should select the data source he wants to get data from. Then the module should read and understand the source.
2. Select information to import: Present the available information in the data source and allow the user to select data for import. Implemented as selecting a column in the data source to be copied to the corresponding column in the DHIS import file.
3. Set up mapping, add default values and finally make the DHIS import file

Basically what the tools does is to rearrange customizable columns in the data source to the DHIS import columns, plus pivoting the data elements and adding necessary mapping and default values. See figure beneath for a simplified example.

Data source example

Facility	Period	Beds used	Doctors
Xai-Xai	May 02	200	35
Delano	May 02	320	43

Simplistic DHIS import file

OrgUnit	Date	Data element name	Value	IntMax
Xai-Xai	May 02	Beds used	200.00	0.00
Xai-Xai	May 02	Doctors	35.00	0.00
Delano	May 02	Beds used	320.00	0.00
Delano	May 02	Doctors	43.00	0.00

Figure 19: Showing how the tool would rearrange, add default value and pivot the data elements in the data source to the left, into the simplistic DHIS structure on the right.

Java was selected as the development language due to my experience. Since DHIS is made with Visual Basic my selection of Java is from an integration viewpoint a bad language to develop in (Microsoft are not happy about Java nowadays, java virtual machine is not included in Win XP for instance), but I had no experience with Visual Basic, and given the time I had I found it difficult to start developing in Visual Basic. Ease of use was important as the tool were to be used by inexperienced users; therefore I selected to have a graphical user interface. This was possible due other less important factors such as hardware requirements and operating systems.

8.2.1 Wrapping (able to open multiple data sources)

In order to gain access to old data the tool needed to read and understand multiple data sources, or different formats, such as:

- Excel spreadsheets
- Structured text files
- Databases
- Custom file formats (epi6)

The 3 first above were selected on basis of the files existing in Mozambique; while the last one was added on request from the HISP software coordinator. In the future new data sources, or file formats might be found in other countries, therefore I had to try to make a solution which is easy to extend with new data source support. Given the amount of different data sources and the possibility for several new ones, I needed some kind of a common way to integrate the data sources with my tool. So that it would be unnecessary to make a new tool for new data sources. As such I was searching for a standardized way of connection different data sources to my tool.

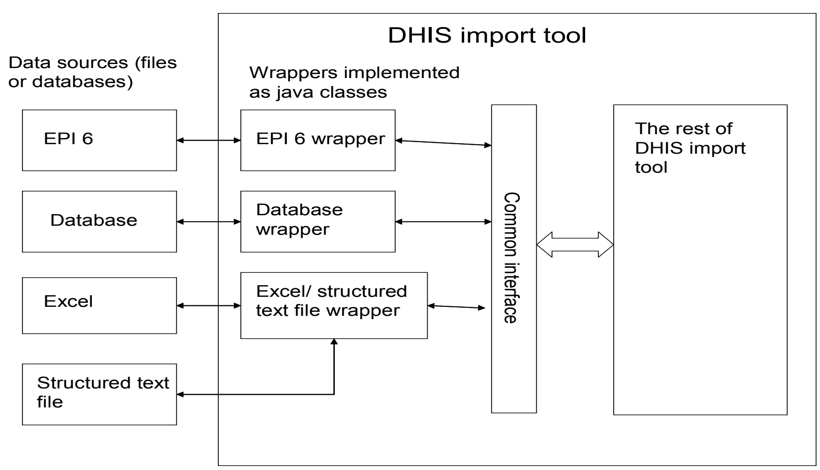


Figure 20: Overview of the wrapper solution. The DHIS import tool only needs to sees the common interface as each wrapper implements the interface.

I selected a wrapper based solution; a wrapper is layer of higher abstraction. In my case it was implemented as a wrapper class for each type of data source (i.e, database, Excel etc.), consequently there are as many wrapper classes as there are different types of data sources. Each wrapper class implements a minimum set of functions which gives access to the content of the data source. This gives the rest of my tool a standardized set of functions which could be applied to all data sources, or a common interface.

This solution was implemented by defining an interface called WrapperInterface, declaring the minimum set of required functions. Then defining an abstract class called Wrapper which implemented the WrapperInterface. The actual wrapper classes are then defined as sub classes of the abstract wrapper class. Actual Implementing of the WrapperInterface is done in the sub classes where the functions are implemented according to the type of data source they are wrapping. Such as that the DatabaseWrapper class has an open functionality which opens a database, setting up the connection. While the open functionality in TextFileWrapper opens a file. As such the abstract class, Wrapper, does not actual implement any parts of the WrapperInterface.

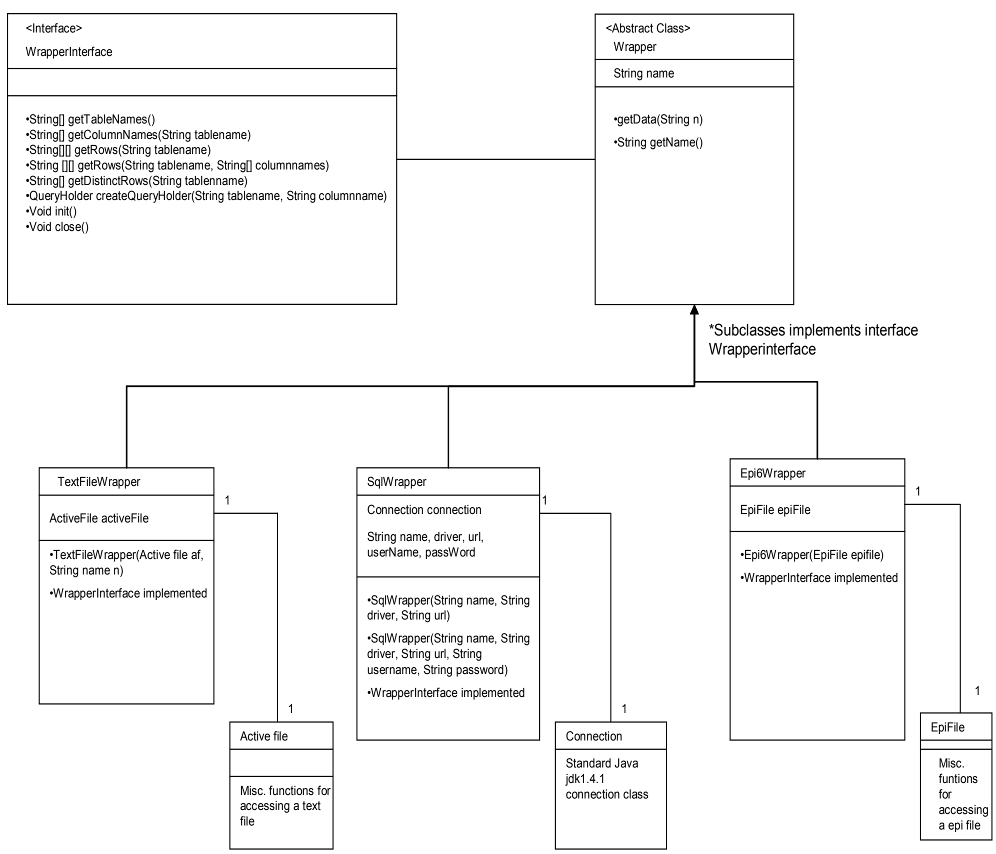


Figure 21: Figure showing how the Wrapper solution is implemented in Java with classes and interfaces. Note that the WrapperInterface functions are implemented and defined as in the WrapperInterface in the subclasses, but it is left out on the diagram due to space.

As usually the solution was more complex than described here, I found out that I needed more classes and interfaces to make it work. A problem was that data selected for import into DHIS could not just be read into memory and kept there until the actual import file was made, cause that would mean that the tool would use lots of memory for large databases. Instead each time the user selected a column of data to import, the correspondingly wrapper class made a QueryHolder, which basically stored the necessary information to get the data. For instance for a database wrapper, a QueryHolder would be a SELECT sentence not executed until the actual making of the file. While for the textfile wrapper it would be a class holding an array of the values from the file, with functionality to return the values, this as I did not have the time to make a read on the fly functionality for text files. However, this would be possible to add later on by editing the QueryHolder for text files. The QueryHolder is implemented in the same way as the wrappers. Consisting of an interface, QueryHolderInterface which is implemented in an abstract class, QueryHolder, which has several sub classes where the interface is actually implemented. As such each wrapper class has a corresponding QueryHolder class implementing the QueryHolder interface, such as SQLQueryHolder and TextQueryHolder.

There are several advantages of using a wrapper based solution, but most important for me is that it is easy to extend and cater for new types of data sources. It is only necessary to make a new wrapper class implementing the WrapperInterface for each new type of data source. Then make a QueryHolder for the data source. The rest of the tool would then be able to use the new wrapper and connect to the data source. A disadvantage is that since the interface must be the minimum of common functionality, all specific functionality for each different data source is made very inaccessible. The only way to gain access to it is by adding more complexity to the code by using class casting. The minimum set of functions making out the WrapperInterface came about after making the different wrappers and testing them. Wherever such a solution would be practicable in more complex environment is difficult to evaluate, for instance if the data to import had been unstructured or structured differently.

EPI6Wrapper

"EPI is a public domain software package designed for the global community of public health practitioners and researchers. It provides for easy form and database construction, data entry, and analysis with epidemiologic statistics, maps, and graphs". Ref: EPI homepage at http://www.cdc.gov/epiinfo/about.htm. EPI is used in South Africa and several other countries, therefore the HISP software coordinator wanted to have a tool which could import data from it into DHIS. Note that I only made a wrapper to support the so called EPI 6 files.

All EPI 6 files have a header with information about the elements, first there is a number indicating number of data elements (elements here is the same as columns in databases). After that each element is defined on a line, such as element name, display colour, and important for me, the position of the actual data in the following data rows after the header. Thus, firstly my wrapper reads the header data and stores it, then parses the actual data values based on the data copied from the header. As with text files all data is copied into arrays. For the user the procedure is easy as there is no necessary user input, it is only necessary to select and open the file then proceed to the next stage.

The added support for EPI 6 file at the later stage only took 1-2 days to make, this due to the fact that I only needed to make three classes. The EPI6Wrapper, an EPIQueryHolder and a class called EpiFile. The tool does not need any altering as long as the new classes support the WrapperInterface and QueryHolderInterface.

SQLWrapper (used for Databases)
To read databases I selected using JDBC (Java DataBase Connectivity) and ODBC (Open DataBase Connectivity by Microsoft). ODBC offers connectivity to databases based on C; the JDBC is Java's extension offering the ODBC services platform independently.

JDBC.ODBC is a set of classes offering a common interface to most databases (each database type have a wrapper which implements the functionality). Since JDBC.ODBC is a common interface supporting all database functionality I had to be careful with what functionality in JDBC.ODBC I based my program on. This as all databases might not offer it. Java throws an exception when functionality in JDBC.ODBC is invoked which is not offered by the database currently connected to. An example of this is when I worked with dbase ver. III, I wanted to get the foreign keys, but dbase ver. III does not support foreign keys, so once functionality asking for foreign keys where invoked my tool cast an exception. Due to this and the fact that the tool was to be used for lots of old databases I choose not to have support for advanced[22] database functionality. My database wrapper was made very simple with the same functionality as the text file wrapper.

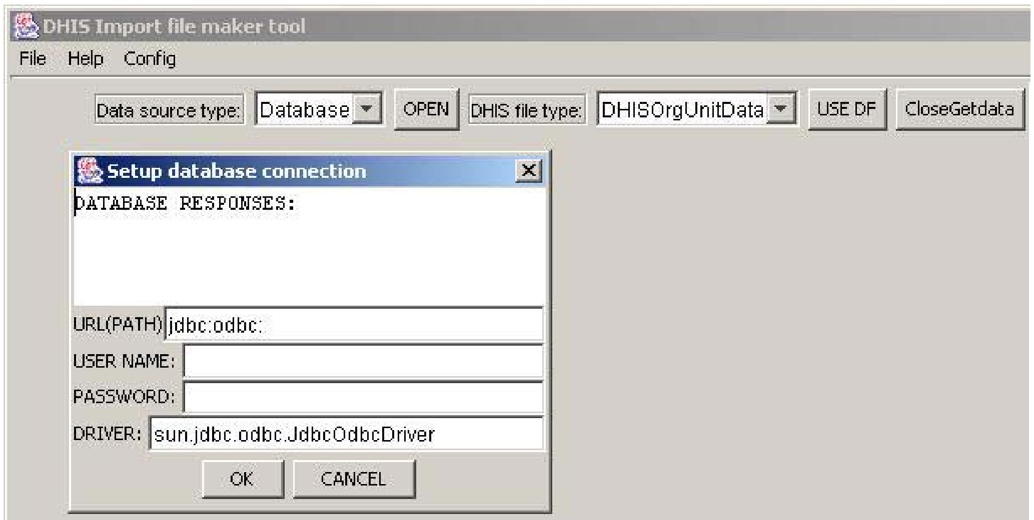
- Connect
- Get a minimum of Meta data (basically table and column names) in order to allow for selection of data to import
- Get all values the user wants from the database

Actual implementing of the SQLWrapper was quite easy after understanding JDBC.ODBC and mostly consisted of writing several SQL sentences which JDBC.ODBC sent to the database. For the user the procedure to connect to a database is easy:

[22] Foreign key is not advanced functionality today, but I also left out everything else such as triggers, checks etc.

1. Make a data source (an operating system specific procedure)
2. Give the necessary information to DHIS import tool to connect (the database name and eventually user name and password)
3. If all is correct the meta data is obtained and the user can proceed to the next stage

The SQLWrapper can also connect to spreadsheets if they have been set up as data sources. This due to the fact that there are MS ODBC drivers made for it.



Picture 11: Screenshot from the dialog used to configure database as a data source. In the background is the standard display used by the user to select type of data source.

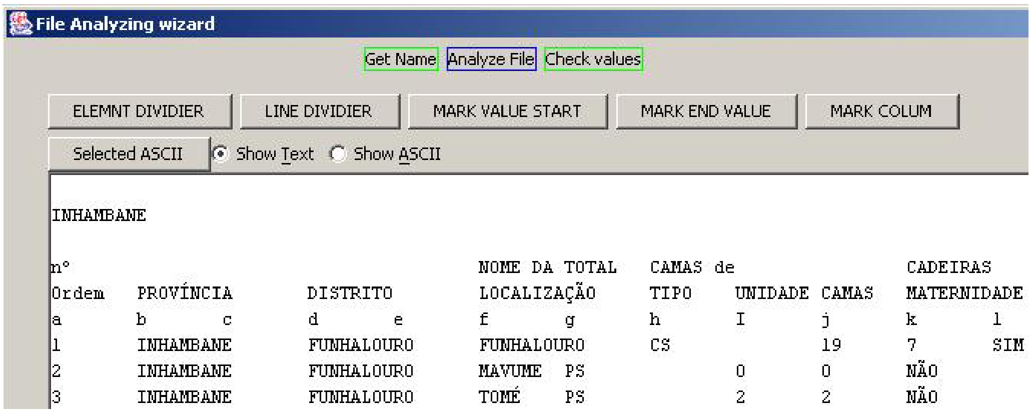
TextFileWrapper (used for Excel and structured text files)
The fact that Excel is a proprietary format from Microsoft made it difficult to connect to it with Java. I searched the web for packages/classes which would allow my Java program to automatically read Excel files, but I did not manage to find any I could use. The ones I found were either expensive or not complete. Instead I made a general text parser which was able to read Excel files saved as text files, this also solved the problem with structured text files. The text files although needs a strong structure to be able to be read and understood by the parser; basically they need to be on a database table format that is.

- Column names in the first line with a separator between the names
- The values in the following lines with a separator between the elements
- No garbage inside the structure, note that there can be garbage before and after the structure, but not inside the columns names and rows.

The TextFileWrapper opens a text file and parses out the information based on user inputs. The procedure is:

1. Select and open the file

2. A wizard is started which display the file and asks the user for the necessary information for the file to be parsed, line and element separator, where the column names are, start and end of the values. See picture below for screenshot.



Picture 12: Screenshot from the tools text parser, showing an Excel file saved as text about to be parsed.

3. The TextFileWrapper then parses the file and copies the data into arrays, column names are copied into one array and the values into an 2 dimensional array[23]
4. The DHIS import tool displays the result and asks the user to verify the result
5. If all is correct the user can proceed to the next stage

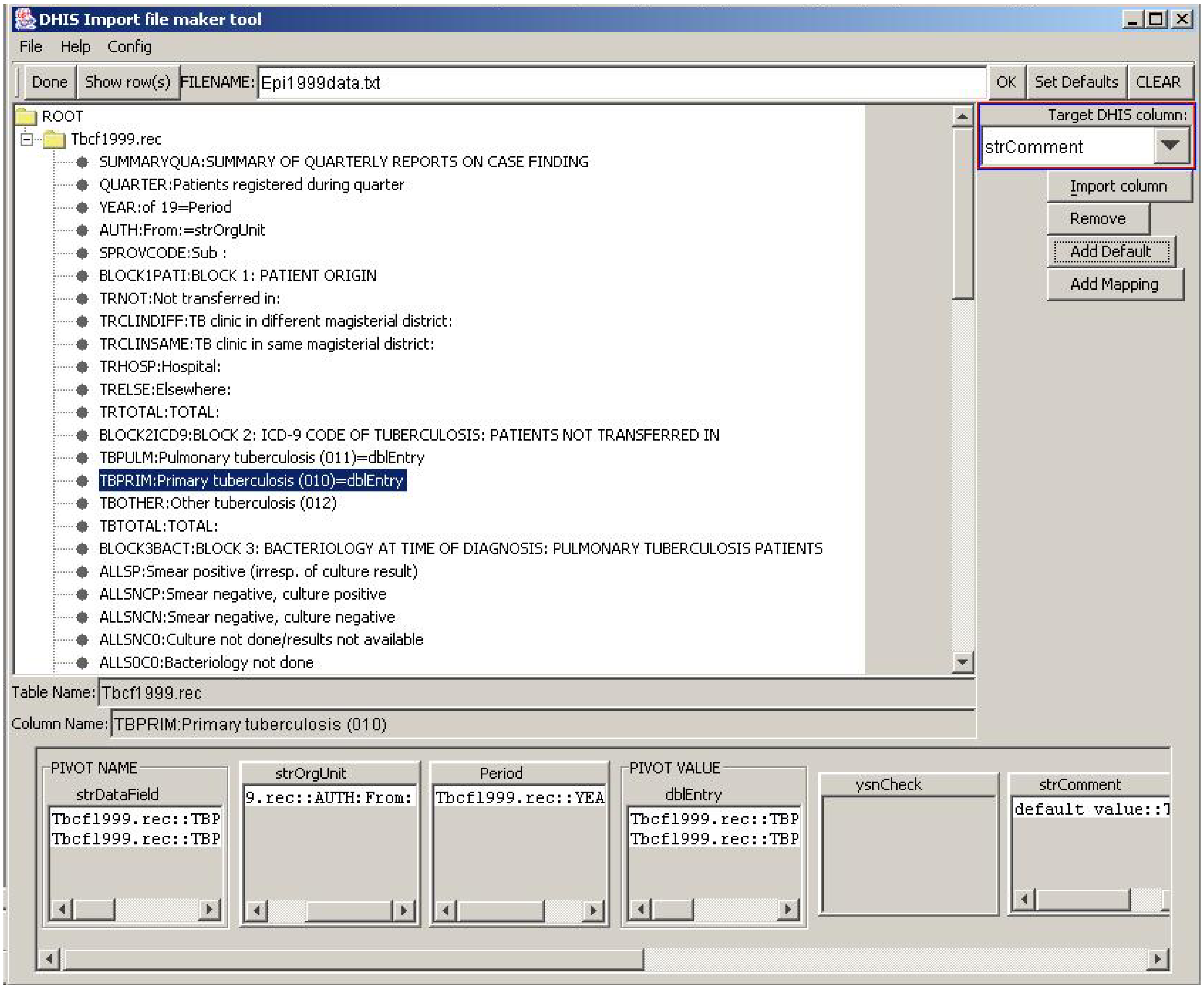
The user entered information for parsing the file, such as element separator is stored in a text file type object. This is saved and read from file once the tool starts and ends. Consequently, similar files types can be parsed by simply opening them and selecting the corresponding text file type.

Note that this is not a perfect solution for spreadsheets consisting of several sheets as only the first sheets are possible to parse. However, the user can parse the other sheets by saving each of them as text files. Note a sheet is not the same as a page; sheets can consist of unlimited pages. In Excel different sheets are selected with the tabbed panes in the bottom.

8.2.2 Selection of data to import

After a data source is opened and the selection of the type of DHIS import file has been done (only monthly data supported). A new screen is shown, see picture 12 on next page.

[23] All data copied is stored as String, though all wrappers have the possibility of returning objects for further expansion of the tool if wanted.

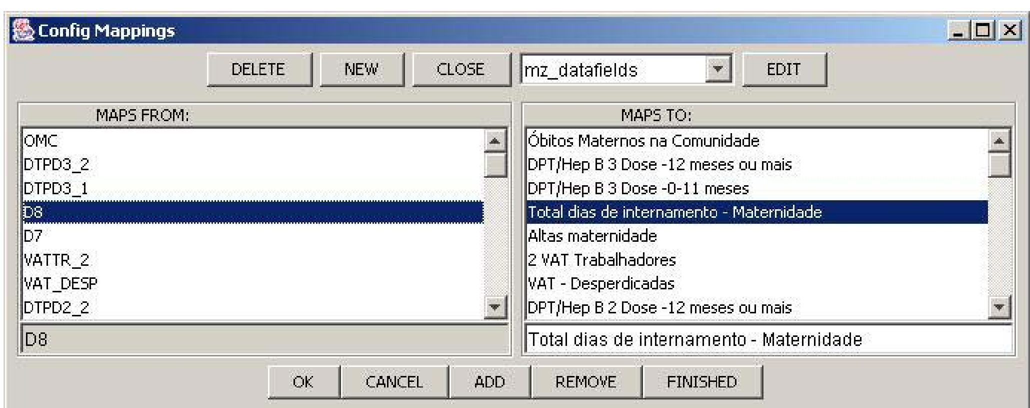


Picture 13: Screenshot of the DHIS import tool main work window, here data for import are selected from the tree to the left, then mapped and eventually default values are added, in the end the import file is named and made.

All the boxes in the bottom of the screen (strDataField, strOrgUnit etc.) makes out the DHIS import file, one box for each DHIS import file column. Selected columns of values from the data source are copied to the corresponding column in the DHIS import file, as shown by the boxes containing text in the screenshot. These are DHIS import file columns which have had data copied to them from the data source. Not that specific selection of data from the data source only relates to selection of entire columns of data, and not individually rows of data. The tree on the top left display the tables and underlying columns from the data source. In this case the data source is an EPI file called "Tbcf1999.rec". In the data source tree all nodes except the root node is a table, underlying leafs are columns in the table. In the example above there is only one table as EPI files consist of several columns in one table; in such cases the table get the name of the file. If this had been a database there would have been several tables under the "Root" level, each having their columns as leafs. The selection of data to move to the DHIS import file is done by highlighting a column in the data source tree (i.e. clicking it, see the above example). Then setting the combo box “Target DHIS column” to the DHIS column you want the data copied to. Then click the “Import column” button. If everything is done correctly a DHIS column in the bottom screen gets a line with the name of the table and the column name. As shown with for example the DHIS column “Period” have had table “Tbcf1999.rec” column “Year:of 19” moved to it. Also note that the source column “Year:of 19” have added “=Period” to show the movement. Only a single column can be moved at a time and not entire tables. Furthermore, users must be sure that the values in the data source column are of the same type as the target DHIS import file column. For example, moving a String to the DHIS column “ysnCheck” would generate errors when the DHIS import file is made. Due to that “ysnCheck” only accepts Integer (0 or 1). To undo a movement select a text element from anyone of the DHIS import file columns list and click “Remove”. To view the content of the data source select a column or a table and click “Show Row(s)”.

8.2.3 Set up mapping and add default values

As explained before the DHIS import file might contain columns which have no corresponding columns in data source. Remember the only required values in the data source are Datafield, OrgUnit, Period and Entry. If this is the case the tool can add default value to any of the DHIS columns, such as 0 for “ysnCheck”. To do this set the combo box “Target DHIS column”, to the DHIS import file column that should have a default value. And click “Add default” and follow the instructions on screen. Or click the “Set Defaults” to set all the extra DHIS columns to standard default values, such as last edit time are set to the current system time. The default values are also shown in the DHIS import file columns list, as shown with the “strComment” in the above example having a default value.



Picture 14: Screenshot of the mapping configuration dialog from the tool

After all the DHIS columns have been set its time for the mapping of values into the respective DHIS names, or the DHIS data dictionary. While implementing and using the tool in Mozambique I quickly learned that DHIS already stored the target mapping. Therefore I extended my tool with the possibility of fetching original DHIS names from the DHIS database. Due to my object oriented approach this was a trivial modification[24].

In order to map the "Target DHIS Column" must be set to the DHIS column to map, then click "Add Mapping"; this launches several new prompt windows and work windows with the following flow.

1. Select an old mapping or make a new
2. If an old mapping is selected the program initialized a mapping screen with the mappings selected. If new mapping is selected the program makes a structure to store the new mapping and initialized the mappings with nothing. The mapping is basically two list displayed opposite each other. The left list is a list of the original data source values, and the right is what they currently map to, if they do not have any mapping then the original values is shown, i.e. left equals right. See picture 14.
3. The user can select a value to map by clicking it and then entering to what it maps, or select a value from the corresponding DHIS names which are displayed as Combo Boxes.

When the DHIS import file is properly set up it must be built and saved. This is done by entering a filename in the topmost text field and clicking "OK". If everything is ok the file is made and stored in a folder called "importfiles" located in the folder the tool is installed. The screen then initialize back to the original screen. If finished the user should select "Done" and quit the program. When the program quits all mappings and text file types are written to disk, these are also read when the program restarts.

For more about the tool see the CD containing Java SDK 1.4.1 and the jar file (a jar file is a Java virtual machine executable file). Also see the user manual in appendix I and on the CD which also explains how to start it.

During my prototyping and use of the tool in Mozambique I noted several extensions which would have been nice, some also necessary to have, these are explained in the following section.

8.3 Import of data

I will in this section describe my findings in the process of importing the monthly data and related processes. An important part of this process is the flexibility of DHIS. Can DHIS be adapted to the new organization, both its

[24]Due to my limited experience with large programs the code and structure got better in the later classes I made. And the classes that display the data source tree were made late. As such in order to fetch customized columns content from DHIS I just needed to create a new data source tree with DHIS as the data source. Note the actual tree is standard Java Swing class extended with functionality, no low level graphical coding has been done.

report and organizational structure? Therefore I am separating this section in two main parts, Initializing of DHIS and import of reported data.

8.3.1 Initializing of DHIS

The HISP team in Mozambique have been active for 3 years. And have during this time trained the health staff and installed DHIS on several computers in the districts and provinces. Included here is translation of DHIS to Portuguese and making a data dictionary for Mozambique. In addition lots of the infrastructure data have been entered into DHIS, focusing on the three test provinces: Gaza, Niassa and Inhambane. During this process there have been added a dummy layer to the organizational structure in DHIS for Mozambique, this as DHIS is looked to 5 layers, while Mozambique have 4 layers, facility, district, province and national.

A mapping was necessary as facilities in SISC are identified by a unit and a district code, while DHIS uses the facility name. As a result of the mapping discrepancies and inconsistency between the infrastructure data already in DHIS, and the SISC infrastructure were discovered. The discrepancies were mutual, i.e. DHIS stored facilities not in SISC and vice versa. Inconsistencies were found for the facility properties stored in both DHIS and SISC. As a consequence, some of the HISP team members started on a process of obtaining a correct infrastructure in DHIS representing the reality for the Gaza province, basically obtaining a list of all health facilities in the Gaza province. However, as will be showed in the chapter “Investigation of existing systems”, the infrastructure is stored in all computerized systems investigated; Spreadsheets, SISC and BES. And there are discrepancies and inconsistency between these systems. Therefore, the only solution to find a correct infrastructure was by close work with the provincial staffs, and it took two evenings works with the SISC responsible person in Gaza to compile a correct infrastructure in DHIS for Gaza province. Once concluded several ghost facilities in SISC were discovered, a ghost facility is a facility which is closed or merged, but still is stored as being active. However, when all SISC facilities where mapped to the new DHIS infrastructure, several of the supposedly SISC ghost facilities were still reporting data in SISC up to that date. For instance in Gaza there is a facility called “Mapai-estacao”, which have been reporting data from January 1997 to February 2002 (i.e. up to this date). When asked why these facilities were reporting and what they did in the system no one seemed know. Later on a visit to a district headquarter it was discovered that there were discrepancies between the new DHIS infrastructure for that district, and the facility list stored at the district, including wrong names and /or missing facilities, wrong type of units (post/center). This indicates bad communication between the district and the province, and that the province does not have a good knowledge of their underlying structure. However, someone at the province headquarter might still have a paper list, but this seem unlikely due to that the HISP team worked there for several weeks and did not find any. Consequently in order to obtain a correct and complete view of the infrastructure all districts at least must be visited. Note there might also be discrepancies between the stored infrastructure at the district and the data at the actual facilities. However, this has not been investigated.

The problem of multiple stored infrastructure data was also observed in Inhambane provincial headquarter where they had the infrastructure stored in an Excel file in addition to the existing systems also storing the infrastructure such as BES. Note the population of the infrastructure data in DHIS is also a part of data import. But I did not focus on that part therefore it was done manually.

8.3.2 Import of monthly data

Import of monthly data from SISC Gaza was begun after obtaining what was believed to be the correct infrastructure in DHIS, i.e. the list from the provincial headquarters. Both my general tool and a specific tool were used, in addition another HISP team member visited Mozambique a while after my stay and made an on they fly solution by using existing tools. I will first discuss the problems associated with the actual import of data, then presents findings about the different tools.

The SMI and PAV data was selected for import as it represented the minimum of what all facilities were reporting in SISC. It is also deemed as the most important. Data from SISC facilities deemed as not active today were left out as the newly compiled DHIS infrastructure list only stored the active facilities. Only data from 2000, 2001 and up to March 2002 was imported, older data was left out. The actual year values in the SISC rows were sometimes found to have strange values; therefore it was selected to use the name of the database as the year. A common problem in all SISC tables was the excessive amounts of 0 as SISC uses 0 as default value. But this was impossible to solve without checking all the original paper forms, therefore the 0 was imported into DHIS as well. Mapping of the infrastructure was fairly easy as the unmapped SISC facilities were left out and deemed as ghost facilities. The mapping of the SISC elements into the DHIS data dictionary proved difficult, although the actual decoding of the SISC column had been done. The reason for the problems was that the SISC data elements were found to be ambiguous, in addition to this there were several strange reporting rules and a general lack of documentation.

The SMI data in table B06, B07 and B08 was the easiest to import as it is aggregated to the DHIS level, which is pr. facility pr. month. Therefore it was not necessary to alter the DHIS or the SISC data. However, as will be presented in the chapter “Investigation of existing systems” the form B06 has been entered with some strange values. Such as all visits in January counted as first, while in the other months it is separated between first and second visits, or follow up visits. It was decided to not try to clean up this as it would require much work. Including travel to all the district/facilities to get to the correct data in the paper forms. This as the reporting routine of reporting all visits as first in January, while separating between first and second visit in other months could only be solved by visiting the facilities and gaining access to the original reported data. Though, the error might actually steam from routines in the facilities, therefore impossible to correct. Given the fact that there is more than a 100 facilities in Gaza alone, some having very bad roads or no roads would make this a huge undertaking. Other wrong data in SISC could have been corrected automatically. An example of this is a poor breakdown of children in one of the forms when entered into SISC. More specifically the SISC element "1-4 years" is entered as form element "0-11 months" + form element "1-4 years". Then form element "0-11 months" is entered as SISC element "0-11 months", this is bad as the group "1-4" now contain "0-11 months" and "1-4 years". This could have been automatically corrected by making an SQL sentence on the SISC original data. DHIS Visits 1-4 years = SISC column (Visits 0-4 years) – SISC column (Visits 0-11 months). This was not done and the data was instead mapped to DHIS element "Visits 0-4 years", leaving the "Visits 1-4 years" empty. It was left as it was due to that there were no requests to fix it, and it could have been dangerous to fix it if the users knew about the bad age group, and were taking actions to compensate for it. This error is due to poor SISC entering routines and will be looked into in the next chapter

Once it came to the PAV data in table A04 it was more difficult. Firstly, table A04 have data on pr. district, pr month basis. This meant that DHIS would not be able to import the data as it only allows facilities to report data. This was solved by creating a dummy facility in DHIS for all districts in Gaza called "PAV_District name" to which all PAV data was mapped. Secondly, form A04 has changed significantly, while the SISC database has remained unchanged. This has lead to ambiguous data elements and discrepancies between the rows (see section Chapter "Investigation of existing systems"section "Forms have changed and SISC lack of flexibility"). Therefore it was decided that in July 2001 all district had changed to the new form. This meant that there were two data elements mappings for A04, one pre July 2001 and one past July 2001. However, this did not solve the problem perfectly as the shift in July 2001 obviously was not the case. Furthermore; the districts had coped with the change of forms differently, leaving different columns empty.

8.3.3 The different tools

There have been used three different methods in import of data into DHIS in Mozambique up to this date (December 2002).

1. My general tool
2. Lungo's specific tool
3. An ad-hoc solution made with existing tool (such as SQL)

Looking firstly at my and Lungo's specific tools. Once finished developed the actual import process using the two tools is quite similar. However, the development time for the specific tool was much shorter as it had to cater for less file formats and problems, 3 weeks by a skilled Visual Basic student, compared to 3 months for the general tool, partly due to the author's lack of experience in developing GUI and large Java programs. The specific tool was made in Visual Basic which easily integrates with other Microsoft formats, while integration Java with Microsoft formats proved more difficult. Furthermore, defining the functionality for the specific solution was easier as it was discovered as the research went ahead. Short cuts could be made in the code to cope with particular problems, for instance having the year in the DHIS import file pre filled out as the name of the database. Short cuts such as these are difficult to make in a tool which is to cope with a large variety of different problems. However, actual usage of both tools requires users who fully understand the back end of the source systems.

In order to make a general tool in a feasible time frame I had to set several prerequisites regarding the data sources, as explained in the manual. For instance the fact that a facility is identified by two columns in SISC made it necessary to make a view on the SISC columns, combining the two columns into one, this as the general tool only supported movement of single columns. This meant that almost all data imported from SISC had to be pre processed before it could be imported. While the specific solution did not need this amount of pre processing, although all data had to be imported into an Access database but this process is simple compared to making SQL view or other source editing.

2-3 months after my research all data was re imported into DHIS as there were discovered gaps and discrepancies between the SISC and DHIS data. After investigation, I learnt that the discovered errors were primarily the missing records from the unmapped SISC facilities. The fact that the HISP Mozambique team so soon wanted all data transferred, in contrast to the decision made to only import data from active facilities. Shows how unclear the requirements were during the project, furthermore, the decision not to import data from unmapped facilities had been little assessed. The re import had be done by the HISP software coordinator using SQL quires and an anti-pivoting tool, a general tool made for the DHIS. All records for Gaza from 1997 to 2002 where imported, and the unmapped records were put in a temporary table for further investigation. This process also discovered some extreme values entered into DHIS several years ago, which probably was forgotten and not removed. However, once DHIS was populated and used it surfaced and made the data bad.

8.4 Summary

In order to enable automation of parts of the import process a tool was developed which supported making DHIS import files. The tool connects to multiple data sources and is easy to extend to cater for new types of data sources, in addition it allows for customized selection and mapping of data and adding of default values for DHIS. An actual project of importing data from SISC was described; problems with inconsistency, discrepancies, ambiguous data and poor reporting routines in existing systems surfaced during the project. This combined with difficulties in obtaining a correct infrastructure in DHIS made the import project time consuming. The import of SISC data allowed before hidden data to surface triggering an analysis into the data described in the next chapter.

9 Empirical work, Investigation of existing systems

In this chapter findings from research into four of the existing systems are presented. The first and most extensive is SISC; an important part of SISC research is analysis of before hidden stored data compared to information from interviews and observations, this in order to understand the use of SISC. Afterwards I am presenting my findings from BES, a program specific system. Then the use of spreadsheets and its content is presented. In the end I will present findings about the SIMP program, an integration tool currently prototyped within the Ministry of health.

9.1 SISC and its use

Most of the inner details of SISC have already been explained in chapter 7, "Decoding the existing systems", the actually import process described in chapter 8 allowed the before obscure data to become available. In this section I am going to look at actual use of SISC and the data stored.

9.1.1 Planned use of SISC

In this section I will describe how I believe SISC was planned to be used, I say believe because SISC is 10 years old and it is difficult/impossible to find information about how SISC was planned to be used.

Provincial

The forms are to reach the province each month and be entered into SISC. Tools in SISC should be used to control that all districts are reporting. In addition SISC tools should be used for analysis and management of the health care.

Communication between province and national

Each quarter the province is to make a backup disk of its recent data and send this disk to the national level.

National

The national level is to receive a disk quarterly from all provinces the recent provincial data; this data is imported into the national SISC database. At the national level they have the ability to make quarterly analysis and reports, even though the data stored support the ability to make monthly analysis. In addition to provide the national level with data, the national SISC database is to function as a back up for the provincial databases.

9.1.2 Actual use of SISC

In this section I will describe how I have found that people are actually using SISC. It is based on interviews and visiting in the Gaza province and Ministry of health.

Provincial level (Gaza and Inhambane)
The province receives the paper forms from their district at the beginning of the following month, and then dedicated SISC information staffs enter the data into SISC. When entering data the only control observed is that the entry staffs are checking that the values they enter are the same as the values on the forms, there were no signs of checking the validity of the data. Presumably, due to bad knowledge about SISC the functionality of finding district which has not reported is not used. In general the SISC staffs seemed to have very little knowledge about SISC. When asked to "Reindex" the database, the staff had to be told how to do it as it had never been done. "Reindex" is something that was to be done regularly according to staff at the national level. I also heard rumours that when ending SISC after data entry, it forces the user to make a floppy backup; this was supposedly by some users avoided by simply turning the computer of.

An interesting observation was that SISC was installed on the oldest computer observed in the provincial headquarter, with a new computer standing right next to it. This was due to of lack of knowledge and support; staffs do not know how to install SISC on the new computer. Even if it was possible the old data had to be moved. This would require skills as the data is too large to be moved on a floppy, which are the only external storage devices available. For instance moving some tables at a time and manually rebuilding the database at the new computer.

The graphical analyzing tool is not used due to its age. The graphs being made by SISC is very simple and very difficult to export to other presentation software, such as Word, Excel etc. The only way to move it is to save the screen with "print screen", and then move the picture into the presentation software. This requires a certain degree of knowledge and skill with the computer which is not present at the provincial level. This combined with the poor quality graphics and old user interface makes it little attractive to use. Instead, analysis of data is done by printing out the data from SISC on paper; and then re-enter the data in a spreadsheet, usually Excel, on a newer computer.

Communication between provincial and national
The planned communication between SISC province and SISC national were found to be broken down. The disks are not arriving regularly, in the Gaza province I were told that they seldom sent the disk, usually they only sent the print outs from the Excel spreadsheets. The disks were only sent when the national level called and asked specifically for the disks. This was also what I was told from the national. However, one person at the national meant that the

disks were coming regularly and that the national database was complete in SISC. In contrast, print outs from SISC and Spreadsheets from the provincial level are arriving regularly each month at the national level.

National level
The national uses the print outs to renter the data in spread sheets for analyzing purposes and making reports. Even though the disks are not arriving regularly, they do arrive sometimes, leading to a incomplete SISC national database. Note, the data is complete at national level, but it is in spreadsheets and on paper, the SISC database is not complete. As was the case in the province the computer at national level running SISC was very old, looked almost liked it was stored away and rarely being used. The other computers observed at national level were newer. For example, SIMP and BES were running on fairly new computers running Win 2000 or Win XP.

An interesting comment was that the national database was only kept alive and pretended to function from a political point of view; almost everybody knew that it did not work.

9.2 SISC content analysis

I will now present the result of my analysis of the content in SISC. Through this chapter I will concentrate on some databases due to the amount of data. The databases I will analyze are:

- GAZA provincial database year 2000
- NIASSA provincial database year 2000
- National database (all provinces at national level)

I will also only look at data related to vaccination (PAV), child and deliveries, prenatal (SMI) and Maternities (MAT).

The reason for this focus is:

- I had the provincial database for Gaza and Niassa
- Vaccination, child and deliveries, prenatal and maternity is the most complete and important data sets
- The year 2000 seems to be the latest most complete year

My content analyzing will firstly deal with the coverage and the stored infrastructure, after this I will check the data for quality (validation).

9.2.1 Coverage

Gaza provincial database year 2000

As previously explained the infrastructure in SISC stores all facilities and the kind of data they are supposed to report, PAV, SMI and MAT. In order to compute the number of reports necessary for full coverage, you multiple the number of facilities which are stored as reporting the corresponding data set with 12, each reporting once a month. This was done for A04 (PAV), B06 (SMI), B07 (MAT) and B08 (SMI), and then the number was compared to the numbers of actual reports. This is found by counting actual reports, or rows of data in the corresponding tables A04, B06, B07 or B08. When I did this I got a very low coverage (ca. 70 %), this low coverage was caused by several facilities which never reported the data sets they were supposed to report.

CASE: According to table UNIDS Gaza have 78 number of health facilities who should report maternity data (MAT). But table B07 (Maternity) only have monthly reports from 69 distinct health facilities. Consequently 9 facilities are never reporting any maternity data the entire year, wherever these facilities have closed down completely, or closed their maternity facility, or just have stopped reporting is difficult to discover. Experience has showed that to learn about which facilities that actually exists you need to go to the district level.

I checked for SMI and MAT in the infrastructure and compared to the number of monthly reports in B06, B07 and B08 (the respectively forms for SMI and MAT).

- 90 health facilities are registered to report SMI data.
- 78 health facilities are registered to report MAT data
- SMI data set. In 2000 81 health facilities are reporting B06 (SMI), while 83 report B08 (SMI), all numbers lower than the 90 supposedly facilities to report
- MAT data set. In 2000 69 health facilities are reporting B07 (MAT), lower than the 78 supposed to report.

Conclusion is that several facilities who are supposed to report SMI and MAT based on the stored infrastructure data are systematically never reporting it.

I wanted to see if the same was true the other way, are facilities who are not registered to report a kind of data reporting that data? The answer was yes. Health facility “Post De Saude Tuana” (DCOD and CODI = (01, 12), in the table UNIDS (which is storing the infrastructure) is registered as supposed to report Maternity data and not SMI data. But the health facility exists in B06, B07 and B08; B07 is the only Maternity table, B06 and B08 are for SMI data only. I addition I have copies of the respective paper forms from Bilene where Tuana is reporting on B06, B07 and B08, see appendix D, E and F. The conclusion is that SISC does NOT check if a facility reporting is supposed to report the data, furthermore SISC infrastructure is not updated on what the health facilities actually is collecting and reporting.

- Facilities supposed to report are never reporting
- Facilities NOT supposed to report are reporting

I was interested to see if this had always been the case; therefore I compared how the development had been over time. I checked the infrastructure data for facilities supposed to report SMI vs. how many are reporting form B06 each year, the results is shown in the figure below.

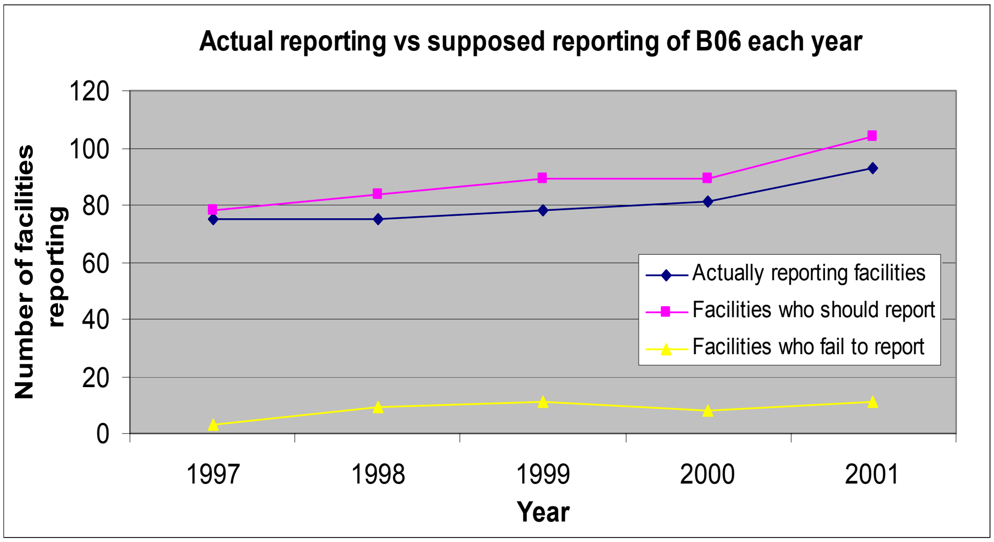


Figure 22: The above graph shows how the actually reporting units evolve over time compared to the infrastructure, "facilities who should report" is facilities who are registred in the infrastructure as supposed to report. Facilities who fail to report are facilities who completely fail to report the entire year.

The graph shows that the amount of facilities who failed to report according to the infrastructure was fairly small in 1997; it has increased thereafter. A hypothesis might be that as time passed, the infrastructure changed in the real world, but SISC was not changed. This would explain the small difference in 1997 while bigger later.

Maybe the infrastructure had never been updated? Note with updated I mean altered their status, as the graph above shows that new facilities have been added. To check for that I compared the infrastructure for Gaza in 1997 to 2002; searching for changes in the infrastructure. My conclusion was:

- There had been added 16 health facilities and removed 1.
- None of the 126 facilities had altered their PAV, MAT or SMI status in 5 years, note that my check would not have captured facilities losing and gaining their PAV, MAT or SMI between 1997 and 2002. I simply checked the database in 1997 vs. 2002.

I can not tell if the small changes in PAV, MAT or SMI status is plausible as I do not know how often these things are supposed to change. But based on actual facilities reporting it looks wrong, it should have been changed for several facilities.

To further check for updates in the infrastructure data I used the population census data, made in 1997 for Mozambique. Had the new census data been moved to SISC? After checking the population count in SISC it was discovered that this had happened. As shown by a distinct population drop in 1997 to 1998 in SISC, the new population count were similar to the population census data. For instance the SISC stored population in Gaza dropped 34 % while Niassa dropped 20 % between 1997 and 1998.

This shows that the infrastructure data is partly updated. And for the continuing analysis I will not use the infrastructure data to find number of reporting facilities. As I believe it is wrong, I will instead use this rule:

If all health facilities reporting in the tables had been reporting perfect each table would have the following number of records: number of distinct health facilities reporting * reporting once a month = Calculated reports in the table below. Note this coverage factor calculating is not affected by facilities never reporting, meaning that the coverage factor is probably overestimated.

Table	Calculated reports	Actual count	Coverage % (calculated)	Coverage % (based on SISC infrastructure)
B06	81*12=972	826	85 %	76 %
B07	69*12=828	694	84 %	74 %
B08	83*12=996	819	82 %	76 %
A04	12*12=144	127	88 %	NA

NOTE: Coverage in Gaza 2000. Note the big difference in coverage between SISC infrastructure data and calculated. The A04 coverage factor only show districts reporting and the real coverage factor might be lower or higher depending upon actually reporting facilities.

As the coverage table shows the coverage figures is not perfect, but regarding amount of resources available it is not very bad I think.

Next I tried to find out if some particular facilities were bad of reporting data in general, and if that could explain the coverage factor. I grouped the facilities reporting form B06 according to how many months' a year they reported. 38 facilities reported for all 12 months, 11 facilities reported for 11 months etc. as the graph in figure 24 shows. Some facilities were very bad of reporting, but most of the missing months come from facilities missing only 3, 2 or 1 month(s). 3 facilities are only reporting for 1 month of the entire 2000, and this pull the coverage factor down. But overall I think it looks okay. As the data was for 2000 and in Gaza it would probably have been influenced by the large flood there in the beginning of 2000. To find out that I compared number of facilities reporting in 2000 to 1999; were there some particular months in one of the years that were distinctively bad?

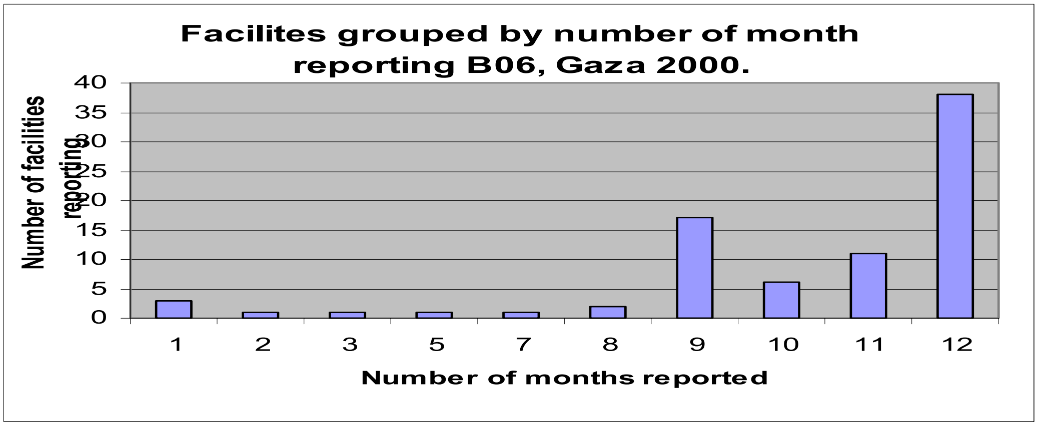


Figure 23: Showing number of reporting facilities grouped by number of months they report

The graph in figure 25 below show that in February 2000 there was a significant drop in reporting facilities, in March and April 2000 it is slowly returning to normal. This fits well with the time of the flood in 2000. The rest of the months in both years are looking fine with no significant bad months. Note that in no months are all 81 facilities supposed to report reporting.

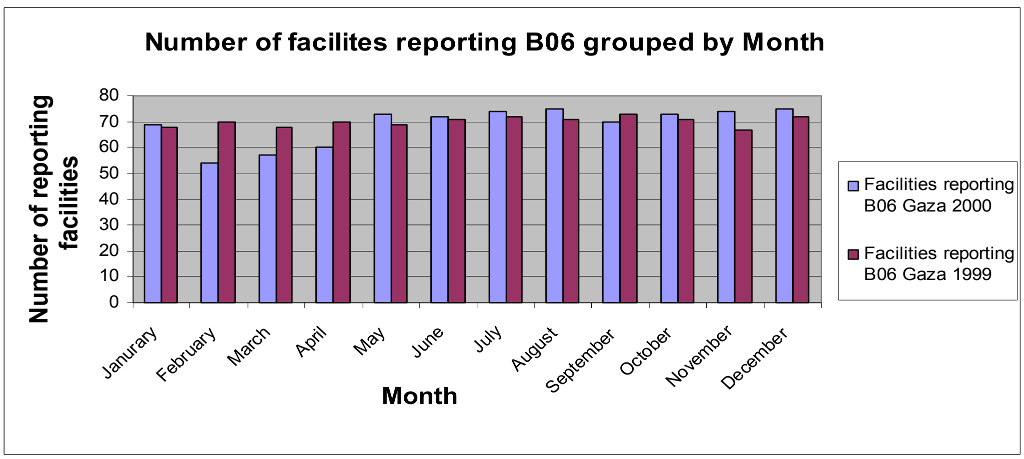


Figure 24: Showing number of reporting facilities grouped by number of months they report, comparing 2000 to 1999

Vaccinations data coverage (PAV)

For vaccinations data in A04 it is not possible to calculate any coverage factor as data is aggregated to district (all health facilities are summoned), which means that details about which facilities are reporting is lost. However, variations in data reported can point to varying coverage, but this is not certain as the variations might be due to other causes, such as bad reporting, extreme values due to external causes etc.

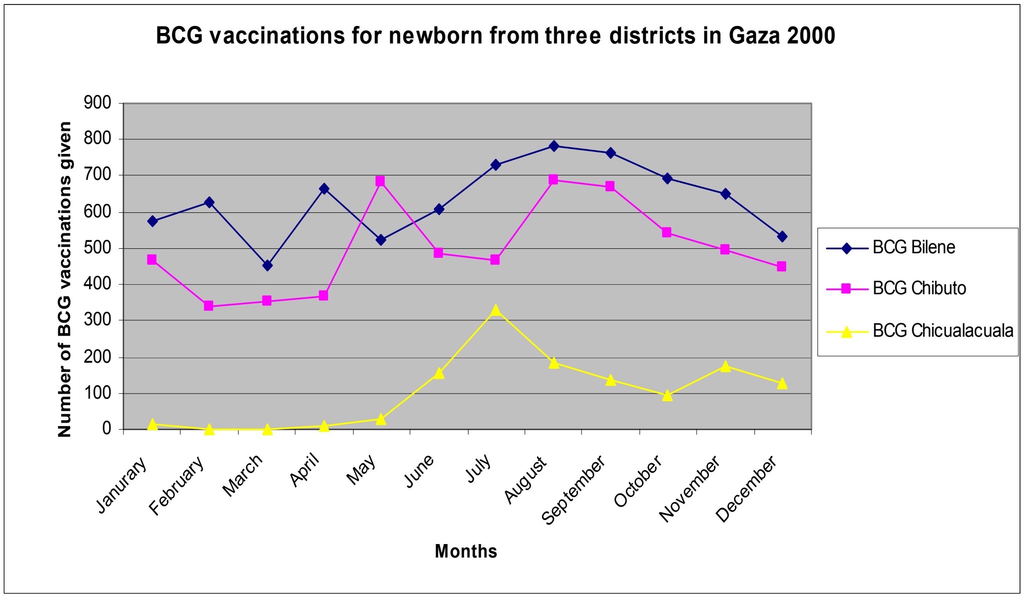


Figure 25: Showing three districts vary with their reporting of BCG for children 0-11 months. The varying between the districts is probably caused by the varying population of the districts, se figure 26 for more.

As the above graph shows, BCG 1 vaccinations vary a lot during a year for all districts. Especially Chicualacuala district looks very bad for the first 5 months; this is probably due to its location on the banks of the Limpopo River which flooded in the beginning of 2000. To investigate this, I checked if any facilities in Chicualacuala district had been bad at reporting form B06 during any months in 2000. But that was not the case, for instance in March there were 4 units reporting form B06, which was above the average units reporting B06 (3.6 each month). Thus I did not find any significant correlation between numbers of facilities reporting form B06 in Chicualacuala district and the low amount of BCG vaccinations given the first months in Chicualacuala. This case illustrates two problems with SISC.

- Aggregation of vaccination data to district in SISC makes coverage control difficult. In order to do a coverage control visits to all the districts would be necessary to check the original tick sheets, one for each day, in the archives, as vaccinations data seems to never be stored pr. month pr. facility.
- Lack of support for null in SISC. Are the 0 vaccinations in February and March caused by actually reporting of 0 vaccinations (due to running out of vaccinations in all facilities for example), or just by facilities or the district not reporting those months due to the flood? For example, the facilities can all report, but the district is failing to report. This is difficult to find out as SISC uses 0 to denote both 0 incidents and no report, or unknown, which should be "null".

The conclusion is that calculating coverage factor for PAV is impossible due to data aggregated to district. However, big variance in data can point to varying reporting health facilities.

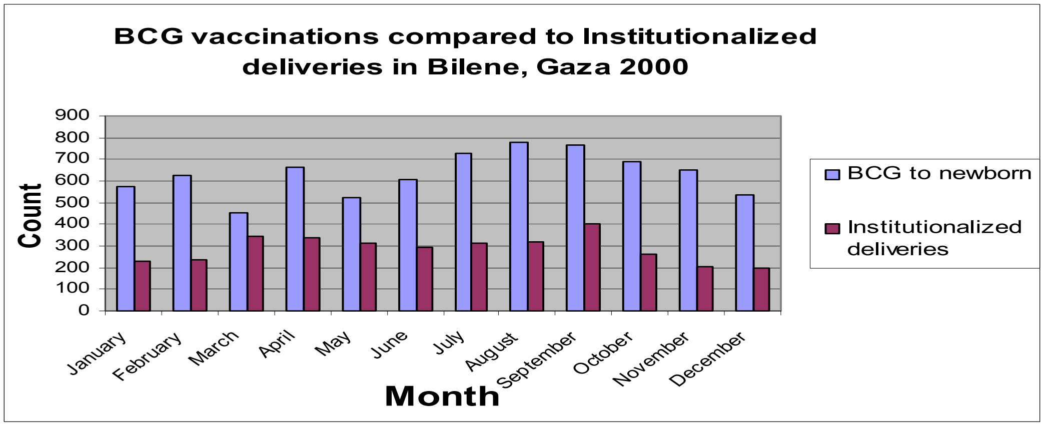


Figure 26: Showing BCG vaccinations (0-11 months) vs. institutionalized deliveries

The above data comparison is difficult, if not impossible to make in SISC, as BCG and deliveries are aggregated differently. However, often this comparison is good to show how many newborn are vaccinated for BCG. As the above example shows there are always given above twice the amounts of BCG vaccinations compared to deliveries in the facilities, this might be due to over reporting of BCG, under reporting of deliveries or the fact that few deliveries are done in the facilities. The latter is most likely the cause as [WHO MZ] states that only 30 % of the births have skilled attendance. In order to check reported BCG against other population sources I checked the [MZ CENSO 97] for Bilene. I found that in 1997 there were reported 4 440 children below 1 year in Bilene. Summon up all BCG to newborn in 2000 in SISC equals 7600. However, as showed in the background chapter the death rate among kids below 1 year is very high. Therefore the number of BCG vaccinations might not be wrong. Also remember the coverage problem regarding vaccinations data.

During these test I have been working with Gaza 2000 which was severely hit by the flood in 2000. In order to check if my findings were particular for that year and province I did some tests on the Niassa database for 2000

Niassa provincial database year 2000

Is has been shown that SISC does not check the infrastructure data when accepting monthly data; therefore I did not check it any more.

Calculated reports = Number of different health units reporting * reporting once a month.

Table	Calculated reports	Actual count	Calculated coverage %
B06	100*12=1200	894	75 %

B07	87*12=1044	771	73 %
B08	93*12=1116	885	79 %
A04	16*12=192	188	97 %

As the table and graphs shows the coverage was lower in Niassa than in Gaza for 2000. I did some similar data analysis in the Niassa database to find bad months or facilities.

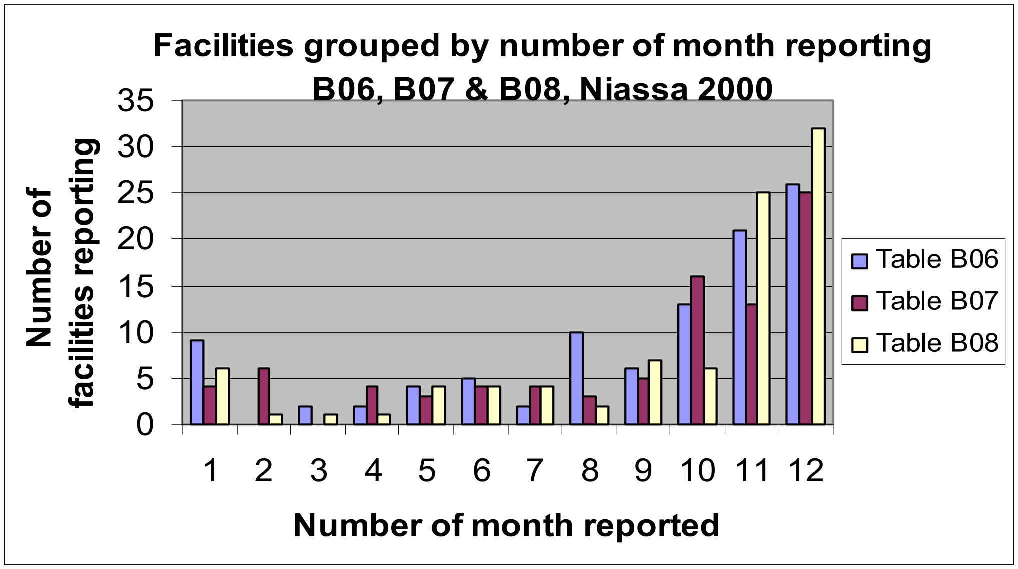


Figure 27: Showing facilities grouped by number of months actual reporting B06, B07, and B08, for comparison to figure 24 and Gaza use B06.

In the above figure the tables B07 and B08 has been added to show that the other tables are quite similar in their reporting. The findings in Niassa are very similar to my findings in Gaza. Most health facilities report 10 or more months in a year. While some misses 1-3 months, but that is not bad regarding the poor infrastructure (roads, telecommunication etc.) and lack of resources. However, note that 8 facilities only report B06 once during 2000! How many is not reporting at all? That is impossible to find out due to the incomplete infrastructure data.

Next I checked to see if there were any particular bad months in Niassa in 2000, the results is shown in figure 29. As shown there were not any particular bad months of reporting in Niassa, this is properly due to that Niassa was not directly affected by the flood in 2000. Note again that in no months are all facilities supposed to report reporting.

As conclusion the coverage factor seems okay when using an ad hoc algorithm to find it, but it is impossible to find the real coverage factor due to missing infrastructure data and aggregation of A04.

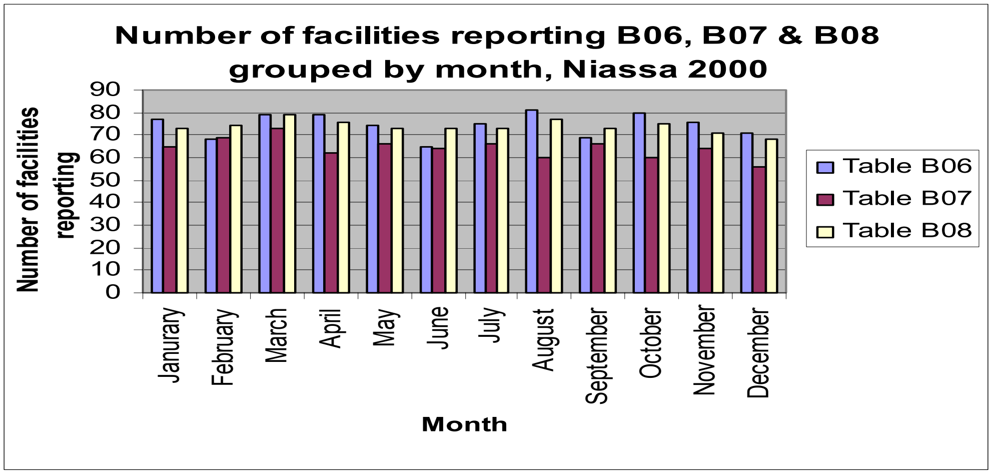


Figure 28: Showing number of facilities reporting a given month, use B06 for comparison to Gaza

Year 2000 bug

I choose the year 2000 because I had the most complete data sets for that year, in addition it was the newest data and most relevant. It was also the year that I got a national annual report for. But in case my findings were caused by the year 2000 bug I did scaled down research for Gaza 1999

Summary of Gaza 1999:

- Calculated coverage is based on number of different health facilities reporting * reporting once a month.
- SISC based coverage is based on SISC infrastructure data.

Table	Calculated reports	Actual count	Calculated coverage %	SISC based coverage %
B06	78*12=936	842	90 %	78 %
B07	69*12=828	701	85 %	75 %
B08	79*12=948	833	88 %	77 %
A04	12*12=144	136	94 %	NA

The coverage was poorer in 2000 than in 1999, but the trend is the same, some facilities are very bad while others are missing 1-3 months. In no months are all the facilities reporting.

Therefore, I have concluded with that the findings for the year 2000 I have presented are probably not related to an eventual year 2000 bug. Even if they are, the findings are still relevant as the problems are going to rid the healthcare as long as they use the current system.

9.2.2 Quality analyzing

In the last section I described and showed coverage problems, or holes in the data; I will now look at the quality of the data actually in the database.

Large amount of 0 and no "null" values

The reason for this is that SISC does not allow for the entry of unknown/not reporting for data elements (null). This mean that elements not reported on a form is converted to 0 when entered into the database. Example:

- A health facility is reporting 1 data element on a form, this value is entered into SISC
- Rest of the elements on that form will become 0 in the SISC database

Thus there is no separating between not reporting data and reporting 0, this leads to poor data such as: "Centre De Saude Melulucas", Gaza, November, 2000 have 17 births, 0 live births, 17 above 2.5 kg and 0 still births! This error is caused and not caught by the following:

- SISC leaves 0 if nothing is entered
- SISC does not control the data for validity
- No or little validity control at province or national level as the errors are in all systems.

Note that if a facility does not report a form at all then there will be a data hole as described in the last section. In addition to poor data, this also means that it is impossible to compute any coverage factor for single data elements.

Invalid data

Some health facilities are reporting data that obviously is not valid. On form B07, you have deliveries ("Partios na Maternidade") and Signed out children ("Altas"). One validity rules here could be, Signed out children<=Deliveries, i.e. a facility can not sign out more children than they have delivered. But when checking the data for occurrences of this the result was not good.

- Gaza for 2000 has 145 reports (B07) of a total 694 where signed out children>deliveries, i.e. 21 % are reporting impossible values.
- Niassa for 2000 has 99 reports (B07) of a total 772 where signed out children>deliveries, i.e. 13 % are reporting impossible values.

To check if these errors were caused by some few facilities who consistently reported wrong, I checked the content for Niassa and Gaza year 2000.

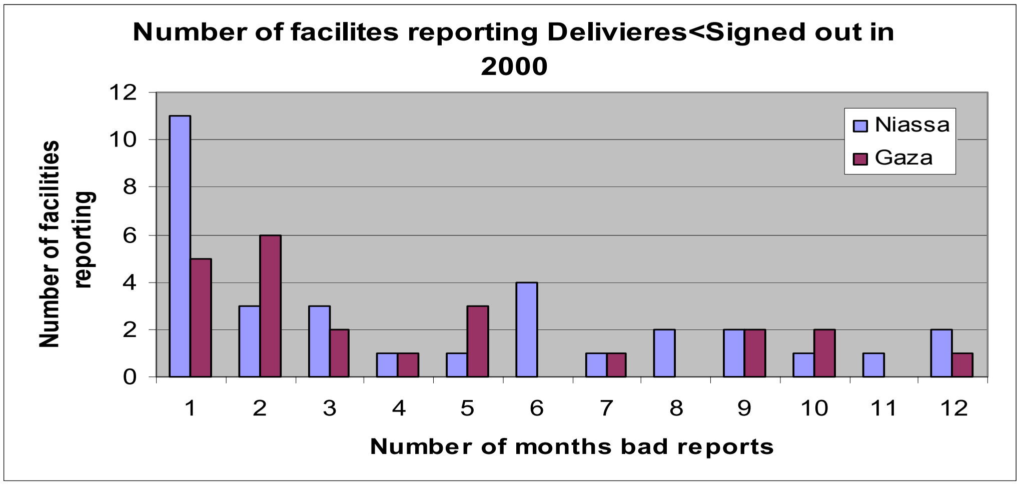


Figure 29: Shows number of units reporting invalid data grouped by number of months they have reported bad data

The graph shows that there are a few facilities who consistently report invalid data. 2 in Niassa and 1 in Gaza have all 12 months in 2000 reported more children signed out then delivered. In addition, the fact that not all facilities are reporting all 12 months must be considered, consequently there might be other facilities reporting invalid data all the months they are reporting. Several facilities report invalid data 1 or 2 months; these might be due to random errors. But again, the lack of obvious validity checks in SISC is leading to poor data being entered and reported. The fact that some facilities are consistently reporting wrong data points to lack of feedback along all levels, as the error is going all the way from facilities to the Ministry of health, where it is lost during analysis because the Ministry only look at aggregated data. For an example of invalid data reported see form B07 in appendix E.

Strange business rules

The districts are reporting the health facilities visits (Consultas) for new delivered babies on form B06. The form is broken down into visits by 0-11 months and visits by 1 – 4 years old children. These are further broken down into 1st. visit and following visits. All the fields are used, but for January ALL visits are counted as first time, there are no following visits in the SISC or the forms, though in the real world there is of course following visits in January. This makes the data look very strange and analyzing difficult. As the figure shows first visits is extremely large in January, while following visits is 0. This error is caused by human's consistently entering wrong data at district level, and the wrong data is going all the way up to the national level, no one are taking actions to avoid it. When asked why it was done it in this way, the response was that it was just something which was done. See appendix E for example on form.

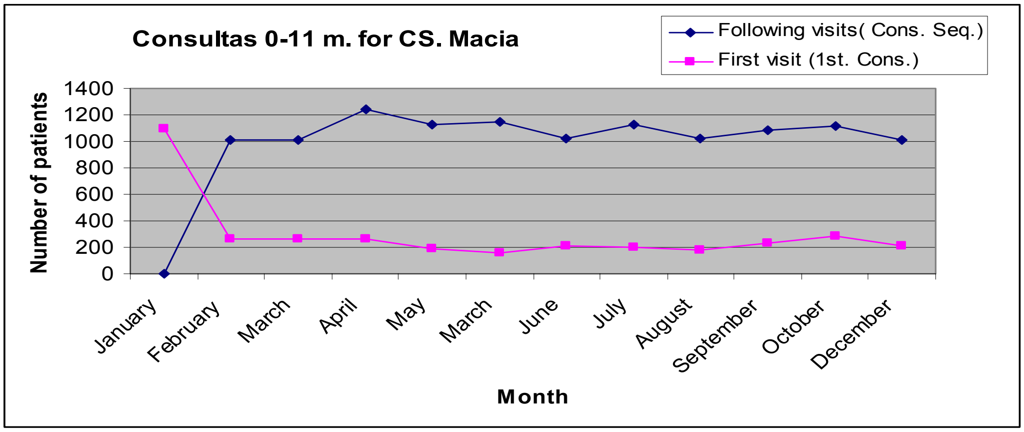


Figure 30: Plotting of visits, note the extreme values in January.

Data about number of mother visits to the health facilities are reported by the districts on form B08. This form has visits broken down into prenatal and postnatal; the prenatal is also further broken down in 1st and second visit. In SISC these visits are stored separated between first and second visit for all months, including January. However, after obtaining form B08 from the Bilene district for January 2001 it was observed that all visits were counted as 1st visit, while the rest of the months separated them. This discrepancy between data on the forms and in the SISC database is difficult to explain, as the form B08, kept at the district are supposed to be a copy of the once sent to the province. But the province must have another copy of the form, or an alternative reporting source. This as it is impossible to separate the values after they have been summed.

Forms have changed combined with SISC lack of flexibility

After SISC was made and deployed in 1994 the developer went home to Europe. Thus there was no one left who could edit the software. This meant that SISC was locked to the world of 1994. But the world has changed and so have some of the forms, particular form A04, vaccinations data, during 2001.

On the form A04, DTP doses 1, 2 and 3 have been extended to also cater for HEP doses 1, 2 and 3, new vaccinations. This change has not been reflected in the SISC system. And the users of SISC seem to have coped with it by entering the DTP data together with HEP. This is possible as DTP is usually given together with HEP. But it is not a good solution, because it makes it impossible to know when the districts started giving HEP in addition to DTP, this as SISC does not separate between HEP and DTP. In addition it is not possible to separate them in the future, a problem if a if a facility runs out of one of the vaccinations, HEP or DTP, or fail to give it, it will not be showed in the reported data as the data elements have become ambiguous. Direct contact with each facility is therefore necessary in order to discover what vaccinations are delivered.

The VAT part of form A04 has also been changed drastically. In the old A04 form you had 5 data elements for VAT, VAT 1, 2, 3, 4 and 5, in the new A04 form VAT 2, 3, 4, and 5 have been merged into a new data element called VAT 2-5. But this has not been updated in SISC, thus SISC still have 5 data elements for VAT. Most districts seemed to have solved this discrepancy between the form and the SISC database by using 2 of the VAT columns in SISC, and leaving 3 empty (with 0). However, this have not been done consistently between the provinces, i.e. some provinces are leaving the last three VAT data elements empty, using 1 and 2 while 3, 4 and 5 are empty. While other leaves the middle three VAT elements empty, 2, 3 and 4 empty while using 1 and 5. As a consequence of this, the SISC database is no longer uniform and elements can not be compared between provinces. Note the same is the case for another set of VAT elements, VATCR 1-5 columns. The changes made to data entry should have been communicated down to ensure that everyone is solving the problem similar, thus enforcing uniform data entry and storage. Based on analysis of the data in SISC, there seems to have been no flagging month when all the health facilities were changing forms, some are using the old forms all the way up to December 2001, while other are switching in July 2001. Also none of these ad hoc solutions have managed to store the new data element on form A04 called “Fully immunized under 1 year!” Thus all vaccinations data is only available on the paper form A04. I have not been able to find any documentation of the changes described above. For actual example of the difference on the forms, see appendix B which is the old and compare it to appendix C, which is the new A04 form.

No flagging day for the change of forms, and the way the provinces have solved the problem of adjusting the new forms to the old database points to lack of feedback. The result of this lack of control and feedback is that decoding the data is difficult; as it is necessary cater for that each province can have solved the problem in their own unique way.

The original form B06 had an error in its design regarding age breakdown of children visits, the first age group is 0-11 months, the second 0-4 years, note the second age groups contain the first. The second age group was supposed to be 1-4 years. This error has been corrected in later B06 forms where the two age groups are 0-11 months and 1-4 year. But the entry clerks of SISC are still entering as it was on the old form, which might be due to that they have been unable in altering the SISC graphical layout of the form, i.e. the name in SISC is still 0-4 years. As a consequence the form elements visits by 1-4 years, and 0-11 months are summed and entered into SISC under the SISC entry 1-4 years. This makes the entry values of 1-4 years wrong. To get the 1-4 years group you need to subtract the values for 0-11 months from 0-4 years. An interesting observation is that doing this error actually increases the workload for the entry clerks, just entering the data as it is would have been the smallest workload. Again, this is an error being done consistently several places; even though the national level is aware it.

There are probably more bad quality data in SISC, but I ended my search here as I am not trained in analyzing medical data.

National database
I have not used much time analyzing the national database as it is obviously incomplete. And most of the people seem to be aware of it. The incompleteness is of several different types:

- Data holes, entire months are missing in yearly databases (Gaza 2001 missing entire September and December)
- The database is completely empty (Gaza 2000)
- Entire database is missing, not even the definition of the database is there (Nampula 2000 and 2001 is completely missing)

The cases in parenthesis is examples, there are more similar errors. In order to verify that no one used the national database I checked the data in an yearly report from the Ministry of health against data stored in SISC provincial and SISC national. The conclusion was that the national level uses the province data, as they get it on paper reports from the province. This raises the question of what is the reason for using SISC today, as data only seems to be entered and then manually being copied out for actual use.

9.3 BES

BES is a computer system used for data entry, report writing, mapping and graphical representation of incidents of diseases for the provinces and the Ministry of health. As was the case with SISC, BES comes in two versions one for national (Ministry) and one for the provinces. The district and health facilities use paper forms for the collecting of BES data. The BES data is a part of SIS.

BES is made by an Engineer stationed in Zimbabwe, whom I have been unable to meet. The first version (1.0) was made in July 1995, since then several new versions have been released, the latest one which I observed in the Ministry of health was 1.6A released in 2000. BES is as SISC based on DOS, but it works better in Windows than SISC, as version 1.4 and later has been made with support for Windows installation, Win 3.1, without the intervention of a technician as the version history claims. BES was easier to understand then SISC as there is more documentation available; it also seems to be much more used than SISC. However, my focus has meant that BES has not been as extensively research as SISC.

9.3.1 BES information flow and storage

Based on my limited findings I have learnt that data is being collected on forms of type C03 in the facilities and hospitals on a weekly basis. The forms are then sent to the province where they are entered into BES. I have been unable to find out wherever the communication between the provincial and national level is computerized or on paper. And I noticed personnel entering data from forms at the Ministry of health; these were dedicated data entry staff, so it was probably not a single incident. In contrast, the BES version history states that in 1995 there was added functionality to make e-mail files for communication with the national level. There could of course be a combination, some provinces uses e-email or floppies, while other sends the forms.

Although the C03 forms are storing data on a pr. facility pr week basis, the data in BES seems to be aggregated to district level. I have been unable in locating any data stored on a facility basis in BES. However, I have only checked the provincial BES version from Gaza as of March 2002. The data is stored in tables with the key being: year, week and district name, the rest of the columns are the data elements. The elements consist of cases and death for different diseases, such as diarrhea; some are further broken down into different age groups. It rather easy to understand the data as the columns name are self explaining. Most of the data is stored in two different formats.

Text files
The text files are highly structured. The first lines declare the columns names, and then the following rows consist of data for a district for a week with a separator separating the elements. The name of the text file determines when the file is from. For example, the file named BEM96_9 contains data for 1996 week 9. This information is also stored in the file. Blank entries are stored as "– ", the few codes used in the files are explained.

Dbase files
There are several dbase tables of varying size containing different data. Some only contains system information, such as which user did what when. Other contains historical data, from 1996 up to today for the province on a pr district pr. week aggregation. As was the case with the SISC dbase, the data is coded and no entry is stored as 0, this in contrast to the text files. Even though the two systems are using dbase, there is no integration between them.

9.3.2 BES user interface and functionality

The user interface of BES is old, but more up to date than SISC as it has support for mouse for instance, a cursor resembling an icon is moved. Furthermore, BES has several more advanced analysis tools, such as GIS, Geographical information system, information maps to enable better analysis. This for diseases is important in order to control and locate the possible source(s). BES also has at some basic functionality for checking data entered for validity, such as:

- Can not enter more deaths than cases

- Can not enter data which is earlier than the current week
- Can not enter data for more than one year before

This might be one of the reasons for that the data quality in BES seems to be better than in SISC.

9.3.3 Use of BES

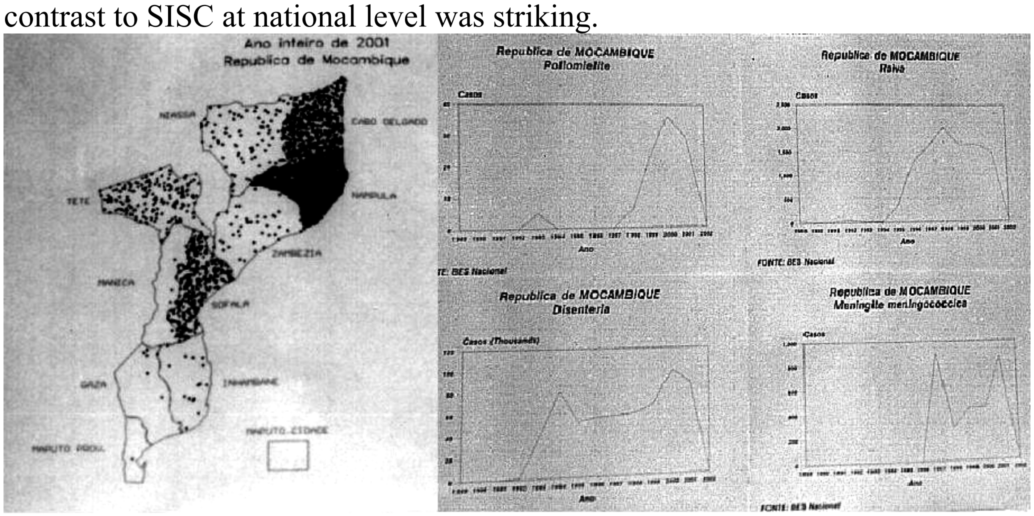
BES was striking different from SISC. BES had a version history which included references to discussions at central level about which data elements to include, alter or remove. In addition there were several bug fixes in the new version, and adaptation to newer hardware and operating systems. For SISC I have not found anything resembling this, there seems to be only one SISC version, the first.

Provincial level (Gaza)

At the province there was a dedicated room for BES, as was the case with SISC; though the BES responsible person did not have enough time to conduct an interview of the length I wanted. The computer running BES was fairly old, and next to it was a newer computer from US aid being used very little as it did not have the BES software installed. This in contrast to the version history which states that BES installation is possible without the need for technicians. I copied the BES software to my own computer in order to enable some analysis of it to be done.

National level

The BES team at national level had several offices with several new computers all running BES, most of them were being used at the time of visit, thus the contrast to SISC at national level was striking.



Picture 15: Pictures taken at BES national office. Leftmost showing location of incidents of Sarampo marked as dots. Rightmost showing incidents of different diseases from 1989 to 2002

I observed that a backup routine seemed to be in place as several floppies were observed in a disc cabinet neatly organized with labelling such as Backup Gaza 2000. The walls displayed several graphs, maps and diagrams made by BES. As observed at the lower levels there were little signs of data understanding, and the skill with the software seemed to be varying.

For instance lots of the graphs had unfortunately selection of the 1 axis periods. The graphs were showing yearly incidents, from 1989 to 2002. Although several graphs had 0, or no values for all years up to about 1995, In the end, in 2002 all graphs dropped to 0 as this was in March 2002, with very little data for 2002 ready, as a consequence all graphs had a step downward trend in the end, see picture 3. Most of these graphs should have been made with a better definition of the 1 axis periods, typically started when data was available, and ended in 2001 as they displayed complete years. None of the graphs was comparing data, for example comparing incidents to population. Other data analysis observed was a map of Mozambique with Sarampo incidents in 2001 plotted. Some provinces were almost black of incidents, while the neighbouring was almost empty, so it could look like that the disease followed the provincial borders, see picture 3. Unfortunately, I was unable in obtaining the national database which would allow me to investigate the source data and search for abnormities. However in total the usage of BES at national level was in striking contrast to usage of SISC at national level, with its hidden away computer.

9.3.4 BES content analyzing

Note that all analyzing of BES have been on the Gaza provincial version, I have concentrated on Malaria in order to give the reader a view of the data quality and how data is stored. This is in no way a complete study. For instance my conclusions should be checked out against other provinces and the national BES database. Note that I have been using the dbase files for analysis as it was the easiest to use. However, in these no reporting is stored as 0.

As figure 32 shows the recent coverage factor for BES in Gaza is very good, in for example 2000 it was 98 % for all districts. Older data, more specifically before 1994, seems to have a bad coverage rate, with some districts reporting below 10 times a year, though it varies considerably as the above graph shows. This could be due to the fact that the software was first released in 1995, with little control of reporting districts before that, or they are not finished with entry of historical data. But most likely it is due to the state of the country in those years, civil war, as the version history also states that the actual data before 1993 represents extreme under reporting. I have been unable in explaining the good coverage in 1989. However, it might be due to training by using historical data once the software was deployed.

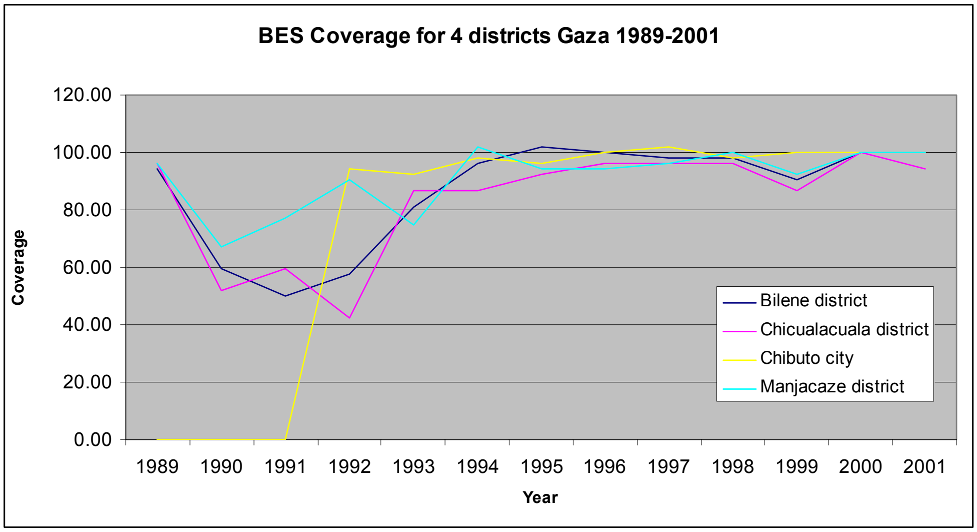


Figure 31: Showing the coverage for 4 districts from 1989 to 2001, the surge for Chibuto City in 1992 is probably because it was defined as a city that year.

As was the case with PAV in SISC, the aggregation of data to district means that the real coverage factor might be lower as facility details are lost. But the following table of missing reports from facilities found in a district headquarter, might give an indication of that the real coverage factor for BES is not as good as the graph in figure 31 indicates.

	BES	BES Received	Missing
Chibuto-Sede	13	13	-
Malehice	13	11	2
Muxaxane	13	9	4
Maievene	13	13	-
Alto Changane	13	11	2
Chaimite	13	-	-
Chipadja	13	-	-

Table 4: List of missing BES reports found in district centre Chibuto/NEP, Source from report made by the HISP team

An interesting fact was that in BES there were 15 reporting units in the Gaza province, 1 provincial hospital, 11 ordinary districts and 3 cities. This is in contrast to SISC Gaza, which only has 12 reporting districts, 1 city and 11 districts. This shows that the infrastructure is represented differently between the different systems. Note this is not error discrepancies, as BES has only selected to separate between district and cities.

Quality analyzing

I will now briefly present findings from my limited analysis of the quality of the data in BES.

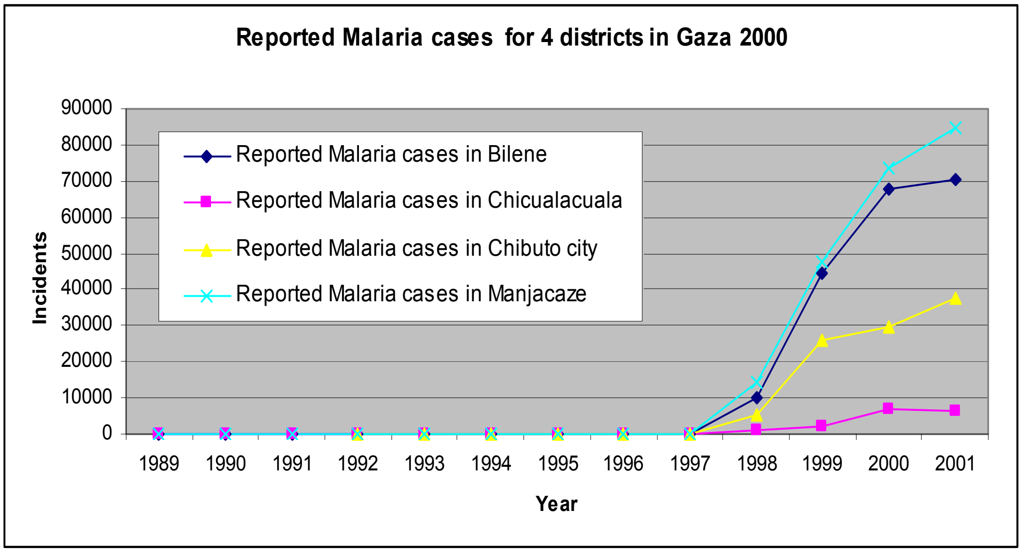


Figure 32: Shows the reported Malaria cases in BES for 1989 to 2001

The graph in figure 32 show that the statement in the version history for BES, that data before 1993 represented extreme under-reporting is correct. However, it could need some specification as the graph shows, its first in 1998 that there are reported malaria cases, this also is the case for the entire Gaza province, i.e. there is no reports of Malaria cases in Gaza before 1998. However, for other diseases such as diarrhea, there are data back to 1989, which looks like being very under-reported.

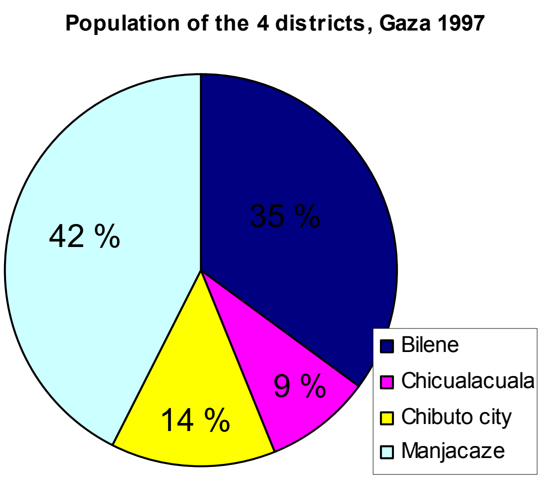


Figure 33: Show the population in the four districts relative to each other, based on [MZ CENSO 97]

The varying amounts of cases between the districts does not necessarily mean that the quality of the data is bad, as the population of the districts also varies considerably as shown in the figure to the left. However, the step increase in reported cases of Malaria over the years looks strange, and might point to that more facilities are reporting. Thus reported Malaria cases might be a function of reported facilities, but this I am unable to verify due to the aggregation.

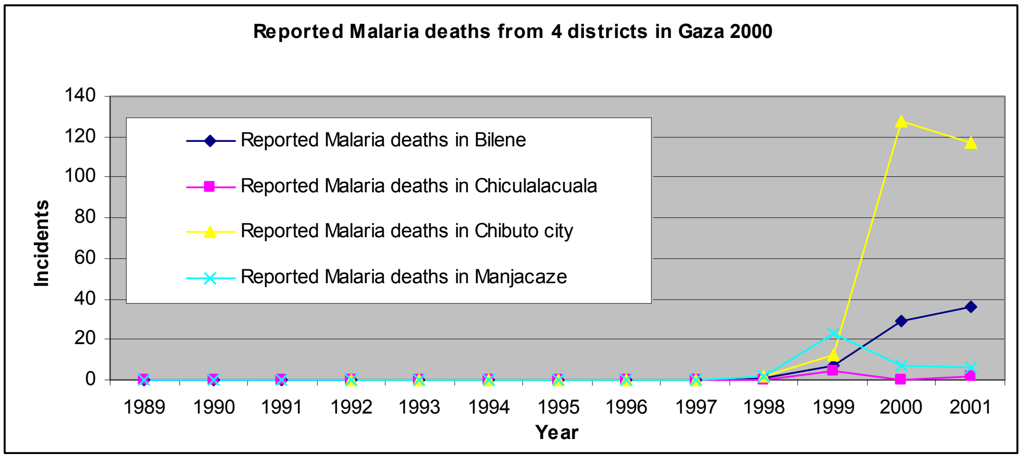


Figure 34: Shows the BES reported actual deaths due to malaria in the 4 districts from the last example. Note the extreme values for Chibuto city compared to the reported cases in the previous graph.

My first impression from comparing the two graphs, figure 33 and 35, was that the death % seemed very low and varied. I therefore calculated the death % for malaria in BES for Gaza 2000. The BES malaria death % was 0.06%; this is well below the UN millennium indicators for Mozambique which are 0.2 % [UN STATS]. However, the UN number is for the entire country and some provinces might be far worse regarding Malaria than Gaza. Or the UN might use another data source. I also noted the high values for Chibuto city compared to the other districts reported cases and population. One cause for this might be that the medical care is better developed in the urban areas and more deaths are diagnosed and reported correctly.

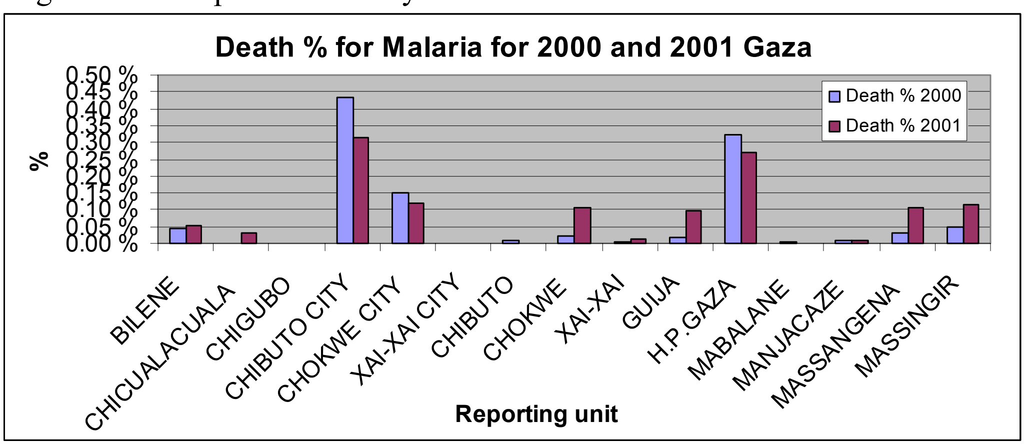


Figure 35: Shows the death % for all BES units in Gaza, note that several districts have 0 % death rate, this is usually caused by 0 reported deaths and several cases.

As the above figure shows this probably is true. As shown both Chibuto and Chokwe city have marked higher death % than the other districts. However, Xai-Xai city have 0 deaths. The other high death % is for the provincial hospital in Gaza, H.P. Gaza, and is probably due to that several serious cases are transferred to the hospital. The large variance between cities and district might imply that there is an under reporting of malaria deaths in the other districts, or that the cities are over reporting. However, it might also be due to that serious sick patients are sent to the cities.

9.4 Spreadsheets

It was discovered that spreadsheets were extensively used at the district, province and national level for several different tasks; the most important for me was the following:

1. Store information about the infrastructure
2. Store and report data, primary for the hospitals and the Health centres (form D03 and D04).
3. Analyze data in other systems, as was the case with the SISC data

The spreadsheets are in Microsoft Excel or Lotus format. My investigations into the spreadsheets were limited, as it seemed as the most time consuming part to analyze and import into SISC.

9.4.1 Spreadsheets use and content

The infrastructure is stored as a list of all facilities, including properties for each, such as number of beds, type. Some of this data is also found in SISC, but there are discrepancies between the two sources, such as:

- Different number of facilities in the districts. SISC has 141 facilities for Gaza in 2002; the spreadsheets list in Gaza (ListaUS) has 109.
- Wrong types of facility (an health centre is stored as a health post and visa versa)

I learnt that when there were discrepancies, the spreadsheets list was the most up to date. The different provinces also seem to use different layouts on their spreadsheets, as was the case with the infrastructure data list in Gaza, named ListaUS vs. one from Inhambane, named PIP-Inhambane. Though, the different names on the list can also point to that it is two different formats for storing infrastructure data, although lots of elements in the PIP-Inhambane[25] and ListaUS were similar. This might indicate that there are different spreadsheets solutions in use storing the same information.

Once it came to checking reported data I had access to rather few spreadsheets, 9 from the Gaza province and one from Inhambane. The 9 from Gaza dealt with hospital and health centre data. I noticed that the form name D04, were part of the names of 4 different spreadsheets in Gaza, this might imply that the solutions in use might be difficult to analyse, as data is spread on several spreadsheets. The spreadsheets from Inhambane also dealt with form D04, but

[25] The PIP files have been made in order to make easier to integrate information I was told.

it looked totally different from the 4 spreadsheets from Gaza. Further complicating the work was that some of the data on the spreadsheets seemed to be stored in BES as well, for instance malaria cases and death. However, the numbers in BES were significantly higher, so it might be due to different data elements. I also learnt that sometimes the print out of the spreadsheets were sent to the national level were they were re-entered instead of the original spreadsheets. The last group of spreadsheets used for analyzing of data I did not see. However, I saw several data analysis and printout of the forms used in reports.

9.5 SIMP in-house development

SIMP is the name of a system made in house to facilitate integration of data from the existing systems. SIMP was started by a developer working in the Ministry of health as an attempt to make a temporary solution to the problem of integrating data from the various systems. It started as a project of optimization of spreadsheets in the provinces, and is currently continued as a prototyping project, having an emphasis on not trying to change too much at once. The objective of SIMP is to enable the ministry of health to collect all experience in on system; the provinces are to have their own versions which they are allowed to customize to fit their local needs. SIMP only sucks data out from external sources and checks the validity of that data semi-automatically. It does not address the collecting of data, and is only a tool for the national and provincial level, not enabling local analysis of data. To date only SISC data is successfully pulled out from the existing systems.

9.6 Summary

Import of SISC data enabled the data to be viewed and analyzed, the conclusion has been that the SISC infrastructure data is poorly updated, with the implication that actual coverage factor is impossible to calculate. SISC has been shown to be non-flexible and unable in evolving with the organization. As a consequence, several ad hoc solutions have been made; these ad hoc solutions are today parts of the reason for the varying data quality. The ad hoc solutions have made the data non-uniform, ambiguous and different from actual reported data on forms. In addition, invalid data is consistently reported which indicate lack of feedback and use of data on all levels. BES was briefly investigated, and was found to be more and better used than SISC. However, data understanding and use is lacking within its users. The development and use of BES were found to emphasis a few central located users. In the end the spreadsheets were looked into. They seem to be used to fill the gaps of the different systems, such as analyzing SISC data, and reporting, analyzing and storing hospital data.

10 Discussion

This chapter offers an analysis and discussion of the empirical findings presented in the three preceding chapters. Its purpose is to answer the research question by discussing the empirical findings in relevance to specific theories.

10.1 Discussing the current Health information infrastructure

In this chapter I will discuss my findings in relevance to the problem:

Study the existing information infrastructure by importing data from the old to the new system. In this way shortcomings of the old systems and potential advantages of the new system, in terms of better data quality, may be studied.

This will be discussed in relevance to the three main topics presented in my theory chapter.

1. Health information systems
2. Legacy systems
3. Information infrastructure

Excessive amounts of theory about health information systems emphasis the need to build a culture for data use at all levels, especially the lower level where health care is delivered [Lippeweld, 2001], [WHO AA], [Potomac, 2001]. A health information system supporting these concepts must be flexible to allow for customized data collection and analyzing, with support for integration of information. Most important, the system must be able to support the different information needs for all the users [Potomac, 2001]. Ultimately, enabling evidence-based health care and supporting local initiatives on all levels, supported by use of data.

As stated in my "Theory, literatures review" chapter, I believe that health information systems may be viewed as an information infrastructure. In the table below I have tried to show how central Health Information Systems concepts relate to some Information Infrastructure concepts taken from [Hanseth, 1998].

Health Information Systems	Information Infrastructure
Flexibility to enable for continuous customized data collecting and analysis at all levels	Enable for new process and routines not all decided at time of deployment
Analysis, presentation and use of data from different sources (integration)	Information must be shared among the large community of different users with

	different needs
Large group of different user with different needs. Ranging from International organizations, such as WHO to clinicians in facilities planning day to day care	Be part of a large heterogeneous network of humans, technological components and institutions
Several organizational routines supported by several systems and programs are in place	Always an installed base, the existing heterogeneous network of routines, processes, systems etc.
Data quality is dependent upon data use. In addition the more users report good data the higher rewards in data analyzing	Interconnectivity, everything is connected and interrelated. Rewards for users increases as the installed base grows, network externalities.

I will now discuss the findings from my investigation of the existing systems, showing how they are failing along one or several health information concepts and, or information infrastructure. Afterwards, I will summarise and show how it can be viewed as failure related to Information Infrastructure concepts.

10.1.1 SISC, use and defects

[Hanseth, 1998] states that flexibility and gradual, incremental development of the individual sub systems is important in an infrastructure. SISC, with its lack of flexibility and resistant to change, is in direct contrast to this. And as so it can be seen as having become unaligned to the installed base, shown by the lack of use and failure in serving any information needs. From a health information system perspective it has also failed, as will be shown in the following section discussing SISC in relation to collecting, transmission, process, analyzing, presenting and use of information which are the characteristics that [Heywood, 1994] and [Lippveld, 2001] is state as important for an routine health information system.

Lack of flexibility in SISC

It has been shown that SISC lacks flexibility, for example, customization, including adding, modifying and deleting of data elements to capture is not possible. SISC is only supporting collection of old predetermined data elements, presumably decided by central levels a long time ago. The lack of evolution and flexibility in SISC is also shown by the following; during implementation of SISC 10 years ago, there were a paper shortage, therefore the Ministry of health only had the possibility to make annual data analysis. Today, SISC still only offers annual data analysis, it has not changed even thought the time specific limitations set 10 years ago have changed. SISC lack of flexibility also seems to influence the lower levels. The data sources for SISC are several data collection forms, which are tightly integrated with SISC. These forms are used at lower levels, and have not changed much since SISC was deployed, and does not support local customization. As such, the heterogeneous user group is not supported by SISC. As a consequence, data is not used much at any levels, if used at all. As [Lippeveld, 2001] states, lack of use of data leads to poor data quality, and as such the lack of data use at lower levels influences the higher levels, poor data for them. Another lock in is described below:

- Needed data is not collected as the system does not support it
- Data is not used because the system is not collecting the needed data.
- The defects in the system data collecting is not perceived unless data is used and its quality assessed

The age of SISC and its poor support for analysis and presentation of data and finally, actual use of information might also be a part in the lack of use of data. However, the organization has managed to circumvent this problem by manually re-entering data in spreadsheets, and then using that for analysis. Unfortunately, most analysis is poor due to a lack of a culture for data use and collection within the organization, which [Potomac, 2001] states are important. However, the circumvention shows that the staffs and the organization want to analyze and use data.

Transmission of data

The horizontally inscription of upward transmission of SISC computerized data is failing as data is transferred by the use of paper and spreadsheets, instead of the SISC disks. As a consequence, the national database of SISC data is found replicated on spreadsheets and paper at the Ministry of health. While the supposedly replicated SISC national database, is full of large data holes. And a complete SISC database is only found distributed to the provinces. However, discrepancies in stored SISC data between the district and provincial level have been found. This might be due to extraordinary business rules in relevance to data reporting. Some of these extraordinary business rules might have come about due to lack of feedback, or downward communication, combined with lack of flexibility in SISC. As shown by how some new data collection forms have readjusted old poor age groups to be more correct and precise. However, the staffs at provincial level now edit the data in the forms to fit the old SISC data elements at data entry. Other indicator toward lack of feedback is the fact that some facilities consistently report invalid data. No one seems to take actions to correct it. They might no even be caught, as it looks like no one check specific health facility data.

Vertically transmission has never been supported, showed by the lack of integration with other systems, and alternatively lack of the possibility to add new data elements to capture. This is unfortunate as theory points out the need for data from several sources to be used in management [Potomac, 2001]. Support for processing of data is also totally absent, as shown by the lack of coverage control and lack of even the most elementary validity and quality checks. These elements are all part of the lack of use of information, as observed in the annual report from the Ministry of Health for 2000, which is built on the data in the systems. The report mostly consisted of printouts of graphs and summoned columns to district and province level. Other, better reports are available, but these are based on other data sources I have been told. That fact is acknowledging the failure of today systems.

Why the failure of SISC

In the paper [Heeks, 1999] the ITPOSMO model of conception – reality gap is used to explain different types of gap between the reality and the system. This argues that the greater the change gaps between current realities and the design conceptions of a new health care information system, the greater the risk of failure. In my case there is a system from 1992 that seems to be locked to the reality in 1992, and its perceived needs. I believe several of the shortcomings of SISC today can be explained as gaps between 1992 and today.

Information gap

SISC is locked to the central information needs from 1992; this is not the same as the central information needs for today, March 2002. This is shown by the change of some of the data collection forms and the ad hoc solutions made to the reporting routines to enable the new forms to match SISC. The ad hoc solutions have been necessary due to SISC lack of flexibility. It seems even minor alternations to SISC require editing the code; such as rename a data element. This is impossible given the lack of source and documentation, and importantly, the lack of software engineer's skills within the organisation. Unfortunately, the ad hoc solutions have led to several problems; some of which could have been avoided by enforcing new standard routines. Problems relating to the new A04 form used for vaccinations data illustrate this. This change has led to at least the following problems:

- Missing data elements: The new data element "Fully immunized less than 1 year" on the new A04 form is not captured at all by SISC
- Ambiguous data elements: The new data element "DTP & Heb B", which are different vaccinations, are now captured as the old "DTP" element, as a consequence the data has become ambiguous in that the SISC element "DTP" now stores a mix between the old and the new element
- Non-uniform data: Data elements in SISC are used varyingly from province to province, making comparison and integration difficult. For example, the VAT elements 2, 3, 4 and 5, 4 elements have been combined into one, VAT 2-5. And today the different provinces are storing the new element in varying SISC columns.

A flagging day and top management could have solved some of these problems. Such as the ambiguous data could have been controlled by a flagging day when all reporting units started to use the new forms. Thereby defining a date when the old "DTP" element changed to become "DTP & Heb B". Enforcing all provinces to use the same SISC elements, leaving the rest empty, could have solved non-uniform data. The missing data element would have been impossible to meaningfully fix without editing SISC. However, one of the empty VAT elements could have been used, but this would again have made the database difficult to understand. The changes on the 3 other forms used were less, and it would have been possible to enforce correct data capture by SISC by management of the database. These problems also show the lack of feedback between all levels.

An interesting question is to which degree the new forms were influenced by SISC. Was it a coincidence that it was possible to make ad hoc solutions in SISC to still allow for data capturing after the changes? Or was it a requirement that SISC should still be able to capture data after the change of forms? The small changes made might indicate that this was a requirement. Besides from fixing small errors regarding age groups on one form, it is only A04 that have changed. However, the new A04 form resembles the old one. Why weren't there added free space to allow for local customization? The question to which degree the new forms were influenced by SISC is interesting as it can indicate to which degree SISC is interconnected to the information infrastructure. But I have not been able to investigate this.

The findings discussed above only relate to the 4 SISC tables and data collection forms that have been extensively researched, and are being used by all provinces today, March 2002. In addition to these 4 tables, there are 4 more small tables that are mostly being used by all provinces today. The rest of the monthly reported tables in SISC, of which there are 8, are used in a varying degree today by the provinces. Most of them are not used at all; some seems to have been used before. However, the corresponding data collection forms are still used. This might also be a consequence of SISC lack of flexibility. The lack of evolution of SISC has meant that new systems have been added in order to fulfil information needs not supported by SISC, such as BES and the spreadsheets. Paving the way to today's fragmented information environment, discussed later in this chapter.

Technology gap

A Mozambique HISP team member told me that the health staffs only believe in paper. With this he meant that they did not trust the floppies or other computerized storage. This might indicate how difficult it was 10 years ago deploying a computer system, or how little have changed in the 10 years since deployment. It also shows that several staffs within the organization has little understanding of computers, and henceforth little thrust in them. Furthermore, it help explain why the electronically communication between the provinces and national level failed, and it could be argued that the SISC was to technological dependent. SISC might have been technological easy to deploy, as it was made to run on plain PC with plain operating systems. The problem today is that SISC is locked to the existing computers, since there are no procedures on how to move it and the database. For example, the only external storage capabilities are floppies, and the databases are large. The consequence of this is that there is often a newer PC not being used standing next to the old PC running SISC, as seen in Gaza and at the national level. And these old PC's today have a tendency to break down. And the infrastructure for fixing computer in the provinces is absent, meaning that the computers must be sent to Maputo to be fixed, as I myself discovered when visiting Inhambane. This might show that today SISC have an large technology gap in that they are totally dependent upon very old computers, which probably will more frequently break down, while at the same time spare parts become more difficult to obtain.

Processes
Findings gaps along the processes conception are difficult as SISC have been deployed and used for 10 years, thereby influencing the organization. Looking at how the situation was 10 years ago is only speculation, but observed was how similar the SISC input screen looked to the original forms. This might have meant that SISC was designed to fit the existing systems and routines. However, today SISC is forcing processes on the users, such as requiring them to alter correct data to fit SISC.

Objectives and values
SISC is made as an upward reporting tool, hence the data collecting and entry staff sees no values in it. It is just something that has to be done. This is also supported by the lack of feedback on reported data from the central levels, as shown by the vast amount of obvious invalid data, some consistently reported 12 months a year. The poor quality of data shows that the central levels are not using the system either. However, SISC is still fulfilling some roles, for example showing the above levels that the employees are working.

Staffing and skills
The existing control functionality and other functionality in SISC does not seem to be used based on interviewing the users and observing their use of SISC. This and other observations points to that computer and data / information skill among the staff are very poor. That is very unfortunate, as SISC requires users with good data skills to enter data due to its lack of validating rules. The lack of skills is discussed later, but this and the lack of a support organization can lead to that SISC can be viewed as a black hole.

10.1.2 SISC the black hole; a legacy perspective

I believe the term "Black hole" is a suitable term to use to denote SISC as it captures its properties nicely:

- Sucking in data: The power of the system is influencing the surrounding staff who regularly feed the system data, this data is consumed by the system and never return

- How does it work: How and why the system work nobody knows, as all documentation and knowledge about it is absent, unexplainable things just seems to happen.

The “black hole” term used here does not relate to [Hanseth, 1998] and the general technical concept of a “black box”.

The lack of documentation is typical for legacy systems as stated by [Bisbal, 1999] and [Robertson, 1997]; in this case it was due to technical and organizational archiving routines. Technical in the sense that the computer storing the documentation was unusable, old laptop without battery, archiving in that the floppies containing the documentation as back up had been lost. In addition, understanding of system details were lacking among most of the staff and administrators. When presented with strange SISC data, the answer was often that it was just happening, no one could explain it. For example, no one could explain the cause for the wrong year values in the database tables. However, the key personnel at the national level seemed to be aware of their lack of understanding of the SISC. While the users in the provinces seem to just feed SISC data because it is required, note this is built upon observations in only Gaza and analysis of data in other provinces. [Bisbal, 1999] and [Robertson, 1997] states that a common property of legacy system is that they are entangled critical systems, which are difficult, impossible to modify; this matches with the findings of SISC. Shown by how it has influenced the reporting routines while at the same time not being able to adapt to the new requirements, i.e. failed to evolve with the organization. The term critical system might not be appropriate for SISC, this as it seems to be very little used except for the entry of data. However, it can be viewed as part of the institutionalizing of collecting and storing of certain data sets. The lack of interfaces in SISC, made the actual import very complex and time consuming, this will be discussed later. Lack of interfaces is also noted as typical property of legacy system in [Bisbal, 1999]. In addition the way SISC is locked to old hardware, which is difficult to maintain also fits well with theory about legacy systems as stated by [Bisbal, 1999].

All in all SISC is performing poorly in its role as a routine health information system; it seems to only be used as a backup system for data at provincial level. However, it is also performing poorly in that function, as data is not stored uniform, and it is difficult to move data out of it. The only purpose of SISC seems to be that of enforcing the data collecting routines. However, the organization views it as important, shown by all staffs, including clinicians, working with feeding it data. In contrast to SISC, there is the newer BES system.

10.1.3 The vertical BES

Unfortunately, I have been unable in doing a comprehensive investigation of BES. But I checked the use of the system and the system. BES seems to function better than SISC, with staff generally seeming much more enthusiastic about the system. The staff at national level was quite large, 10 +, and on the walls there were several print outs of presentations and analysis done with the BES software. This is a sign of actual data use; and data in BES seemed to be of better quality than SISC data, an indication of that data use led to better data quality as stated by [Lippeweld, 2001]. However, data use seems to be lacking in several aspects, but this require a deeper research into the BES programme, including forms, to verify. The version history for BES is up to this data, and shows how data elements have been added to the systems after discussions at the central level. This evolution has been possible because the expatriate developer of the system is still committed to support of the system. This show the low degree of flexibility in BES, as adding a data element is only possible by editing the code. And as such BES functionality will probably freeze once the developer ends his support, unless action is taken. In addition, the version history shows the centralized and vertical nature of BES. As only the rather small group of top users has any saying in deciding data element to collect and eventually use. This fits well with findings in [Lippeveld, 2001]; bottom levels are viewed as data source, with data collection instruments and forms designed by central data people. Furthermore, shown by the fact that all data is aggregated to district level; making local analysis impossible or irrelevant as stated in [Potomac, 2001]. As this require a breakdown of data to the level at individual health facility. Weekly aggregation of data and the varying representation of the infrastructure from SISC show how the system was made as a closed system, not adjusted to the regular monthly reporting of data and the infrastructure in SISC. This in turn makes integration of BES data with other systems difficult. Note also that BES does not have an interface for data export to other systems, showing the absent lack of plans for data integration. This lack of integration is also present in the seemingly most popular application of them all, spreadsheets.

10.1.4 The “power” of Spreadsheets

A finding of my research is the extensive use of spreadsheets in the information infrastructure. In the previous section it was showed how spreadsheets were used to fill gaps in the existing systems, such as collecting, reporting, analyzing and presentation of data.

Consequences of extensively spreadsheets use

Spreadsheets have little support for validity rules in data entry; this means that data control is totally dependent upon the data entry staff, or the ones using the data. However, as shown, there is little use of data and procedures for validation of data seems to be absent. Integration of data in spreadsheets is difficult, if not often impossible, as each spreadsheet can have a unique layout, and often has. This lack of uniform data storage implies that automatically data integration is difficult, if not impossible, and is therefore a costly and highly complex process. As shown by how the SIMP integration solution had not yet managed to pull data out from the spreadsheets. And I believe it never will as the spreadsheets are highly varying and difficult to automatically parse. Spreadsheets support for organized data storage is poor, and as such the organization must have some routines in order to save historical spreadsheets. I was unable to discover if such routines were in place. However, the spreadsheets obtained was all stored in a folder name "Lotus files 2002", indicating that there might be a folder for each year.

In contrast to this you have the database solutions, storing data uniform and ensuring consistency and referential integrity, with eventually customized validation rules.

Note, it is possible to get quality data by use of spreadsheets, for example new spreadsheets have support for checks on data entered. And spreadsheets can build upon data in other spreadsheets by linking, but several spreadsheets linked together are almost impossible to get a god understanding of. One change in a spreadsheet can cascade generate errors in others spreadsheets which can be difficult to locate. In addition, this requires skilled users. To summon, spreadsheets are not appropriate means for storing and analyzing important data. So why are they so popular in the Mozambican context?

Why use Spreadsheets?

It seems there have been made a spreadsheet whenever the information needs were not captured in the existing systems. As shown by how some information on the spreadsheets was originally supposed to be captured by SISC. However, the breakdown and lack of evolution of SISC has led to spreadsheets filling the gaps. In addition to this, there seems to be information needs which never had a system fulfilled by spreadsheets. Why have the spreadsheets proved so popular in filling the gaps?

Spreadsheets are easy and highly flexible; partly because they are building upon paper conceptions, the data and presentation, or formatting is all mixed, and the data can be entered as it is on the form into a similar structure in a spreadsheet. There is no need to think about the bigger picture, or use complex intermediate structure. The high degree of flexibility allows the user to enter whatever he wants, as there are few rules to adhere. For the data entry staffs this can be seen as an advantage in that they do not need to alter the data on the forms at all. Unfortunately, it also means that errors are not caught. The spreadsheets is what [Heeks, 1999] calls a reality-supporting application, they require few preconditions to be imposed in order to be used. Users only need to open a file and then start to enter data without any thoughts for validity or others planned use of the data. Data entered can be used immediately in making simple graphs and figures. Unfortunately, only the data entered is used for this, as integration of data from other sources, including other spreadsheets is difficult. If eventually data from other sources are integrated, it is easiest to copy it, and not link it; a consequence of this is that inconsistency can flourish. So in this context spreadsheets have poor support for ensuring quality data, storing and sharing of data.

Databases can often be seen as rigid; having several rules the user must adhere to, such as unique records, mandatory values, and customized validation rules. The database is enforcing the user to adhere to these rules, and as such the database is often seen as a rationality-imposing application [Heeks, 1999]. Note that a database application can be made as a reality supporting application. The rationality-imposing nature often implies that the user must have customized training in using the application, and good skills. In addition, a database application requires database skills within the organization, as the database and application must be developed and extended. The use of databases requires more training and support than spreadsheets, and as such represents a higher investment than spreadsheets. However, this higher investment has the potential to return a greater return than the spreadsheets, as it can support better information use.

10.1.5 Fragmentation

The organizational fragmentation and following fragmentation in health information systems is documented by many [Azelmat, 2001], [Lippeveld, 2001]. My research shows that Mozambique is no exception. I have chosen to separate fragmentation in organizational and social fragmentation.

Organization fragmentation

My visits to the facilities and districts indicated to me that the fragmentation at the lower levels was less than higher up. At one district centre I learnt that the 3 different staffs working with different data sets, all were capable of doing each other job, and shared the same office. This shows that the organization is committed to local data use, and that local integration and data use is possible.

In the province and national level the organizational fragmentation was more severe, as each health programme had their organizational funding through varying external donors. This again led to the development of different programme specific information systems to address the specific needs. In the Ministry of health building, the systems had different department and floors in the likewise building. This physical separation was also observed in the province. In the end the fragmentation is shown by the lack of information presented based on integration of data from several systems.

Technical fragmentation

As a result of the organizational fragmentation, and the fact that no actual database system has the capacity to meet the information needs, the different organizational structures have all developed their own systems without regard for integration of data. As time has passed the systems have become locked to the computers first installed on, this as there are little technical skills in the provinces and national level available to move the system, or simply reinstall them. In addition, databases are growing, making transfer on floppies of data difficult. Partly due to this, there is now a large variety of computers, Intel 486 to Pentium 2, with different operating systems, Win 3.1- Win XP, in the organization. However, there are other reasons for this such as varied donors and lack of skills. It seems that new computers are installed with an operating system, and then transferred to the health units. The new computers are not adapted to the existing systems or users. A consequence of this is that old hardware is not phased out, or transferred to less critical areas. Phasing out old computers would probably be cheaper as old computers are expensive and difficult to maintain, as broken computers must be sent to the national level to be fixed, due to lack of support for fixing computers in the provinces. The presence of new, little used computers standing next to old ones being used, indicates that this is not a problem of unavailable hardware. But more of resource allocation and management, and the lack of a working support organization for computers and software. This fits well with legacy theory of software locked to old hardware which is difficult and expensive to maintain [Bisbal, 1999] [Roberston, 1997]. In addition, the large variety of computers and operating system influences the development of the existing systems still being developed. For example, in BES ver. 1.3 Nov 96 they had removed an element added only for use on 286 computers. Due to the conclusion that at least none 286 computers were used in Maputo; note that Intel started selling 386 in 1985, replacing 286.

The problem of fragmentation is currently addressed by the Ministry of health with several systems, such as PIP, uniform spreadsheets and SIMP. But these systems seem to only be made for assisting the provinces and national level to integrate existing information for use in planning, and none seem to offer a comprehensive integration of all data. Note that my research into these systems has been limited. The emphasis on the higher levels is also showed in these systems. However, this might be due to that the higher levels are the only levels computerized. The emphasis on central levels combined with that none of the systems seems to address the problem of poor data collected in the facilities, means that eventually rewards in using the systems might be limited. SIMP functionality to validate data, requires large amounts of feedback to actually improve data quality, this as errors will only be discovered monthly in the provinces. And as such the province must provide large amounts of feedback to the districts, and then on to the facilities, this will probably take some time, and might mean that once data analysis actually appears at lower levels, they will be outdated as showed in [Azelmat, 2001]. In addition, integration of data constitutes a large and difficult project as shown by the work gone into getting SISC data during this research. This is at least stored somewhat uniform, in contrast to the spreadsheets. However, SIMP and the various planned integration systems show that the organization is aware of the need to integrate information.

Redundancy and Inconsistency

An effect of fragmentation is often redundancy and inconsistency in data collected and stored [Lippeweld, 2001]. The inconsistency was primarily found between the systems, but there were also some inconsistency within the systems. For my research the main problem was inconsistency in the different systems storing the infrastructure. None of the stored infrastructures were similar, and none were found to be complete and accurate. Findings during the import process also indicate that the provinces might not be updated on the actual number and types of facilities in their districts. This indicates that in order to compile a correct and complete infrastructure list, all districts, and, or facilities must be visited. Inconsistency in stored and used data is shown by different population data being used. In addition, the usage of spreadsheets for analyzing also adds to the redundancy data storeage.

The inconsistency is also shown when inspecting data in international organizations databases. For example, the data element, “Skilled attendance at birth”, is according to the United Nations Development Program 44 %[26] for Mozambique in 2000, while according to WHO it was 30 %[27], unfortunately WHO does not state for which year that number is for, though most number are for 2000 and 2001 in WHO. Other health related data are difficult to finds as the organizations uses different indicators and have number for different years.

Why has this fragmentation come about? I believe that in order to answer this one must look outside the health care organization and into the country in general, more specifically the lack of resources and an educated workforce.

10.1.6 Lack of resources and an educated workforce

Lack of resources and an educated workforce is inherent in developing countries. As stated in my background the illiteracy rate in Mozambique is at 57 %, indicating a low educated people. Resources for public health is also scare, as only 30 Int $ [WHO MZ] are used on public health care pr. citizen annually. This combined with findings that only 2-3 % [Gravitz, 2000] of budget is spent on information system in the public health care, leaves scare resources for information systems. In addition most technology must be bought with US dollars, of which only 9 $ are spent pr. citizen annually.

This lack of resources and the low educated population means that developing information systems are dependent upon external assistance and donors, as stated in [Haga, 2001]. This dependency was observed in the health

[26] http://hdr.undp.org/reports/global/2002/en/indicator/indicator.cfm?File=cty_f_MOZ.html
[27] http://www.who.int/reproductive-health/publications/MSM_96_28/MSM_96_28_table4.en.html

organization with computers having sticker denoting that they had been donated by USAID or another Italian foundation. Unfortunately, the systems and computers are not sustainable, once the donors ceases their support the systems quickly turns into frozen non flexible system, as stated by [Braa, 2002]. And will most likely after some years become legacy systems with its related properties, this due to lack of a support network, as has been shown with SISC. Remember that SISC required editing of code and Meta data if data elements were to be added, removed or modified. As this was not possible therefore modifying was not possible, and as consequence the organization has made several ad hoc solutions in data reporting routines and added spreadsheet in order to computerize data not caught by the systems. Looking farther back in time, it is noteworthy how several tables in SISC are not used anymore. For instance medical stock data seems to be captured in a new system today, leaving the SISC tables empty. The fact that a computer system can function and evolve with external support is shown by BES.

The lack of an educated workforce in developing countries is also shown in how the adaptation of DHIS is dependent upon the South African development team, even when DHIS is being implemented in several other countries, such as Mozambique, Cuba, Malawi and India. However, in India the lack of an educated workforce can not be the reason. So this problem is probably more complex, factors such as the amount of DHIS use must also be considered. However, use of DHIS is again dependent upon skills in data use.

Data skills and use

The staffs at all levels are not well trained in collecting, validating, controlling, analyzing and use of information. Shown by poor quality data consistently collected and transferred from facility level up to national level. These errors are not due to laziness; as doing the error sometimes increases the workload, as shown by staff at provincial level summon up data elements, instead of just entering the data which would have been more correct. Few of the errors actually decrease the workload. The coverage factor and storing of forms in neatly arranged organizers at district level, show that it is possible to collect and store good data. And the manual re-entry of SISC values in spreadsheets shows that there is a desire to analyze data systematically. Unfortunately, the end result is poor due to lack of skills. Another observation in regard to lack of skill is the fact that most of the ad hoc solutions are not documented, and the staffs seem unconscious about them and the consequences. The lack of data use is also shown in SISC.

10.1.7 Vertical top down systems and no feedback

Most systems currently in Mozambique seems to be programme specific closed systems, with no support for collecting more data than necessary for the supported programme, i.e. little flexibility regarding data collection. This is very much inline with the description from other developing countries as described in [Lippeveld, 2001] [Azelmat, 2001] and others. For each new health program, new systems and forms are made almost irrespectively of other existing systems, funded and implemented by an external donor. With the consequence that data collecting is redundant and information infrastructure fragmented. There seems to be little awareness of how much data is actually collected, as all systems are focused on their own vertical closed programme data. The lack of emphasis on the lower levels is also shown in the lack of feedback to the lower levels, as shown by the following:

- Facilities are consistently reporting obvious invalid data
- Change of forms seems to have been very little managed, should been a flagging day and uniform changes of the reporting routines

The lack of feedback is unfortunate as all the information systems are based on good collected data, which is difficult given that the collectors have no tool or skills themselves in validating or verifying data collected. The lack of feedback led to the data collectors losing faith in the collection of data, as for them it seems that no one is using it; as such data use is not encouraged in the organization.

10.1.8 Information infrastructure perspective revisited

There were found faults in all existing systems along several central information infrastructure concepts [Hanseth, 1998]. These failures can partly explain the failure of the entire infrastructure.

Enabling: The systems are not enabling for new process and routines as the functionality and flexibility of the systems is poor. Everything is decided at central level and enforced with none flexible technology.

Sharing: Lack of integration support in all systems makes sharing of information among the larger community impossible. In addition, the top down emphasis led to closed homogeneous data sets, which fail to support the information needs of the users.

Heterogeneous Installed base: The heterogeneous installed base network is in place:

- Humans, such as different users from clinicians to national health planners
- Institutions, such as different donor agencies requiring programme specific information for evaluation of the programs
- Technology, such as several computers and the health information systems

This heterogeneous network is not supported by the existing systems, as the systems seems to primarily be homogenous as defined by [Hanseth, 1998], closed and only supporting the needs of the central level. As a result the systems are little used, and several ad-hoc mechanisms are in place, such as use of spreadsheets to fill gaps. However, the spreadsheets are not well suited for this, as they for instance have poor support for sharing of information. The use of datasheets shows that there is a demand for analyzing and customized data collection, or flexibility.

Interconnectivity and network externalities [Hanseth, 1998]: Data collected at the lower levels are making the basis for decision on higher level, or should. However, the systems used at higher levels influences, or are interconnected, to the design of the data collection forms and routines at lower level. Thereby influencing data collected at bottom level, such as frequency, aggregation and what is collected. This is influencing the bottom levels eventual data use. This situation is sustained by systems implemented years ago, which today is difficult to change due to the large installed base. This is what [Hanseth, 1998] would call a path dependency, past selections have a large impact on future development. In addition, change today is difficult as the systems does not offer any flexibility, and as [Hanseth, 1998] states, flexibility is important in order to avoid and get out of a lock-in, and make a working information infrastructure. A lock-in in this case is that the information infrastructure is unable in changing due to poor flexibility of today systems. Another is that data is not used due to poor quality and wrong data collected. However, this can only change if data is used, as the lack of necessary data elements and poor quality will then emerge. Network externalities are shown in that rewards for users of the systems is dependent upon number of other users. A system more used, both data use and amount of data collected, means that the quality and amount of data is higher, as such more and better analysis is possible.

HISP states that the decisive element in obtaining a properly working information infrastructure in the health care, is by training of staff in data use; systems can only enable data use. In the next section, I will try to identify some constraints and aspects of the current infrastructure. And then discuss advantages of the new system, combined with how challenges might be tackled from an information infrastructure perspective.

10.1.9 The new system, possible advantages

Analyzing of the data stored in the old systems has shown how the existing systems are not supporting information infrastructure and health information systems concepts. This might be an important part of the failure of the entire information infrastructure. And as such my analyzing made possible by the data import has shown the need to change and advantages of the new system, these in addition to making the old data re-emerge. Several key constraints have been pointed out in the analyzing of the existing infrastructure, primary the lack of skills and education. The new system, DHIS, might be viewed as an part of an information infrastructure, as shown by how it will be implemented.

Installed base

As stated by [Hanseth, 1998], aligning to the installed base is important for an infrastructure to be successful. This is a large area of research which I can not hope to treat fully here; instead I am opting for some parts I believe to be important.

Looking at DHIS today it is aligned along some areas, such as the variety of hardware and operating systems environment found in Mozambique. However, proper use of DHIS requires skills and a support network, this seems to be lacking:

- Lack of skills and routines in data collecting, analyzing and use
- No support network for information systems available

These factors and more must be in place for DHIS to function. Using [Heeks, 1999] theory the changes are called gaps, which must be minimized, or eliminated for the new health information system to function. [Heeks, 1999] proposes to minimize the gaps by changing the organization, such as training. While information infrastructure emphasis that the old must be aligned to the new [Hanseth, 1998], and training is a way to do that.

Aligning installed base, training and gain alliances

Training, building and institutionalizing a data culture, emphasising data use at all levels within the organization is the most important part in the aligning phase for HISP. In addition, there must be built a support network within the organization, that can support and modify DHIS, this group should have cross disciplinary skill in software, computers, statistics, general health care and other related skills. This is what [Heeks, 1999] is calling hybridisation agents. The above elements have always been HISP focus [Braa, 2002], as HISP is trying to build sustainable networks [Braa, 2002]. The building of alliances is used by HISP, as shown with how key members of the Ministry of Health are members of the HISP group. This is also supported by the experience from South Africa, where the roll out and implementing of DHIS is not so much described as technical work, but more as work related to enrolling support from key actors, or alliance building [Braa, 2000], and the social problems which comes due to organizational change.

DHIS, flexibility and development

DHIS as a flexible tool is only possible with skills and an established data culture, without it, DHIS might become a non-flexible top down tool similar to the already existing systems, and over the years turn into a legacy system. This as the inscriptions for local data use is weak and flexible enough to be avoided if necessary. Another aspect of DHIS within information infrastructure is that of one application implying an organizational universal standard, and accordingly homogeneity and lack of flexibility as described by [Hanseth, 1998]. This is in contrast to the heterogeneity and flexibility of information infrastructures [Hanseth, 1998]. In addition, DHIS is a closed system, not enabling new applications. However, the heterogeneity, flexibility and closeness can be viewed from several levels. And from the DHIS user view it is allowing sharing of information (openness), information from all facilities can be imported into anyone facility as long as the referential integrity is upheld. Flexibility and heterogeneity support in that it is fully customizable, and ultimately editing of code. But this requires a larger support network; as such DHIS is not a closed system, but a part of a larger information infrastructure. However, as will be shown in the import process discussions, integration of data might require more standardized data collection and storage to allow data comparison. And as such, the move towards DHIS as the one system might led to less flexibility for each health program, but this is again dependent upon how flexible DHIS is.

In these sections several shortcomings of the existing systems were shown. Most important seems to be a lock-in due to the lack of flexibility in the systems deployed years ago. The decisive element in changing the information infrastructure and breaking the lock-in seems to be the training of staff in data use, this in order to build a data culture and a support network within the organization. This would most likely lead to better data quality compared to the poor data quality today.

10.2 Import process

In this section findings from my study into the import process are discussed, its aim is to answer the problem statement of:

Study how data from legacy systems can be made available in a new system

The import process was separated into several processes. Firstly, the data to import had to be uncovered. Then the actual import process was started. The actual import process was very much captured in the following question. Import what data from which system at which level in which order? I have therefore decided to separate the entire process into the following steps.

1. Uncovering data for import, find, read and decode
2. Adaptation of DHIS and setting the prerequisites
3. Selection of data to import
4. Adaptation of source data
5. Eventually data cleaning due to poor data quality

The above process were not strongly separated as believed in beginning, this as for example setting the prerequisites is connected to the import due to various infrastructures representations in the systems. In addition, the obscurity of the data in the legacy system meant that data cleaning was difficult without first importing the data.

10.2.1 Uncovering data for import

Gaining access to the data is one of the first steps in the imports process, and can be separated into three problems.

1. Finding the data
2. Reading the data
3. Understanding the data

As shown, the infrastructure is fragmented. As a consequence finding, reading and understanding all data is a time consuming task as each system have their own unique data storage. This caused me to focus on data in SISC for the actual import, while at the same time investigate BES and the spreadsheets in order to gain a wider perspective on the import process.

Finding the data
The three systems investigated are all storing data on the workstation, and there are no central data servers. The SISC system replication of the database to the national level each quarter had failed. And the most complete SISC database was found distributed to the provincial levels. In addition, SISC data was found on paper in primarily the district level and in spreadsheets in the provincial and national level. However, those sources were not appropriate for data export due to several reasons, such as difficult to automate export and, or lack of complete data stored. Consequently, there was a decision to be made with a trade off between data quality vs. workload.

Reading the data
Being able to read the data was a technical problem, and proved to be the easiest to tackle. The SISC system, which was my focus, used Dbase ver. III that Access, Excel and ODBC (Object database connectivity) supported, i.e. able to open the files and view the data. The data in the other systems were easy read as all used different types of spreadsheets, Lotus, Excel, and text files to store the data.

Understanding the data
SISC data was found coded in such a way that viewing the back end was meaningless without documentation and a mapping for the coded column names. The making of this mapping was a comprehensive process due to the legacy nature of the system, lack of documentation combined with the unavailability of key staff with good technical understanding of SISC. The lack of availability of technically skilled SISC staff in the beginning meant that some extra work had to be done. However, their availability would only have

helped in the mapping of the SISC elements to the papers forms, and not solved the problems related to ambiguous and non-uniform data. In order to decode and verify all mappings of the coded data elements, it was necessary to check the actual paper forms from the district and the actual use of SISC in the provinces. Only after the first decoding of the data was made could the import process start. This import made wrong mapping emerge and new mapping and import had to be done. Only SISC stored data coded beyond comprehension, the other systems data did not need the same kind of work to be understood.

10.2.2 Adaptation of DHIS and setting the prerequisites

Adaptation and setting the prerequisites are two similar projects. In this case, adaptation of DHIS deals with the general set up of DHIS to enable it to be used, such as translation, defining the data elements, storing of the infrastructure data. In contrast, the prerequisites deals with smaller parts in order to allow for database population, such as making sure the data imported have the corresponding reporting facility in the infrastructure list in DHIS.

DHIS adaptation

An important element is the fact that the DHIS has been developed for the South Africa health care organization. And as such the program logic and functional requirements are made for the South African context, therefore the flexibility of DHIS is an important factor, can it be adjusted to fit the Mozambique context. From an actor network viewpoint how strong is the South African inscriptions, and are the existing inscriptions correct for the Mozambique context.

DHIS is currently locked to the organizational structure of the South African organizational hierarchy, consisting of five levels; national, region, provincial, district and facility. Mozambique, has in contrast four levels; national, provincial, district and facility. This has been solved by the HISP team by entering a dummy level in DHIS in Mozambique; all provinces have a dummy region below them, which the districts in that province belongs to, see figure 38. Another problem regarding the stored infrastructure in DHIS is the fact that only the end nodes, leafs, in the organizational hierarchy can report data. This represented a problem in the import of vaccination data in Mozambique, as the vaccination data, PAV, was aggregated to district level in SISC. In order to enable for import of vaccinations data from SISC, there were added a PAV facility for each district in DHIS, storing only the vaccination data. This was made as a direct consequence of the import process, in contrast to the dummy region which was added long before.

I have been told that similar problems have been found in Cuba where they have an organizational hierarchy consisting of six levels. In Cuba the solution in DHIS is to aggregate the lowest level. In addition they have made similar solutions to cater for data reporting from other units than the facilities.

The lack of flexibility in DHIS is probably due to that DHIS have not been made as an international system, but as a closely prototyped system in South Africa. And today is lacking important functionality, or flexibility, to be properly adaptation to new countries.

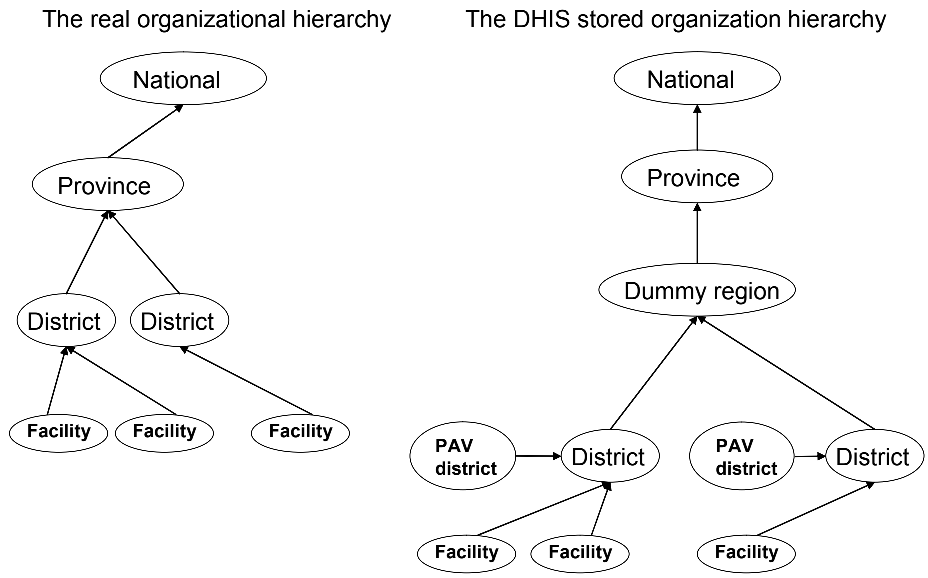


Figure 36: Showing how the organizational hierarchy in DHIS had to be adapted in order to enable for data import in Mozambique. The PAV district is stored as a facility in DHIS, but is actually on the district level, connection between PAV district and district is a one to one.

DHIS internationalizing

The development of DHIS has led to a system which is not separated into the international core part and localized culture dependent part as described by [Yeo, 2000]. The DHIS development team, as shown by the recent released multi language version, is currently addressing this. But this is a highly complex procedure. For example, when translating the text to be shown in buttons the size of the buttons sometimes needs to be readjusted, this can again trigger re organization of the graphical layout. This problem is sometimes further increased when translating to other non Latin alphabet as the letters require other fonts with different sizes. The separation between the international parts and localized parts is not a clear cut and easy one. As of today, the DHIS team is working on enabling various frequencies of reported data, support for data reporting for all organizational units and varying levels in the organizational hierarchy. New demands for flexibility is probably going to emerge, as I my self heard talks from the Ministry of Health in Mozambique that they wanted to store 2 infrastructures, one active and one planned.

In making international software there is a need not only to ad new functionality, but also to separate the existing functionality and systems, this separation might go across existing modules. I believe that making a fully international DHIS is important in order to make it easier to adapt DHIS without making so many ad hoc solutions, which as I have shown have a tendency to survive for a long time in the health organizations. The added international functionality might be in conflict with the original DHIS principles. For instance only allowing the end nodes to report data was done purposefully in order to force the organization to let all data pass through the district, this partly in order to enable data use at all levels. This represents a classic conflict between the flexibility for the organization, and enforcement of wanted new routines as shown in both [Heeks, 1999] and [Hanseth, 2002]. However, in the context of importing data there was no way to change the existing data, i.e. de aggregation is impossible in most cases. The fact that the PAV districts have now been added can lead to that DHIS is not going to alter the routine of aggregating vaccination data in Mozambique.

The prototyping approach used by HISP in developing DHIS in South Africa might have made it difficult to make an international tool, as the emphasis might have been on to fast provide new solutions, and as such the overlying design might have been less emphasized. [Hanseth, 2002] states that small and modular solutions made for specific problems is the best way to start the work on an information infrastructure, where the modularity is supposed to easy enable new functionality to be added. This has been possible to use in South Africa through prototyping. However, now DHIS has grown large and modifying it is not so easy anymore. As such during deployment of DHIS in other countries, it might be difficult to offer the same small solutions for specific problems.

Recently, the South African government has made a statement that they intent to use free, open source software whenever possible[28]. This could for HISP be what [Hanseth, 1998] calls a bandwagon to jump on, because HISP wants to make DHIS platform and software vendor independent, i.e. run on both Linux and Windows with different relation databases. The process might also be used to facilitate the internationalization of DHIS. As implementing platform independence would most likely be a reengineering one. Due to the fact that new methods and languages have to be used. Note that platform independence must include all of Microsoft operating systems, this as experience has shown that the ability to run on multiple Microsoft operating systems combined with various office packs to be important in gaining users acceptance. As it allows DHIS to be installed on the existing systems, or aligned to the installed base [Hanseth, 1998]. However, today this strong dependency is making DHIS expensive, as Microsoft licenses have to be bought in order to use DHIS. Thereby making the claim that DHIS is free relative. It is this dependency that the platform independence hopes to remove, which also is an important part of the South African goal of free and open source software.

[28] The move toward free open source is supposed to pervade through the IT business and society in South Africa as shown by the citation from http://www.bday.co.za/bday/content/direct/1,3523,1266306-6099-0,00.html "Our ultimate goal is to stimulate the birth of companies and an entire industry based on open source software".

Setting the prerequisites for DHIS

Setting the prerequisites mostly consisted of checking the mapping from SISC data elements to DHIS data elements, including checking that the infrastructure was complete and accurate as each reported data element maps to a facility, or organizational unit. The prerequisites were therefore tightly integrated to the actual import process and attempts at importing made problems surface.

Discrepancies between DHIS infrastructure and the SISC infrastructure were found once the mapping of SISC monthly data was started. Discrepancies were also found when checking the DHIS infrastructure against other stored infrastructure data, such as spreadsheets. DHIS contained facilities not in SISC and the spreadsheets, and vica versa; some of the discrepancies were due to different names used on facilities, but not all. In addition, some facilities were registered with different attributes, such as DHIS stored it as a health centre, while SISC stored it as a health post. These discrepancies were solved after extensive collaborative work with staff at provincial level with local knowledge. The staffs were necessary as the stored infrastructure at provincial level also contained discrepancies, and as such none of the information systems could be trusted in containing a complete and accurate infrastructure list. The new compiled DHIS infrastructure led to that several SISC facilities were not having a correspondence facility in DHIS, which again led to that data for these facilities were not being imported. Therefore, setting the prerequisites had an influence on the selection of data to import. Note that this is one of the reasons why migration of data is not seen as a good solution to preservation of data, as only the active data tends to be migrated, leaving out purely historical data [Rothenberg, 1998]. In this case much of the unmapped data was probably data from closed facilities. As discussed in the adaptation section above there was also the need to add a dummy PAV facility for each district in DHIS.

The existing data dictionary in DHIS was found not to have all the corresponding SISC data elements, this as the DHIS dictionary had been made with the assumption that the data elements on the forms where the same as the data in SISC. And as my research has shown, this is not the case as the reporting routines, or lack of routines has led to ambiguous, invalid data etc. However, the data dictionary was not edited and the problem was eventually dealt with in the data cleaning procedure.

10.2.3 Selection of data to import

As stated, only SISC data was selected for import. However, the question of import of what from which level remains. An important factor when selecting data to import was workload vs. good information, no doubt that the forms from the health facilities would represent the best information, but that would require manually entering of all information, thus a very high workload. The other alternatives were the computerized SISC data at provincial and national level.

The national level were supposed to have all the SISC data for the country, but actual inspecting showed that it contained several large data holes, such as missing entire years for some provinces. The provincial level only has local SISC data, i.e. data for the likewise province, but with less data holes. However, the quality was poorer than in the forms at district and province level. A decision to import SISC data from the provincial SISC database was made, as it was the most uniformed computerized data, and most complete. And as such it represented the easiest to automate export from. For example, the spreadsheets were only used to analysis data and as such did not necessarily contain all SISC data. Only data for the last 2 years were selected for import, leaving data from 1992 to 1999 in the SISC database. As such recreating historical infrastructure was not important. However, data left in SISC would decay if, or when DHIS is replacing SISC, therefore a manual were made for the Ministry of health on how to save the old data, see appendix F.

Regarding data for BES, I have been unable in comparing provincial data to national data, and cannot therefore conclude at which level the data should be pulled out from. The provincial BES data contains all historic data, from 1996 up to today for that province.

10.2.4 Source data adaptation, transformation

The representation of SISC source data had to be adapted in order for my tool to map the facilities. This due to the fact that the facilities in SISC were identified by two columns, district id and unit id, which together uniquely identifies a facility in a district. In contrast, my import tool only allowed the mapping of a single column. Therefore a SQL view was made on the SISC source data that joined the two columns into one, so that this new column could be mapped. Similar task had to be done for the dates in SISC as SISC represented date as month and year in two columns. Note that the above adaptation does not edit the data in itself, only how it is represented, in contrast to what has to be done to the BES data as discussed below.

Eventual import of BES data would have needed at least one source data adaptation. As BES weekly data would have to be aggregated to monthly as DHIS only accepts monthly data reports. In addition, the BES infrastructure would have needed to be represented in DHIS. This could be done either by adding the BES districts as dummy facilities in DHIS, i.e. facilities only storing BES data, as was done with the PAV data from SISC. However, this would not allow for fully data comparison between BES data and SISC data in DHIS, as some of the BES districts are separated into district and cities. The other solution to this, could have been to combine the cities only found in BES into their districts, for instance Chokwe city only found in BES would then have its data merged into the BES Chokwe district. This in order to align with the infrastructure in DHIS based on the SISC data. In this solution the BES data could be connected to the PAV_district, which would have been renamed to better indicate the data stored there. This would allow for at least some data comparison in DHIS, and would probably have been the best solution as it would not add any more dummy units to the organizational hierarchy, and allow for easy comparison of data. However, it would constitute a change for the BES organization as they would have to work with a new infrastructure.

Regarding the spreadsheets it would probably have needed large amounts of layout editing in order to be imported automatically, but not so much source data adaptation as it is on monthly basis pr. unit, where unit is everything from a facility to a provincial hospital.

10.2.5 Data cleaning

As shown, the existing data was of poor quality. This had to be set up against that several of the routines resulting in poor data quality was done intentionally, and therefore data was used with the knowledge of what it actually represented. As such cleaning it might be dangerous However, some of data was cleaned.

- Monthly reported SISC data from facilities not part in the newly compiled infrastructure were left out
- Monthly reported SISC data with obvious wrong year was corrected

In addition to this there were necessary to do some cleaning of the vaccinations data due to the change of the form A04. This was ad hoc cleaned in that it was decided that all data from after July 2001, was given with the new A04 vaccinations form. This ad hoc solution was used because otherwise it would have required an extensive study of each single record in order to verify which form they were using. However, the facilities had not necessarily switched forms when they started giving the new vaccination. My guess is that they were using up the old forms irrespectively of wherever they were giving “DTP” or “DTP & Heb B”. Consequently, cleaning the data any better than the ad hoc solution might be impossible. The solution chosen was a trade off between invested resources and correct data.

No more data cleaning was done, partly due to it being costly in resources, as most cleaning would have made it necessary to check the paper forms. Secondly, due to the obscurity of the data in SISC some of these errors were only found after the actually import. And as such the import process and cleaning of data was a cyclic process, were the import of data triggered new discoveries about the state of the data, which could influence cleaning of data.

The later import process done by the HISP software coordinator included the data for the unknown facilities as well, based on that all data should be imported and then analyzed to find the errors. Such as discover wherever the unknown facilities are same as an old one but with a new name, is a duplicate, is a closed down facility or a ghost facility added for training but not removed.

10.2.6 Prototyping in action

Due to the unknown requirements for the import tool I opted for prototyping. My first prototype was made in Norway, this was later partly scraped once arriving in South Africa and working with the HISP software coordinator and Softcraft. The new version supported the ability make a DHIS import file with several data elements, which was deemed as necessary. However, most experience was gained in Mozambique as I participated in an actual import project. Some of the experience was added as functionality to the tool during the research, such as perfecting of the DHIS data dictionary and facility names to facilitate the mapping process. However, most of the experience was not added to the tool, as I did not have the time. Some new findings in relation to the tool are presented here. New solutions and other possibilities are presented in the next section dealing with the use of legacy theory in changing an information infrastructure.

During and after making the tool I have found HISP as an organization to be separated in what they want. The HISP software coordinator wanted a general tool to be used for populating by technically skilled users in several countries. While the Mozambican team, or parts of it, wanted a more automatically specific solution, as the following request after seeing the tool shows.

"I want a simple executable file, that could be installed in the provinces and should automatically make a DHIS import file with SISC data once run."

This also included that the user input should be minimum, preferably none. As the tool were to be used by the health staff. This very much resembles a gateway and is discussed later.

After the tool was made I posted in on the web at my homepage with an English user manual. HISP in Mozambique, Cuba and India were notified of its existence and where it could be downloaded from. However, the tool seems to have been little used, from Cuba I heard that the HISP students there did not have so much time to import data, and what was done was specifically for Cuba. An Indian from HISP India was in touch with me, but never replied after the initial mail consisting of the URL to the program and user manual, emphasising that I wanted feedback. Another HISP team member in Mozambique is using the tool to import data. The fact that the tool is only used in Mozambique touches an important aspect I believe. The complexity of the existing systems makes it difficult to make a simple user friendly general tool in a short time. In addition, the specific tool was not that much more automatically and still required skilled users with knowledge of the old systems. My experience tells me that it seems to be difficult to automate parts of the process in order to enable non-skilled user to do it semi automatically. Either it is done almost fully automatically by the use of gateways. Or an expert does it by the use of ad hoc solutions, and general software tools, such as SQL. As happened in this case, the HISP South Africa software coordinator later came and imported data by the use SQL and general tools. However this requires more research I believe.

Extending of the existing tool

The prototyping used enabled additional functionality to be identified, such as allowing several data source columns to be joined into one DHIS import file column, typically necessary regarding SISC facilities as they were identified by two columns.

More importantly, I think that adding an XML parser would be the biggest contribution. This is based on the functionality in the latest Excel version, Offices XP, of saving files as XML[29]; this combined with the supported file formats in Excel, such as Excel spreadsheets, Lotus spreadsheets, Dbase files. Make an XML parser a valuable tool, as it could enable import from a large variety of file formats, or data sources currently in use in the health sector. For example, from Cuba, I have heard that systems with FoxPro database as backend and general spreadsheets were used. This data might be more easily imported with the use of an XML parser.

XML support in my import tool would enable connecting to several more different data sources than it does today, in addition the text parser would be unnecessary as Excel contains a text parser for structured text files. The XML parser would most likely use a second party vendor for the actual XML parsing, while the functionality of selecting and mapping of the data in the source data to DHIS import file would have to be made. Part of the decision of not to make XML support in the beginning was the emphasis on making a tool with little user editing of files. Accordingly I first opted for the solutions which did not need any pre possessing of the files by the users. Another part was that I opted to start and make something small, which grew into today's tool. Now afterwards, I think that the XML solution would have been the best as it was using much more finished technology, and as the saying goes "Experience is cheapest second hand".

More functionality requirements were identified. Some examples are:

- Editing the representation of data in the data source, for example the tool should recognize data and allow the user to select in which format it should be stored in the DHIS import file. For example, data as "Feb 02", should be possible to automatically store as "February 2002"

[29] Microsoft is known for not always following standards and the XML might also be none standard. However it would properly be possible to find some walk around, secondly the XML parser would properly be some second party vendors which might already take this into account.

- Better text parser and add more functionality to the JDBC.ODBC drivers possibility to read Excel files
- Functionality to use copy and paste from different opened data sources into a text buffer in the import tool.

Due to the complexity and varying degree of different systems making an import tool to parse all the different storage systems is a large task, way to big for a cand. Scient. research.

10.2.7 Evaluation of the import process

The entire import procedure was complex and resource demanding, this due to the nature of the legacy systems combined with the extensive fragmentation of the systems and routines. For instance, finding and compiling a complete and accurate infrastructure list in a province took 3-4 weeks in total, and later findings at district level indicate that it might not be complete and accurate. In addition, ambiguous and non-uniform data made the mapping, cleaning and importing of data a long and complex process. As a consequence importing data from a system require a unique and extensive study into the system at all levels, this as each level is not necessarily updated on the actual situation in the levels below. And as such travel will probably be a time consuming part of an import process.

The fragmentation of the systems has made the representation of data different, as it is mapped to different concepts, or context as shown in the list below:

- Different infrastructure units (districts vs. district and city)
- Different frequency of time (weekly vs., monthly)
- Different administration level (facility vs. district).

And the import process showed how this variation of aggregation influenced eventual data integration. Looking at vaccination data, it is clearly shown how in inappropriate it is to aggregate the data to district level, as it makes data comparison to other data difficult. This is also the case with BES data, which has weekly frequency and is aggregated to district level. This makes integration of data in analyzing difficult, even after eventual import into DHIS. For example, BES data might have had to be aggregated to monthly frequency and new districts in order to allow for comparison to data already in DHIS, and as such simply importing data would not automatically enable integration. And as such, a comprehensive integration of data might require changes in today data collecting and reporting routines.

The fragmentation has also led to various infrastructure representation used, and it must be decided which to use? In this case the BES infrastructure might be the best as it is the newest and the system most used. However, it does not store facility data, and therefore is not complete. Eventual integration of existing data from different systems requires more than simply importing it, needed is investigation of all systems. This to decide several fundamental elements, such as which infrastructure representation to use, or if several are to be used. An eventual new overall infrastructure which supports all existing data can only be made after all systems on all levels have been investigated. This might only be relevant to DHIS due to the lack of flexibility in DHIS, as it does not allow weekly reporting and the use of different infrastructures. However, a system consisting of several different infrastructures would be very complex, as it would need functionality to compare data between different infrastructures. And this might prove to be even more difficult to get sustainable. But this discussion is outside the scope of this thesis.

Other factors adding to the complexity was that several versions of the new DHIS database were used, and these versions became inconsistent. Actually knowing who had the latest database proved difficult. Therefore the different databases had to be merged in the end. This shows the need for a strong version control, as new findings seems to constantly be done, sometimes in settings no one would expect. Large amounts of time were spent testing DHIS after data import to make sure the imported data was correct. And that the dummy units were correctly positioned in the hierarchy, this fits well with the theory that large amounts of time are being used on testing the target system [Bisbal, 1999]. In this case most testing was not for to errors in the new system, but testing the data actually imported. This is a long process as showed by how 3 months after the research several old reported data were discovered in DHIS Mozambique, entered a long time ago as yearly sums.

My general tool with the longest development time was the tool that required the most editing of presentation of the data. As almost all data imported by my tool had to the presentation format changed, typically by making a view in SQL. This, in contrast to the specific solution, that used the source data without editing the representation, as the SQL views could be hard coded. In addition, the HISP software coordinator afterwards re imported all data by the use of SQL. This as HISP now wanted all facilities and relating data, including the ones that were not mapped. DHIS was then a complete replication of the SISC database at the time of import. This indicates lack of requirements for the project and the need for good requirements.

10.2.8 Contributions, Data available

An important contribution of the data import, has been that data which before was unavailable, is today available. The SISC system that before was a black hole has been opened and its data is available. Vaccinations, maternity prenatal and children and health data for 3 provinces for the years 2000, 2001 and partially 2002 was imported. My research into this data has led to several findings about the state of the installed base as discussed in the preceding section, and shown in the findings chapter. These findings would not have been possible without the import process. As this work triggered and enabled the data analyzing to take place. In addition, the research showed how difficult it is to integrate the existing systems, which is important as the Ministry of health is currently working on different integration solutions.

What was not done?

Only some data was imported, as the process of actually importing data was very complex and time consuming. Very much in line with what [Bisbal, 1999] is stating. All data investigated seems to be possible to import, though some need large amounts of pre possessing and data source adaptation. Preserving of the data was partially fulfilled; only recently data was moved. However, one of the back ends for an existing system were decoded and documented. And a manual was made on how to save the data. Preservation of data that has been moved to DHIS is dependent upon DHIS, and wherever DHIS gets an archival format saving ability. Note with archival format saving ability I am referring to a format that saves the data with enough Meta data to allow the data to be understood without the software [Rothenberg, 1998]. However, the import at least temporarily saved some data.

In these sections it was discussed how data from legacy systems could be made available in a new system. Consequently importing of data was done from one system; however the legacy nature of the fragmented systems made the task complex and time consuming. Each system requires large investments of time and resources in order to be fully decoded. Integration of data requires that all systems and data must be investigated in order to decide on the infrastructure and frequency of data reporting. If there is to be only one infrastructure, or one frequency of reporting, then there must be some transformation of the data during the import.

10.3 Legacy theory used in implementing an information infrastructure

In this section I will discuss the problem statement of:

Analyse how to enable the implementation of a new infrastructure with the use of legacy theory to examine the interrelations between the old and new systems

Legacy theory, more specifically target database population, is used in relation to replacing the systems, while information infrastructure is used to invest the larger change.

10.3.1 DHIS deployment from a legacy viewpoint

The HISP teams want to deploy DHIS, a complete new system to replace most of the old ones. Thus redevelopment is what HISP wants. DHIS is going to make substantial changes to the legacy systems functionality, thus making a large disruption to the existing operational and business environment; this is what [Bisbal, 1999] is referring to as Big Bang. The HISP team wants redevelopment as the existing business rules regarding collecting and use of medical data today in Mozambique is poor on all levels [2Braa, 2001], [Braa, 2001] and this thesis. The wrapping and migration approach as descried by [Bisbal, 1999] is deemed inappropriate in this setting, as they both transfer old business rules.

HISP is aware of that trying to do the complete switch at once will probably fail due to the size of the organization and the interconnectivity. This means that the change is large and constitutes a large coordination effect, which a large change of failure as referred to in [Heeks, 1999] and [Hanseth, 2002]. Therefore I have participated in a small step of pulling data out from the existing systems and showing how the existing systems are failing. From a legacy deployment perspective I am providing a temporary solution providing valued data and business logic to the new system linked from the old legacy systems [Bisbal, 1999] [Hanseth, 2002].

For the actual transition from old to new systems, [Bisbal, 1999] are mention three methods. Cut and run that is closing the old system and turning on the new. This is not feasible for the HISP team due to the high element of risk; if the new system is not working properly it will have severe consequences for the health organization. And the fact that the size of the health organization means that the change process would be a large and highly costly one, because of the need to deploy large amount of computers and train large parts of the organization to use DHIS. This is also making a flagging day or a short switching period impossible. The second transition method, parallel operations are too costly, both in money and resources; it would require vast amounts of resources and a large organizational support team to have two systems operational. The third transition method, the phased interoperability, seemed to be more appropriate. However, it requires gateways, which are difficult to make due to the large amounts of different legacy systems with no interfaces to other systems, this is discussed later. Therefore a combination would most likely prove most appropriate. For example, using the cut and run method at a component level, replacing one or more system at a time with no integration to the existing systems, at the time of replacement the old system data is moved to DHIS with my tool. However, continuous integration, which is part of the phased interoperability would no be supported.

10.3.2 The infrastructure perspective

I have before described how I believe that the entire information system is an information infrastructure and that several of the failures of the systems are due to lack in central infrastructure perspectives. How do HISP and DHIS relate to the changes in Mozambique in an information infrastructure perspective?

[Hanseth, 1998] describes three generic strategies for changing an information infrastructure; these are evolutionary, daring and radical. As shown there is no strong central authority, nor is the change to DHIS and inscribed processes a small change which can be done on a flagging day. This eliminates the daring and radical change, because as [Hanseth, 1998] states that a central authority to enforce an easy change must be in place in order for them to be successful. The evolutionary change strategy is therefore seen as the most appropriate. This is a self reinforcing process, the word self means that it can not be controlled, only influenced, and once started it can grow fast. This means that HISP must try to enable and support the change, try to pull in the general direction, as the large organizational change must come from within. This is a part of the HISP policies as shown by their emphasis on cultivation [Braa, 2002], and organizational change. These aspects of change are also found in [Heeks, 1999] as change agents supporting organizational change.

In information infrastructure flexible technology is important in order to avoid a lock-in, and eventually get out of one once entered [Hanseth, 1998]. In the previous chapters I have shown that the existing systems are not flexible, and that there seems to be a lock in. In order to evolve, or change, the lock in must be broken and there are several ways to do this, but having flexible solutions, or technology is one of the most important as stated in [Hanseth, 1998]. However, today systems offer little flexibility; therefore they must be replaced in order to heighten the change for the lock in to be broken. And as such the first aim of changing the information infrastructure might be to replace the existing systems, this in order to enable support in the information systems for future organizational change, or change of the infrastructure. Selected theory argues that the smaller the change, the higher change for success [Heeks, 1999], [Hanseth, 1998] and [Bisbal, 1999]. This means that the strategy most likely to be successful in replacing the existing systems is to completely adapt DHIS to today systems and organization. However, this might not legitimate the invested resources, as part of replacing the systems is organization change. As stated in both [Heeks, 1999] regarding health information systems and [Bisbal, 1999] regarding migration old legacy systems. But after DHIS have been accepted and aligned to the installed base the change process can begin, as DHIS offer the flexibility to support the change. This is very much the same as [Bisbal, 1999] advocates in the migration approach regarding legacy systems; first implement the new system without added functionality, once operational start to ad functionality. However, this approach is difficult in this context as the new system, DHIS, is already finished developed. Secondly, the fragmentation of the existing systems, both in data collected, infrastructure representation and analysis possibility makes it impossible to completely adapt DHIS. And making DHIS flexible enough to support this variety of functionality is probably not feasible and / or possible. An alternative to implement DHIS in the organization without causing a large disruption to the installed base is by the use of gateways.

Gateways
Gateways could be used to connect DHIS to the existing systems data storage in the province and national level. Each system could have a gateway which semi-automatically transferred data between the old system and DHIS. [Azelmat, 2001] states that routine fresh data is important if data is to be used in management. Gateways properly made and supported might provide routine, continuous and integrated data into DHIS from the existing systems. And as such gateways might trigger actual usage of data from the existing systems. This eventual usage of data would most likely make it clear to the organization that the state of data is poor, and thereby lead to action to improve the quality of data. As stated in several papers that data use lead to better data quality [Lippeveld, 2001] [Braa, 2002]. This would mean that the wanted change would start from inside the organization, and especially the province level which is seen as the information hub [Braa, 2002]. The gateways could be seen as a way to quickly provide wanted specific solutions to everyday problems which [Hanseth, 1998] state is a good way to initiate the work on an information infrastructure. HISP could then focus on routines, training and change (organizational) which would enable local initiatives based on customized data collected and information usage, which would be possible as DHIS would support this. Ultimately building an organization where action based on data is initiated at all levels without orders from the levels above.

In the beginning of this project I thought I was making a gateway, as it allows data stored in the old systems to be accessible in DHIS. However, experience showed that it was not automatic enough to be sustainable; it requires a user with good knowledge about DHIS and the old existing systems. A gateway usually means a black box which transparency allows several otherwise incompatible systems/networks to communicate with each other [Hanseth, 1998]. In addition to this my tool only allows for data transfer from the old systems to the new, and as such consistency is not supported. A scenario of using DHIS to replace some systems while using the import tool to integrate with other still function systems would most likely fail. This based on the needed skills and complexity, such as manually ensuring consistency between the two systems. And as such the tool is not enabling a stepwise incremental replacement strategy as described by [Hanseth, 1998]. My tool will most likely be used once DHIS is deployed to get the historical data into the DHIS database.

Gateway example
The large variety of systems in the Mozambican health care means that making gateways is a large project; requiring systems developers and participation from the users and administrator of the old systems, this due to the fragmentation of the systems and lack of interfaces. Each system needs their own gateway which will most probably be manually controlled, i.e. administrator must run an executable file which makes a DHIS import file. This file must then manually be moved by the use of floppies to the DHIS computer and imported into DHIS.

Making gateways work both ways in order to ensure consistency would be difficult, this due to the lack of interfaces in the existing systems, data can be pulled out from them, but pushing data back into them is another question. In addition, DHIS would most likely need some new interfaces and new functionality to enable alterations to the data to be replicated back to the old existing systems. Therefore, I believe that making gateways both ways might not be feasible. One way gateways, from the old systems to DHIS might be more suitable. As data is used and analyzed in DHIS, feedback could be given back to the respective programmes, and the systems. The various programmes could then send the feedback further down and ultimately down to the bottom levels. However, this feedback would require a large change in today routines. Unfortunately, using DHIS by the help of gateways will not increase the flexibility in the lower levels, such as data collecting, and thereby limit possible data analysis. Actual data collecting would still depend upon the existing routines and systems, such as the forms. These have been shown to be poor and little flexible. In addition, much of the data collected is aggregated to various reporting frequency and organizational unit, leading to loss of facility details and lack of support for integration of data. This a gateway would not change, therefore, analysis might prove worthless for the local levels, as stated in [Azelmat, 2001] as facility details are lost. The needed and important change would only be possible once the old systems and their routines are decommissioned. Therefore, the gateways would only be temporary as used in legacy theory [Bisbal, 1999], in contrast to gateways in information infrastructure, which becomes permant parts of the information infrastructure [Hanseth, 1998].

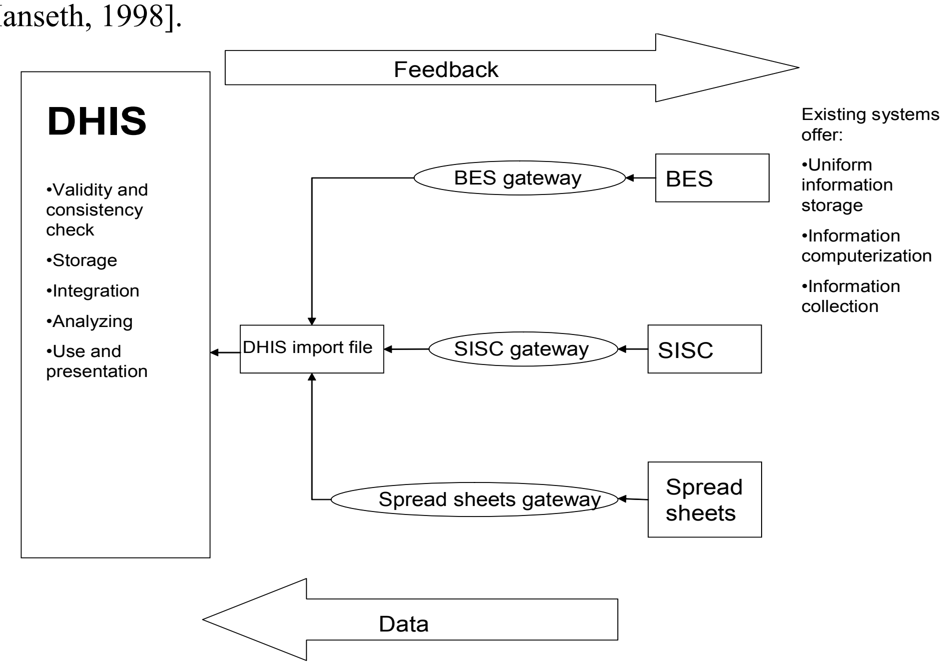


Figure 37: Showing an example of a gateway solution. Data transferred to DHIS for analysis and use, feedback going back.

The fact that most theory states that running two systems parallel is a highly complex and error prone scenario [Bisbal, 1999] [Sommerville, 1994] must also be considered. This in addition to the administration of the gateways and the migration process, which [Bisbal, 1999] states is very difficult. This means that most likely a comprehensive migration approach would not be possible. A compromise of replacing some of the poorly working systems while using one-way gateways for the ones, which are working, might be a solution, leaving out the problems of ensuring consistency. This could then enable a slow incremental change to heighten the change of success [Hanseth, 1998] [Heeks, 1999]. But the large change, which HISP is trying to make, would still constitute the largest work, including training on all levels and flexibility on all levels to enable usage of the skills learnt in training. Only a new routine flexible system backed up by skilled user and a support network can provide this.

In this section it was discussed how legacy theory could be used to make the change process to a new information infrastructure easier for the organization. This as a legacy system might often be part of a larger information infrastructure as it is entangled into the business. However, there are several different approaches to this. One, which was discussed, was the use of gateways to allow for continuous backward compatibility and integration of data. This might lead to actual data use again triggering organizational change.

10.4 Summary

This research has been possible and was started as a direct consequence of my population of DHIS with data from existing system by a developed import tool. The import process was resource demanding and complex due to the legacy nature of the existing system. The legacy characteristic of the existing system meant that data was hidden and not available for the organization; my import enabled the data to be viewed and analyzed. This analyzing was used to evaluate the existing systems in relation to health information systems and information infrastructure theory. Findings have been lack of sharing and integration of data, lack of flexibility and lack of enabling for new process in relevance to data collecting, processing, analyzing and use of information. There were found several indications of that it would be possible to change to a working system, or information infrastructure. Finally, I analyzed the change process to DHIS with legacy and information infrastructure theory and found my import tool to have several limitations, primarily lacking sustainability. However, legacy theory used in conjunction with the changing of an information infrastructure can be used as a part of the aligning phase. The key factor though, turned out to be the training and building of a data culture and a support network within the organization.

11 Conclusion

The result of my study was that old hidden data was made available and the new routine health information system, DHIS, was populated. This enabled the analysis of data and a larger investigation into the existing systems, identifying important faults in these in relevance to information infrastructure theory and health information systems theory. I will now in three sections write my conclusions in relevance to the three problem statements from the introduction.

Analyse how to enable the implementation of a new infrastructure with the use of legacy theory to examine the interrelations between the old and new systems

The import process performed in Mozambique has been a part of the larger change processes of a new routine health information system. The idea of using a general data import tool as a gateway between the new and the old systems has proven to be unsuccessful due to the complexity of the process and the non sustainability of the actual tool, and as such the tool failed to function as a gateway. If successful, it might have allowed continuous data import, thereby triggering usage of the new system and data in management without replacing the existing systems and reporting routines. An implication of this could have been that a stepwise approach to the larger change process might have become possible. Where the first step would be to replace today's none flexible systems at certain levels, with a flexible system supporting change, and afterwards each system on the other levels could be replaced one at a time. As such, the first step in a replacement process can be to make the technology flexible, with the smallest possible changes for the installed base. Other replacement routines such as cut and run and parallel operations were deemed as too complex to be possible in the organization. A valuable aspect found after import, where that the data stored in the existing systems could be analyzed and potentially demonstrate the need for a new system. That was done and a report was made and delivered to the "Ministry of health". This might be part of triggering an eventual change in the future.

Study how data from legacy systems can be made available in a new system

The legacy systems properties of lack of documentation and lack of integration to other systems made this task complex and time consuming. The first tasks of finding, reading and decoding the data in the legacy system required extensive research into how the systems were used and the technical aspects of the systems. The system replication of the database to the national level was found not to be working, and a complete database at national level was only found in spreadsheets and on paper. Therefore, the databases selected for import was the distributed provincial ones, as these were the ones that were most uniform and complete. This decision was a trade-off between quality of data imported, and resources invested in the import process. The SISC national database would have been the easiest to import, but had the poorest data quality. The best quality data found on forms in the districts would have required manually re-entering of data. The decision to import from the provincial databases means that it is necessary to get all provincial databases at the provincial centres in order to import the complete national data. During this research, data from 3 provinces were imported.

Due to lack of flexibility in the systems, non-uniform data storage and ambiguous data elements had been introduced. In addition, there existed inconsistency regarding stored routine reported data between the levels. The infrastructure data stored in the legacy system was not complete and accurate; and inconsistency within the stored infrastructure was found between various systems and different levels. Compiling of a complete and accurate infrastructure was only possible by working tightly with provincial staff, as none of the systems were found to store a complete and accurate infrastructure. However, the list was not correct as later research into the district showed; and the best infrastructure data seems to be on paper forms in the districts. The entire import process of selection, cleaning and importing data was a cyclic process, as actual data import made errors in the source data surface, which again influenced selection and cleaning of data. If data from other systems were to be imported, and eventual integrated, it would have to be transformed in order to be adapted to the rest of the data. This due to the fragmentation which have led to varied routines and process, which now has to become more standardised, or homogenous as they move towards one system and integration. For instance different aggregation levels, frequency of reporting and different representations used for the infrastructure implies that integration of data is difficult. However, the amount of standardization is dependent upon the new system flexibility. Several incidents of inconsistency would also have to be solved. And as such a fully import and integration of data from all systems requires a large project, where all systems had to be investigated in order to decide the following at least

- Which infrastructure to use, one or several
- Which frequency of data reporting to use, one or several
- Which organizational aggregation to use

Basically, the actual import process of data consists of a large process where the existing systems and data must be cleaned. This cleaning consists of investigation of all levels where the system is used, including levels without computers.

Study the existing information infrastructure by importing data from the old to the new system. In this way shortcomings of the old systems and potential advantages of the new system, in terms of better data quality, may be studied.

The information infrastructure was found fragmented into centralized vertical program specific systems, which were functioning in a highly variable degree. Within this the routine health information system was performing poorly and totally lacked integration support. The systems are not flexible partly due to lack of sustainability, and are today preventing evolution of the organization, as modifying the systems is impossible. As a consequence, the information infrastructure seems to be a in a lock-in due to non flexible systems which are preventing evolution, this lack of flexibility makes it difficult to get out of the lock in. In order to cope with the change in information needs and evolve several ad hoc solutions have been made, including several spreadsheets based solutions. However, none are addressing the main problem, the poor quality of the data collected and little data use, this due to lack of skills in data use at all levels. In addition, some of the ad hoc solutions have led to non-uniform and ambiguous data storage. As information infrastructure, the systems were not enabling and not supporting the different information needs within the organization and the heterogeneous user groups. The systems are connected to the organization, enforcing routines and process on the lower levels, and stifling any local flexibility. Data is reported and never used or provided any feedback. The data transfer routines are poor and data quality deteriorates successively as data is transferred upwards. Lack of integration has led to that data is not shared and there is redundant data collection and storage. All legacy characteristics were found, running on obsolete and expensive to maintain hardware, resistance to evolution and change, critical for the business and the lack of integration. However, despite the poor quality of the data and the information infrastructure, findings indicate that changing to a working system is possible because:

- There is a desire to analyze and use data as observations and the ad hoc solutions shown
- Data can be methodically collected and stored as this is done today
- The need for integration and change is perceived by the organization as shown by various solutions made to provide integration support

However, the new system must be supported by training in data use and the building of a support network within the organization to enable flexibility and empowerment of the lower levels. There must be a large organization change.

11.1 Further work

In general the process of replacing several fragmented information systems with one integrated system needs more work. A part of this is that of importing data from the eventual existing fragment systems, and then to try and integrate this data, and not just store data in a new system. A specific part requiring research is I believe spreadsheets, can the XML supports in new spreadsheets make it easier to semi-automatically extract data from them.

12 Acronyms and word used explained

Below is a non-comprehensive list of acronyms and word used in this thesis that I think is necessary to specify:

BES: is a programme specific system used in Mozambique to capture data related to dangerous diseases.

Data element: A data element is used to denote the simplest form of data, for instance "Live births" is a data element

Data set: Date set in my master refers to a set of data items/elements, typically being collected. In a country there are usually several data sets, for example:

- WHO data set
- Local district data sets
- National data set

DHIS: District Health information system, the software package for entering, collating, validating and reporting routine monthly data developed by HISP in South Africa

Epidemiology: is the study of how health and disease are distributed in populations and factors that influence or determine this distribution.

Health facility: A health post or a health centre, the bottom health stations responsible for delivering of the day to day medical care.

Health unit: Any health related organizational unit

Hospital data: Is used to denote data collected at the hospitals

HISP: Health Information Systems Programme, A team of people and organizations working with health information systems. For more see www.hisp.org.

MAT: The name used to denote maternity data set, which is a part of the SMI data set

OrgUnit: Organization Unit data, data about the health facilities and the organization of the healthcare.

PAV: The name used to denote the vaccination data set in SIS and SISC

PH: Public health, the science and art of promoting health, preventing disease, and prolonging life through the organized efforts of society [WHO].

PHC: Primary health care, day to day care done by health facilities employees, for example vaccination of all children. Typically healthcare in developing countries is divided in PHC and Hospital data. DHIS is only supporting PHC

SMI: The name used to denote maternity (including deliveries), prenatal and mother and child data set in SIS and SISC

SISC: is the routinely health information computer system being used in Mozambique

SIS: is a collection of data collection forms, procedures for reporting and aggregation data from the health facility, to the district, province and national level within the Health care in Mozambique, almost all information in the health care is part of SIS

SIS.D: Name used for DHIS in Mozambique, note I am always referring to DHIS as DHIS

RHINO: Routine Health Information Network Organization (RHINO), see http://www.cpc.unc.edu/measure/rhino/rhino.html

Legacy information systems or LIS: Typically old information systems which have become entangled in the running of the business, usually it is difficult, impossible to extend. In this thesis Legacy systems will be used to denote Legacy Information Systems

13 Bibliography

[Azelmat, 2001]
Mustafa Azalmat & Michael Edwards "A Decentralized Information System for the Monitoring and Evaluation of Maternal and Child Health/Family Planning Program Performance" in *Issues and innovation in routine health information in developing countries* RHINO workshop, March 14-16, 2001, available at: http://www.cpc.unc.edu/measure/rhino/theme1/morocco_paper.pdf

[Braa, 1995]
Braa J., Monteiro E., Reinert E "Technological transfer vs. technological learning: IT infrastructure and health in developing countries", Information Technology for development, Vol. 6, No. 1, 1995, IOS Press, pp. 15-24

[Braa, 2001]
Jørn Braa, Joao Mavimbe "A survey of the health information system in Niassa – 2001 Findings and suggested improvements". Not published

[2Braa, 2001]
Jørn Braa, Esselina Macome, Joao Carlos Mavimbe, Jose Leopoldo Nhampossa, Joao Leopoldo da Costa, Bonifacio Jos, Aureli Manave, Antonio Sitoi . "A study of the actual and potential usage of information and communication technology at district and province levels in Mozambique with a focus on the health sector". The Electronic Journal on Information Systems in Developing Countries, EJISDC-ISSN 1681-4835, Vol. 5

[Braa, 2000]
Braa, J. & Hedberg, C. "The struggle for Developing District Health Information Systems in South Africa" IRIS 23, 1:25

[Braa, 2002]
Braa, J., Monteiro, E. & Sahay, S. (2002a): 'Networks of action: sustainable health information systems across developing countries', submission for special issue on action research, MISQ.

[Bisbal, 1999]
Jesus Bisbal, Deirdre Lawless, Bing Wu and Jane Grimson "Legacy information systems: Issues and directions" IEEE Software, September/October 1999

[Budde, 1991]

Budde et al. Chapter 1"What is prototyping" and Chapter 4 "Prototyping". "Prototyping an approach to evolutionary systems development", ISBN: 3-540-54352-x

[CIA, 2002]
CIA World Fact book www.cia.gov Visited December 2002

[DAVID]
David, Pail A. "Clio and the Economics of QWERTY"

[Dick, 1993]
Bob Dick "You want to do an action research thesis"
available at: http://www.scu.edu.au/schools/gcm/ar/art/arthesis.html

[Feldman, 1981]
Martha S. Feldman and James G. March "Information in Organizations as Signal and Symbol", Cornell University

[Gravitz, 2000]
Debbie Gravitz "Health care and the Net"
http://www.redherring.com/mag/issue83/mag-healthnet-83.html, visited June 2002

[Haga, 2001]
"Conditions for health information systems in South Africa" Ulrikke Berdal Haga August 2001, Cand. Scient. Thesis University of Oslo, Institute for Informatics

[Hanseth, 1997]
Ole Hanseth, Eric Monteiro and Morten Hatling "Developing information infrastructure: The tension between standardizations and flexibility", Techonology and Human values, Vol. 11, No. 4, Fall 1996, p. 407-426. SAGE Periodicals Press

[Hanseth, 1998]
Ole Hanseth & Eric Monteiro "Understanding Information Infrastructure" Unpublished book, http://www.ifi.uio.no/~oleha/Publications/bok.html

[Hanseth, 2002]
Ole Hanseth "From systems and tools to networks and infrastructures – from design" to cultivation. Towards a theory of ICT solutions and its design methodology implications" Unpublished article, available at http://www.ifi.uio.no/~oleha/Publications/ib_ISR_3rd_resubm2.html

[Heeks, 1999]

Heeks, R., Mundy, D. & Salazar, A. "Why Health Care Information Systems Succeed or Fail" in Information Systems for Public Sector Management, Working Paper Series, Paper No. 9.

[Heywood, 1994]
Heywood, A., Campbell, B. & Awunyo-Akaba, J. (1994): "Using information for action: A training manual for district health workers". Royal Tropical Institute (KIT), Amsterdam, the Netherlands.

[Holy Grail]
"Integrated care pathways – the Holy Grail?"
http://www.medrecinst.com/resources/ehealth2k/manual/clinician/TBssheph.pdf, visited June 2002

[INEMZ]
Instituto De Estatistica (National statistics bureau of Mozambique)
http://www.ine.gov.mz visited 2002 December

[Lippeveld, 2001]
Theo Lippeveld "Routine Health Information Systems: The Glue of a Unified Health System" in *Issues and innovation in routine health information in developing countries* RHINO workshop, March 14-16, 2001, available at:
http://www.cpc.unc.edu/measure/rhino/keynote/keynote_paper.pdf

[MZ CENSO 97]
Mozambique Population Census 1997, for more se [INEMZ]

[McDonald 1993]
McDonald, C.J. (1993) "ANSI's Health Informatics Planning Panel (HISPP) - The Purpose and Progress." in De Moor, G.D.E., McDonald, C.J., and Noothoven van Goor, J.(eds.). (1993) Progress in Standardization in Health Care Informatics. IOS Press, Amsterdam.

[NHP]
African National Congress "A National Health Plan for South Africa" in A National Health plan for South Africa, African National Congress 1994 pp.1-91

[Norad MZ]
NORAD homepage about Mozambique, visited 2002 December
http://www.norad.no

[O'Brien, 1998]
Rory O'Brien "An overview of the Methodological Approach of Action Research" in *"Theory and practice of action research"* from Universidade

Federal da Paraiba available at
http://www.web.net/~robrien/papers/arfinal.html

[Potomac, 2001]
Rhino "The Potomac Statement on Investment in Routine Health Information in Developing Countries" available at:
http://www.cpc.unc.edu/measure/rhino/statement.pdf

[Robertson, 1997]
Paul Roberston "Integrating Legacy Systems with Modern Corporate applications", Communications of the ACM, Vol. 40, Issue 5, May 1997

[Roemer, 1992]
Milton I Roemer "National health systems of the world: the issues.", Dec 1992, Oxford University press ISBN 0195078454

[Rothenberg, 1999]
Jeff Rothenberg "Avoiding Technological quicksand: Finding a viable technical foundation for Digital Preservation", A report to the Council on Library and Information resources, ISBN 1-887334-63-7

[Sommerville, 1995]
Ian Sommerville "Software engineering", fifth edition, ISBN 0-201-42765-6

[Smith, 1991]
Mark Easterby-Smith, Richard Thorpe and Andy Lowe "Management research an introduction", ISBN: 0761972846

[Ulrich, 2002]
William Ulrich "Legacy systems must support key business initiatives", Computerworld magazine March 27, 2000

[UNDP]
United Nations development program, http://www.undp.org , visited 2002 December

[UNDP HDR2002]
United Nations development program Human development report 2002 available at: http://www.undp.org/hdr2002, visited 2003 January

[USAID MZ]
US Aid Page about Mozambique
http://www.usaid.gov

[UN STATS]

United Nations Statistics Division Millennium Indicators visited January 2003, available at: http://unstats.un.org/unsd/mi/mi_goals.asp

[WHO]
World health organization http://www.who.int or http://www.who.int/reproductive-health/publications/ visited 2002 December

[WHO MZ]
World health organization statistics for Mozambique http://www.who.int/country/moz/en/ visited 2002 December

[WHO AA]
World Health Organization Alma Ata declaration, International Conference on Primary Health Care, Alma Ata, USSR, 6-12 September 1978 http://www.who.int/hpr/archive/docs/almaata.html visited 2003 January

[WHO, INDICATORS]
Indicators to monitor maternal health goals visited 2003 January
Available at: http://www.who.int/reproductive-health/publications/MSM_94_14/MSM_94_14_table_of_contents.en.html

[WILSON]
Randy Wilson, Jon Rohde, Ruben Puchert, Calle Hedberg. “South Africa’s district health information system: Case study from Eastern Cape Province” Unpublished

[Yeo, 2000]
Alavin W. Yeo. “Are usability assessment techniques reliable in Non-Western Cultures?”. The Electronic Journal on Information Systems in Developing Countries,
EJISDC-ISSN 1681-4835 Vol. 3

14 Appendix A

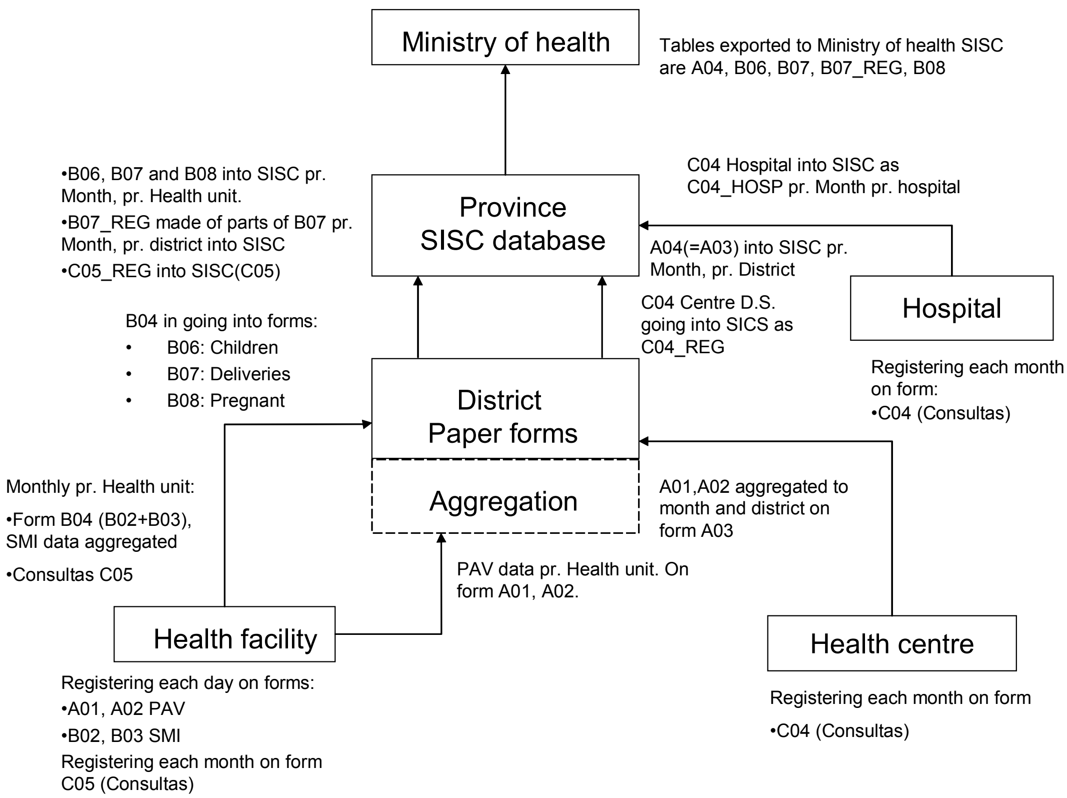
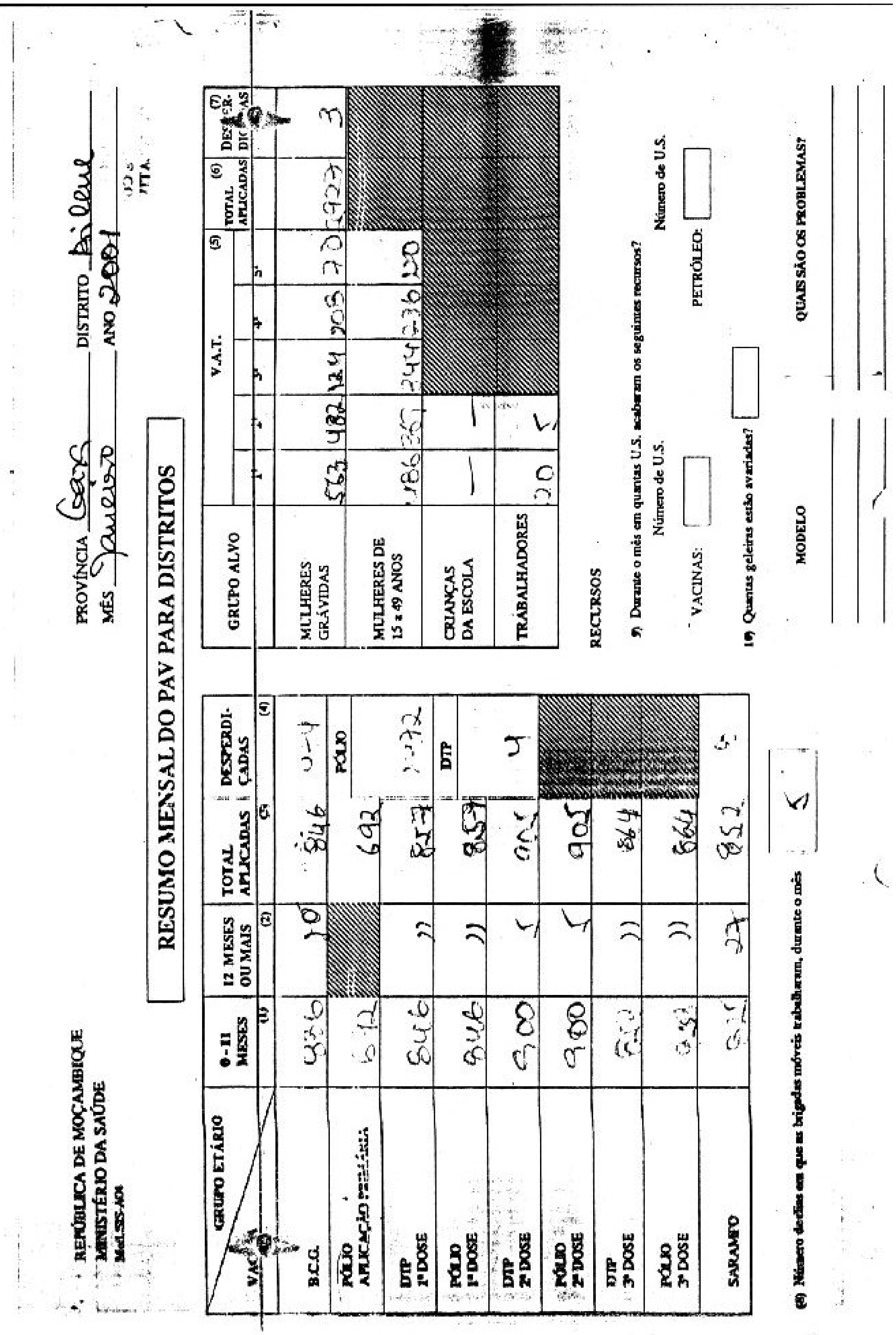


Figure 38: Figure showing the flow for most of the used SISC data set / elements today

15 Appendix B

Form A04 old



REPÚBLICA DE MOÇAMBIQUE
MINISTÉRIO DA SAÚDE
Mod.SIS-A04

PROVÍNCIA Gaza DISTRITO Bilene
MÊS Janeiro ANO 2001

RESUMO MENSAL DO PAV PARA DISTRITOS

GRUPO ETÁRIO / VACINA	0-11 MESES (1)	12 MESES OU MAIS (2)	TOTAL APLICADAS (3)	DESPERDIÇADAS (4)
B.C.G.	836	10	846	0-4
PÓLIO APLICAÇÃO PRIMÁRIA	692		692	PÓLIO
DTP 1ª DOSE	846	11	857	2472
PÓLIO 1ª DOSE	846	11	857	DTP
DTP 2ª DOSE	900	5	905	4
PÓLIO 2ª DOSE	900	5	905	
DTP 3ª DOSE	853	11	864	
PÓLIO 3ª DOSE	853	11	864	
SARAMPO	825	27	852	5

(8) Número de dias em que as brigadas móveis trabalharam, durante o mês 5

GRUPO ALVO	V.A.T. (5) 1ª	2ª	3ª	4ª	5ª	TOTAL APLICADAS (6)	DESPERDIÇADAS (7)
MULHERES GRÁVIDAS	563	482	129	203	70	1927	3
MULHERES DE 15 a 49 ANOS	186	367	244	236	120		
CRIANÇAS DA ESCOLA	—	—					
TRABALHADORES	20	5					

RECURSOS

9) Durante o mês em quantas U.S. acabaram os seguintes recursos?

Número de U.S. Número de U.S.

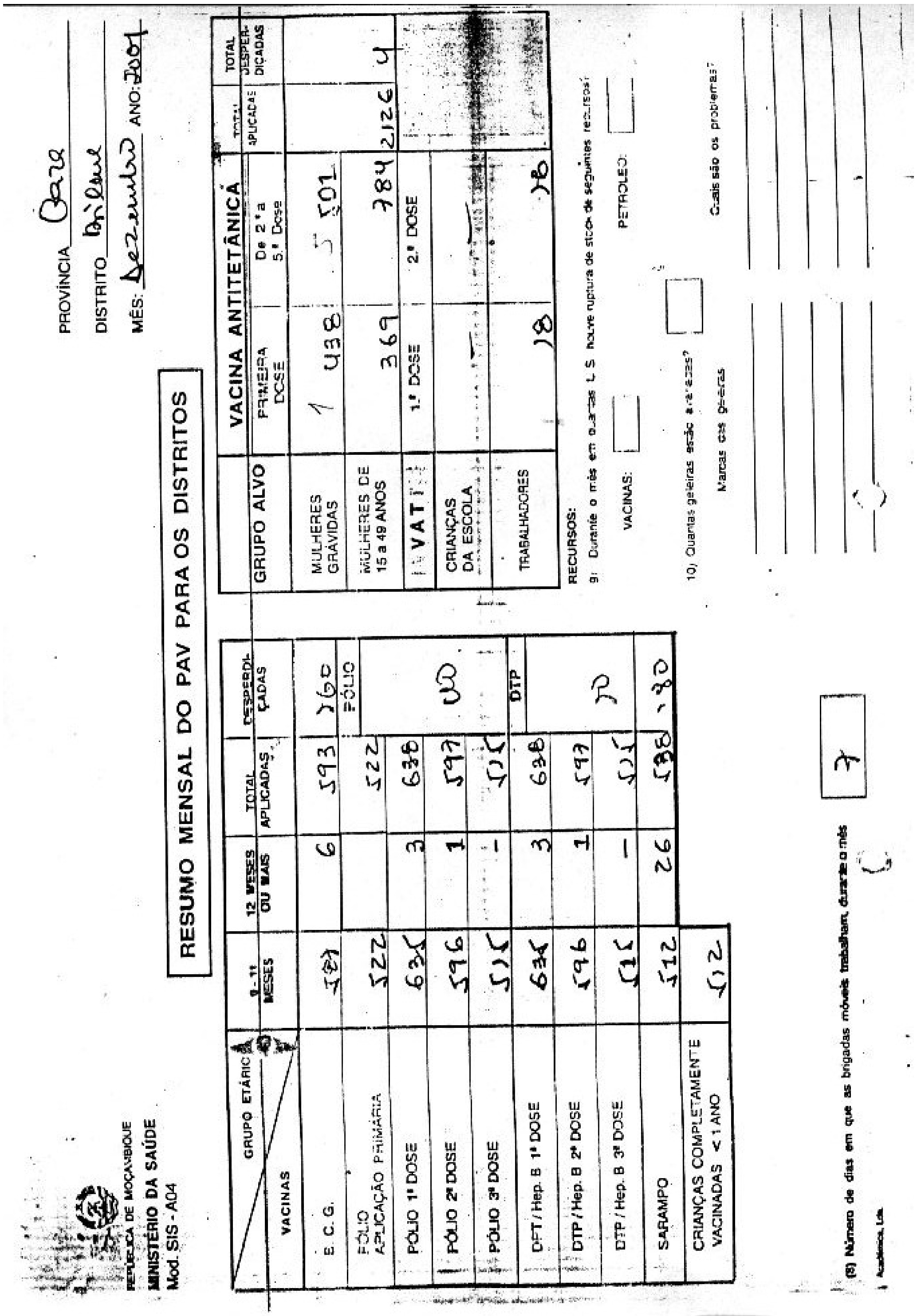
VACINAS: PETRÓLEO:

10) Quantas geleiras estão avariadas?

MODELO QUAIS SÃO OS PROBLEMAS?

16 Appendix C

Form A04 new



REPÚBLICA DE MOÇAMBIQUE
MINISTÉRIO DA SAÚDE
Mod. SIS - A04

RESUMO MENSAL DO PAV PARA OS DISTRITOS

PROVÍNCIA Gaza
DISTRITO Bilene
MÊS: Dezembro ANO: 2001

GRUPO ETÁRIO / VACINAS	0-11 MESES	12 MESES OU MAIS	TOTAL APLICADAS	DESPERDIÇADAS
B. C. G.	587	6	593	160
PÓLIO APLICAÇÃO PRIMÁRIA	522		522	PÓLIO
PÓLIO 1ª DOSE	635	3	638	
PÓLIO 2ª DOSE	596	1	597	110
PÓLIO 3ª DOSE	515	-	515	
DPT / Hep. B 1ª DOSE	635	3	638	DTP
DTP / Hep. B 2ª DOSE	596	1	597	70
DTP / Hep. B 3ª DOSE	515	-	515	
SARAMPO	512	26	538	80
CRIANÇAS COMPLETAMENTE VACINADAS < 1 ANO	502			

GRUPO ALVO	VACINA ANTITETÂNICA PRIMEIRA DOSE	De 2ª a 5ª Dose	TOTAL APLICADAS	TOTAL DESPERDIÇADAS
MULHERES GRÁVIDAS	438	501		
MULHERES DE 15 a 49 ANOS	369	784	2126	4
VAT	1ª DOSE	2ª DOSE		
CRIANÇAS DA ESCOLA	-	-		
TRABALHADORES	18	18		

RECURSOS:

9) Durante o mês em quantas L.S. houve ruptura de stock de seguintes recursos?

VACINAS: PETRÓLEO:

10) Quantas geleiras estão avariadas?

Marcas das geleiras

Quais são os problemas?

(8) Número de dias em que as brigadas móveis trabalharam, durante o mês: 7

17 Appendix D

From B06

REPÚBLICA DE MOÇAMBIQUE
MINISTÉRIO DA SAÚDE
Mod. SIS - B06

Provincia Gaza
Distrito Bilene
Mês Janeiro Ano 2001

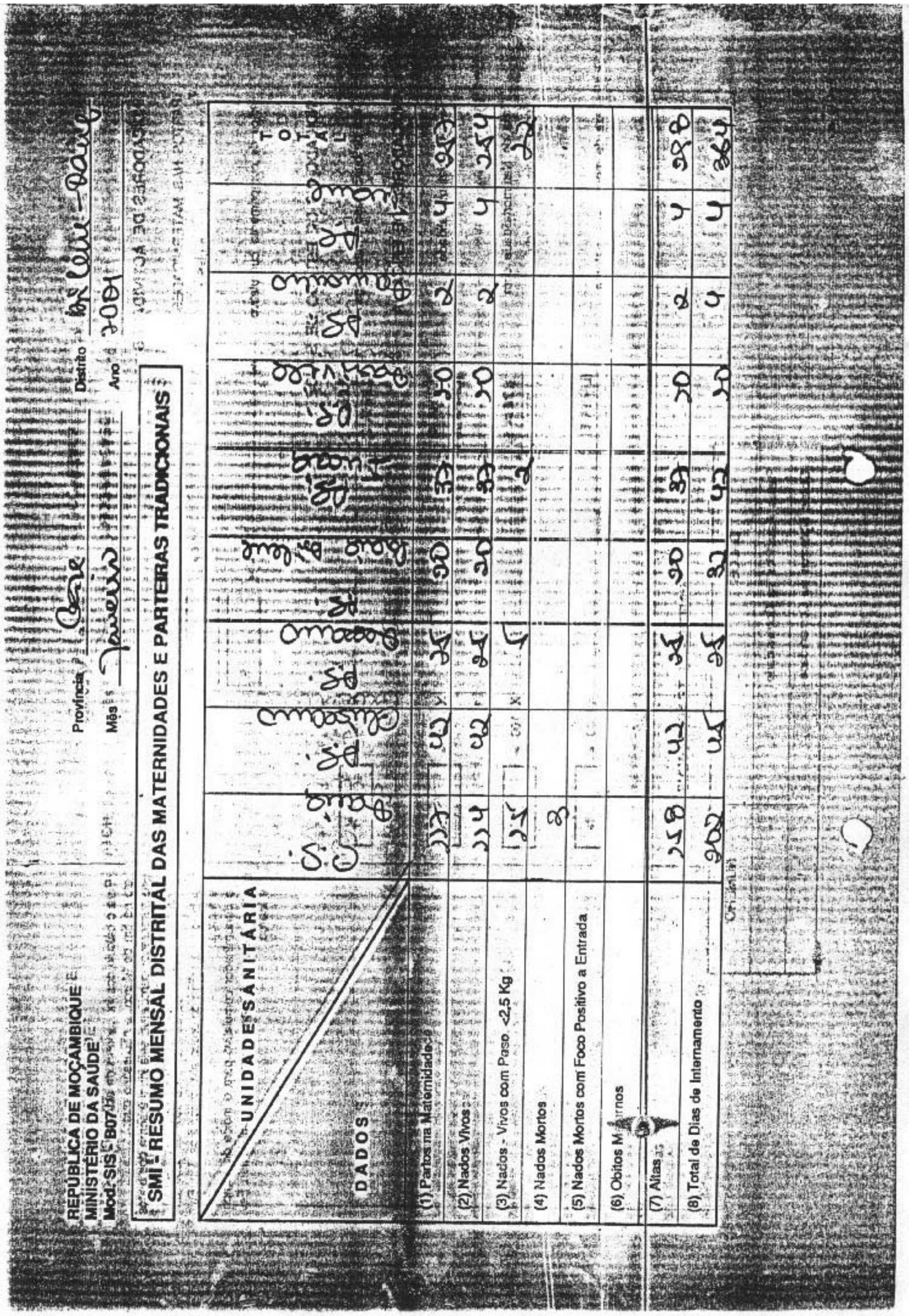
RESUMO MENSAL DISTRITAL - SMI CONSULTAS 0 - 4 ANOS/VIGILÂNCIA NUTRICIONAL

UNIDADE SANITARIA (1)	CONSULTAS 0 - 11 m (2)			CONSULT. 0 - 4 ANOS (3)			CONTROLE DE CRESCIMENTO (4)			
	1ªs Cons.	Cons Seg.	TOTAL	1ªs Cons.	Cons. Seg.	TOTAL	Bom	Mau	TOTAL	% Mau
C.S. Macie	2357	—	2357	2025	—	2025	1621	382	2003	19%
D.S. Chissano	2014	—	2014	753	—	753	1252	72	1324	5%
D.S. Messano	261	—	261	84	—	84	233	5	238	2%
D.S. Praia Bilene	291	—	291	221	—	221	335	4	338	1%
D.S. Incaia	464	—	464	241	—	241	673	7	680	1%
D.S. Mazivila	286	—	286	26	—	26	286	5	291	2%
D.S. Damonuho	220	—	220	96	—	96	113	3	116	2.5%
D.S. Tuane	57	—	57	73	—	73	134	2	136	1%
TOTAL	3540	—	3540	2435	—	2435	4647	480	5127	9%

RECURSOS: Número de U. S. em que acabaram os Cartões de Saúde durante o mês

18 Appendix E

From B07



REPÚBLICA DE MOÇAMBIQUE
MINISTÉRIO DA SAÚDE
Mod. SIS - B07

Província: Gaza Distrito: [illegible]
Mês: Janeiro Ano: 2001

SMI - RESUMO MENSAL DISTRITAL DAS MATERNIDADES E PARTEIRAS TRADICIONAIS

DADOS / UNIDADE SANITÁRIA	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	TOTAL
(1) Partos na Maternidade	227	42	25	20	37	20	2	4	[illegible]
(2) Nados Vivos	214	42	25	20	37	20	2	4	254
(3) Nados - Vivos com Peso <2,5 Kg	25		5		2				32
(4) Nados Mortos	3								
(5) Nados Mortos com Foco Positivo a Entrada									
(6) Obitos Maternos									
(7) Altas	258	42	25	20	37	20	2	4	288
(8) Total de Dias de Internamento	202	45	25	32	42	20	4	4	364

19 Appendix F

Form B08

REPÚBLICA DE MOÇAMBIQUE
MINISTÉRIO DA SAÚDE
Mod. SIS - B08

Província Gaza
Distrito Bilene
Mês Dezembro Ano 2011

RESUMO MENSAL DISTRITAL - SMI CONSULTAS PRÉ-NATAIS/POST-PARTO/PF

UNIDADE SANITÁRIA (1)	CONS. PRÉ-NATAIS (2)			1ªs CONS. POST-PARTO (3)	PLANEAMENTO FAMILIAR					
					DIU		PÍLULA		INJECTÁVEL	
	1ªs Cons.	Cons. Segs	TOTAL		Nº de Novas Utentes (4)	Nº de DIUs inseridos (5)	Nº de Novas Utentes (6)	Nº Ciclos Distribuídos (7)	Nº de Novas Utentes (8)	Nº Doses Aplicadas (9)
C.S. [illegible]	178	608	786	57	—	—	41	151	89	271
P.S. Chissano	148	280	428	35	—	—	61	103	5	54
P.S. Messano	46	94	140	8	—	—	19	33	19	23
P.S. Praia-Bilene	130	304	434	15	—	—	4	30	10	68
P.S. Incaia	115	156	271	39	—	—	10	35	13	35
PS Mazivila	31	70	101	3	—	—	17	38	—	—
PS Nhamanho	12	12	24	—	—	—	—	—	—	—
PS Tiame	12	15	27	—	—	—	—	—	—	—
TOTAL	732	1479	2211	157	—	—	152	430	136	451

RECURSOS:
Número de U.S. em que acabaram os seguintes recursos durante o mês.

PÍLULA	DIU	INJECTÁVEIS	PRESERVATIVOS	LUVAS
não	não	não	sim	não

20 Appendix G

Yearly 2000 reports from Ministry de Saude
Borrowed report from Ministry of Health. Report consists mostly of tables and graphs from Excel or Lotus.

On Page 1: " INDICES DE COBERTURAS E DE UTILIZACAO, 2000(Janeiro-Dezembro\'00)"
Page was showing PAV info. That is vaccination, table is A04, and columns checked are BCG1, DTPD1_1 and DTPD3_1.

Case: GAZA provincial database year 2000:
The National database for Gaza 2000 was empty (no rows). I checked the report against the provincial database.

Checked BCG in report:
- Report total Gaza=47.555.
- Provincial database sum (BCG1) =47.555.

Checked DTPD1_1 in report:
- Report total Gaza=49.047.
- Provincial database sum (DTPD1_1) =49.047

Checked DPTD3_1 in report:
- Report total Gaza=41.496.
- Provincial database sum (DPTD3_1) =41.496

Conclusion: All the three values in the report were correct in respect to the SISC provincial database, no check done against national because it was empty.

Case: Maputo_p (provincial) national database year 2000:
Checked against database at national level, full monthly range (1-12)

Checked BCG in report:
- Report total Maputo Provincial=21.542.
- National database sum (BCG1) =21.542

Checked DPTD1_1 in report:
- Report total Maputo Provincial=30.531.
- National database sum (DPTD1_1) =30.531

Checked DPTD3_1 in report:

- Report total Maputo Provincial=28.534.
- National database sum (DPTD3_1) =28.534

Conclusion: All the three values in the report were correct in respect to the SISC national database, no check done against provincial because I don't have it

Case: Niassa national and provincial database year 2000
Checked against database at national level, not full monthly range (1-9)

Checked BCG in report:
- Report total Niassa=49.388.
- National database sum (BCG1) =36.378
- Provincial database sum (BCG1)=49.388

Checked DPTD1_1 in report:
- Report total Niassa=41.666.
- National database sum (DPTD1_1) =32.130
- National database sim (DPTD1_1)=41.666

Checked DPTD3_1 in report:
- Report total Niassa=32.354.
- National database sum (DPTD3_1) =25.617
- Provincial database sum (DPTD3_1)=32.354

Conclusion: All the three values in the report were NOT correct in respect to the SISC national database, but they were correct in respect to the provincial one.

On page SMI- TOTAL DE CONTACTOS, Janeiro a Dezembra de 2000
Concerns SMI, deliveries. Table checked is B08. Checking total Consultas Pre-Natais, that is the sum (CP1) + sum (CS1)

Case: GAZA provincial database year 2000:
Checked Total de Consultas Pre-Natais in report
- Report total Gaza=193.088
- Provincial database sum (CP1) + sum (CS1)=193.088

Case: Maputo_p (provincial) national database year 2000:
Checked Total de Consultas Pre-Natais in report
- Report total Maputo Province=124.561
- National database sum (CP1) + sum (CS1)=124.561

Case: Niassa national database year 2000
Checked Total de Consultas Pre-Natais in report

- Report total Niasse Province=175.330
- National database sum (CP1) + sum (CS1)=120.983
- Provincial database sum (CP1) + sum (CS1)=175.330

Conclusion: Same as in the first case, the first two elements were correct, when discrepancies between national and provincial it is the provincial which is used.

21 Appendix H

Version: 1.2.
Date: 2003.04.22
Last edited by: Tore
Note: Some changes to text and layout made in 2003, this in order to make it more understandable for an outsider in order for it to be used in my Cand. Scient. thesis.

Plan for collecting and saving SISC database

This document contains information on what the Ministry of health in Mozambique can do to get and save the information in the SISC database currently being used at national and provincial level.

Why getting the data

The field study conducted in the period 2002.04.10 - 2002.05.17 has shown that the SISC national database is incomplete. I believe that the state of the national database is such that it is better to go to the provinces and get all the data there than to try to reassemble the national database. The reason for this is that the SISC national database is incomplete in several different ways.

1. Entire year folders are missing for several provinces.
2. Entire year databases are empty for several provinces.
3. Several databases are incomplete, several months are completely missing.
4. Some information is NOT available at all in the national database-

A project to try to complete the existing database will still contain just as much work because you will still need to get much old data from most of the provinces.

Some data may only be available in the National database; this data will most likely be the earliest year (1992-1996). But it should be fairly simple to merge that data in because it is complete years.

What to do

The process is divided in three parts:

1. *Get the databases*
2. *Merge the national with the provincial*
3. *Decode the columns names and save the database in a well know format*

Get the databases:

Travel to each provincial center and get the local SISC database. To do this you need to bring:

- A Zip drive with a parallel cable (do NOT bring one with USB as the PC in the provinces do not support USB)
- Drivers for the zip drive on a floppy (drivers should support Win95, Win98 and Win 2000. Win95 is the most important)
- 3 zip disk a 100 MB

OR

- Several floppy discs (60)
- A floppy disk with WinZip or other compression tool

In the province find the computer running SISC.

If you are using a Zip drive:

- Install the driver for the zip drive
- Copy the entire SISC folder to a Zip diskette, remember to put it in a folder with the province name as it name. I.e. “Niassa” data go to a folder called “Niassa SISC data”.

If you are using floppies:

- Install WinZip on the SISC computer
- Enter the folder of SISC
- Zip all the folders named (1992, 1993 … 2002). Copy the Zip file(s) to the floppies; you might need to use several floppies depending upon the size of the database. Name the Zip archives with the province name.

Note: The health facilities list in SISC is not complete. You might also want to clear out the names and which are active, this is a large job witch must be done at the provincial level because that’s where they know it. This is also a fairly complex job because the information is stored in the UNIDS table and the forms tables are mapping to it. So changes made to UNIDS table must also be reflected in the respective forms tables.

Merge the provincial with old national

The file structure for SISC at province is as follows:

1992
1993
1994
…2002

Each folder contains the database for that year (see “Content of SISC for information about the database”).

The file structure for SISC at national is as follows:
Niassa
- 1992
- 1993
- … 2002

Gaza
- 1992
- 1993
- etc.

Each folder contains the database for that year.

Now copy all the provincial databases to a hard disk on a computer. Structure the provincial data so that it is equal to the national database as described below.

Make a folder named SISC_data:

For each province database:

1. Make a folder with the provincial name in the SISC_data folder
2. Copy the provincial data into that folder, the provincial data is the folders with names as years. The rest of the files are for the SISC program and not necessary.

Now compare the provincial databases with the national one, if a provincial database is missing some year folders then copy the year folders from the national database, you might also check the content of the provincial database in case some are empty (see “How to view SISC database”).

When this is finished you should have the following file structure

SISC_data:
- Gaza
  - 1992
  - 1993
  - …2002
- Niassa
  - 1992
  - 1993
  - .. 2002
- …. The rest of the provinces

Decode columns names and save the data

The decoding of the columns names should be done now. It is impossible to parse and understand the data if the columns are not decoded. And we have been unable to find this coding documented anywhere. The only way to decode is by using the forms, and if the forms changes, then the data content will be lost (forever?).

Open each of the tables/files in Excel and change to column name to a name with more meaning. Here you may need to use the corresponding forms or use the mapping we used (see "Mapping"). For the year you must change the entry to the correct year, the year value is usually wrong, the correct one is the folder in where the database is stored. For the UNIDS and DISTS you may want to map the Unit name and District name into the tables to allow for easier viewing of unit name. DCOD and UCOD are mapping into the UNIDS and DISTS tables respective.

Note the column called SNUM is not important as it is a System column, several of the columns in the REG files are also system columns.

You may also want to convert the entry values to number, as this makes it possible to make graphs directly from the file. In XP, mark all the values, then click the exclamation mark and select "Convert to Numbers".

After the columns have been given new names save the table as Excel spreadsheet. This makes the data available for others, if you using Excel XP then you might want to save as XML to. XML is a new data format widely supported, and would allow people who are not using Microsoft technology also to view the data.

The files you now have made should be stored a place were they are available for people who wants to look at them.

Resource allocation

It is impossible for me to guess how much work it is to travel to the provinces. But the work of renaming the columns and reorganization the data should take 2 weeks to complete I think.

How to view SISC databases

Find the database you want to view, that is move to a year folder under a province. In this folder you have several files, but mostly they consist of the following types:

- *dbf: The file containing the actual data
- *cdx: System file.

The dbf files can be opened in Excel. Start Excel and select “Open”, set file type to dbase. OR right click the file in Windows and select “Open With”, then select Excel. You are now able to view the file, or table.

Content of SISC

The files in SISC are of the following types.

- **Table A04 is PAV data pr district pr. Month**
- **Table B06 is SMI data, children consultancies pr. Health unit, pr. month**
- **Table B07 is SMI data, deliveries (Maternidades) pr. Health unit, pr. Month**
- **Table B08 is SMI data, pregnant consultancies pr. Health unit pr. Month.**
- **Table B07_REG is SMI data, Deliveries pr. District pr. Month.**
- **Table C04_REG is Total consultas pr. health unit (Centro De saude).**
- **Table C04_HOSP is Total consultas pr. Hospital (unit of type HR/G)**
- **Table C05_REG is total consultas pr. Health unit (Posto De saude).**
- **Table J0 and J0_REG inventory, beds and health units**
- **K01 and K0_REG human resources and inventory**
- **C05 and C04 is Medical kits, use etc.**
- **Table DISTS is the district list for the province**
- **Table UNIDS is the health units list for the province**

Mapping

Mapping I used. “?” Is a sign for that the SISC columns has not been mapped.

Table B08:
PF12 = No de DIUs Inseridos
PF11 = No de Novas DIUs Utentes
PF32 = No de Doses de Injectávels Aplicadas
PF31 = No de Novas Injectávels Utentes
CP2 = 1as Consultas Post-Parto
CS1 = Consultas seguintes Pré-natais
CP1 = 1as Consultas Pré-natais
PF22 = No de Novas cíclos de Pilula distribuídos
PF21 = No de Novas Pilula Utentes

Table B07:
D8 = Total dias de internamento - Maternidade
D7 = Altas maternidade

D6 = Óbitos maternos
D5 = Nados - Mortos com Foco Positivo à entrada
D4 = Nados Mortos
D3 = Nados - Vivos com peso <2,5 Kg
D2 = Nados - Vivos
D1 = Partos na maternidade

Table B06:
CS2 = Consultas seguintes 1-4 Anos
CS1 = Consultas seguintes 0-11 meses
CCM = No Mau Crescimento 0-35 meses
CP2 = 1as Consultas 0-4 Anos
CP1 = 1as Consultas 0-11 meses
CS2 = Consultas seguintes 0-4 Anos
CCB = No Bom Cresimento 0-35 meses

Table B07_REG
PAC =?
OMC =?

Table C04_HOSP
CONSULTAS = Consultas

Table C04_REG
CONS_CENTR= Consultas
CONS_HR= Consultas
CONS_DIST= Consultas
QTD_APE=?
CONS_APE=?

Note, as the new system didnt seperate Consultas on Centro De Saude etc. Therefore we used this mapping. Consultas for a centro De Saude is obvious Consultas for a Centro De Saude and NOT Posto De Saude.

Table C05_REG
CONSULTAS = Consultas
QTD_APE=?
CONS_APE=?

Table A04:
OMC = Óbitos Maternos na Comunidade
BCG2 = BCG - 12 meses ou mais
BCG1 = BCG - 0-11 meses
POLD3_2 = Pólio 3 Dose - 12 meses ou mais
POL_DESP = Pólio - Desperdicadas
POLD3_1 = Pólio 3 Dose - 0-11 meses

PAP = Pólio applicacão primária (0-11 meses)
POLD2_2 = Pólio 2 Dose - 12 meses ou mais
POLD2_1 = Pólio 2 Dose - 0-11 meses
PAC = Partos na maternidade
DTP_DESP = DTP - Desperdicadas
POLD1_2 = Pólio 1 Dose - 12 meses ou mais
POLD1_1 = Pólio 1 Dose - 0-11 meses
SAR_2 = Sarampo - 12 meses ou mais
SAR_1 = Sarampo - 0-11 meses
BCG_DESP = BCG - Desperdicadas
SAR_DESP = Sarampo – Desperdicadas
VATCR_1 = 1 VAT Crianças de escola
VATCR_2 = 2 VAT Crianças de escola
VATTR_1 = 1 VAT Trabalhadores
VATTR_2 = 2 VAT Trabalhadores
VAT_DESP = VAT - Desperdicadas

A04 BEFORE JULY 2001
VATMG_1 = 1 VAT Mulheres grávidas
VATMG_2 = 2 VAT Mulheres grávidas
VATMG_3 = 3 VAT Mulheres grávidas
VATMG_4 = 4 VAT Mulheres grávidas
VATMG_5 = 5 VAT Mulheres grávidas
VATMA_1 = 1 VAT Mulheres de 15 a 49 anos
VATMA_2 = 2 VAT Mulheres de 15 a 49 anos
VATMA_3 = 3 VAT Mulheres de 15 a 49 anos
VATMA_4 = 4 VAT Mulheres de 15 a 49 anos
VATMA_5 = 5 VAT Mulheres de 15 a 49 anos
DTPD3_2 = DPT 3 Dose -12 meses ou mais
DTPD3_1 = DPT 3 Dose -0-11 meses
DTPD2_2 = DPT 2 Dose -12 meses ou mais
DTPD2_1 = DPT 2 Dose -0-11 meses
DTPD1_2 = DPT 1 Dose -12 meses ou mais
DTPD1_1 = DPT 1 Dose -0-11 meses

A04 AFTER JULY 2001
VATMG_1 = 1 VAT Mulheres grávidas
VATMG_5 = 2-5 Dose VAT Mulheres grávidas
VATMA_1 = 1 VAT Mulheres de 15 a 49 anos
VATMA_5 = 2-5 Dose VAT Mulheres grávidas
DTPD3_2 = DPT/Hep B 3 Dose -12 meses ou mais
DTPD3_1 = DPT/Hep B 3 Dose -0-11 meses
DTPD2_2 = DPT/Hep B 2 Dose -12 meses ou mais
DTPD2_1 = DPT/Hep B 2 Dose -0-11 meses
DTPD1_2 = DPT/Hep B 1 Dose -12 meses ou mais
DTPD1_1 = DPT/Hep B 1 Dose -0-11 meses

Special note for table A04: During the 2001 the forms have changed, this means that several of the elements are ambiguous today. This change should have happened around July 2001. But observation of the values shows that several districts did not change in July 2001, and there are no flagging months (a flagging month means that all district or facilities changed form that month), the different district are changing in different months. Columns affected are:

DTPDX_X, Before July 2001 these columns only stored DTP, after July 2001 they also stored Hep.

VATMG_1, VATMG_2, VATMG_3, VATMG_4 and VATMG_5
The VATMG_2 up to VATMG_5 have been merged into VATMG_2. After July 2001 the columns VATMG_1 and VATMG_2 have been containing all the VAT information. While VATMG_3, VATMG_4 and VATMG_5 have been left empty. Note that some provinces might have instead chosen to use VATMG_1 and VATMG_5, leaving VATMG_2, 3 and 4 empty. They same has also happened with the VATMA columns in A04.

Note. Some other forms have also had minor changes to the data elements, to learn more about this check a report made by the HISP team in 2001.

22 Appendix I

DHIS Import tool user manual

Author: Tore Skobba (torech@ifi.uio.no)
Version: 3.2
Last edited: 2003-02-20

Table of contents

1 Installing & use

1.1 Prerequisites

To run the DHIS import tool you need at least Java runtime 1.1.4. To get the Java
Runtime go to: http://java.sun.com/getjava/download.html and follow the instructions on the page. Note some computer already has Java runtime installed, if you do have Java make sure you are running Java 1.1.4.

The tool ONLY works when the DHIS have been set up with the correct infrastructure that is ALL the organizational units have been registered in DHIS. Furthermore the data dictionary must have been made. This is order so that the imported data can map to the correspondingly right values.

The user of the tool should known DHIS good; furthermore he/she should have the user manual ready and read the sections about importing data from other systems!

1.2 Installing

Simply unzip all files in zip file "dhisimportver1_1.zip" to a folder where you want to store the program.

1.3 Starting it

There are two ways to start the tool:

1: Go to the folder where you unzipped dhisimportver1_1.zip and click on the "dhisimport.jar" file, note the .jar file requires that the system know to use "java" to execute Java byte code.

2: Start the command prompt. CD to the folder you installed to, folder contains dhisimport.jar. Write "java –jar dhisimport.jar".

Note: Make sure you have installed Java runtime 1.1.4!! Note NONE of Microsoft's releases contains the Java runtime 1.1.4 as of 2003 January!

1.4 Special notes:

The program has been tested under Linux but it is not function perfectly there because inconsistency in Sun's java technology (it is unable to read config. files).

2 Use

In this chapter I will explain how to use the DHIS import tool. The Process can be divided in 3 parts.

- Opening a data source: Select the data source you want to import data from, and then assisting the program in understanding the data source.

- Select which type of DHIS import file to make, simple select if you want to make a monthly data import file or import file for organizational unit's data.

- Making the DHIS file, select which data you want to import, add default values and map values to DHIS names.

2.1 Opening a data source

A data source is a file or databases were the information you want to import is stored. An example of a data source is an Excel spreadsheet containing monthly data reports from different health units. Note you can only work on ONE data source at a time. There are currently three different data sources the programs understand.

- EPI 6 File: Data files made by EPI 6. They usually have a .rec suffix.
- Structured text file: A plain structured ASCII text file. For example an Excel spreadsheet saved as a text file.
- Database: Any database witch you have drivers for, note most databases are supported by the Microsoft ODBC driver with DHIS import tool support. So in 99 % of the incidents you don't need to think about drivers.

To select a data source type use the combo box in figured below, first select with type of data source, then click open.

Picture 16: Screenshot of the data source type functions

Now the process is different depending upon the type of data source you selected.

2.1.1 Database

To open a database you must first make a **data source in the operating system**. Note the data source here is NOT the same as in my program. The data source here is so that the operating systems understand the file! But when the operating system understands the file so does the Dhis import tool. Note that the data sources you make in the operating system are kept when Dhis import tool quits. You must manually remove them!

Making an data source in Windows
To make a data source click the ``Start" button and go to ``Settings.Controll Panel.AdministrativeTools.Data sources". For more help regarding data sources refer to ``Microsoft help" installed on your system. Remember the name of the data source you make.

Making an data source in Linux/Unix/Mac etc.
If you are running Linux, UNIX or Mac ask the local guru.

Open the data source in DHIS import tool
Append the data source name in the URL (PATH field), append means add the name AFTER "jdbc:odbc:" DO NOT EARSE IT. Changing the existing text is only for professional users. You can add Username and password if that is necessary for the data source, note that the driver field should NOT be changed unless you know what you are doing, it you change it by error, then click cancel and retry to open a database.

When finished press ok. If the windows closes then connection was ok, else check for errors and retry. This completes the opening of a database.

2.1.2 EPI 6 File

Simple locate the EPI 6 file, select it and then click ok.

2.1.3 Text Files & Excel files

To open a structured text file you must first check to see if the file is structured in a way so that the dhis import tool can understand the file. The file must have an element divider dividing the values and column names; furthermore all the records must have the same number of columns. There must also be no garbage in the tables.

When you edit data files make sure you are editing copies and NOT the originals!!!!!

Below here are examples of files NOT structured enough; note it is Excel for example only.

There must be no garbage mixed within the values/columns!

This file contains information about AIDS in Norway

	CityName	Number of incidents	Number of deaths
Ok	Oslo	523	4
Ask Tim	Bergen	267	2
Ok	Stavanger	152	0
FALSE	Trondheim	98	1

This file needs to be cleaned up by a human being before being used. The first column must be cleaned; there are values with no column header there!

This file contains information about AIDS in Norway

CityName	Number of incidents	Number of deaths	
Oslo	523	4	Ok
Bergen	267	2	Ask Rolf
Stavanger	152	0	Bad
Trondheim	98	1	Good

Same as the previously example, here the garbage is at the end column, and it must be removed. There must be no garbage at all mixed with the values, or garbage within the records.

Note there can be "garbage" in the ROWS before and after the values

Check this file Mark

This file contains information about AIDS in Norway

CityName	Number of incidents	Number of deaths
Oslo	523	4
Bergen	267	2
Stavanger	152	0
Trondheim	98	1

Some of this values look strange get feedback!

This is fine because the garbage in the beginning and end of the file is not mixed in with the values!

Each value and column header must have ONE cell

City	Number Of	Number of	Death causes		
Name	incidents	deaths	TB	Diarrhea	Other
Oslo	523	4	2	0	2
Bergen	267	2	1	0	1
Stavanger	152	0	0	0	0
Trondheim	98	1	0	0	1

This file would NOT be okay, because the column header is a result of several cells! And the program will not be able to parse the correct names out. To fix you would need to add the column names together. Se example down here:

City Name	Number of incidents	Number of deaths	Death TB	Death Diarrhea	Death other
Oslo	523	4	2	0	2
Bergen	267	2	1	0	1
Stavanger	152	0	0	0	0
Trondheim	98	1	0	0	1

Note all column names have one cell, furthermore the death causes have been moved down to each sub column.

Example of a typical structured file DHIS import tool can understand

This file contains information about Diarrhea in Norway
City;Number of incidents;Number of deaths
Oslo;1023;0
Bergen;452;1
Trondheim;;
Stavanger;258;0

The second line contains the column names, the rest of the lines after line 2 is the data with ``;" as element divider. Note that for Trondheim you don't have any data but the element dividers are still there.

If a text file is not structured enough then you must go in and structure it yourself, or find another way to import the data (not using Dhis import tool).

Excel files to text files

To easily be able to open an Excel file in DHIS import tool you would need to convert the file to a text file.

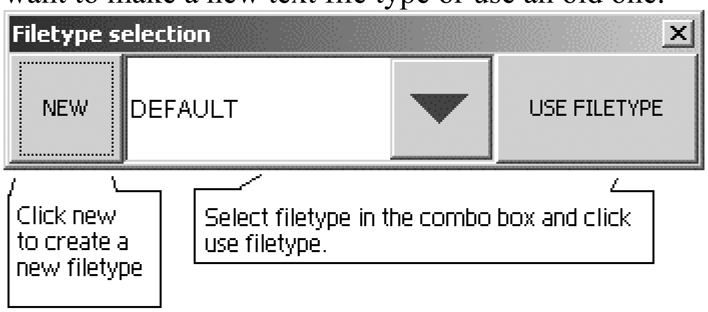
1. Open the file in Excel
2. Restructure it if it is not structured enough. Typically having the column names in the first row of cells, then the next rows of cells contains the values.

3. Save it as a text file, using ``TAB" or any other sign as the separator (element divider).

You may also create a data source of the Excel file but this demands that the structure is PERFECT! Just like a database.

Selecting the file
When you are ``sure" that the file is structured enough open it by locating it, selecting it and click ``ok".

Select Text file type
After you have selected the file, this dialog appears, here you are asked if you want to make a new text file type or use an old one.



Picture 17: Screenshot of the File type selection

Text files types is storing your previously work on file types (so that the programs learns what you have done before), but if this is your first time using text files in DHIS import tool you must select ``NEW".

When you press new the program starts a file analyzing wizard. If you select an old one the program tries to understand the file with the text file type selected. If it is unable you must analyze the file yourself, see the next sections.

Analyze first step
First task is simply to enter the name of the file type then press ``ok" (NOTE YOU MUST PRESS OK AND NOT ONLY CLICK ENTER ON KEYBOARD!) and then ``next". The name of you enter will be the name of a new file type witch stores all the information you do in the next step. This is useful if you have very many similar text files you want to import data from. Then you only analyze one file once and use the text file type for the rest of the files (instead of clicking new you just select the filetype and click “USE FILETYPE”). Note this requires that the files ARE built up exactly similar! Only the records might be different.

Analyze second step
Analyze the file. Here you must tell the program how the file is structured.

Button “Element Divider”: Click and then enter the string separating the values and column names. Tip: Mark and copy the text between the values and paste it into the input prompt.

Button “Line Divider”: Used to enter the string separating the lines. NOTE line divider is default “\n” and should NOT need to be changed. Only for very particular files should this be changed.

Button “Mark value start”: Mark out where the values start in the text field and click ``Mark Value Start".

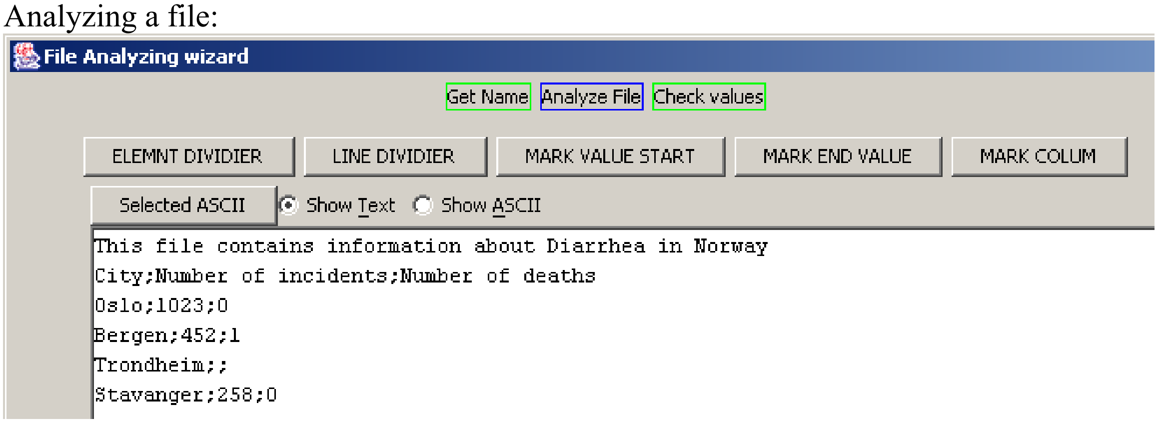
Button “Mark end value” Mark out where the values end and click “Mark value end", usually the end of the file.

“Mark Column" Mark out the entire columns names and press “Mark Column".

The rest of the buttons are to help with the parsing. You can use the radio buttons (“Show Text”, “Show ASCII”) to display the entire file with its ASCII values separated with |, or just mark out parts of the files and click “Selected ASCII”.

After you have finished analyzing click ``next". If the program still misses some information it will tell you what it lacks and won’t allow you to proceed. Note there may still be errors because the information the program has is wrong!

Analyze example
Analyzing a file:



Click “Element divider” and enter “;” in text field, click ok.
Mark out “Oslo” and click “Mark value start”.
Mark out “Stavanger;258;0” and click “Mark value end”.
Mark out “City;Number of incidents;Number of death” and click “Mark Column”.
Line divider is NOT used because the file already uses \n as line divider, as you can see because each line have a line carriage (i.e. not one loooong text).

Then hit next.

Analyze control
Simply look at the values and check to see if they look correct. NOTE DO NOT CHANGE THE VALUES AS CHANGES HERE ARE STORED IN THE PROGRAM,
BUT NOT IN THE ORGINIAL FILE! If everything looks correct then press "next" else press "back" and try to reanalyze the file.

If you now have finished everything, remember that for similar files use the text file type to automatically parse them!

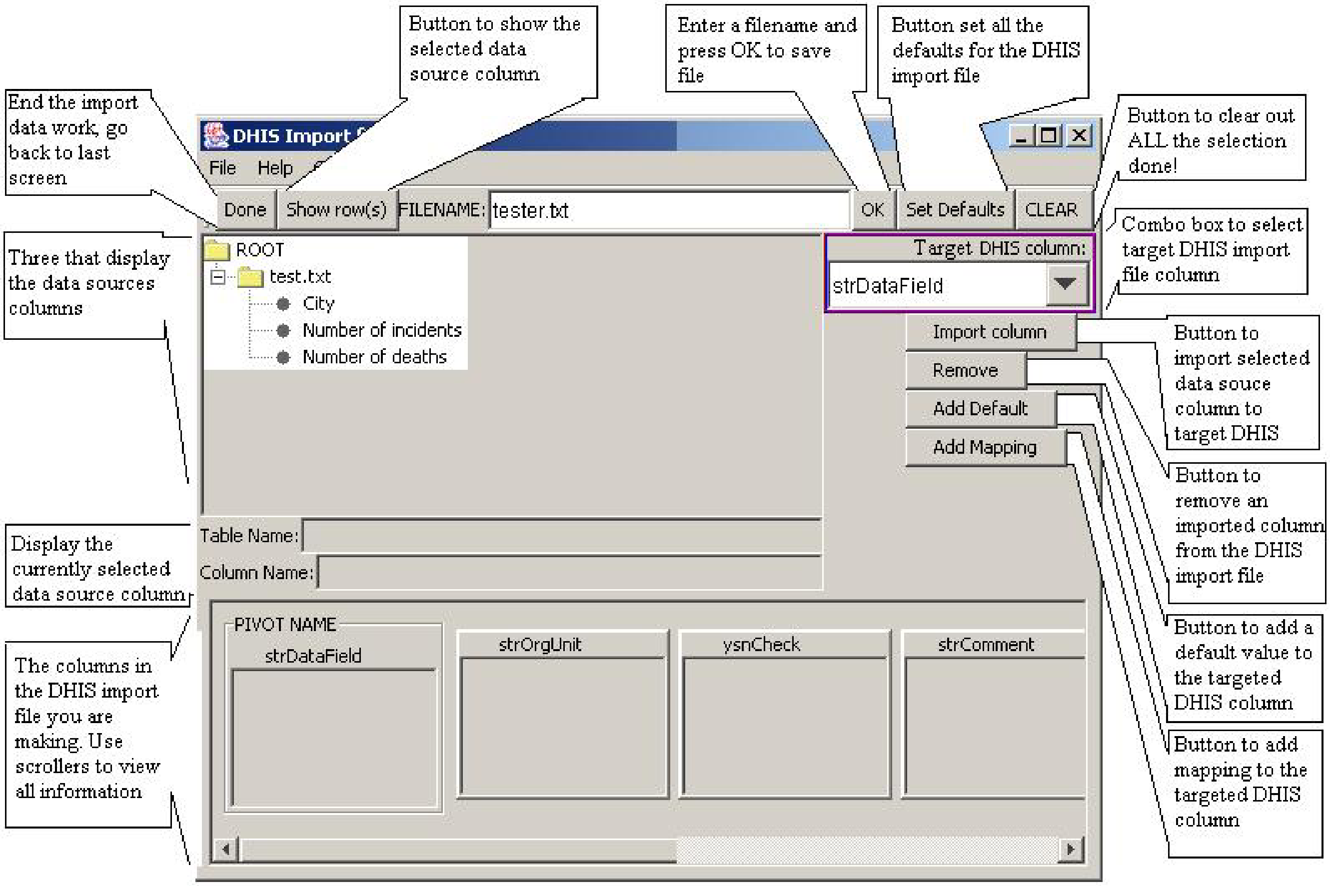
2.2 Selecting type of DHIS import file

The last section showed you how to open a data source, now that you have opened a data source (and parsed it if necessary) you must select what type of DHIS import file you want to make, simply select what type of DHIS import file you want to make in the DHIS type combo box and click the ``USE FT". Note that now only monthly data works. Org. unit. Data is NOT implemented.

2.3 Making DHIS import file

Now it is time to select which data to import into DHIS. This phase can be divided into three parts.

1. Select which column you want to import (typically the data elements you want, the reporting org unit and the date)
2. Set the default values and name the DHIS import file
3. Map the values



Picture 18: The screen where you make the DHIS import file

After you have selected which type of DHIS import file you want to make you will see this screen. The tree on the top left display the columns from the data source, when you click on one of the columns then table name and column name display the name of the column. When you have selected a column in the data source you can view the values by clicking “Show Row(s)”. And import the column into the DHIS import file which is displayed at the bottom (for description of what the different DHIS import file columns are check the DHIS Monthly data import file structure section or the DHIS manual).

2.3.1 Selecting column from data source for import/removing

The tree on the top left display the tables and underlying columns from the data source. In this case there is a text file called “test.txt” which is the data source. As supported text files only consist of several columns there is only one table with the same name as the file. If this had been a database there would have been several tables under the “Root” level. The selection of data to have in the DHIS import file is done by selecting a column in the data source by clicking on it, note the highlighting in the tree. Then setting the combo box “Target DHIS column” to the DHIS column you want the data moved to, and the click

"Import column" button. If everything is correct you will see that the targeted DHIS import file column box gets a line with the name of the data source and the column name, also note that the column name in the data source tree appends the DHIS column name. It is only possible to move a single column and not entire tables, furthermore the user must make source that the values in the data source columns are of the same type as the DHIS import file column, i.e. moving a String to the DHIS column "ysnCheck" would generate errors when the DHIS import file is made due to that "ysnCheck" only accepts Integer (0 or 1). To undo a movement select an element from anyone of the DHIS columns list and click "Remove". If the user needs to view the content of the original data source it is possible by selecting a column or the entire table and clicking "Show Row(s)".

For DHIS monthly data files you would import all the data elements to dblEntry, the date would go to Period, the reporting facilities name would go to strOrgUnit. The rest of the DHIS columns would usually be some default values.

2.3.2 Set default values and name the file

A Default value is a standard value that is filled out for all the rows in that column. For instance ysnCheck has a default value of 0. This means all new data entries have 0 in ysnCheck.

Default values are necessary because most data source does not store the same information about each data element as DHIS does, so you will need to fill out the rest with default values.

There are two ways to fill out default values.

Method 1 (Slow)
Select a DHIS import column in the "Target DHIS column", then click the "Add Default" button. Enter the default value in the text field that will be displayed. The default value must be of the same type as the DHIS import column (targeted). For example ysnCheck must have an integer! Any letter or special sign will be refused by the software.

Method 2 (Fast)
The fast method was made because usually you want to enter the same default value several times, therefore you can click the button "Set Defaults" which set default values to most of the DHIS import columns to the typically default values. If you want to change some of them afterwards, for example strUser, then remove the default value for strUSer and then add the new default value you want strUser to have (typically your name).

2.3.3 Map values

The representation of the values in the original data sources are usually not the same as in DHIS, sometimes values and column names are coded, For instance each org. unit is given a number as an identifier instead of their original name. While in DHIS each org. unit is stored with its original name as the identifier. If this is the case then you must tell the software what the original data source values map to.

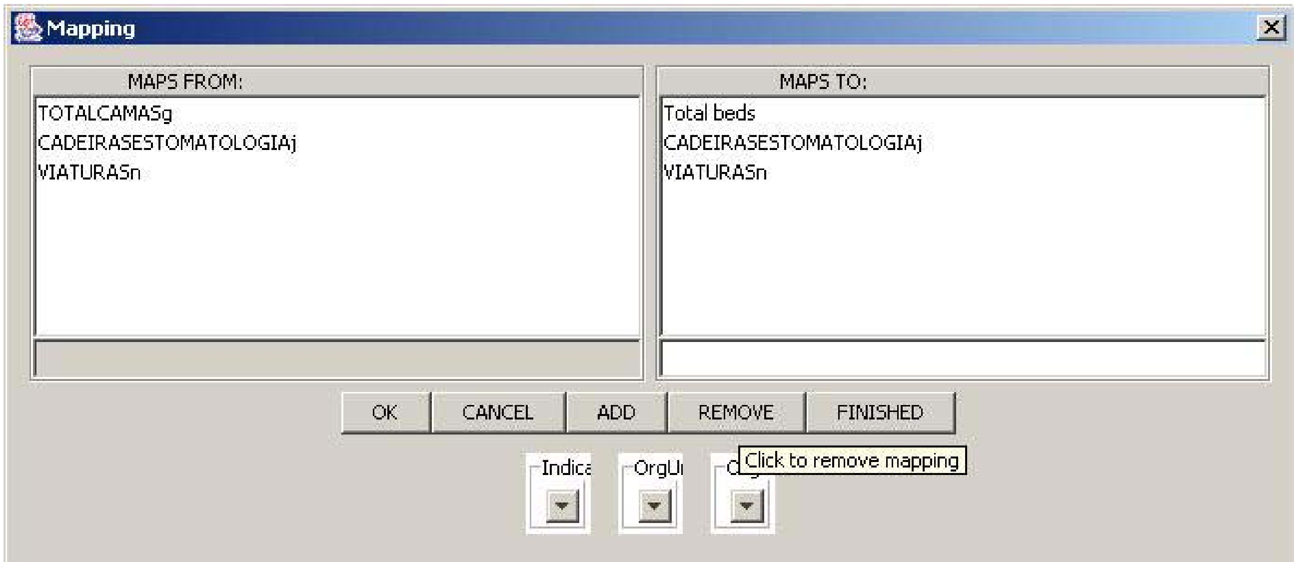
Since the mapping is from the original data source to the names in DHIS you may fetch the values in DHIS you want to map to, this will make the mapping procedure much easier, note you may only do this if you are on the computer which have DHIS installed or the DHIS database. To do this:

- Go to the “Config” menu
- Select “Get DHIS names”
- Enter/append the name of the DHIS data source (check the data sources on the computer) in the “URL (PATH)” text field. Take care not to remove any of the text there!
- Now locate the DHIS names you want to map to! In 99 % of the cases it well be:
    - Table=“tblMonthlyDataField”, column=”strDataField”
    - Table=”tblOrgUnit”, column=”strOrgUnit”
    - For each of these click “Get Column”. When you have gotten both of them, or the ones you want to map click “Done”.

Now you can start the mapping. Select which DHIS Column values you want to map. Typically in DHIS Monthly data file you would map strOrgUnit, strDataField and maybe Period. To begin mapping select the column you want to map in the “Target DHIS column”, then click “Add Mapping”. The following dialog appears:

Picture 19: Screenshot of the mapping selector

Either make a new mapping by entering a name in the text field or select an old one in the combo box. Note all the mappings you make are stored, so give it a good name, such as mappingfromSTDdatabase if you were mapping from a database called STDdatabase. The mappings are very often used several times! After this is done the mapping dialog appears.



Picture 20: Mapping dialog. Here TOTALCAMASg is mapped to Total beds, while the rest of the values are not mapped(they are mapping to the same as they are).

The screen above shows the mapping dialog. To map:

1. Select which value you want to map by clicking in the left most list ("MAPS FROM:").
2. Enter what it maps to in the right most text field. Then click "OK" or "CANCEL" if you clicked on it by an error. Note: You must click "OK" and not just press return!

Use "ADD" if you want to add a completely new mapping (that is a mapping which is not in MAPS FROM). "REMOVE" button removes a mapping, displayed by showing the same value in MAPS FROM and MAPS TO. "FINISHED" finishes the mapping dialog.

The tree combo boxes in the bottom of the dialog are the values you may have gotten from the DHIS database. These are only there depending upon if you got names from the DHIS database. To use them simple select which value you want to map, then select which value it maps to from the combo box, then press ok, that value is automatically filled out in the maps to text field.

2.4 Finished

Now that the file has been properly set up it is time to save it, write the filename you want it to have in the textbox at the top. Then press ok. If everything was okay you should now have a DHIS import file in the folder "importfiles" under the folder where you installed the DHIS import tool.

2.5 Special notes regarding mapping

To maintain the mappings, select config menu, then "New/Config mappings". Play around with the mappings, all actions should be easy to understand.

3 Special notes

3.1 Bad code

Due to some bad code by you need to restart the application from time to time, it does not free up all resources properly, and I don't know why.

3.2 Dbase ver. III.

I had problems using the tool with Dbase ver. III (In Mozambique). The problem was that the ODBC drivers didn't allow several quires at the same time to the database. The problem was solved by creating a temporary Access database and importing all the tables into that database, then making a data source for the Access database and using the DHIS import tool to connect to it.

3.3 Multiple columns is an data element

In Mozambique the key to a health facility was a combination of "District code" and "facility code". Thus in order to import data from facilities into DHIS the column strOrgUnit in the DHIS import file needed to contain both District code and facility code (two data elements from the old database). The way I solved this was to make a view/query in MS Access which selected both district code and facility code and the rest of the columns.

Example of Query from MS Access

```
SELECT DCOD+UCOD AS UNITCODE, *
FROM B08;
```

Where DCOD (District code) and UCOD (Unit code) is columns in table B08

3.4 DHIS Monthly data import file structure

This information is taken from DHIS help files.
Exporting data from other Information systems into the DHIS export format (ASCII)

The District Health Information Software (DHIS) from HISP uses a relatively simple ASCII text format for exporting and importing data. Other systems – e.g. electronic patient record systems – can relatively easily export their data in anonymous/aggregated form into this format. A DHIS export file looks as follows (example is from Abaqulusi Health District in KwaZulu-Natal):

```
"strDataField","strOrgUnit","Period","dblEntry","ysnCheck","strComment","intMin","intMax","ysnDisplay","strUser","dtmChanged"
"Adrenaline 1/1000 (1ml) vial","kz Vryheid Mobile 3","2001/08/01",10.00,0,,0.00,0.00,1,"jabu","2001/09/18 11:57:38"
"Amoxicillin 125mg/5ml suspension","kz Hlobane Clinic","2001/08/01",4.00,0,,0.00,0.00,1,"jabu","2001/09/18 09:35:50"
"Amoxicillin 125mg/5ml suspension","kz Swart Mfolozi Clinic","2001/08/01",75.00,0,,0.00,0.00,1,"Jabu","2001/09/07 10:59:58"
"Antenatal booked","kz Friesgewacht Clinic","2001/08/01",168.00,0,,0.00,0.00,1,"jabu","2001/09/18 09:16:49"
"Antenatal booked","kz Mondlo 2 Clinic","2001/08/01",193.00,0,,0.00,0.00,1,"jabu","2001/09/18 11:49:33"
"Antenatal out","kz Friesgewacht Clinic","2001/08/01",168.00,0,,0.00,0.00,1,"jabu","2001/09/18 09:16:49"
"Attendance at the health education sessions","kz Bhekezulu Clinic","2001/08/01",3.00,0,,0.00,0.00,1,"jabu","2001/09/18 11:47:27"
"Attendance at the health education sessions","kz Bhekumthetho Clinic","2001/08/01",4.00,0,,0.00,0.00,1,"jabu","2001/09/18 12:00:43"
"BCG at birth","kz Bhekezulu Clinic","2001/08/01",10.00,0,,0.00,8.00,1,"jabu","2001/09/18 11:47:27"
```

Note the tree key aspects of this format:

1. The first line contains the names of each field, and each subsequent line represent one DHIS data record (tblMonthlyData)
2. It is a comma-delimited file, and double quotes (“”) is used to denote TEXT values.
3. DATE values are in the format “yyyy/MM/dd”.

Also note the following about each field exported:

1. “strDataField” contains the NAME of the Data Element. If the equivalent Data Elements in the external information system is called something different, you must use a lookup table to TRANSLATE those Data Element names into the DHIS names before exporting to ASCII.

2. “strOrgUnit” contains the NAME of the Facility. If the same facility is called something else in the external information system, you must use a lookup table to TRANSLATE those Facility names into the DHIS names before exporting to ASCII. If the DHIS data file uses a provincial prefix, do not forget to include that in the translation.

3. “Period” is a date – in format yyyy/MM/dd – use to represent the reporting period. The DHIS is ALWAYS using the FIRST DAY of such a period to represent the period. In other words, “2000/02/01” represents February 2000 with monthly data and “2001/07/01” represents 2001 Third quarter with quarterly data.

4. “dblEntry” is the entry value of the record, represented as a number with 2 decimals.

5. “ysnCheck” represent the value in the “check-it” tick-box in the DHIS (values used for true/false in Access is 1 for True and 0 for False). NOTE: You can safely use a default of 0 for this field.

6. “strComment” represent the value in the “comment” field in the DHIS. NOTE: You can safely leave this blank, but you COULD also consider inserting a standard sentence here, like “Imported from TBSYS” or “Imported from CardReader system”.

7. “intMin” and “intMax” contain the minimum and maximum range values. This is basically values only used in the DHIS, so you can safely use default values of 0 for these two fields. NB: if you don’t use 0 (zero) for both, at least make sure that intMax is greater than or equal to intMin for all records!!

8. “ysnDisplay” contains the value in the YsnDisplay tick-box in the DHIS range table (tblMonthlyDataFieldRange). Since you are obviously capturing and exporting data for these data elements, use a default of 1 (True) for this field.

9. “strUser” contains the NAME of the user that captured or last edited the record exported. If the source system keeps such data, its values can be put into this field. If not, use a default value that identifies the user somehow (e.g. in the example above, we could have used “Abaqulusi” or “KZN” as a default user name).

10. “dtmChanged” contains the DATE of capture/last editing. Again, if the source system keeps such data it can be exported into this field. If not, use the function “Date()” or equivalent to put in the current date (when the data was exported) to the DHIS.

So each line from line 2 and downward represents one data record, and each line have 10 fields only.

WARNING:

1. Do NOT include any empty lines at the end of the export file
2. Avoid duplicates in the export file – the combination of strDataField, strOrgUnit, and Period must be unique.

The quickest and easiest way to test an Export file is to import it into the DHIS – the DHIS will reject all funny records or duplicates.

4 Definitions

DHIS Import: The program I have made
DHIS file: Import files for DHIS: MonthlyData and OrgUnitData
DHIS file columns: Means a column in a DHIS file, example the DHIS file "MontlyData" has the DHIS column "IntMin"
DHIS values: Means if not otherwise specified the values in the DHIS columns
Data Source: The data source you want to import into DHIS files
Get Data (GD): Another name for data source
Data Source Columns: The columns in the data source
File type: Text files of similar types, meaning that the same rules can be applied to them to parse out the information, the set of rules is called File type.

5 Code comments and docs

The src is available and you are very free to do whatever you want to do to it. The copyright added in some classes are just standard which I haven't bothered to remove. The docs are made with java doc and are viewed with a standard html viewer.

5.1 Files used

Files used by the application are stored in the following folder:

- importfiles: The DHIS import files the application makes
- config: The config files, typically how the DHIS files are built up. Read the code and read me file in the folder for more information
- stored: Store system files which are storing the mapping and text file types you have made. For the file structure check the code.

23 Appendix J

CD containing the tool made during this thesis, as:

- Java virtual machine executable file, a jar file
- Source code as .java text files
- Standard Java doc Html documentation of the code and classes used.

Also included are the user manual for the tool, Java JDK 1.4.1 runtime and SDK environment for both Windows and Linux operating systems (1.4.1 or newer Java version is mandatory to run the tool, in order to check which Java version installed do “java –version” in shell).