The Impacts of Legacy Information Systems in Reporting Routine Health Delivery Services: Case Studies from Mozambique and Tanzania

Juma Hemed Lungo
Dept. of Computer Science
University of Dar es Salaam
Dar es Salaam, Tanzania.
jlungo@udsm.ac.tz

and

José Leopoldo Nhampossa
Dept. of Mathematics and Informatics
Eduardo Mondlane University
Maputo, Mozambique
leopoldo@nambu.uem.mz

Abstract
The awareness of the importance of effective health information systems (HIS) has increased substantially and is reflected by many ongoing efforts of HISs reform in many developing countries. However, the one mostly mentioned obstacle for health information systems reform is *Legacy Information Systems* (LIS). The impacts of LIS in the reporting of routine health delivery services were studied in a participatory action research using case study sites in Tanzania and Mozambique. LIS impacts are on the process of introducing changes (reforms) on the HIS, and on everyday functioning of the HIS. LIS were determined to cause poor quality of health data, incomplete reporting of health data, and burden to health workers. The study recommends the Ministries of Health to relinquish the LIS. A demonstration on extracting and loading of locked health data on LIS to new health information software using extraction transformation and loading (ETL) software was performed.

Keywords: Legacy information systems, extraction transformation and loading systems, installed base, health information systems reform.

1.0 Introduction: the importance of health information systems

The world health organization (WHO) has long identified health information systems as critical for achieving health for all. A report of a WHO meeting (1987) clearly links improved health management to improved health information systems as it argues that, "of the major obstacles to effective management, information support is the one most frequently cited" (WHO 1987). The rationally for addressing health information systems (HIS) is, HIS generate information in order to inform health planners and decision-makers on what is happening at health delivery facilities. In this way, health information systems improve health management and health management is a pre-requisite for good health delivery services.

Health information systems are also seen as social systems implemented in the health sector. Braa *et al.* (1999) describe health information systems as complex systems because they tend to be deeply embedded in social work practices. Braa (1999*)* argues that, "working with data and information within the health sector includes filling in forms and registers, collating data into aggregated forms, statistics and reports and the reporting of these to higher levels are tasks that make up important aspects of most health workers' jobs" (Braa *et al.* 1999, p.4). Health information systems are complex systems and are social systems as they tend to be deeply embedded in social working practices. Lippeveld and Sapirie advise a successful way of designing and implementation of health information systems, by arguing that "the success of a health information systems reform depends not only on technical improvements but also on in-depth understanding of political, socio-cultural, and administrative factors" (Lippeveld and Sapirie 2000, p.249).

Despite the potential that health information systems have, in practice the collection, compilation, analysis, and reporting of health data are riddled with major problems, especially in the context of developing countries. Most health care providers in developing countries equate information systems with filling endless registers, collating, and compiling health data, conducting minimum data analysis, and sending out reports without receiving adequate feedback (Sauerborn and Lippeveld 2000). Furthermore, the data received are often not helpful for health management decision-making because they are incomplete, inaccurate, untimely, obsolete, and unrelated to priority tasks and functions of local health personnel (Braa *et al.* 2001; Sauerborn and Lippeveld 2000; WHO 1987). In other words, information systems tend to be data driven instead of action driven (Sandford *et al.* 1992). Wilson and Smith (1991 cited Wilson 2000) suggest that, "the creative use of microcomputer technology is one of the most promising means of improving the quality, timeliness, clarity, presentation, and use of relevant information for primary health care" (Wilson, 2000, p. 199). Recent experience (Braa and Hedberg 2002; Wilson 2000; Wilson *et al.* 2001) attests to the potential for using computers in health information systems.

Many bottlenecks in the development and implementation of effective IS have been identified by many researchers from different developing country contexts and includes the centralized and fragmented character of services, lack of coordination, poor quality and use of information, and the complex organizational context (Avgerou and Walsham 2000). Poor focus on the development of local expertise on the part of donor initiated projects and the tendencies of neglecting of social and organizational issues are cited as factors contributing to the problem of ineffective implementation of IS in developing countries (Lippeveld et al. 2000; Litlejohns et al. 2000). Implementation of IS in developing countries is a complex and very challenging task as the process demands not only a transfer of technology but also the introduction of the culture that go with the system. As Heeks (2002) citing Shields and Servers (1989) points out what is transferred is not only machines, hardware, software, skill and knowledge but also the attitudes, the value systems together with the social, political, and cultural structures. While it may be relatively easy to transfer the technical artefacts, it is far more complex to "transfer" the socio-cultural context to other settings. Braa et. al. 1995 concludes that like all other technologies, IT is also context sensitive and ensuring technological learning is crucial to its successful transfer to developing countries.

2.0 Legacy Information Systems

Hanseth 2002 describes a theory which views information systems as parts of larger infrastructure, which comprise of heterogeneous components that are integrated through standard interfaces to provide shared open resources to a community of users. The emphasis of Hanseth 2002 on shared resources contrasts very much with the concept of ISs, which are normally used relatively independently and constitute private properties.

Information Infrastructure evolves over time as new infrastructures are designed as extensions and improvements of existing ones. In turn, the new or improved elements have to link with the old and what is described as the existing installed base (Hanseth 2003), which heavily influences how the new elements can be designed. As the installed base grows, its development and further growth become self-reinforcing, both enabling and constraining further development. Successful development of information infrastructure requires, first, the creation of such a self-reinforcing process, and second, managing its direction.

Many of today's computer systems, used in applications ranging from corporate accounting to air traffic control, were created decades ago, and over the years were patched and fine-tuned to perform their jobs. Sommerville (2001), argues that,

> "*Many computer software in large information systems remain in use for more than 10 years and are still business-critical, that is, the business relies on the services provided by the software and any failure of these services would have a serious effect on the day-to-day running of the business*" (Sommerville 2001, p.582).

Sommerville describes legacy information systems as "socio-technical computer-based systems, that include software, hardware, data, and business processes" (Sommerville 2001, p.583).

Legacy information systems are typically too slow, unreliable, and inflexible for handling new, more diverse and demanding tasks. Unfortunately, the functions of these systems are very difficult to understand, and their replacement with a new and efficient designed system seems virtually impossible. Replacing a legacy information system is a risky business strategy for a number of reasons (Sommervile 2001): there is rarely a complete specification of the legacy information system. The original specification may have been lost. Therefore, there is no straightforward way of specifying a new system, which is functionally identical to the system that is in use. Business processes and the ways in which legacy information systems operate have been designed to take advantage of the software services and to avoid its weaknesses. If the system is replaced, these processes will also have to change, with potentially unpredictable costs and consequences. Important business rules may be embedded in the software and may not be documented elsewhere. New software development is itself risky, so that there may be unexpected problems with new system. It may not be delivered on time and for the price expected.

In describing problems of running legacy information systems, Sommerville (2001, p.583) points to the following expenses in changing legacy information systems: Different teams have implemented different parts of the systems. There is, therefore, no consistent programming style across the whole system.Part or all of the system may be implemented using an obsolete programming language. It may be difficult to find staff who have knowledge of these languages and expensive outsourcing of system maintenance may be required.System documentation is often inadequate and out of date. In some cases, the only documentation is the system source code. Sometimes the source code has been lost and only the executable version of the system is available. Many years of maintenance have usually corrupted the system structure, making it increasingly difficult to understand. The data processed by the system may be maintained in different files, which have incompatible structures. There may be data duplication and the data itself may be out of date, inaccurate, and incomplete.

The research questions were,

> "***what are the impacts of legacy information systems in the reporting of routine health delivery services in the health sector?. We further wanted to study the nature of legacy information systems in the health sector in developing countries and to find out on how do these legacy systems impact upon: the process of introducing reform, and the everyday functioning of the health information systems***".

Thinking about taking action to leverage legacy systems, Chislenko (1995, pp. 2-3) has advised five techniques: *Parallelism and Specialization* where the increased responsibilities of a legacy information system are divided among a number of old systems. The work is substantially improved as individual systems are optimised for performing particular tasks and relieved from other duties. *Redundancy* where several systems work in parallel then the result is compared to make the output more reliable. *Wrapping* where the layers of the system that cannot be understood are left alone while the others are replaced. *External aids* technique deals with providing the legacy system with necessary resources, preprocessing them for the input, and performing some tasks the old system is not good at. Finally, *replacement of parts* technique in those cases where the structure and function of some of the part of the system is well understood. The part can then be directly replaced with its improved equipment (Chislenko 1995, pp.2-3).

Although the above approaches by Chislenko (1995) proved useful in updating many computer systems, these have proved to be a temporary solution and sometimes magnify the problem. In his paper labelled *Reengineering work: do not automate, obliterate*, Hammer (1990) argues,

> "*It is time to stop paving the cow paths. Instead of embedding outdated processes in silicon and software, we should obliterate them and start over … use the power of modern information technology to radically redesign our business processes in order to achieve dramatic improvements in their performance*" (Hammer 1990, p.104).

The best option is to replace the legacy information systems with new systems. This is because it is risky to run legacy systems as outlined in the earlier discussion and because since legacy systems were developed in old technologies, as time goes, the hardware and software will fail. However, replacing the legacy systems is also a risky activity as it was presented in the earlier discussion, but this will ensure the sustainability of the organisation, as the new systems are implemented in modern technologies. While developing a new system to replace the legacy one, the most risky aspect is to loose organisation data collected for several years. The question is how the vast amounts of data locked in legacy systems can be secured and migrated to the new system.

2.1 Migrating data from legacy information systems to a new information system

Instead of adding patches to the old system as discussed in section 2.4.1, a guaranteed solution is to implement a new system and migrate all the data from the legacy system to the new system. This is common practice in building data warehouse systems as these systems aim at creating an enterprise reservoir of data, that is, integrate all operational systems and store their data in one place, the data warehouse. The process of migrating data from one system to another has a known technical terminology *Extraction, Transformation and Loading (ETL)* (Microsoft 2000).

While ETL can be done manually through "copy and paste" for a simple problem, it is impossible to migrate data from one database to another manually. The alternative is to automate the ETL processes by developing an application software system. Microsoft (2000, p.2) outlines four distinct functional elements of an ETL system: extraction, transformation, loading and meta data whereas; the ETL extraction element: is responsible for extracting data from the source system. During extraction, data may be removed from the source or a copy made and the original data retained in the source system. The ETL transformation element: is responsible for data validation, data accuracy, data type conversion, and business rule application. It is the most complicated of the ETL elements. The ETL loading element: is responsible for loading transformed data in the target system, and the ETL meta data element: is responsible for maintaining information (meta data) about the movement and transformation of data. It also documents the data mapping used during the transformations.

Developing an ETL system seems to be the most feasible solution for leveraging legacy database because it gives users an opportunity to implement new technologies, without worrying of losing their data. In this study we have attempted to develop a software that has extracted vast amount of health data from the old computer database of the health information system of Mozambique to their new and modern computer database.

3.0 Health information systems reform

There have been many problems reported on the performance of health information systems in developing countries (see e.g. Braa et al. 2001; Lippeveld et al. 2000; McLaughlin 2001; Simwanza and Church 20001). In the 1990s, many developing countries have been engaged in restructuring their health information systems. In Tanzania for instance, the former President of the United Republic of Tanzania His Excellency Ali Hassan Mwinyi in a speech presented on 8th June 1990 was quoted as saying, "we need to improve our health information systems in order to enable individuals and the government to make sound decisions based on correct information" (MoH 1993, p. i). Similar health information systems reform efforts have been reported from South Africa, Zambia, Uganda, and Mozambique (see Braa and Hedberg 2002; Mwaluko et al. 1996; Simwanza and Church 2001). The important health information system reform in Mozambique occurred after independence in 1975. However, those efforts were deteriorated by the 16 years (1976 - 1992) civil war. Mwaluko *et al.* describe socio-economic reforms in Mozambique that, "with the peace agreement in 1992 and the democratic elections in 1994, the new government designed an economic, and social programme, emphasizing National Reconstruction and Rehabilitation of Economic and Social Infrastructures" (Mwaluko *et al.* 1996, p.4). Peace time in Mozambique has made it possible for people to re-establish effective communication with the population in areas formerly cut off by the war and to collaborate with the communities in planning and implementation of programmes intended to help them to improve their lives.

While restructuring health information systems, many countries focus on decentralising their systems to empower the lower levels in the HIS hierarchy. According to Muquingue *et al.* (2002), national health information systems are built up from the informational activity carried out in multiple, minuscule, often hierarchically insignificant points in the geographical structure of a country; these points are generally districts. The administration structure of many developing countries includes the community (village), district, provincial and national levels. The national health information systems in many developing countries have been strongly based on Primary Health Care (PHC) and the district becomes the most appropriate level for co-ordinating top-down and bottom-up planning, for organising community involvement in planning and implementation, and for improving the co-ordination of government and private health care (WHO 1987). Being the information and physical hub between the community and the national health information system, the district consists of a large variety of interrelated elements that support the health system in a specific geographical area. A district includes the health care workers and facilities, up to and including first and second referral hospital levels (Amonoo-Lartson *et al*. 1984).

4.0 Application of IT in HIS

As many HIS reflect reform processes, the drive for the reform coincided with a revolution in information and communication technology (ICT). The computer has knocked on the doors of even the most reluctant Ministries of Health, like the computer import ban of Tanzania (Splettstoesser and Kimaro 2000). With this state of the art of technology combined with pressure from the computer industry, most HIS restructuring is featured by computerisation to a certain degree.

The computer database system of the Ministry of Health in Mozambique (known by its Portuguese acronym SISDB - Sistema de Informação de Saúde Base de Dados) was developed and deployed at the national and provincial health directorates in 1992-1994. The database was developed on "dBase III" relational database management system (DBMS). In Tanzania, the first national computer database of the Ministry of Health was developed in Mid 1990s on "dBase IV". However, due to cumbersome operations of that dBase IV database, the Ministry of Health in Tanzania decided to scrape it away and introduce a new database developed on Microsoft Access 97 (known by its Swahili acronym MTUHADB - Mfumo wa Taarifa za Uendeshaji wa Huduma za Afya Database) in the year 2000 (Lungo 2003a). MTUHADB is installed in all Regional Medical Offices in Tanzania.

In this study, the threats of LIS in the reporting of the routine health delivery services in the HISs were examined using case study sites from Tanzania and Mozambique. The authors are pioneers and active members of a broad participatory action research, Health Information Systems Project (HISP), in their respective countries of origin. HISP started in South Africa in 1996 and has spread to other countries including Mozambique, Tanzania, India, and Cuba

(Braa *et al.* 2003). In Mozambique, HISP started in 1999 where the DHIS software is piloted in thee provinces, Gaza, Niassa, and Inhambane. A memorandum of understanding between the University of Oslo and university of Dar-es-Salaam in collaboration with Ministry of Health in Tanzania was signed in July 2002, which form the contractual agreement of the implementation of HISP project in Tanzania.

HISP is featured by open source software, District Health Information Software (DHIS), designed for being used at district levels as a health data analysis tool (Braa and Hedbarg 2002). The argument is that districts should be empowered to be able to analyse and interpret health data. For the districts to be able to do so, HISP argues for a strong decentralisation of the health information systems to the district and sub-district levels. In addition, a computer-based database system should be implemented at the district level to facilitate better storage, analysis, and dissemination of health data. This is important because in many developing countries, Tanzania and Mozambique being examples, their respective health information systems use paper-based databases at the district levels.

5.0 Research Methodology

This study follows under the paradigm of "Action Research Paradigm". Action research has been typified as a way to build theory, knowledge, and practical action by engagement with the world in the context of practice itself (see, e.g. Kock 1997; Whyte *et al*. 1991). Dick (2002, p.1) explains an action research as a research approach, which has the dual aims of action and research:

- action to bring about change in some community or organisation or program;
- research to increase understanding on the part of the researcher or the client, or both.

In this study, five phases of action research approach are adopted, as Baskerville and Wood-Harper (2002, p.133) argue that, "the most prevalent description of action research details a five phase, cyclical process which can be described as an 'ideal' exemplar of the original formulation of action research". This ideal approach first requires the establishment of a 'client-system infrastructure' or research environment. Then, five identifiable phases are iterative: (1) diagnosis, (2) action planning, (3) action taking, (4) evaluating, and (5) specifying learning. The key assumptions of action research are that "social settings cannot be reduced for studying and that action brings understanding" (Baskerville 1999, p.7).

Diagnosing: Baskerville and Wood-Harper (2002, p.134) describes diagnosing phase as it "corresponds to the identification of the primary problems that are the underlying causes of the organisation's desire for change". Diagnosing involves self-interpretation of the complex organisational problem to develop certain theoretical assumptions about the nature of the organisation and its problem domain.

Action Planning: After the diagnosing phase, researchers and practitioners then collaborate in the next activity, action planning. The discovery of the planned actions is guided by the theoretical framework, which indicates both some desired future state for the organisation, and the changes that would achieve such a state. The plan establishes the target for change and the approach to change.

Action Taking: This phase implements the planned action. The researchers and practitioners collaborate in the active intervention into the client organisation, causing certain changes to be made.

Evaluating: After the actions are completed, the collaborating researchers and practitioners evaluate the outcomes. Evaluation includes determining whether the theorised effects of the action were realised, and whether the effects relieved the problem. Where the change was unsuccessful, some framework for the next iteration of the action research cycle (including adjusting the hypothesis) should be established.

Specifying Learning: The knowledge gained in the action research (whether the action was successful or unsuccessful) can be directed to three audiences (Baskerville, 1999).

> First, the restructuring of organisational norms to reflect the new knowledge gained by the organisation during the research, and second, where the change was unsuccessful, the additional knowledge may provide foundations for diagnosing in preparation for further action research interventions. Finally, the success or failure of the theoretical framework provides important knowledge to the scientific community for dealing with future research settings (p.14).

Action research aims to contribute both to the practical concerns of people in an immediate problematic situation and to the goals of social science through joint collaboration within a mutually acceptable ethical framework. The ideal domain of action research is therefore revealed in three distinct characteristics (Baskerville and Wood-Harper 2002 p.136) of the approach: The researcher is actively involved, with expected benefits for both researcher and the research client, in this case the district medical officer. The knowledge obtained could be immediately applied. The research is a cyclical process linking theory and practice.

The action research approach was important in this study because we wanted to attempt to extract health data from old systems so that health systems can migrate to new and modern computer systems with their historical health data.

6.0 The research design and methodology

The research design strategy was to perform case study in multiple sites. The study was conducted in two countries, Tanzania and Mozambique. In each country, the study involved several health units as case study sites. Three-health information software were studied. The software are the national computer database system of the Ministry of Health in Mozambique (known by its Portuguese acronym SISDB - Sistema de Informação de Saúde Database), District Health Information Software (DHIS), and the computer database system of the Ministry of Health in Tanzania (known by its Swahili acronym MTUHADB - Mfumo wa Taarifa za Uendeshaji wa Huduma za Afya Database). These are described as follows:

6.1 Sistema de Informação de Saúde Database (SISDB)
SISDB is the national computer database system of the Ministry of Health in Mozambique installed at the National Health Information System (NHIS) section of the Ministry of Health and in all Provincial Health Directorates in the country. SISDB was developed and implemented at the national and provincial levels since 1994. SISDB was developed on "dBase III" database management system (DBMS).

6.2 District Health Information Software (DHIS)
DHIS is an open source software from South Africa, which was introduced in the health information system in Mozambique through the Health Information Systems Programme (HISP) since the year 2000. The DHIS has many functions such as maximum and minimum ranges, validation rules, data definitions, indicators, report generator, a number of modules ranging from PHC to hospital and TB, organisational unit infrastructure, and annual surveys. DHIS was designed to capture health data at the district level of the health information system. In Mozambique, the DHIS is currently being piloted in Niassa, Inhambane, and Gaza provinces.

6.3 Mfumo wa Taarifa za Uendeshaji wa Huduma za Afya Database (MTUHADB)
MTUHADB is the national computer database system of the Ministry of Health in Tanzania. MTUHADB was developed on Microsoft Access 97 database management system and was installed in all Regional Medical Offices in Tanzania in the year 2000.

In this study we analysed the design and implementation of the computer databases in both countries, through the use of self-administered questionnaire s while conducting interviews with health workers. As it is discussed in section 7.0 (results), the health information systems in Tanzania and Mozambique were observed to be reluctant in introducing important health reforms approaches including implementing modern computer database systems, because the current old systems they run has vast amounts of data.

To address a way of leveraging the health systems from legacy systems, we developed Extraction Transformation and Loading (ETL) software and demonstrate how the software works. Through that extraction transformation and loading (ETL) software, we managed to extract and load health data from 1999 to March 2002 in Mozambique from the SISDB to the DHIS.

6.4 The Extraction, Transformation and Loading Software

To extract data from one computer database to another, you need to know, what is exactly to be extracted from the source database system. Health facilities have a defined list of health data elements that health workers record their instances. For example, number of deliveries per month, number of out patients visited the health facility per month, etc. Thus, these data elements are constant values; the health data is the count (data entries) of occurrences of these health data elements. A complete routine services reporting record could be "February 2003, rural hospital Chicumbane, number of deliveries is 39". This record has four main parts: period, place, data element name, and data entry value. Period is 'February 2003', place is 'rural hospital Chicumbane', data element is 'number of deliveries', and the data entry is '39'.

Data element name, health facility/organization units (place), and period are pre-defined values, thus, they exist in both databases, SISDB and DHIS. To extract health data from one database to another we need to transfer the "data entry" values from the source database to the target database. To do so, we need to map the three parameters: "place" where the data was recorded, "period" when the data was recorded, and "name" of the data element and the health facility from the source database to that of the target database. After mapping those three parameters, the next step is to copy the data entry values from the source database to the target database.

Major discrepancies between SISDB and DHIS

Date formats: SISDB database stores date into two data fields in number format while the DHIS stores date filed in one column and in the date format. For example, SISDB has "ANO (year)" column and "MES (month)" column, e.g. February 2000 is stored as follows:

ANO	MES
2000	02

While the DHIS stores the same date as follows:

Period
February 2000

Naming of data element system: SISDB uses short name of health facilities, like "cs Alto Changane", while the DHIS uses long name like "Centro de Saúde de Alto Changane". The SIS database uses abbreviation symbols to record the data field (data elements), while the DHIS uses full name, see Table 1.

Table 1: Names of health data elements in SISDB and DHIS

ID	SISDB Data Field	Related Paper Form	DHIS Data Field
8	CCM	B06	No Mau Crescimento 0-35 meses
9	CP1	B08	1as Consultas Pré-natais

Source: Comparative analysis of SISDB, DHIS, and SIS paper forms

Different data field values that represent the same thing: The DHIS records the name of the health unit, while SISDB records the code of the health unit. In addition, DHIS records more details of the health unit than those given in the **SISDB.**

Table 2: Data fields and names of health facilities in the DHIS

strDataField	strOrgUnit	dblEntry
1as Consultas 0-11 meses	Centro de Saúde de Alto Changane	247
1as Consultas 0-11 meses	Centro de Saúde de Alto Changane	290
Consultas seguintes 0-11 meses	Centro de Saúde de Alto Changane	65

Source: Analysis of the DHIS export file, April 2002

Table 3: Data fields and codes of health facilities in the SISDB

DCOD	UCOD	CP1
10	01	1248
10	02	695
10	16	583
10	08	286
10	10	159

Source: Analysis of SISDB data files, April 2002

Where, 'DCOD' is the district code, 'UCOD' is the health unit code and 'CP1' is the name of the health data element called "1as Consultas 0-11 meses". The default value for SISDB is 'zero', while that of DHIS is 'null'. The unit of analysis of SISDB is the provincial level, while the unit of analysis of the DHIS is at district level. The difference is that with SISDB, data are recorded by district, while in DHIS data are recorded by health facility.

The ETL software is built on homogenous architecture, as it involves only a single source system and a single target system. Data flows from the single source of data through the ETL processes and is loaded into the target system, as shown in the Figure 1.

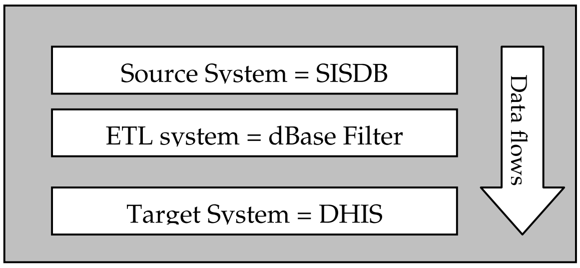


Figure 1: Database Filter system architecture

The ETL system connects the two databases using connection "bridges", and maps from the database to the ETL system. This enables the ETL to copy the data from the source database (SISDB) to its temporary data files, and transforms the data into a format that the target database (DHIS) understands. To load the data to the DHIS, the ETL system creates texts and Microsoft Excel files. The DHIS imports the data to the DHIS database. Figure 2 presents how the ETL system integrates the DHIS and the SISDB databases.

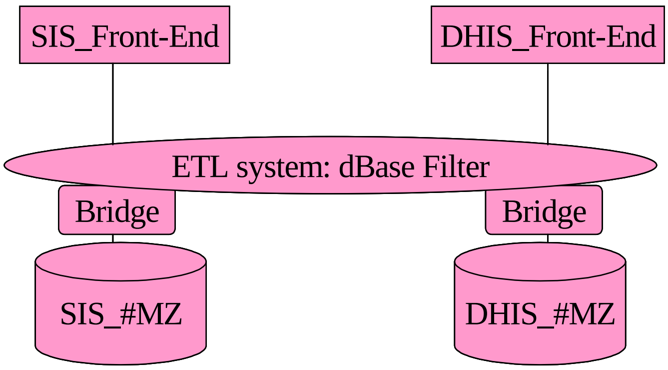


Figure 2: SISDB and DHIS integrated by Database Filter ETL system

This ETL system consists of four distinct functional elements: Extraction, Transformation, Loading, and Meta data

Extraction: The extraction element is responsible for extracting data from the source system. This ETL software connects to the SISPROG and DHIS databases at the same time, then copies the data, leaving the original data in place.

After connecting to the SIS database and the DHIS, the program presents screens that allow users to map health units and health data elements from that of SISPROG to DHIS. Figure 3 presents health data elements from SISDB and DHIS, so that users can match the two lists

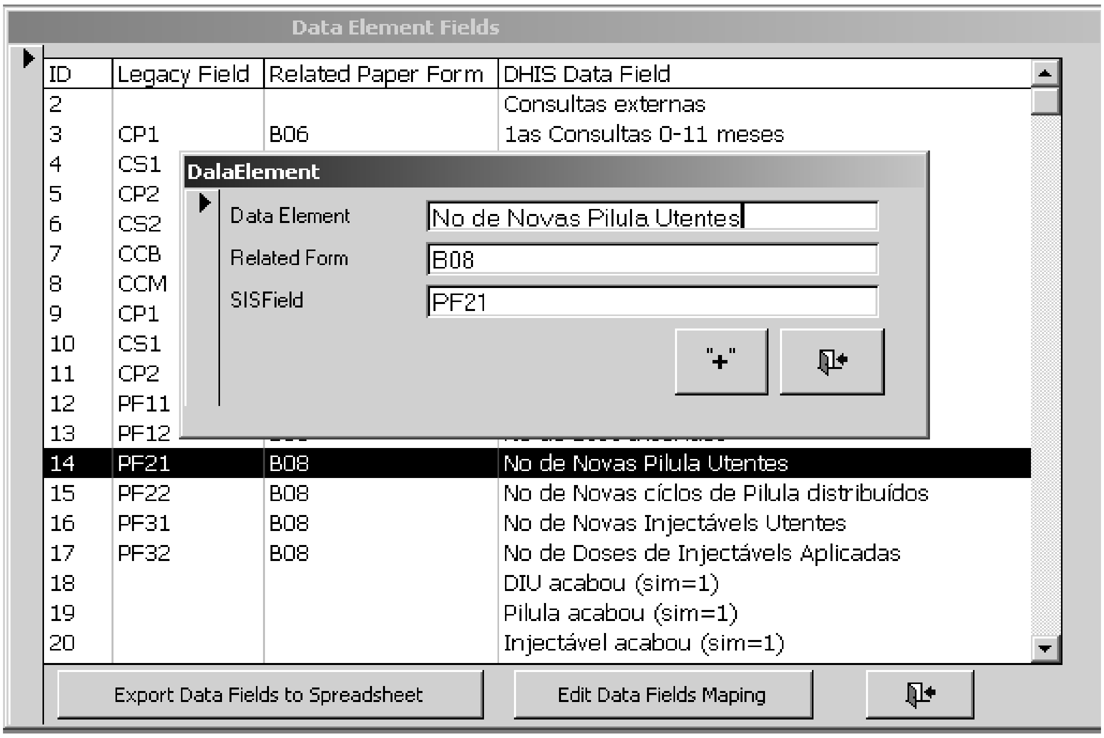


Figure 3: Mapping data elements from SISDB to DHIS

When the mapping of data elements is ready, the next step is to load the data in temporary data files using a command menu "Update Data File" (see Figure 4). The subsequent commands in Figure 4.6 have the following functions:

- Select Data Field – allow the user to select the data elements to extract data
- Create DHIS File – this command performs the transformation functional element of the ETL system.

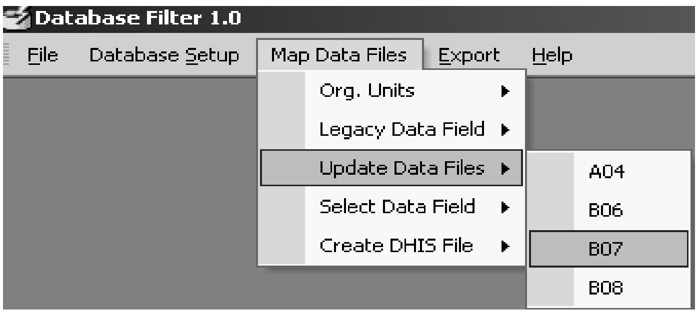


Figure 4: Extracting data modules

Transformation: The major tasks performed here are data validation, data accuracy, data-type conversion, and business rule application.

Data validation: This is important to enforce data integrity. Table 5 presents the DHIS import file headings. Since SISPROG has no data entries for every column, default values were added to the columns that cannot receive data from the source database.

Figure 5: DHIS Import file headers

strDataField	strOrgUnit	Period	dblEntry	strComment	intMin	intMax	ysnDisplay	strUser	dtmChanged

Data accuracy: To ensure that Boolean data fields contain appropriate values, "YES or 0" values from SIS are converted to "TRUE" and "NO or 1" values from SIS are converted to "FALSE" in the 'ysnDisplay' column of Table 5

Data type conversion: This is to ensure that all values for a specified field are stored in the same way in the target system (DHIS) regardless of how they were stored in the source system (SISPROG). Thus, the major task here is to convert the "dates" and 'names' data types used in SIS to conform to the DHIS data types. Figure 6 shows how SIS date format was converted into DHIS date format.

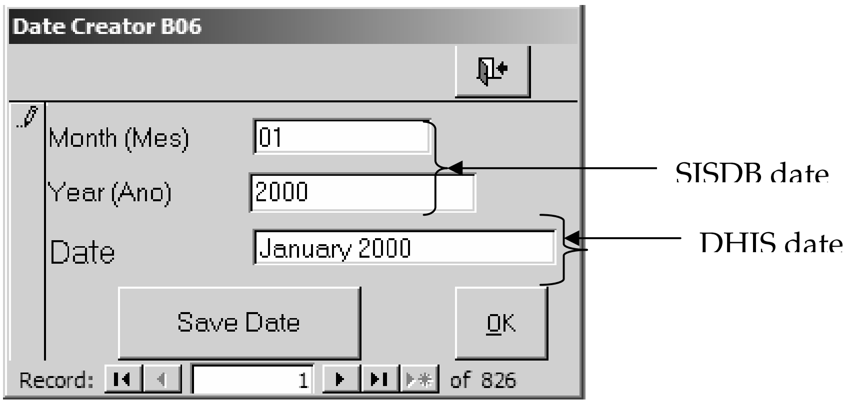


Figure 6: Date creator

Business rule application: The default values for SISDB is 'zero' while that of DHIS is 'null'. However, the business rule in HMIS is that a health unit can report 'zero' occurrence of health data element in a certain period. Since SISDB has zero as its default values. This is contradicting in that, either the health unit did not report or it has reported zero values. However, for my interest, the SISDB data will be imported as they are, that will help in assessing the quality of the data. Figure 4.8 presents extracted and transformed SISDB data in the Database Filter ETL software system.

Loading data: Extracted and transformed data is saved in a text file formatted as a DHIS import file. This resulting text file then is loaded into the DHIS using the import module of the DHIS software. The data extracted can also be exported to Microsoft Excel file, however the DHIS can only accept data formatted in 'text file'.

Meta data: The ETL system meta data functional element is responsible for maintaining information (meta data) about the transformation of data and documents the data mapping used during the transformation. This allows users to save their data elements and health units mapping, so that they can retrieve the mapping, instead of mapping it again in the future. For example, if one maps all health units from Gaza province, then Inhambane province, it should be possible to see the mapping of Gaza again.

7.0 Results

7.1 General features

The databases studied are uncompleted: The SISDB was expected to computerise 10 data collection paper forms of the Ministry of Health in Mozambique, however; only four forms were computerised fully. The MTUHADB has many uncompleted reports and data entry forms. For example,
It is not possible to review the entries of data entered through the computer form F002. There is no report out of F002 available. Form F005 does not allow the data entry of drug kits for other health facilities except "Local Government owned health facilities", while MTUHA policy requires private facilities to report.
In addition, in MTUHADB, there is mismatch between the paper forms data field with that of the database capture form.

The databases are without software specifications: The Ministry of Health in Tanzania did not receive the MTUHADB systems specification and the source codes. As a result, the Ministry of Health has to consult the software vendor for any changes to be made in the MTUHADB. The SISDB was developed a decade a go. At the time of this study, no system design specification and the source codes found. The consequences of lack of source codes and because of being developed long time ago, the vendor of SISDB is no where to be found and thus no one was able to reprogram further the database in order to accommodate the dynamic changes.

Date formats: The two databases store date as text and in two data fields. That is, "May 2003" would be stored as 05|2003. For example,
In the MTUHADB form F004, dates have to be entered as a text, this implies that is not possible to sort the data according to the date and users cannot perform operations related with dates.

7.2 Poor design of the legacy information systems lead to incomplete reporting

This study indicates that, the newly introduced health data elements after the computer database come into operations are not incorporated in the computer databases. This is because the databases were designed in such a way that, the health data elements are "hard coded" and thus to introduce a new data elements, users need to re-program the databases. Since health workers are not capable of reprogramming the computer databases, they do not record those newly introduced data elements in the computer database. From Table 1, it is seen that the data elements are the column head of the table, thus to introduce new data elements one have to add a new column on the table.

Table 4: MTUHADB data file structure

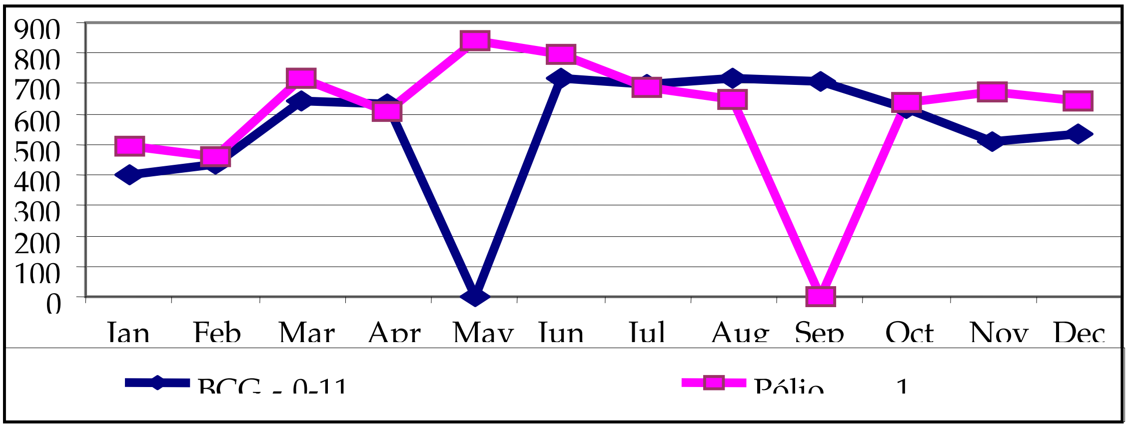
Data element	Record 1	Record 2	Record 3
Record_ref	[illegible]	[illegible]	[illegible]
Region Code	04	04	04
District Code	04	03	01
Year	2000	2000	2000
Quarter	1	1	1
Number of Health Facilities	45	60	47
Number of Loc Gov HF and Gov [illegible]	22	38	37
Number of HF with Service area	20	49	39
Number Health Facility Reports	12	39	34
Number [illegible] commodities reported	2	286	168
Amoxyciline Tablets	266	541	111
Benzylbenzoate Emulsion	120	257	148
Benzylpenicilline Injection	146	223	202
Chloroquine Tablets	189	292	64
Chloroquine Injection	234	484	520
Chloroquine Syrup	58	81	98
Co_trimoxazole Suspension	362	767	352
Co_trimoxazole Tablets	216	790	240
Doxycycline Tablets	133	343	100
Ergometrine Injection	168	425	186
Fe(2+)/folic acid_tablets	193	221	51
Lidocaine Injection	117	106	132
Mebendazole Tablets	88	345	158
Metronidazole Tablets	140	199	42
Oral Rehydration Salt sachet		10	96
Oxytetracycline eye ointment	63	71	65
Paracetamol Tablets	221	653	244
Procaine Penicilline fortified Vial	178	269	119
Water fot injection	154	83	70
St_aminoplylline	45	12	5
St_benzylbenzoate_emulsion	21	1046	7900
Chloroquine syrup	51	455	223
Chlorpromazine injection	163	665	422
Ephedrine(HCL) tablets	8	19	16707
Oral Rehydration salt sachet	4306	8312	6307
Children<5 Weighed			0
Children<5 Weighing <60%			0

Source: MTUHADB data file (Truncated), HMIS Section of the Ministry in Tanzania, August 2002

In Tanzania, while the database was put in place, the official first round of malaria drug was "Chloroquine". After several researches, it was revealed that Chloroquine are not effective enough to cure malaria, thus the Ministry of Health banned the use of Chloroquine and introduced "Fansider" in the year 2001. As it can be seen in Table 1, until the time of this study in 2002, the MTUHADB was not updated. It still has Chloroquine columns and has no column for Fansider. As a result, health workers are not able to record Fansider information in the computer databases and some health workers record Fansider information in the Chloroquine columns, which create more confusion. In Mozambique, while the database was introduced, there was "DPT" vaccination. In the year 2000, the Ministry of Health banned DPT vaccination and introduced new data elements, "DPT/Hep B" and "Vitamin A supplements". Until the time of this study, the SISDB was unable to record information for "DPT/Hep B" and "Vitamin A supplements" because there was no one to reprogram the SISDB in order to introduce the new columns for the two data elements.

7.3 Poor design of the LIS degrades the quality of the health data.

This study indicates further that, poor quality of the health data is contributed by the poorly designed computer databases. The SISDB has "zero" values as default values. That means, if a health unit did not report, the database put zero automatically for the corresponding data elements. As a result, health planners are not able to tell whether there was a "zero reporting" or "non-reporting". Figure 1 and Figure 2 show how the zero values cause different interpretations for decision-makers and health planners.



7.4 Old technologies used to develop the LIS impose burden to health workers.

The SISDB is designed in dBase III DOS version and is running on Microsoft Windows 95. The operation of the database is slow, the computers where the database is installed have very limited memory, and most are 16 MB RAM and 700 MB hard disk capacity. Despite the existence of more powerful computers, at the time of this study the Ministry of Health had no installation files for the SISDB, thus it was not possible to install SISDB onto the new computers. Figure 3 shows photos of two health offices equipped with two computers each, one for the old database (the old one), and the new one for secretarial purposes.

Figure 2: Despite the existence of new powerful computers, the legacy systems are running on the old computers.

The old computers cause burdens to the health workers for them have to take longer time to accomplish some tasks that could have been performed in short time with powerful computers. For example, it takes up to 8 minutes for the computer to start. If users want to include the health data in their reports in word processors, they have to print the data, and then re-type them on their reports because the computers are not able to copy-paste large amount of data. In addition, if they want to transfer the data to another computer (sharing data), they have to use diskettes or ZIP drives because the computers have no CD writers, however not all health offices have ZIP drives thus the common method is to use diskettes. To generate a customised graph of the health data, users used to print the data and retype them in spreadsheet software like Microsoft Excel. This is because the grapgy generated by the SISDB cannot be copied to word processor programs because is generated in MS DOS program.

7.5 Data locked in LIS can be migrated into new suitable information systems

In this study, we prototyped a computer database, District Health Information Software (DHIS) at the district level. The aim was to demonstrate how the DHIS can be used as a health data analysis tool at the district level. In our prototype, we decided to populate the DHIS software with real data from the field. To do this, we had two options: to key in data from the completed paper forms or to develop software in order to extract data from the SISDB to the DHIS. We decided to develop an extraction transformation and loading (ETL) software. The ETL was designed successfully and we managed to populate the DHIS with health data captured in SISDB from 1999 to April 2002.

8.0 Discussion

8.1 Reflections of the findings

Our reflections on the findings is that, the poor design of the LIS was a result of old technologies during the implementation of those computer systems, the structure of the health information systems (vertical reporting systems), and limited computer skills of the health workers.

8.1.1 old technologies

The LIS were implemented in old computer technologies (old version of database management systems and old computers (486 family)). This reflects the reality that during those days, these technologies were the available technologies.

8.1.2 The structure of the health information systems

The health information systems were structured in such a way that, all major health data analysis and interpretations were centralised at the national levels, so the focus were to compile the data at the provincial level and send to the national level. In those days, the unit of data analysis was at the provincial level and the districts were seen to have very limited capacity in analysing data. As a result, the computer systems were installed at the provincial level and not at the district levels.

8.1.3 Limited computer skills of the health workers

The designers of the systems focused on that users have very limited skills and thus a suitable system for them is one that is fixed so no changes will be required. As a result, they designed systems that cannot accommodate the dynamic changes of the health information systems, like the introduction and deletion of different health data elements.

8.2 The need for changes

Now days there are computers that are more powerful and the computer technologies have substantially advanced. The health information systems reform focus is the districts (WHO 1987). This also implies that the districts need to be empowered in order to analyse health data. Thus, there is a need for changes.

8.3 Implementing a new computer database: lessons learned

As this study was performed under the HISP project, its main goal was to shed light on problems of the old computer databases and to pilot the new one, the DHIS. The study, focused on the progress of the HISP project in order to identify learning lessons that could be sensitively applied in other developing countries. The lessons to share in replacing the LIS and introduce new ones are that (1) the new databases need to be extensively demonstrated to health workers. In this study, we populated the new database with real health data and conduct on site and workshop trainings for health workers, and (2) deliberate political negotiations between the collaborating project and the Ministries of Health need to be in place. This is because technical solutions are not convincing enough the Ministries of Health officials.

9.0 Conclusion

In this study, we are calling the attention of the Ministries of Health to review the performance of their computer databases. This is because the old information systems are contributing to the underperformance of the health information systems. We further conclude that, in relinquishing the old information systems, a deliberate effort should be introduced to secure the health data locked in the old systems. In this study, we managed to extract data from SISDB designed in 1994. While selecting new computer database to be used in the health information systems, sustainability should be the main guiding principle. We further conclude that the DHIS is suitable software for health data analysis.

10. References

Braa, J. and Hedberg, C., 2002. The struggle for district-based health information systems in South Africa. *The information society*, 18 (2), 113-128.

Dick, B., 2002. *Thesis resource paper: you want to do an action research thesis?*[online]. Available from: http://www.scu.edu.au/schools/gcm/ar/art/arthesis.html, [Accessed 27 December 2002].

Hanseth, O., 2002. From systems and tools to networks and infrastructures - from design to cultivation. Towards a theory of ICT solutions and its design methodology implications. Working paper, University of Oslo

Hanseth, O., and Monteiro, E., 1996. "Information Infrastructure Development: The Tension between Standardization and Flexibility" In Science, Technology & Human Values, Vol. 21 No 4, Fall 1996, 407-426.

Hardon, A., Boonmongkon, P., Streefland, P., Tan, M.L., Hongvivatana,T., Geest, S.V.D., Staa, A.V., Varkevisser, C., Chowdhury, M., Bhuiya, A., Sringeryuang, L., Dongen, E.V. and Gerrits, T., 2001. *Applied health research manual: anthropology of health and health care*. Netherlands: Het Spinhuis Publishers.

Heeks, R., 1998. *Information age reform of the public sector: the potential and problems of IT for India* [online]. Institute for development policy and management, University of Manchester. Available from: http://www.man.ac.uk/idpm/idpm_dp.htm [Accessed 21 January 2003].

Lippeveld, T., Sauerborn, R., and Bodart, C., eds., 2000. *Design and implementation of health information systems*. Geneva: World Health Organisation.

Lungo, J. H., 2003a. Data flows in health information systems: an action research studyof reporting routine health delivery services and implementation of computer databases in health information systems. Thesis (M.Sc.). University of

Oslo.

Lungo, J.H., 2003b. Legacy information systems in health management information systems: a case study from Mozambique. *In: The public health 2003 conference on health systems, development, and equity,* Holiday Inn Eastern Boulevard 24-26 March 2003. Cape Town: Public Health Association of South Africa and Health Systems Trust, 149.

McLaughlin, J., 2001. Using health information to sustain support for health reform in Africa. *In*: *The RHINO workshop on issues and innovation in routine health information in developing countries,* The Bolger Center, Protomac, MD, USA 14-16 March 2001. Arlington, VA 22209, USA: MEASURE Evaluation, JSI Research and Training Institute, 71-73.

MoH-TZ, 1993. Health management information system (HMIS) implementation plan *1993-1996*. Dar es Salaam: Ministry of Health in Tanzania.

Mwaluko, G.M.P., Neuvians, D. and Pazvakavambwa, B., 1996. *Health sector reforms*in Eastern and Southern Africa: a training manual for district health workers. Harare, Zimbabwe: Biomedical Research and Training Institute.

Simwanza, A. and Church, M., 2001. Establishment of an information culture: achievements and challenges in Zambia. *In*: *The RHINO workshop on issues and innovation in routine health information in developing countries,* The Bolger Center, Protomac, MD, USA 14-16 March 2001. Arlington,VA 22209, USA: MEASURE Evaluation, JSI Research and Training Institute, 227-234.

Sommerville, I., 2001. *Software Engineering,* 6th edition. Essex CM20 2JE, UK: Addison-Wesley.

Splettstoesser, D. and Kimaro, F., 2000. Benefits of IT-based decision-making in developing countries. Electronic journal of information systems in developing countries, 3 (3), 1-12.

Wood-Harper, A.T., and Davies, L.J., 1990. Worldview boundaries and informationsystems development: cross-cultural issues. *In*: S.C. Bhatnagar, and N. Bjørn-Andersen, eds. *Information technology in developing countries*. North-Holland: Elsevier Science Publishers B.V. and IFIP, 63-77.

Wood-Harper, T., and Bell, S., 1990. Information systems development for developing countries. *In*: S.C. Bhatnagar, and N. Bjørn-Andersen, eds. *Information technology in developing countries*. North-Holland: Elsevier Science Publishers B.V. and IFIP, 23-39.

Wood-Harper, T., Antill, L. and Avison, D.E., 1985. *Information systems definition: the multiview approach.* Oxford: Blackwell Scientific.