The Integrated Model for e-Health Data Collection and Sharing under Distributed Environments in Tanzania

Vitalis Ndume,
NM-AIST, P.O.Box 447,
Arusha,Tanzania

Yaw Nkansah-Gyekye,
NM-AIST, P.O.Box 447,
Arusha, Tanzania

Jesuk Ko
Department of Healthcare Management,
Gwangju University, Korea

ABSTRACT
This paper reports on a research aimed at developing an electronic data collection and integration tool for heterogeneous data. The article reports on modeling of a generic tool for collecting e-Health data using relational database. It also explores a methodology of collecting Verbal Autopsy data using Mobile devices. The V-Model approach was deployed in developing the Dynamic Link Library. The document engineering approach was adopted to analyze both the property and conceptual model of the information context of the forms. The skipping pattern algorithm was developed and sixteen mapping controls were designed to handle any data format in the Dynamic link library. The model was prototyped using Verbal Autopsy forms. The model is found to be useful in collecting e-Health data using mobile devices. The design of the Dynamic Link Library was loosely coupled to allow adaptation to others digital data collection in biomedical field. In addition, sixteen control classes in the Dynamic Link Library were able to handle more than 250 variables in a Verbal Autopsy form.

General terms
Software development

Keywords
e-Health, electronic data, data integration, data sharing, and health repository

1. INTRODUCTION
Due to the need of using ICT for achieving the Millennium Development Goals in the health sectors, many attempts have been made by low income countries to integrate health data yet little results is seen [1, 2]. Electronic data collection has become a golden rule in the healthcare service arena [3, 4]. The adoption of e-Health records and web services as software infrastructure is an extremely important prerequisite in the biomedical fields [1].

Even though there is much innovation on hardware devices [5, 6], building software application for collecting e-Health data remains a challenge to information system developers [7-9]. Many health system software applications are often developed without specifying how they will interact with existing or future health information systems [2, 5]. Furthermore, a data management system that supports centralized management of data generated at Institutional level, must accommodate diverse types of data from different clinical domains [10]. These problems can be solved by developing a Dynamic Link Library (DLL) which is independent of the back-end database. We call this library Survey-Pro. The goal of designing the Survey-Pro is to provide a higher level technical framework that allows for customization and flexibility of software application for e-Health data collection. In addition, Survey-Pro enables a researcher to capture data from the point-of-care to the central database with minimum effort. It is argued in [5, 7] that use of dynamic flexible tool in collecting data using mobile devices such a Smartphone provides a shared value for the common good. Moreover, handheld devices offer portable and remarkable access to clinical data and timely information at the point of care.

This paper reports on a research aimed at developing a generic electronic data collection and integration tool for heterogeneous data as one step forwards for e-health data integration. The article reports on the ongoing work of modeling a generic tool for collecting e-Health data using relational database. It also examines a methodology of collecting Verbal Autopsy (VA) data using mobile devices. After all, the proposed system is capable of synchronizing data from the point-of-care to the central database.

The extensive development of medical information requires an infrastructure that uniformly integrates the distributed and heterogeneous collection of e-Health records to deliver value-added information to healthcare [1]. It is argued that implementation of data integration and sharing has many challenges; including lack of technical standardized framework for data collection [3], heterogeneity of data sets and autonomist of data [11, 12]. With the growing number of researches in health facilities, scientists require a tool that makes data sharing easy, and accelerates large scale analysis as well as access of information "anytime, anywhere". It is claimed in [13-15] that developing a standardized tool and exchange software for heterogeneous data sets still remains a challenge. Despite the fact that many of the data sets have neither a standard format nor global information model, scientists need to work with data from a variety of sources. Lack of standards in data collection, integration, and exchange is therefore an obstacle in biomedical data sharing [16]. It is claimed in [17] that a solution that eliminates the need for human involvement in data retrieval would save valuable research time and make possible automated workflow implementable.

Organization: Section II gives a brief on the design approach and technology choices used in the development of the DLL. Section III enlightens on the methodology of administering VA. It also gives an account of the proposed architecture of the DLL and its core schema. The main contribution of the paper is on section III. Section IV explains the design for heterogeneity of hardware and screen size. Section V describes the data

synchronization techniques between and within computer machines. Finally, conclusion is drawn in section VI.

2. METHODOLOGY

The V-Model approach [18] was deployed in developing the Survey-Pro. The model focuses on the verification and validation of various forms used to collect health data in different projects. The document engineering approach [19] was adopted at the lower level to analyze both the property of each form and conceptual model of the information context of the forms. The Survey-Pro was programmed with VB.Net and the back-end database was constructed using SQL Server 2008 R2. The default front-end database was SQL Server Ce. The Survey-Pro model was prototyped with both VA forms and Household Economic Survey (HES). The system was tested using tablet PCs including VILIV and SAMSUNG.

Information concerning method of collecting electronic data, database schema, storage mechanism and technical factors associated with electronic data collection [20] were gathered using questionnaires. The primary data were collected in order to identify factors associated with EDC. Questionnaire was administered in order to measure intensity and strength of the factors associated with EDC as well as specifying the means of collecting, processing and sharing data between collaborators. The respondents were in two groups: the first group was made up of research scientists and principle investigators, and the second group consisted of software developers and system administrators. The correlation factors associated with EDC is beyond the scope of this paper.

Analysis of the database schemas in seven branches of Ifakara Health Institute (IHI) and at National Institute for Medical Research (NIMR) was also studied. Interview was conducted with data centralization team in order to understand better the challenges in data sharing.

3. RESULTS AND DISCUSSION

3.1 Methodology of Administering Verbal Autopsy

The main purpose of VA data collection is to analyze health in the community with the goal of determining individual cause of death and community/specific mortality fraction in a population without vital registration system. The information is gathered through interviews with family, or friends or caregivers of the deceased. Thereafter the interpretation is done by a coder. The Interpretation of VA data provides an opportunity for health planners, policy makers, and epidemiologists to understand better the patterns and implications of mortality in the community.

A questionnaire is administered to obtain health data which later on are used to ascertain the cause of death when a death event is reported. A baseline census is usually the source of data. A baseline census is conducted initially to provide a denominator of the population. Within the enumeration area, individuals are registered in their respective households. Any member who intends to stay in the house for more than six months is registered. A community integrated system is defined such that whenever there is a vital event within the community, that event is reported by a key informant (KI) using a mobile phone. Once an event is verified, a complete event registration is done using a death event questionnaire as shown in Fig. 1.

Data for cause of death are critical input in formulating good public health policy. However, data need to be collected reliably and interpreted consistently to serve as a global indicator [21]. For those countries with no vital registration, VA is a reliable method that is commonly used to study the pattern of cause of death. Regardless of the methodology and tool used, the process of collecting, interpreting and processing VA data is very involving and uncertain [8, 22, 23]. It is pinpointed in [22] that rigorous validation of VA procedure is needed to establish confidence in the data collection. Additionally, in order to understand the operational characteristics of VA in the population under study and to identify misclassification patterns, a controlled method of information collection is indispensable.

Furthermore, the significance of collecting VA is to improve country and regional global health information. The VA information is vital for public health, decision making, health sectors reviews, planning and resource allocation as well as program monitoring and evaluation [24]. Also, the cause of death statistics is useful to understand which disease kill and how many people die [24-26]. Collection of cause of death requires strong collaboration between the ministry of health, department of civil registration, and national bureau of statistics as well as the health research institutions.

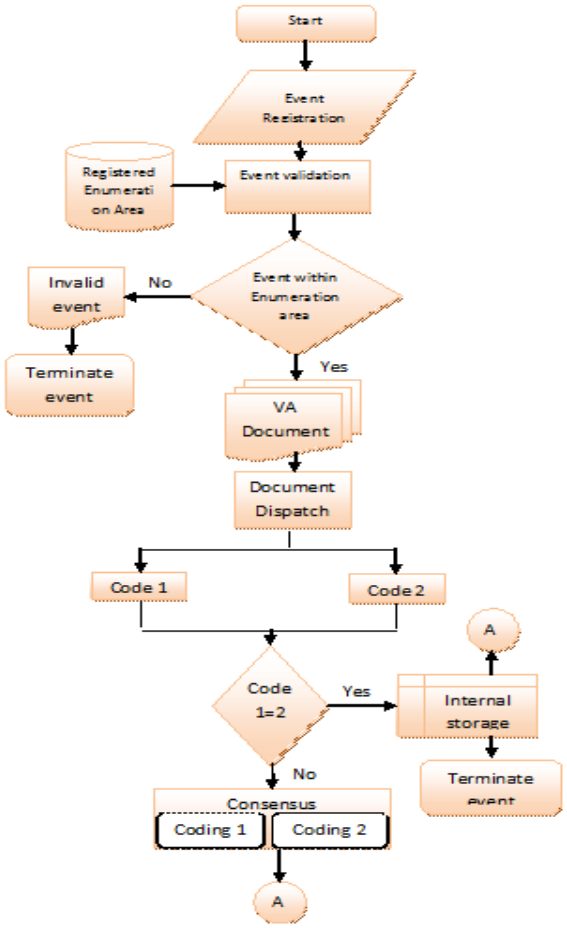

Fig. 1: Algorithm for collection, validation and coding of VA

3.2 The Dynamic Link Library Model

In order to construct the DLL model, entities were categorized into two main parts: primary and secondary. The primary entities/relations are bound together with DLL while secondary relations may be defined by the scientists based on the project needs. The primary part of the DLL contains four relations namely; "Questions", "Possible Responses", "Skipping Rules", and "Final Answer". Secondary relations include, but not limited to, Event, Household, Members, and Regions, District, Wards as well as Enumeration areas. The Event relation is the core within secondary relations and it is used to trigger the action on the DLL. The data model called Entity-Attribute-Value (EAV) is used to achieve the capability of the system. The proposed Entity Relation (E/R) diagram is generic enough to accommodate database operations such as deletion, addition and update. Nevertheless, the table schema is convenient for database integrity and positively ensures query efficiency. The principle applied in the schema design supports interface-level as well as implementation-level. The schema description of the DLL is presented in Fig. 2.

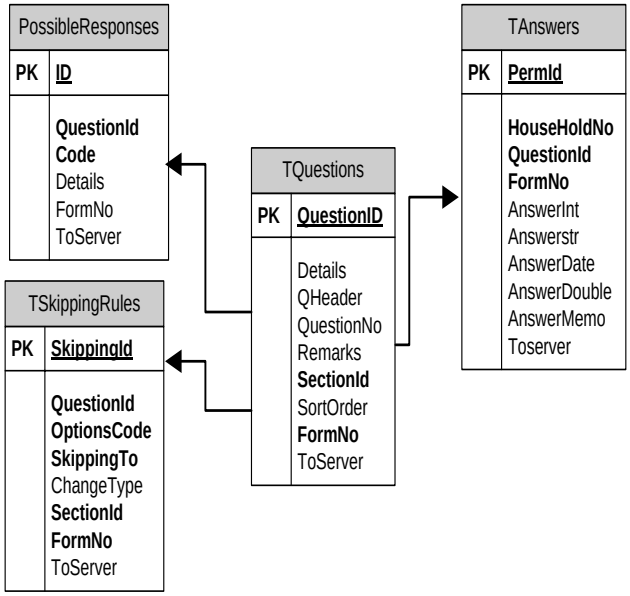

Fig. 2: Generic schema of the primary part of the DLL

3.3 Query Processing

The Survey-Pro uses a normalized relation tables to support subset of the T-SQL queries. The system supports selection (σC), Projection (π) and join (∞) of relation on primary key. In order to display the information on the screen the general query execution strategy is to join the major two tables that are T_Questions, and T_PossibleResponces. The T_SkippingRules is a look up table consulted on choosing the possible answer. The system makes use of the property index of each table based on the query requested. The indexes are also the identifier allowing the identification of the tuple being processed. For each row, the primary key is concatenation of the form being processed, section number of the form and question identifier. For T-Skipping Rules and T_Possible Responses the system add concatenation with Response ID. This approach is effective when the result is output into a non-normalized final table for statistical analysis. In the final result table that is T_Answers, additional identifiers are added so that there are composite keys to identify the object. For example, in case of verbal autopsy deceased person is a household member having a hierarchical Permanent Id. The permanent Id is concatenated from region code, district code, ward code, enumeration code, house code and then member code. Such Id makes easy the identification of the object in a large set of database and simplifies query processing.

3.4 The system Architecture

The proposed system architecture is divided into three layers as shown in Fig. 3. The complete set up of the system is made up of the source layer, the mapping layer and the external layer. The source layer and the mapping layer are the major part of the DLL. The external layer is considered as a subsystem communicating with the DLL. The two layers, that is, source DBs and mapping layer are encapsulated, allowing the actor to interact with mapping layer without knowing the source DBs. The explanation of each layer is as follows:

i) Source DBs layer: This lower module concentrates on achieving efficiency on handheld devices for data collection. The system can be configured using any lightweight database that the user can easily configure. SQL server Ce was used in the pilot. MS Access, MYSQL, SQL Server and Visual Fox-Pro are also set as alternative databases. MYSQL is still under test. The Source DBs provides list of questions and possible answers, as well as the property of the question and skipping rules. A virtual table is created on the fly when the source database is loaded. A mapping rule is therefore extracted from the virtual table. The use of a virtual table increases the efficiency of the system. This approach is known as chunking. Chunking is a process which groups information into a large, more meaningful unit, thus minimizing the demand on working memory and therefore reducing memory load.

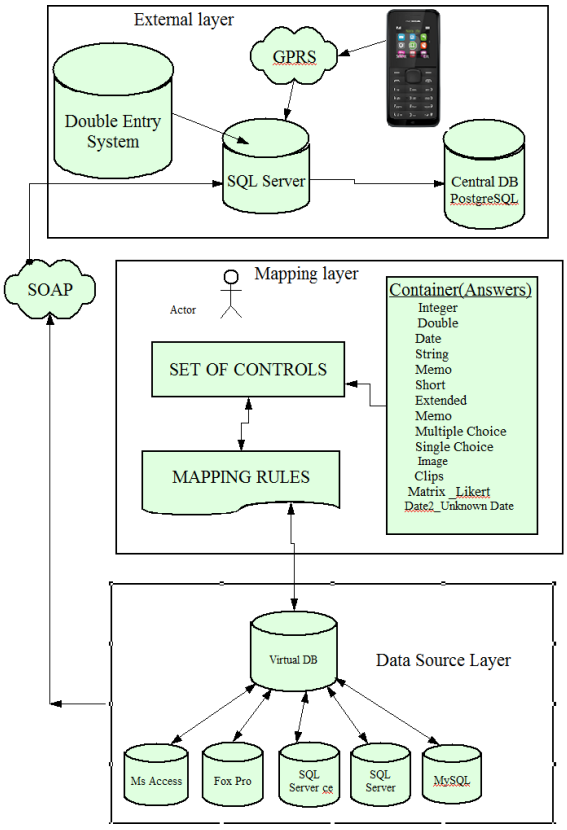

Fig. 3: The proposed system architecture

ii) Mapping layer: The mapping layer contains a set of controls and mapping rules as well as container that inherit property of the "AnswersType" which is set inside the control. That means, control has only one variable named "AnswersType". Each of the controls is set as public overidable in order to allow inherited variable to override the control. The role of mapping rules is to mediate between the virtual table and controls.

The skipping pattern algorithm is defined in the mapping layer. The algorithm uses questions, list of possible responses and question's property as an input. The skipping pattern is a simple table that indexes the questions and its behavior such as data type required, index, "skipping from" and "skipping to". It also stores information on the system response to each answer. The algorithm for skipping rules is presented in Fig. 4 below.

```
Skipping Pattern algorithm
**********************
Input:  Questions
        Possible Answers
        Property of question
Output: Index of next question
        Appropriate control class to display
Start:
        Load questions
        Mark first Question as Index I
        Mark last question as Index N
  While Index I < > Index N
       Read property of index I
       Load control for display
       Locate Index of next Question
       Index I = Index of next Question
  Loop
  End:
     Save records
```

Fig.4: The skipping pattern algorithm

On selecting the appropriate responceses, the virtual table consults the skipping pattern to determine the response to the question. Once the response is determined, the appropriate user control is loaded by the algorithm. The control is bounded by the set of containers which are responsible for translation of data type. The code for inheritance between control and container is shown in Fig. 5 below. During report generation, the virtual table uses predefined services such as store procedures and functions to generate the customer reports. This approach enables system developers to add more user controls, or more data types for any new databases without redefining the schema. Separation of primary and secondary entities enables programmers to concentrate on the core activities of the DLL. The secondary entities can therefore be added without affecting the architecture of the DLL.

```
Protected Sub getAnsewer(ByVal answer As Object, ByVal
MyQuestioID As String)
      TempBindingSource.Filter = "QuestionID = '" & MyQuestioID &
"'"
 If TempBindingSource.Count > 0 Then
         Dim dr As DataRowView = TempBindingSource.Current
         dr.BeginEdit()
  Select Case Type
  Case AnswerTypes.AnswerInteger,
AnswerTypes.AnswerImageAnswer,
    AnswerTypes.AnswerClipAnswer, AnswerTypes.AnswerLikert,
    AnswerTypes.AnswerMultipleChoice2 :     dr("AnswerInt") =
answer
 Case AnswerTypes.AnswerExtended, AnswerTypes.AnswerTwoText
   dr("AnswerStr") = answer
 Case AnswerTypes.AnswerString,
AnswerTypes.AnswerMultipleChoice,
    AnswerTypes.AnswerDate2 :          dr("AnswerStr") = answer
 Case AnswerTypes.AnswerDate :         dr("AnswerDate") = answer
 Case AnswerTypes.AnswerYESNO :     dr("AnswerInt") = answer
 Case AnswerTypes.AnswerMemo :      dr("Answerstr") = answer
 Case AnswerTypes.AnswerStringContainer :  dr("AnswerStr") =
answer
 Case AnswerTypes.AnswerYESNODN :       dr("AnswerInt") =
answer
 Case AnswerTypes.AnswerSingleChoice  :    dr("AnswerStr") =
answer
 Case AnswerTypes.AnswerClipAnswer  :      dr("AnswerClips") =
answer
 End Select
    dr.EndEdit()
    CurrentQuestionId = QuestionID
    CurrentAnswer = answer
End If
 End Sub
```

Fig. 5: Source code for inheritance of data types from answer types

iii) **External layer:** The external layer is a subsystem that allows scientists to share data in a centralized access. Currently the communication between the systems is by SOAP which allows data synchronization by means of serialization. A module that deals with double entry system for data collected by this paper was developed and can be configured separately. A mobile system is also a subsystem used to collect vital events for VA. The mobile system is set using OpenXData. This is an open source application running on a simple NOKIA mobile phone with MIDP 2.0. In the external layer the Central DB (PostgreSQL) is independent project managed by other team.

3.5 Sensitivity Analysis of the DLL

The DLL was tested using various VA forms and the HES forms used to measure household economy. Also, the baseline form in the Demographic Surveillance System (DSS) and various forms from the Neonatal Vitamin Supplements Project were used in sensitivity analysis. All these forms impose various requirements on the screen display and the data types. Moreover, they have different rules for their skipping patterns. The sensitivity analysis results show that the DLL can be applied to many research projects with minimum customization.

3.6 Heterogeneity of Hardware and Software

Hardware heterogeneity requires software to be developed and run on different operating systems on different hardware platforms. It is asserted in [27] that the proper selection of computer software that matches hardware component is a key

aspect in developing computer system for biomedical data. Hardware challenges for electronic data collection include design for screen size and operating system. Hardware options for electronic data collection include PDAs, Tablet PCs, Ultra Mobile Computers, as well as Mobile phones. The approach therefore, was to create an interface that displays one question on the screen at a time. This is far better for small screens. The design of the interface that mediates the interaction of people with devices is a crucial characteristic of the overall human computer interaction design. It is demonstrated in [28] that visual interface is a central component of interaction design. Also, it is explained in [10] that among the reasons for the success of information system is their ability to lower technical barriers for nontechnical users. In that respect the proximity, similarity, continuity and closer principles were applied in the design of the interface. With proximity principle "Next" and "Save" buttons were grouped together away from "Close" button and placed in the footer of each screen. On the other hand, continuity principle was applied in designing scrollable interface to connect elements which are disconnecting as shown in Fig. 7 and Fig. 8. During the process, techniques of design focused on user cognitive psychology in order to enhance usability and acceptance of the system. For example, a generic error message was designed to guide the user instead of allowing the system shut down. Some output screens of the program execution are described below:

i) Question manager: This is a module for managing questions and possible responses. In this module a user provides questions details, adds or modifies questions, links a question with possible answers, and manages possible responses.

ii) Skipping pattern: The module for skipping pattern is shown in Fig. 6. The user specifies behavior of the question, such as skipping to/from, indexes, and change type. These rules are used by the virtual table to manage how a mapping should work when a user provide an answer.

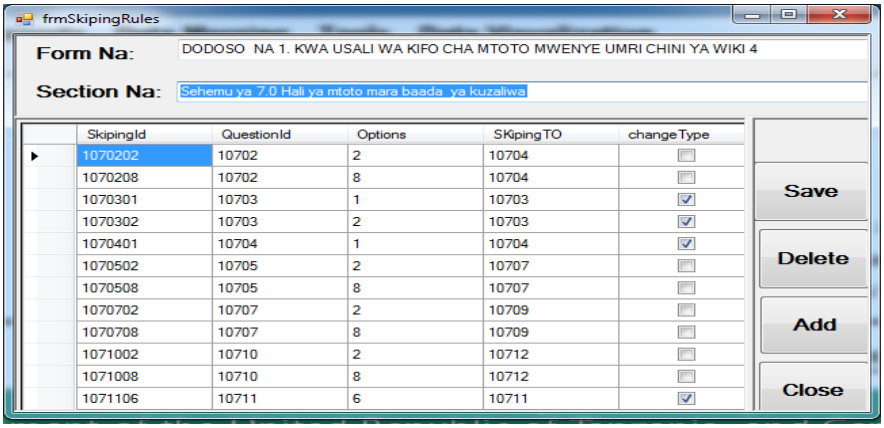

Fig.6: Module for skipping pattern rule

iii) Single choice: These are typical questions that require an enumerator to select only one answer among the given options as shown in Fig. 7. When the user provides an answer a code of the option is stored in the database.

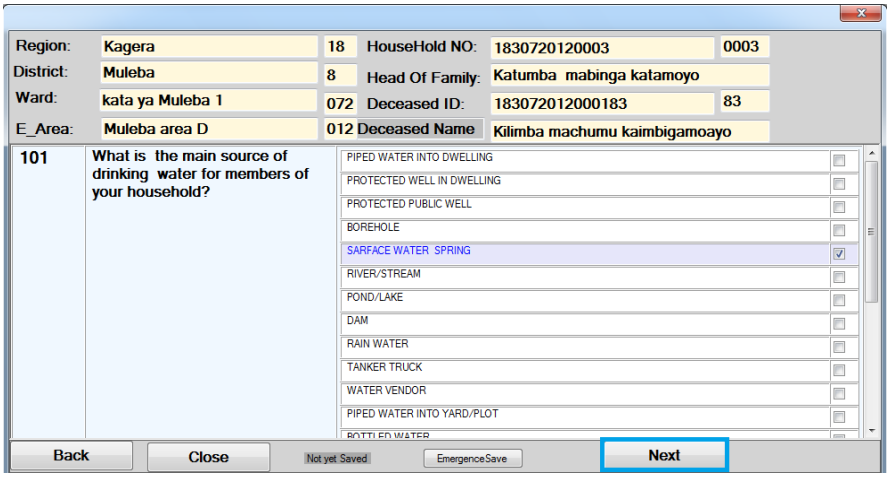

Fig. 7: Control for single choice

iv) Date type: In a semi structured data collection, date is the most difficult parameter to deal with. The date was categorized in two types; those which only take date format and those which accept unknown values of the date. For a known date, the date picker control was used otherwise the "99/99/9999" format was used. This approach enables the user to identify which part of a date is known and which part is unknown.

v) Numerical and float: These questions require the user to fill in numerical values. They are very common in HES.

vi) Multiple answers with single choice: These are typical questions with a hierarchical nature where each statement is treated as a separate variable and is stored separately. However, each variable has the same data type. This type of control can span over "Yes", "No", and "Don't know" as shown in Fig. 8.

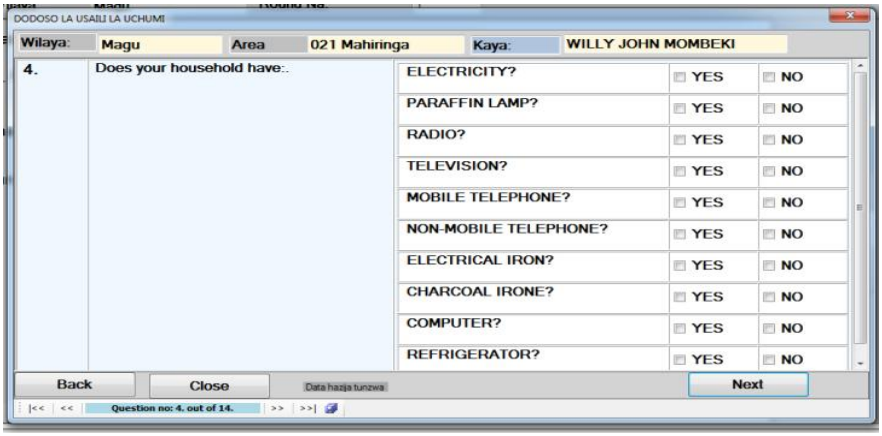

Fig.8: Multiple answers template

vii) Likert Scale based questions: These are matrix based questions that seek to determine the respondent opinion on some aggregated items. A range of 5 scales which are commonly used to measure patients opinion on the quality of hospital care is presented in Fig. 9. This matrix allows a range of aggregated opinions to be presented easily.

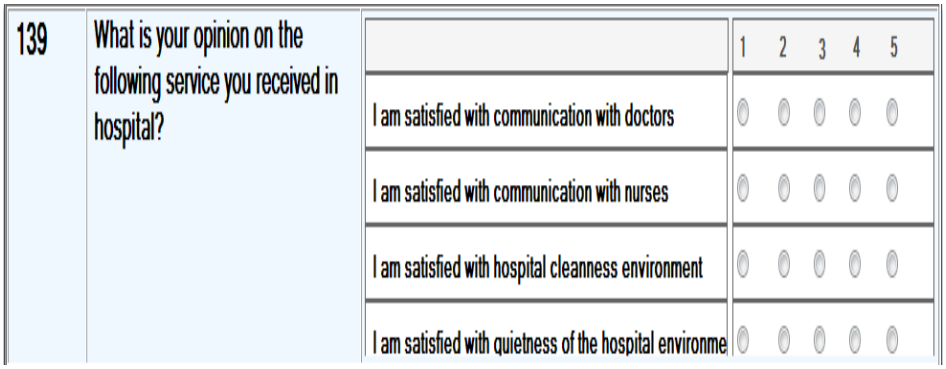

Fig. 9: Likert based question example

It is understood that the quality of a software design is difficult to measure, but some quality aspects that are commonly used included robustness, understandability, extendibility, modularity maintainability and reusability. Therefore, in this study it is argued that the separation of the system into three layers will significantly contribute to the following quality: First, the underlying components of the system to behave as expected, even if the system is used in different circumstances than the original implementer intended. Second, the mapping and the source layer can be extended to accept the new form of data and new algorithm without disrupting existing package. And last, the system components that are tightly coupled are grouped together logically and obey common naming conversion and protocol. The fundamental concern in the system is usability.
Considering, broader classes of potential users with data management, starting from enumerators at lower level, data clerk to data managers, top managements, as well as non IT experts, an ease of use is essential in the software design. Navigation, use of color quick helps need to be addressed in the system. The approach in the interface design emphasizes the ease of use so that the system could be accepted at different level of users.

3.7 Data Synchronization

Data synchronization mechanism was deployed using SOAP protocol. The wireless modem was used to transfer data between the central server and user databases. In order to handle heterogeneity, a library which maps data type of two debases was developed. The synchronization between the enumerator database and supervisor database was done by a merging script installed in the supervisor's computer.

The occasional application connection (OAC) was used in order to avoid real time connection. This technique stores data on the computer temporarily before transmitting it to the research centre. Once data is collected from the field it is first merged into the supervisor's PC, and when there is a connection data is pushed to the staging database.

3.8 Challenges of TCP/IP Protocol in Data Synchronization

In a distributed environment, synchronizing data with TCP/IP is very challenging because it needs handshaking and a continuous connection between the client and central server. Insufficient

timeout and expiration mechanism leading to brute-force and replay back is experienced in TCP/IP protocol. The Serialization techniques for data transfer between the client and server was deployed. The objective of serialization is to find non-serial schedules allowing transactions to execute concurrently without interference, thereby producing a database state that mimics serial execution. Each transaction is executed consecutively without any interleave operation from other transactions. A schedule S consists of a sequence of operations from a set of n transactions $\{T_1, T_2, ...T_n\}$ subject to the constraint that the order of operations for each transaction is preserved in the schedule. Thus, for each transaction T_i in schedule S, the order of operations in T_i must be the same in schedule S. In each transaction only a single row is sent in a serial schedule. Once a record is successfully updated in the central database it is blocked from the client side and then the next record is sent. This approach of scheduling of the transactions ensures that only new record is transacted at any synchronization process.

4. RELATED WORKS

Electronic data collection is an increasingly popular research methodology. Therefore, many tools have been designed to capture data online. Most available are web based tool restricting the researcher to develop questionnaire over the internet. Some of that tool include: SurveyMonkey [29], Zoomerang [30], KwikSurveys [31], SurveyGizmo [32], LimeSurvey[33], freeonlinesurvey [34] FluidSurvey [35], QuestionPro [36] asw well as surveyPlanet [37].

The above mentioned tools are online surveys, requiring the researcher to design the research tool online and making the responder to fill the forms online or by email. This limits their use to the environment with internet access only. Such tools when used in research methodology suffer from measurement errors, technology difficulties, responses rates as well as representative of the sample [38]. In addition, they use Hyper Text Markup Language (HTML) technology, restricting the interoperability. The Survey-Pro proposed is designed to bridge this gap; it is a dynamic link library allowing the researcher to have a research tool on hand and provide choice of database for data storage. It allows the programmer to add new modules as required, and enables the data manager to merge data using an embedded merging tool. Furthermore, many controls are available allowing questions to be presented in different formats.

5. CONCLUSION AND FUTURE WORK

At the time of writing, 16 mapping controls and four database systems were designed. A library that maps the heterogeneity of the data types between the databases was developed. The mapping controls were able to handle 250 variables in VA form. The proposed system is capable of synchronizing data from the point-of-care to the central database. The architecture provided is generic enough to hide the implementation level of the database. The main achievement at this time is deployment of a skipping pattern algorithm in the DLL. The design of this generic library is loosely coupled to allow adaptation to digital data collection in the field of biomedical science. Future work involves developing techniques for transferring data across an environment with limited resources using Web service and SOAP technology.

6. ACKNOWLEDGMENT

This work is undertaken as part of research in improving health data collection at the Ifakara Health Institute (IHI). The initial financial support was granted to the researcher by the IHI under the scheme of Post Graduate scholarship. The additional financial support was granted by the Commission for Science and Technology through the Nelson Mandela African Institute of Science and Technology (NM-AIST). We are grateful to the NM-AIST scientific committee and Institution review board of IHI for granting a permission to undertake this research in health research institutions. Collaboration between NM-AIST and IHI is also highly appreciated. All other institutions involved in this work have been essential, and we are particularly grateful to all IHI staff who have contributed in gathering system requirements.

We also acknowledge that part of this work which was presented in Pan African international Conference of Information Science, Computing, and Communication.

7. REFERENCES

[1] I. Amarouche, *et al.*, "Electronic health record data-as-a-services composition based on query rewriting," *Transactions on large-scale data-and knowledge-centered systems IV*, pp. 95-123, 2011.

[2] H. Mwanyika, Lubinski,D.,Anderson,R. Chester,K.,Makeme, M., Steel.M., and de Savigny.D., "Rational system Design for Health information system in Low-Income countries. An enterprise Architecture Aproach," *Journal of Enterprise Architecture*, vol. 7, 2011.

[3] K. Atalag, *et al.*, "Putting health record interoperability standards to work," *electronic Journal of Health Informatics*, vol. 6, p. e1, 2010.

[4] C. Bertolini, *et al.*, "Towards a formal integrated model of collaborative healthcare workflows," *Foundations of Health Informatics Engineering and Systems*, pp. 57-74, 2012.

[5] N. Palmer, *et al.*, "Raven: Using Smartphones For Collaborative Disaster Data Collection," in *Under Submission To The Intl. Workshop on Information Systems for Crisis Response and Management (ISCRAM 2012)*, 2012.

[6] Z. A. Rajput, *et al.*, "Evaluation of an Android-based mHealth system for population surveillance in developing countries," *Journal of the American Medical Informatics Association*, vol. 19, pp. 655-659, 2012.

[7] R. Chaudhri, *et al.*, "Open data kit sensors: mobile data collection with wired and wireless sensors," in *Proceedings of the 2nd ACM Symposium on Computing for Development*, 2012, p. 9.

[8] J. E. Haberer, *et al.*, "Challenges in using mobile phones for collection of antiretroviral therapy adherence data in a resource-limited setting," *AIDS and Behavior*, vol. 14, pp. 1294-1301, 2010.

[9] J. Shang, *et al.*, "A pervasive Dietary Data Recording System," in *Pervasive Computing and Communications Workshops (PERCOM Workshops), 2011 IEEE International Conference on*, 2011, pp. 307-309.

[10] C. A. Brandt, *et al.*, "Metadata-driven creation of data marts from an EAV-modeled clinical research database," *International journal of medical informatics*, vol. 65, pp. 225-241, 2002.

[11] B. Beran, *et al.*, "Web Services Solutions for Hydrologic Data Access and Cross-Domain Interoperability," *International Journal On Advances in Intelligent Systems*, vol. 2, pp. 317-324, 2009.

[12] J. T. McDonald, *et al.*, "Heterogeneous Database Integration Using Agent-Oriented Information Systems," DTIC Document2000.

[13] I. Letunic, *et al.*, "SMART 4.0: towards genomic data integration," *Nucleic acids research*, vol. 32, pp. D142-D144, 2004.

[14] J. Lim, *et al.*, "Concurrent data access in mobile heterogeneous systems," 1999, p. 10 pp.

[15] Z. Zhao, "Data Quality-Oriented Data Integration in Peer-to-Peer System," in *Hybrid Intelligent System,2009*, Shenyang, 2009, pp. 419-422.

[16] R. L. Richesson and J. Krischer, "Data standards in clinical research: gaps, overlaps, challenges and future directions," *Journal of the American Medical Informatics Association*, vol. 14, pp. 687-696, 2007.

[17] M. Piasecki, *et al.*, "Development of an Information System for the Hydrologic Community," in *9th International Conference on Hydroinformatics HIC 2010*, , Tianjin, CHINA, 2010, pp. pp. 4353-4359,.

[18] N. M. A. Munassar and A. Govardhan, "A Comparison Between Five Models Of Software Engineering," *IJCSI International Journal of Computer Science Issues*, vol. 7, pp. 94-101, 2010.

[19] R. J. Glushko and T. McGrath, *Document engineering Analysing and Designing Document for Business information & Web service*: MIT Press, 2005.

[20] A. Maedche and S. Staab, "Ontology learning for the semantic web," *Intelligent Systems, IEEE*, vol. 16, pp. 72-79, 2001.

[21] C. J. Murray, *et al.*, "Validation of the symptom pattern method for analyzing verbal autopsy data," *PLoS Medicine*, vol. 4, p. e327, 2007.

[22] E. Fottrell, *et al.*, "Revealing the burden of maternal mortality: a probabilistic model for determining pregnancy-related causes of death from verbal autopsies," *Popul Health Metr*, vol. 5, 2007.

[23] S. O. Oti and C. Kyobutungi, "Research Verbal autopsy interpretation: a comparative analysis of the InterVA model versus physician review in determining causes of death in the Nairobi DSS," 2010.

[24] E. Fottrell and P. Byass, "Verbal autopsy: methods in transition," *Epidemiologic reviews*, vol. 32, pp. 38-55, 2010.

[25] WHO, "Health System in africa, Community Perception and Perspective," Brazzavile, The Report of a Multi-Country Study2012.

[26] WHO. (2013, Health statistics and Health information system: " World Health organization ". *(May 1st, 2013)*. Available: www.who.int/healthinfo/

[27] C. Armstrong, *et al.*, "Exploration of Game Consoles as a legitimate computing platform for in-the-field biomedical data acquisition and management," *African Journal of Information & Communication Technology*, vol. 7, p. 8, 2013.

[28] A. Cooper, *et al.*, *About face 3: the essentials of interaction design*: John Wiley & Sons, 2012.

[29] SurveyMonky. (2004). *SurveyMonky, Create Survey Get Answers*. Available: https://www.surveymonkey.com/

[30] Zoomerang. (2014, 26/05/2014). *Zoomerang Online Survey*. Available: https://www.zoomerang.com/

[31] KwikSurvey. (2014, 26/05/2014). *KwikSurvey Free online and Questionnaire tool*. Available: http://kwiksurveys.com/

[32] SurveyGizmo. (2014). *SurveyGizmo, Survey Builder for every one*. Available: http://www.surveygizmo.com/

[33] LimeSurvey. (2014, 26/05/2014). *LimeSurvey - the free and open source survey software tool!* Available: www.limesurvey.org/

[34] FreeOnlineSurvey. (2014, 26/05/2014). *Free Online Survey*. Available: http://freeonlinesurveys.com/

[35] FluidSurvey. (2014, 26/05/2014). *FluidSurvye-Online Survey Software at its best* Available: http://fluidsurveys.com/

[36] QuestionPro. (2014, 26/05/20014). *Questionnaire Pro , OnlineResearch made easy*. Available: http://www.questionpro.com/

[37] SurveyPlanet. (2014, 26/05/2014). *SurveyPlanet Pro: Powerful tools for our power users.* Available: https://www.surveyplanet.com/

[38] D. H. Granello and J. E. Wheaton, "Online data collection: Strategies for research," *Journal of Counseling & Development*, vol. 82, pp. 387-393, 2004.