JHIDC Health Informatics in Developing Countries

www.jhidc.org
Vol.4 • No.2 • 2010

A Survey of Health Data Models in Sub-Saharan Africa

Adebayo Idowu and Rotimi Adagunodoi

Department of Computer Science & Engineering
Obafemi Awolowo University
Ile-Ife
Nigeria

Corresponding author:
paidowu@oauife.edu.ng
paidowu1@yahoo.com

Abstract

Globally, diseases has been a treat to the health of people living in different part of the world both developed and developing nations but there are some diseases that affect developing nations than developed countries of the world such diseases are HIV/AIDS, malaria and tuberculosis.

With the use of a reliable and flexible health data model for development of an electronic health records system, high quality real time health information would be available to the health care providers, researchers, decision makers and all health stakeholders.

In this paper, we present a review of the two existing health data models developed for two countries in sub Saharan Africa region. We started with a brief review of three globally leading health data models. We present the review of the model based on strengths and limitations and finally provide a comparative analysis of the two health data models based on the structure, functionality of the models.

Introduction

Developing nations most especially sub Saharan Africa are characterised with one or more health problems. Deadly diseases such as HIV/AIDS, tuberculosis, and malaria have become one of the most destructive epidemics in both developed and developing nations of the world. Although, the public health sector of developed nations have computer based health information systems. This makes monitoring of the distribution and spread of these diseases easy. Public health researchers, especially epidemiologists, are currently battling with these deadly diseases.

Sub-Saharan Africa is one of the most affected regions in the world; however, computer-based health information systems are not currently used in almost all the countries in the region for controlling, monitoring and management of these epidemics. One of the most common and incurable of all these diseases is HIV/AIDS and sub Saharan Africa hosts the highest number of people living with HIV/AIDS (Idowu, et al, 2009). The need to address the problem of HIV/AIDS both in Nigeria and Kenya gave birth to the health data models we reviewed in this paper.

The aim of this paper is to present a review of these two health data models developed for both Kenya and Nigeria in order to identify the strengths and limitations of the models and finally compare the two models based on their structure, and functionality. The identification of the strengths and limitations of the models would make it easy for any researcher to have good understanding of the models in terms of merits, lapses, coverage, among other things. It would also allow the researchers to be able to adopt, modify and extend the models easily.

Data model is an important and essential aspect of software design, and implementation; but unfortunately is most unexciting aspect of the task. It is used to represent a process and a system in order to have understanding of a subject area so as to be able to analysis, design, and implement a system. In spatial world, geometry (such as point, line, and polygon) and geometry collections with their attributes are used to depict data model in the real world. A data model is a diagram that shows and describes the activities and entities along with the relationship and association of information needed in a domain. Data model diagram is design to show the relationship between different data required in a particular domain. According to Business Intelligence Solution in 2004, a data model is a formal structure representation of real world entities, focused on the definition of an object and its associated attributes (Business Intelligence Solution, 2004).

Health data model, which is what this paper focus on, is a model that "provides information about important foundations of data to be captured which can be transformed into meaningful information to support many different uses across the health system" (Canadian Institute for Health Information, 2001). Health data model is a model that allows users to be able to capture efficient, consistent and

Referencing this article

Idowu, A. and Adagunodoi, R. (2010). A Survey of Health Data Models in Sub-Saharan Africa [Electronic Version]. *Journal of Health Informatics in Developing Countries, 4(2)*, 11-19, from http://www.jhidc.org/index.php/jhidc/issue/view/10

reliable data and information which can be use to support decision making and management of health care delivery system.

With the use of a reliable and flexible health data model for the development of a electronic health records system, high quality real time health information would be available to the health care providers, researchers, decision makers and all health stakeholders. Reliable health data model would enable the development of a reliable and efficient national health database and electronic health record system which could be use for health surveillance, monitoring and management.

The aim of this paper is to review Conceptual Data Model for Disease Surveillance, Monitoring & Prediction in Nigeria and AMPATH Medical Record System Data Model developed for Nigeria and Kenya respectively in order to identify the strengths and limitations of the two models. It is hope that with this paper, researcher that are interested in developing nations most especially sub Saharan Africa would be able to have good understanding of the models along with their strengths, limitations structure, and functionality. With this understating of the model, the researchers would be able to improve on them in order to develop health data model that would be comprehensive enough to address the limitations of the two models.

We started the review of the models with a brief overview of the three leading health data models so as to present the scope of the leading health data models which may serve as a useful information for any researcher that want to work on the sub Saharan Africa health data models. We conclude the paper with comparative analysis of the model based on structure, and functionality of the models.

Health Data Models

Leading Health Data Model

There are quite a number of health data models globally. In this paper, we briefly review three leading health data models namely: Health Level Seven (HL 7), Digital Imaging and Communication in Medicine (DICOM), and Good European Health Record/Open Electronic Health Record (GEHR/OpenEHR). HL7 has give birth to other health data model such as Public Health Conceptual Data Model (PHCDM), and Canadian Conceptual Health Data Model (CHDM) which are not discussed in this paper.

HL7

The HL7 is an American National Standard Institute (ANSI) accredited standard for health care specific data exchange between two or more software applications. The word “Health Level Seven” was derived from the Open Systems Interconnection (OSI) top layer protocol (Level 7) for the health environment. The need for HL7 arises from the fact that healthcare centers, physicians, hospitals and laboratories globally needs a system that would be used to send and receive health related data on daily basis. The HL7 is the language developed to solve this problem.

Some of the aims of HL7 are to develop a data model that allow exchange of medical data using a standardised format in order to ensure that the data is globally understood and organised (HL7, 2008); and to collaborate with other health data models and healthcare information system users so as to have a supportive and compatible health data model in order to make sure that HL7 data model meet the world requirement. HL7 captures information about comprehensive public health system (HL7 News, 2006) and is widely use, and adopt by other health data models (CDC, 2000; CIHI, 2001)

DICOM

In 1983, the American College of Radiology (ACR) and the National Electrical Manufacturers Association (NEMA) agreed to develop a data model that can be use for communication of digital image information. Digital Imaging and Communication in Medicine is a data model for the communication of images which was developed to handle exchange of digital data and information between medical imaging equipment and other system.

The aims of DICOM are to have a data model that can allow creation of diagnostic information database that can be accessed by various devices distributed across different locations; and “promote communication of digital image information, regardless of device manufacturer; facilitate the development and expansion of picture archiving and communication systems that can also interface with other systems of hospital information”(DICOM, 2007),

DICOM is used in departments such as radiotherapy, radiology and other related departments that required high quality image processing and viewing. The major limitation of this data model is that it focuses on medical imaging.

GEHR/OpenEHR

The GEHR/openEHR data model started in January, 1992 as a European Union project called Good European Health Record. The GEHR project ended in 1994. Australia later joins the project which led to the change of the name from Good European Health Record to Good Electronic Health Record.

The project led to establishment of open source foundation in 1999 and the name openEHR was adopted and currently the project is being maintained by the openEHR foundation. The openEHR foundation is “an international non profit organisation, online community whose aims is to promote and facilitate progress towards electronic healthcare records of high quality, and to support the needs of patients’ clinicians everywhere” (Eichelberg et al, 2005).

Some of the aims of the GEHR/OpenEHR are to make available formal specification of requirements for representing and communication electronic health record (EHR) information, EHR information architecture, models and data dictionaries; and maintain open source reference implementations. The main contribution of GEHR/openEHR is the introduction of “archetype” concept to the existing health data models. An archetype is “a computable expression of a domain content model in the form of structured constraint statements based on some reference model (Beale and Heard, 2005).

Apart from the models discussed above, there are other health data models but some of them are developed for a particular country use or for a branch of health care system such as disease surveillance, and medical imaging. HL7 is the only health data that capture information about comprehensive public health system (HL7 News, 2006). The model is widely use, and adapt by other health data models (CDC, 2000; CIHI, 2001) but it often too comprehensive to be used in any healthcare delivery system without modification.

Having briefly discussed the leading health data models, this paper presents the two health data models from sub Saharan Africa. Conceptual Data Model for Disease Surveillance, Monitoring & Prediction in Nigeria and AMPATH Medical Record System Data Model were developed for Nigeria and Kenya respectively. This paper presents the review of these models based on scope, strengths, and limitations.

Sub Saharan Africa Health Data Model

In this section of the paper, we present the reviews of the two health data models (AMPATH Medical Record System Data Model and Conceptual Data Model for Disease Surveillance, Monitoring & Prediction in Nigeria) developed for two sub Saharan Africa countries that is Kenya and Nigeria.

AMPATH Medical Record System Data Model (AMRS)

AMRS data model was developed in response to the need to address the growing number of people suffering and dying from HIV/AIDS in Kenya. In 2001, the partnership between Indiana University School and Moi University School of Medicine Eldoret, Kenya led to the development of Mosoriot Medical Record System (MMRS) which was adopted to serve as an HIV/AIDS database.
In 2003, WHO, USAIDS and other international organisation asked the developers to scale down the database to meet the needs of AMPATH and this led to AMRS data model (Mamlin & Biondich, 2005).

The aim of this data model is to meet the immediate need for electronic medical record system to help scale up HIV/AIDS prevention and treatment. In addition, it aims to reduce critical human errors and support the research necessary to guide future efforts.

The model comprises of seven classes namely Encounter, Patient, Observation, Concept, User, Order and Form.

Encounter Class

This class describes places where the health care takes place. This class capture information about name, description and address of the location. It also provides information about the city, state, zip code, longitude and latitude. It gives information about different type of encounters between healthcare provider and patient as shown in figure 1. It captures information about encounter type, encounter date and time among others.

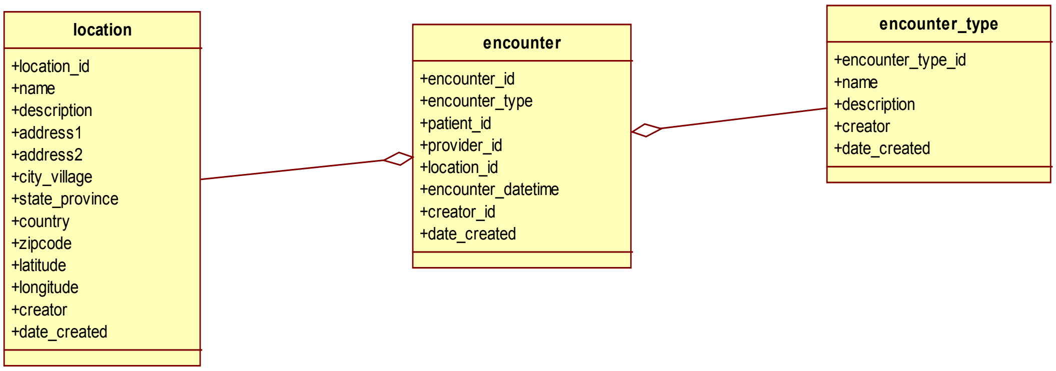
Patient Class

The patient class as shown in figure 2 below captures comprehensive information about the patients. The class has four major sub class namely tribe, patient address, patient identifier, and patient name. The patient address provides information about where the patient reside and the geographical information about the address such as longitude and latitude. Patient name sub class gives information about the patient given name, family name among other things while the patient identifier provides more information about patient. The tribe sub class gives information about the tribe of the patient. All these information are necessary for health care providers in order to have full background information about patient.

Observation Class

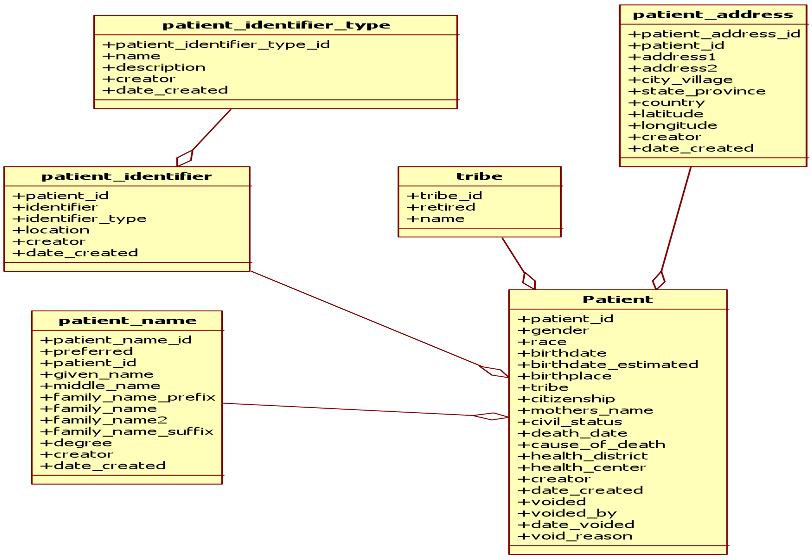
The observation class provides information about all clinically measured or observed things such as weight, platelet count, cardiac examination, etc as shown in figure 3. The observation class reference other class such as order, concept, patient, location and encounter through order identifier, concept identifier, patient identifier, location identifier and encounter identifier. It captures information about date and time observations were made.

Figure 1
Encounter Class



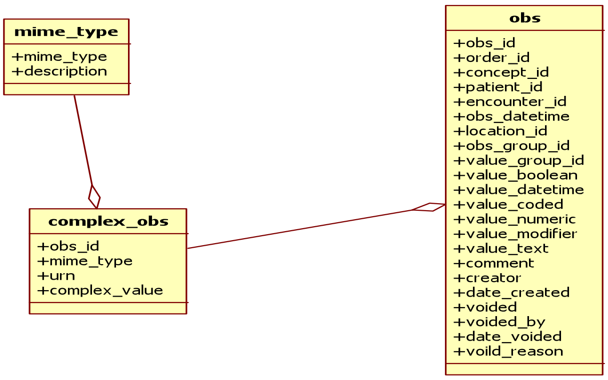
Source: Mamlin BW and Biondich PG 2005. AMPATH Medical Record System (AMRS): collaborating toward an EMR for developing countries

Figure 2
Patient Class



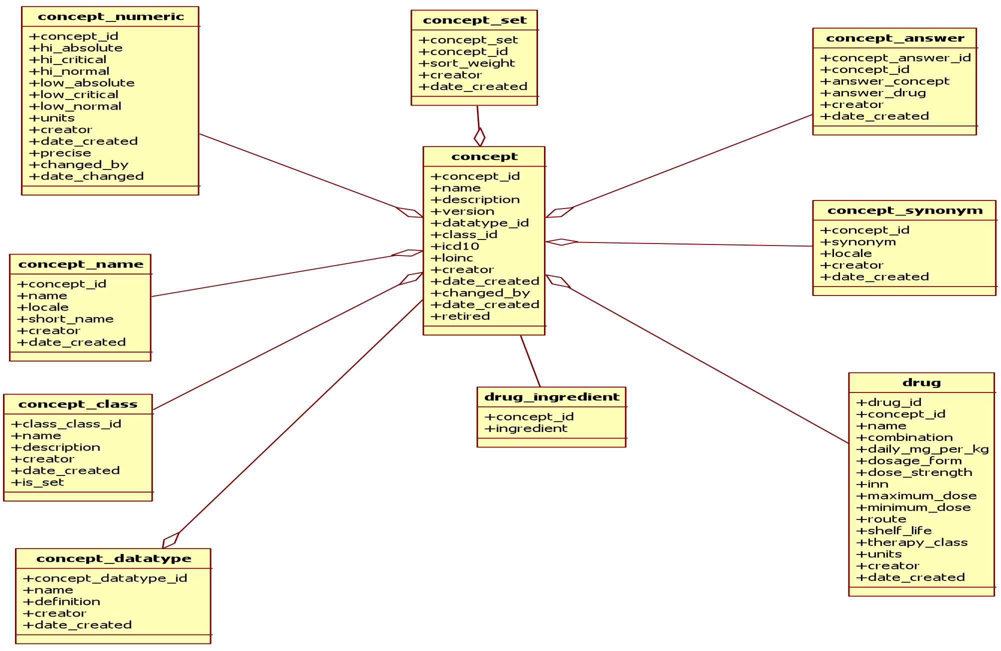
Source: Mamlin BW and Biondich PG 2005. AMPATH Medical Record System (AMRS): collaborating toward an EMR for developing countries

Figure 3
Observation Class



Source: Mamlin BW and Biondich PG 2005. AMPATH Medical Record System (AMRS): collaborating toward an EMR for developing countries

Figure 4
Concept Class



Source: Mamlin BW and Biondich PG 2005. AMPATH Medical Record System (AMRS): collaborating toward an EMR for developing countries

Concept Class

The class provides dictionary of all medical terms used within the model. As shown in figure 4, concept class capture information about concept answer, concept set, concept name, concept synonyms, concept class, concept data type with drug and drug ingredient. In the drug sub class of the class, information about drug dosage and dosage strength with the ingredient of the drug are all captured.

User Class

The user class provides information about the user of the system, its roles and privileges the user has. It gives details information about the names of the users, the user's name and password as depicted in figure 5.

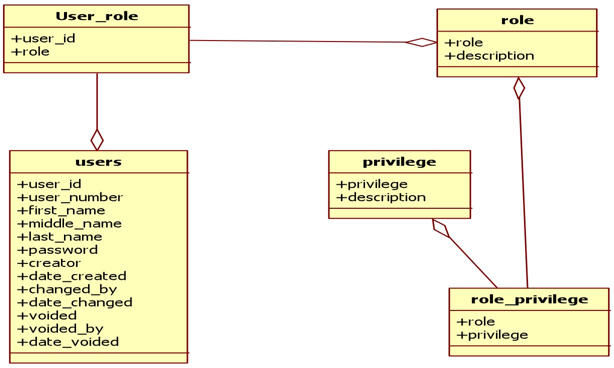
Order Class

The order class as depicted in figure 6 below provides information about drug order and type of order. It capture information about drug's name, drug inventory, dose, quantity, expiring date and the date of commence of the drug among other things.

Form Class

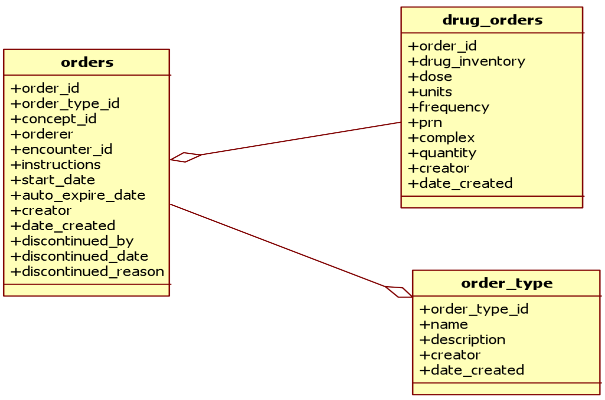
The form class gives information about field type, field answer, form and form field as shown in figure 7 below. The class captures information about field type, field name among others.

Figure 5
User Class



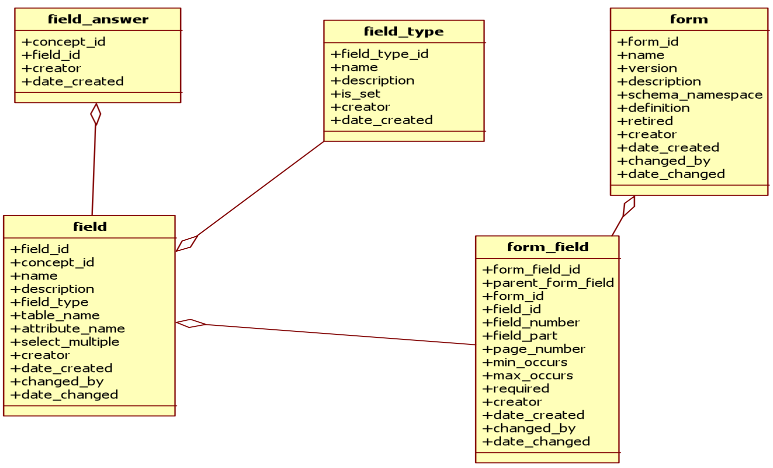
Source: Mamlin BW and Biondich PG 2005. AMPATH Medical Record System (AMRS): collaborating toward an EMR for developing countries

Figure 6
Order Class



Source: Mamlin BW and Biondich PG 2005. AMPATH Medical Record System (AMRS): collaborating toward an EMR for developing countries

Figure 7
Form Class



Source: Mamlin BW and Biondich PG 2005. AMPATH Medical Record System (AMRS): collaborating toward an EMR for developing countries

Strengths and Limitations of AMPATH

The strengths of the model are that the model is already implemented using MySQL, Plone and Python; and the model makes use of coded data so as to allow flexibility and scalability of the model. The limitations of the model identified are that the designer of the model supposes to make use of class relationship in order to link classes in the model together. For example the patient address under patient class is not necessary since encounter capture information about location. Class relationship should be use to establish links between the classes together. In addition, there are some features that are present both in patient address and location such as longitude and latitude, this may lead to redundancy when the database if sully developed.

The model covers HIV/AIDS diseases surveillance only and focuses on Kenya alone though that is what the model is out for as depicted in table 1 below. The model lack spatial feature and it may be difficult to adopt it for the use of public health without including one or more major classes.

Conceptual Data Model for Disease Surveillance, Monitoring & Prediction in Nigeria

This model was developed as a means of having HIV/AIDS disease surveillance system that could be used for surveillance and monitoring of HIV/AIDS in Nigeria.

The idea of having a disease surveillance system for HIV/AIDS in Nigeria was conceived in 2005 at the Department of Computer Science & Engineering, Obafemi Awolowo University, Nigeria but the development of the disease surveillance data model did not commence until January 2007 and not in Nigeria but in England though by a Nigerian researcher from Obafemi Awolowo University, Nigeria and two British researchers at Aston University, Birmingham, United Kingdom. The research was sponsored by Commonwealth Commission.

The aims of the research are of two fold namely immediate and future aims. According to Idowu et al in 2009, the immediate aims are to identify and document the detailed information needed in order to develop effective diseases surveillance, monitoring and management information system; and to enhance the existing health data model by providing support for flexible spatial data. The future aim is to develop well structured diseases surveillance and monitoring system in Nigeria, so as to have adequate epidemiological data to effectively monitor, and prevent diseases

The model comprises of three core components or subject areas which provide required data for development of the disease surveillance database. The components are party, location and health activity. The model is quite unique because of its spatial feature which is not well spell out in any of the existing health data models. This spatial feature allows the use of geometry such as point, line and polygon.

Party Component

The party component contains information about a person, and groups of people that are of interest to the health system such as physicians, epidemiologists, public health workers, hospitals, among others. It also captures information about relationships within the parties. Figure 8 depicts the party component of the model.

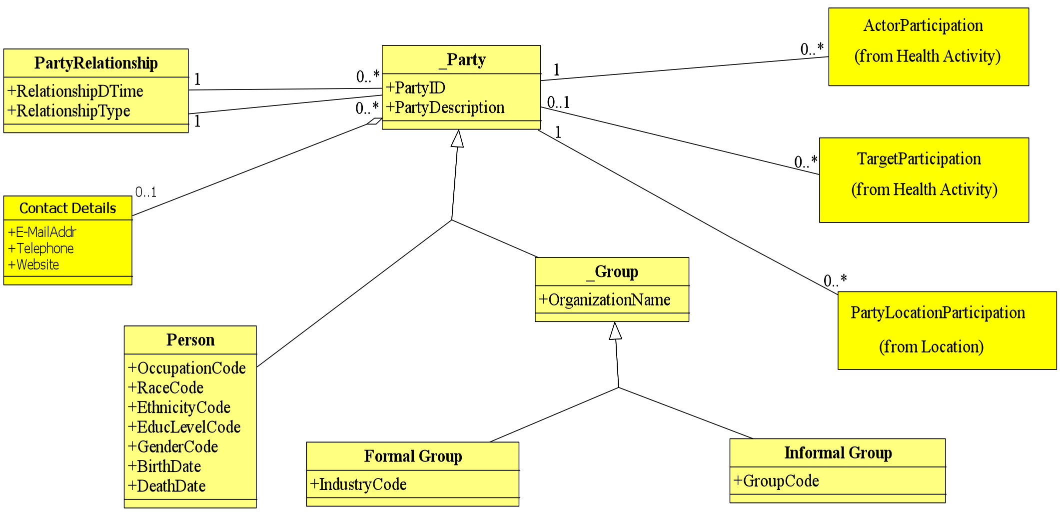
Health Activity

The health activity component captures all the activities that involved in disease surveillance. It contains information about all the activities that occur between patient and health provider. The core health activities in this data model include observation, diagnosis, laboratory test, treatment, and intervention and these are depicted in Figure 9.

Location Component

The location component contains information about the addresses and spatial features associated with the other two core components (Party and Health Activity). Figure 10 below show different types of geometry that can be used to represent location in the model. Location may be used to represent the position of a range of parties and activities including, for example, hospitals, buildings, cities, or local government areas where patients reside.

Figure 8
The party component



Source: PA Idowu, et al 2009. A Conceptual Data Model for Disease Surveillance, Monitoring and Prediction in Nigeria

Figure 9
The health activity component

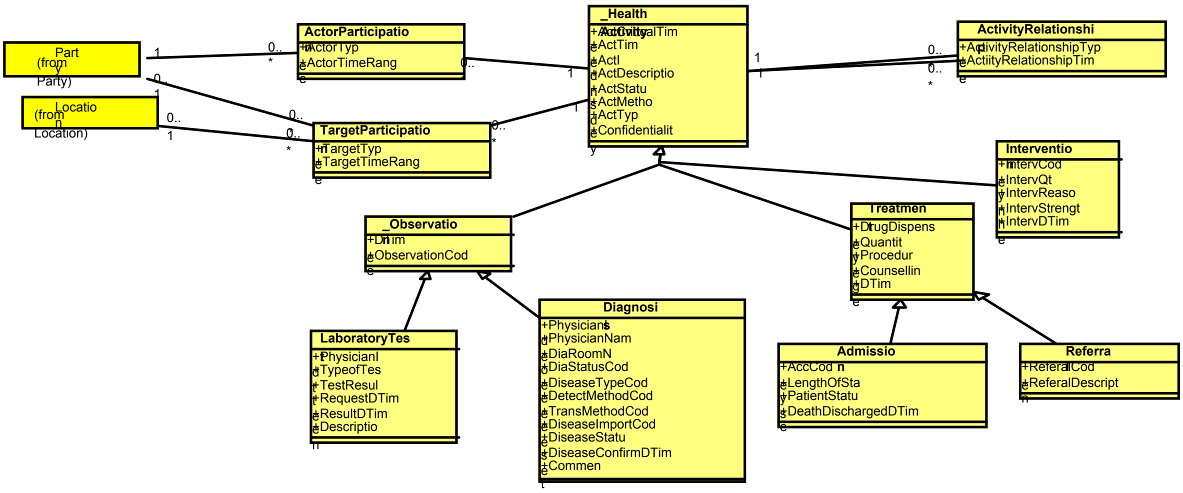
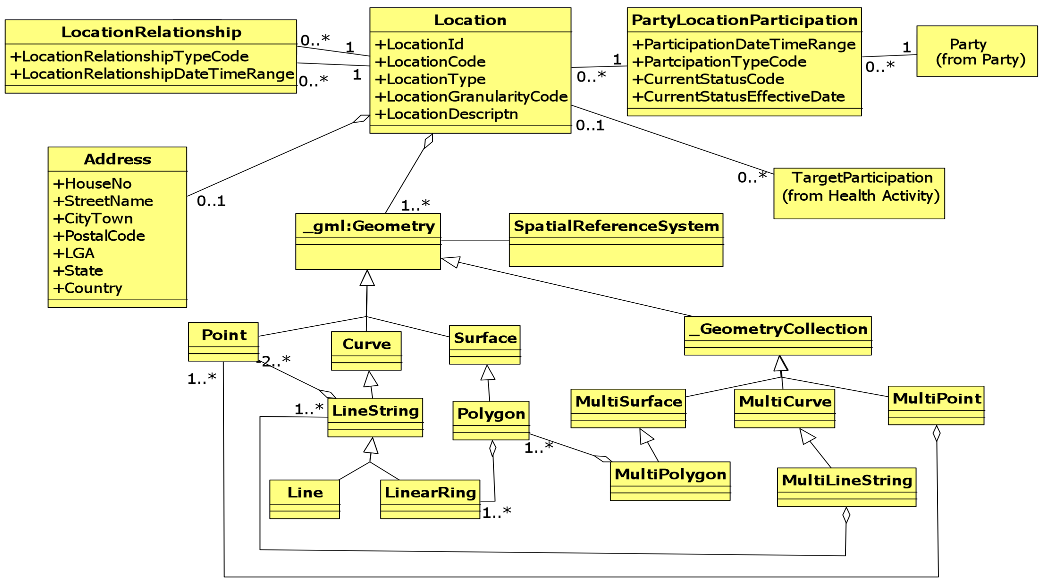


Figure 10
The location component



Strengths and Limitations of Nigeria Health Data Model

In order to develop the model, the data modellers make sure that there are no misrepresentations of attributes within the model. All the attributes required for each of the classes and sub classes are used. The model is flexible, because the model can cope with possible changes and accommodate new data. Codes were used in order to give room for flexibility and extensibility; the model uses HL7 codes so as to allow interoperability of the model with other models most especially those that use HL7 codes.

In addition, the model is simple and easy to understand. The system developer can easy pick the model and use it to develop a disease surveillance system without any problem. Also, the model was developed using UML in order for the system developer to be able to visualize the model. With this visualization feature of the model, users of the model would be able to comprehend the model.

The flexible spatial feature of the model is strengths of the model because it supports the use of geometry such as point, line and polygon. It also makes use of HL7 codes which allow it to support interoperability. In addition, the model was developed using open source data modelling tool and finally it has a complete data dictionary which would make it possible for people to have good understanding of the model in case there is a need to extend, modify and adapt it by other researchers in the future.

The limitations of the model are its inability to cover the whole public health activities though the model was designed for Nigerian diseases surveillance only. The model is not yet implemented. Table 1 below present the summary of the two models based on scope, strengths and limitations.

Table 1
Scope, strengths and limitations of the model

Source	Year	Country	Scope	Strengths	Limitations
AMRS	2005	Kenya	to capture efficient and accurate clinical information	• Implemented • Flexible and scalable	• Not interoperable • Focused on disease surveillance that is HIV/AIDS only • Focused on Kenya alone • Make use of commercial software • Lack spatial feature
NDSS	2009	Nigeria	• to identify and document the information needed to develop diseases surveillance • to provide support for flexible spatial data	• Spatial feature • Support interoperability • Uses open source software • Flexible and extendable	• Focused on disease surveillance • Focused on Nigeria • Not implemented

Comparative Analysis of the two Models

In this section of the paper, we present the comparative analysis of the models based on their structure and functionality. The structural aspect of the model compares the classes or components in each of the models. The AMRA has seven classes namely patient, encounter, observation, concept, user, form and order while Nigerian disease surveillance data model has three classes which are party, health activity and location as discussed above. The classes in each of the model are almost the same but with different nomenclatures.

In AMRA patient class is like party class in Nigerian disease surveillance data model while encounter and observation classes in AMRA data model is health activity class in Nigerian disease surveillance data model. Order class that deals with the handling of drug is not available in Nigerian disease surveillance data model.

Functionality Analysis of the Models
The analysis of models in terms of functionality is based on coverage, interoperability, & spatial support and market relevance.

Coverage, Interoperability and Spatial Support of the Models
Table 2 summarizes the coverage, interoperability and spatial support of the two data models. It shows the country in which the data model is developed for and which of the models support interoperability. It also show which of the models support spatial features which could allow users to use geometry such as point, line and polygon.

In term of coverage both AMRA and Nigeria disease surveillance data models covered Kenya and Nigeria health sector respectively. AMRA data model does not support interoperability because it does not make use of any features such as code from any existing international or leading health data models while the Nigerian disease surveillance data model makes use of HL7 codes. The code would be useful if there are new diseases that need to be added to the existing disease, the disease code and name would be included in the database having in mind the fact that the data model was originally designed for HIV/AIDS.

Nigeria disease surveillance data model make use of spatial feature and is the first health data model to introduce spatial feature to the existing health data models and this make the model unique from other existing health data models but AMRA data model does not support spatial feature.

Analysis of Market Relevance
Table 3 summarizes the market relevance of the two data models. It shows whether or not the data models are commercialised, available as a final specification, whether the models are implemented or not and whether the implementation is in use or not. In addition, It shows whether the models make use of open source software or not and whether the models focus on disease surveillance or not.

Table 2
Coverage, Interoperability and Spatial Support of the Models

	Coverage	Support interoperability	Support spatial feature	Strengths
AMRS	Kenya	No	No	• Implemented • Flexible and scalable
NDSS	Nigeria	Yes	Yes	• Spatial feature • Support interoperability • Uses open source software • Flexible and extendable

Table 3
Market Relevance of the Models

	Commercialised	Final specification available	Implemented	Implementation in use	Open source model	Focussed on disease surveillance
AMRS	No	Yes	Yes	Yes	No	Yes
NDSS	No	Yes	No	No	Yes	Yes

The two models are not commercialised and have final specification available. They are both focussed on disease surveillance, in fact, the two models was designed for prevention and surveillance of HIV/AIDS in Nigeria and Kenya. Nigerian disease surveillance data model have no implementation in use nor implemented but it was developed using open source software and intends to make open while AMRS was develop using proprietary software. Nigerian disease surveillance data model used Star Unified Modelling Language (StarUML), open source software in order to develop the model and, postgreSQL to develop the physical database.

Conclusion

The evaluation of the two health data models in sub Saharan Africa shows that the two models have almost the same structure though the terms used in each of the models are not the same. The models have no commercialised version and have final specification. While the two data models focused on disease surveillance instead of public health as a whole. The two data models are developed for HIV/AIDS surveillance and monitoring. In addition, the two models are developed for the use of a specific country that is Kenya and Nigeria. These are the two major limitations that are common to the models. In the future, it is hope that researchers that are interested in sub Saharan Africa would carry out research that would lead to the development of public health data model, which could be used in almost all the countries in sub Saharan Africa and beyond.

Acknowledgements

We hereby acknowledge the referees, Iyanuoluwa, Iyinoluwa and Titilope and those that contributed to this paper.

References

Beale T. and Heard S. (2005). The open HER common information model. Available from: www.openehr.org/releases/1.0/architecture/rm/common_im.pdf [Accessed 26 May 2008]

Business Intelligence Solution, (2004): Data Model, Data Standards and XML, January, 2004 Available from http://www.knowldege-integrity.com/columns/bi200401.htm [Accessed May 8, 2009]

Canadian Institute for Health Information (2001). *Conceptual health data model version 2.3*. Available from: http://secure.cihi.ca/cihiweb/en/downloads/infostand_chdm_e_CHDMv2_31.pdf [Accessed 17 April 2009]

Centres for Disease Control and Prevention (2000). Public health conceptual data model, *US department of health and human services public health service*, Available from www.cdc.gov/nedss/DataModels/phcdm.pdf [Accessed 3 May, 2009]

DICOM (2007). Digital imaging and communication in medicine Part 1: Introduction and Overview, *National Electrical Manufacturer Association*. Available from: http://medical.nema.org/dicom/2007/07_01pu.pdf [Accessed 22 Mar 2009]

Eichelberg M. Aden T. Reismeier J. Dogac A. and Laleci B.(2005). A survey and analysis of electronic healthcare record standard, *Association for Computing Machinery*. 37(4) p. 277-315

HL7 (2008). *Health level seven*. Available from http://www.hl7.org/ [Accessed 4 May, 2009]

HL7 News, (2006). HL7: On the front lines of interoperability, The official publication of Health Level Seven Inc, January 2006.

Mamlin BW and Biondich PG (2005). AMPATH Medical Record System(AMRS): collaborating toward an EMR for developing countries. AMIA symposium proceedings pp 490-494.

Idowu PA, Cornford Dan & Bastin Lucy (2009): A Conceptual Data Model for Disease Surveillance, Monitoring and Prediction in Nigeria. Proceeding of Second International Conference on Health Informatics, pp 443-449, January 14-17, 2009, Hotel Gaia Melia Porto, Porto, Portugal.