



AUTOMATING THE IMPORT OF HEALTH AND DEMOGRAPHIC SURVEILLANCE SYSTEM DATA FROM ODK TO OPENHDS USING MIRTH CONNECT

Iwara I. Arikpo

Department of Computer Science, and
CRHDSS UNICAL, Directorate of Research
University of Calabar
Calabar, Nigeria

Ideba M. Ekinya

Department of Computer Science, and
CRHDSS UNICAL, Directorate of Research
University of Calabar
Calabar, Nigeria

Abstract: Health and demographic surveillance systems (HDSS) are community-based platforms that collect longitudinal data on core demographic and socioeconomic indices as well as key health indicators at regular intervals within defined geographical populations. These systems were previously paper-based. However, the advent of electronic data collection and processing with heterogeneous technologies has posed a problem for data validation and integration. This paper describes the development of ETL channels for automated import of HDSS data from ODK to OpenHDS. The setting was the Cross River HDSS, University of Calabar, South-south Nigeria with a surveillance population of 37,808. Data collection by Fieldworkers was done with Android smartphones running the ODK Collect, while the Supervisors used Android tablets devices to review data prior to submission to an intermediate cloud-based ODK Aggregate server. The validated data was serviced by the OpenHDS and resided on MySQL server maintained locally at the Data Centre. Channels were developed in Mirth Connect to read field data from ODK, validate against the OpenHDS repository and insert onto MySQL database, rejecting inconsistent data when found. Twenty-six channels were developed, tested and deployed for import of data onto the OpenHDS. With contextual modification, this tool can be extended to other health and demographic surveillance systems (HDSSs) under the INDEPTH Network.

Keywords: mirth connect, mobile devices, data collection, extract transform load, database, OpenHDS, server

I. INTRODUCTION

Health and demographic surveillance systems (HDSS) are community-based platforms that collect longitudinal data on core demographic and socioeconomic indices (births, deaths, marriages, migration, and social amenities) as well as key health indicators at regular intervals within defined geographical populations called demographic surveillance areas. These systems are established in settings where the health management information systems (HMIS) are weak or near-absent. There are 49 HDSS sites from 19 countries in Africa, Asia and Oceania, with 32 of the sites in sub-Saharan Africa [1]; [2].

Due to their rigor, various data collection platforms are deployed in health and demographic surveillance system to sustain the longitudinal nature of their data systems. Traditional paper approaches of data collection were common, until recently when electronic methods are gradually taking over. Some of the electronic tools include ODK (open data kit), FormHub, OpenXData, Magpi (formerly Episurveyor), Microsoft Data Gathering (formerly Nokia Data Gathering), EpiCollect, and REDCap. For HDSS sites that are using non-paper approaches, data forms are designed with electronic tools, most of which are open source, thereby making the process of data collection easy and seamless, reducing the challenges experienced in the manual and paper-based data collection methods [3]. Despite the reliefs engendered by electronic data collection in HDSS operations, there is the problem of transmitting data of dissimilar structure and formats from mobile devices

and validating same against a central database for analysis and reporting. The goal of this paper is to describe the use of Mirth Connect in developing an extract, transform and load (ETL) tool that automatically imports community-based data from ODK to the core HDSS repository, the OpenHDS.

The first effort at ETL (extract, transform, and load) technology was the EXPRESS system developed by [4]. It was the first experimental prototype data translation system intended to act as an engine that produces data translations, taking as input data definitions and conversion of nonprocedural statements. Although the system suffered major drawbacks in the later years, the focus on data integration and construction of wrapper-based mediator was a primitive ETL scripting for exchange of data amongst integrated database systems [5].

Currently, there are several open source and proprietary ETL tools offering a range of application features based on user targets. Some of these include DataStage, Microsoft's SQL Server Integration Services, Mirth Connect, Pentaho, Oracle Data Integrator, and Fussion HumMer.

Mirth Connect (now NextGen Connect) is an open source, platform-independent solution for integration of data between healthcare applications (<https://www.nextgen.com/products-and-services/integration-engine>). It has been described as one of the most well-established interoperability frameworks along with Open Health Tools Integration Platform [6]. Mirth Connect is considered one of the best choices because of its

increasing popularity, ability to handle scalable message volumes, and low-cost licensing.

Mirth Connect works by creating data channels that get the data from a source, validates, transforms, and loads it onto one or more target destinations. Its open source licence, platform-independence, extensibility as well as native support for multiple standards and network infrastructure seems to be one of the most important reasons for its popularity [7]; [8]. Mirth Connect can realize the mutual communication amongst different forms of medical and health data, such as JSON, CDA, HL7 message, SOAP, XML, and so on, to realize the intercommunication of data between heterogeneous systems, serving as an integration engine in this respect [9]; [10]. Within the healthcare domain, Mirth Connect can be deployed to integrate heterogeneous information into a desired format [11].

Mirth Connect has been severally and successfully used as an ETL in health projects, including the Health Information System Programme (HISP), India to interface between two systems for the transfer of data [12]; the Capstone Project where Mirth Connect was used as integration between laboratory information system (LIS) and EMR – electronic medical record system [13]; a C-CDA (consolidated clinical document architecture) project where Mirth Connect was used as a middleware to transmit different form data between systems [9]; Help4Mood Project where Mirth Connect was deployed to interact with different data subsystems giving different priorities for messages in queues [14], just to mention but a few.

II. MATERIALS AND METHODS

A. Study area

The Cross River Health and Demographic Surveillance System (Cross River HDSS) operates two research cohorts located within the southern senatorial district of Cross River State in South-south Nigeria, with a combined population of 37,808 persons in 9,452 households as at 2018 (about 47.4% of which are rural dwellers). The first is a rural cohort located in the Akpabuyo Local Government Area (LGA) of the state and the second, an urban cohort located in Calabar Municipal, the state capital (Figure 1). The rural and urban sites are further delineated into 46 and 43 contiguous Enumeration Areas (EAs) respectively. Calabar Municipal, with an area of 142.74 km², is located on 4° 58' 28.056" North and longitude 8° 20' 29.9328" East. The city majorly comprises of the Quas and Efiks ethnic groups. Similarly, Akpabuyo LGA, spanning an area of 813.68 km², [15] is located on 4° 57' 0.4248" North and longitude 8° 23' 35.9088" East. In addition to the Quas and Efiks found in Calabar Municipal, Akpabuyo LGA also has the Efuts ethnic group. Both sites are situated in the tropical rain forest belt of southern Nigeria, with an annual rainfall in the range of 2500mm to 3000mm and mean annual temperature of 30°C [16].

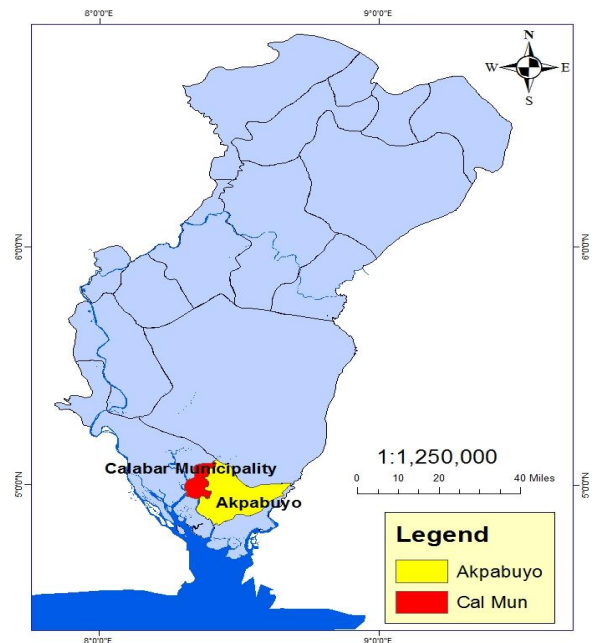


Figure 1. Map showing urban and rural cohorts of Cross River HDSS

B. Study design

The authors have been actively involved in the operations of the Cross River HDSS, University of Calabar, Nigeria, since 2011 when the first surveillance round was conducted (predominantly with mobile technology). Figure 2 shows the conceptual model of a health and demographic surveillance system with the OpenHDS running the core data system; while Figure 3 presents the architecture of the OpenHDS software. Data Access Object (DAO), Domain, Service, Web, and Report are the main modules of OpenHDS core architecture that provide the minimal functionality needed to run a longitudinal data system.

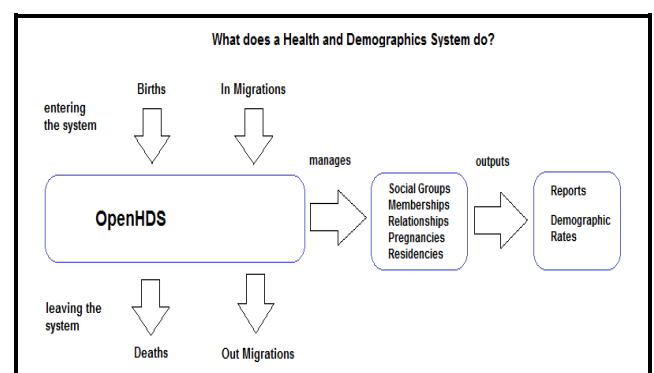


Figure 2. Conceptual model of a HDSS [17]

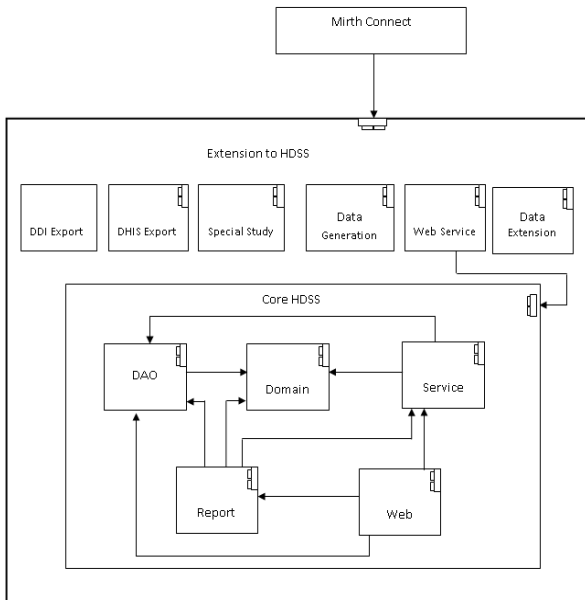


Figure 3. OpenHDS application core architecture [18]

The architecture of Mirth Connect is presented in Figure 4. Mirth Connect channel has one source connector used to pull data from database, file system or File Transfer Protocol (FTP) server and always ends with the destination connector. Every channel must have at least, but not limited to one destination connector. This depends on the number of applications to be integrated. The source and destination connectors can be chained together for additional compound logic, and may also have one or more filter and transformers. The channels developed in this study using Mirth Connect utilizes one of the physical data extensions to the core HDS module (web service) to transmit data to OpenHDS using a web service call.

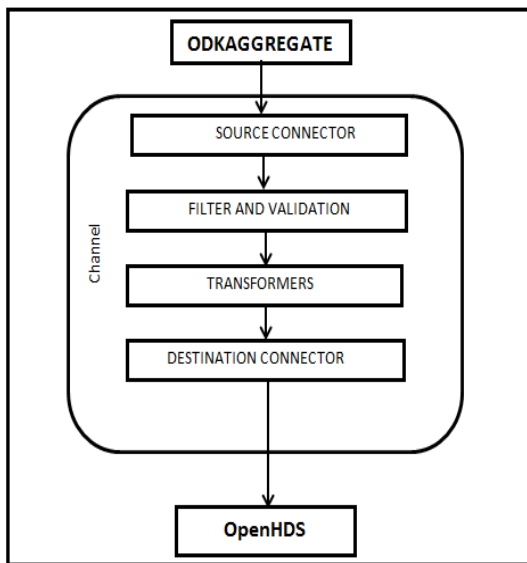


Figure 4. Mirth Connect channel-based architecture [11]

The use case diagram in Figure 5 illustrates the users' interactions with the system. As shown in the diagram, a user known as the Field Worker visits the field site and

interviews eligible respondents from each household within the surveillance area. Using ODK Collect mobile client running on Android phones, the fieldworker collects data and fills the mobile form instance of that event. At the close of work each day, the fieldworker uploads the finalized forms to the ODK Aggregate data repository hosted locally on the CRHDSS server. Based on startup settings by the system administrator on Mirth Connect Fieldworker channel, the uploaded records are moved to an intermediate repository called Mobile Interop server. The submitted forms with data are then downloaded and previewed by the Data quality supervisor (DQS) using the Mobile Helper client (an application used by quality supervisors to view data).

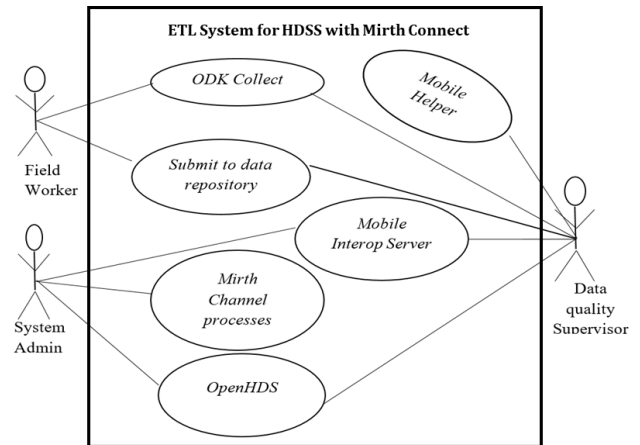


Figure 5. Use case diagram showing users interaction with the system

The Mobile Helper hosts all forms awaiting preview and validation. Thereafter, the supervisor reviews the data collected by fieldworkers on ODK Collect before uploading all finalized and saved forms to ODK Aggregate server. The cycle continues until all validated records are extracted, transformed and loaded onto the MySQL backend database of the OpenHDS. All validated forms by the DQS (Data quality supervisor) are differentiated from the fieldworker's form with a unique data field called the "DERIVED_ID". At this point, since the supervisor has submitted the form, Mirth Connect now attempts to insert the validated data onto the OpenHDS database. If this operation fails, the form is sent back to Mobile Interop server where it can be downloaded by the supervisor at a later time to fix the validation failure, otherwise, the OpenHDS database is updated with the new data.

Activities of the Data quality supervisor include, checks on the destination application (OpenHDS) in order to correct queries and inconsistency in the fieldworkers' submissions. On the other hand, the system administrator has superuser access to each level of the application and does all the necessary configurations and troubleshooting. Data collection tools were all designed with ODK 1.4 (<https://opendatakit.org/>). The system was implemented using smartphones running the Android 7.0 (Nougat) operating system. The mobile devices were used for field data collection. The local server was also setup on a Dell

PowerEdge T410 server running Microsoft 2008 R2 Server Operating System.

III. RESULTS AND DISCUSSION

Some concurrent activities that formed the major implementation tasks were repeated throughout the implementation process. The completed data collection tools released at each version (cycle) of the implementation were: Visit Registration form, Out-Migration form, Migration-In form, Pregnancy Registration form, Birth Registration form, Death Registration form, Social Group Registration form, Location Registration form and Vaccination.

The next stage of the implementation was the development and setup processes of the Mirth Connect, Mirth channels, Mobile Interop server, Mobile Helper, OpenHDS and ODK Collect applications. Borrowing a leaf from [19], that innovative ideas do not start from the scratch, Mirth Connect was used to build the channels for seamless import of data from the ODK Aggregate server to the OpenHDS application running MySQL database server at the backend. The robustness of Mirth Connect is its ability to use filters in processing messages of varying kinds and sources. Filter was built into a message processing mechanism and it is useful for determining whether the message should be processed or not.

A total of 26 channels were developed, out of which three channels (Fieldworker channel, Supervisor channel, Sender channel) were developed for the processing of each data form. The two other channels, Form-Submission and Form-Complete, were developed for sending of forms to the configured Mobile Interop server. The other task of channels includes enforcing data integrity and loading of the data onto OpenHDS database.

The deployment of channels was done in the following order: Form-Submission, Form-Complete, Field-Worker, Sender, and Field-Supervisor channels. The Field-Supervisor channel's initial state was configured to "stopped". This allowed for forms to be processed in an orderly manner, thereby minimizing the failure rate of form processing. The completed forms from Fieldworkers are uploaded onto ODKAggregate server as shown in Figure 6. At this stage, the fields derived From Uri, Supervisor Status, and Proccesed By Mirth are blank, while the Validation Failed field is false. This is because the forms are yet to be processed by Mirth Connect.

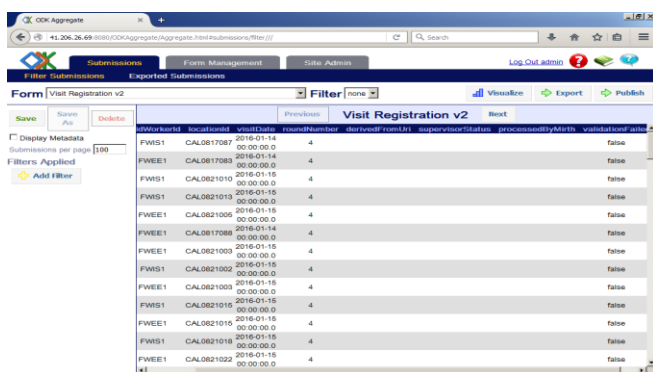


Figure 6. Form submission from Fieldworkers

Figure 7 is a screenshot of ODK Aggregate server indicating that the *processedByMirth* column is updated to "1" (true) meaning that Mirth Connect has processed the Fieldworker form and moved it through the Form Submission channel to Mobile Interop server for the Data quality or Fieldworker Supervisor to download and preview at some point in time. The screenshot of Figure 8 shows that the Data quality supervisor successfully downloaded the form from Mobile Interop server, performed data quality checks, thereafter resubmitted to ODKAggregate. This results in the *derivedFromUri* and *SupervisorStatus* columns being populated.

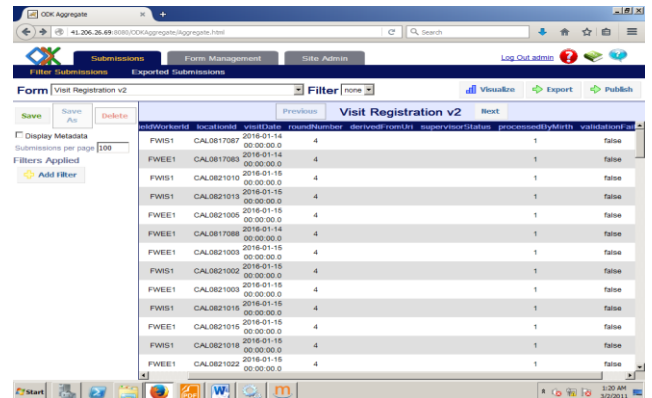


Figure 7. ODKAggregate showing Field Worker form processed by Mirth

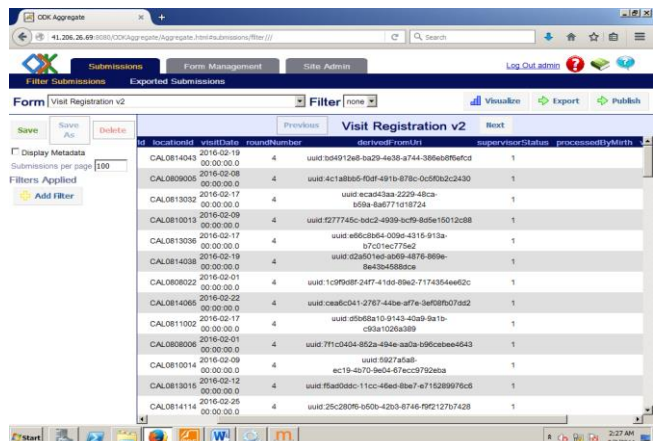


Figure 8. ODKAggregate preview after form re-submission by supervisor

After resubmission by data quality supervisors, Mirth Connect now begins the ETL process of loading the data onto the MySQL database of the OpenHDS. Figure 9 is a screenshot of ODKAggregate server showing populated columns for *derivedFromUri*, *SupervisorStatus*, and *ProccesedByMirth*. This is because the supervisor has downloaded, previewed and resubmitted the form earlier submitted by the fieldworker.

The *ValidationFailed* column updates with *false* and *true* status once Mirth Connect processes the forms (Figure 9). The *false* status indicates that the form that has been downloaded, previewed and resubmitted by the supervisor did not fail validation, while the *true* status shows that the Supervisor form failed validation and could not be imported onto the OpenHDS database, but moved into Mobile Interop server for a re-download and correction by the Supervisor.

At this point, the form is downloaded with error code showing the cause of the validation failure (Figure 10).

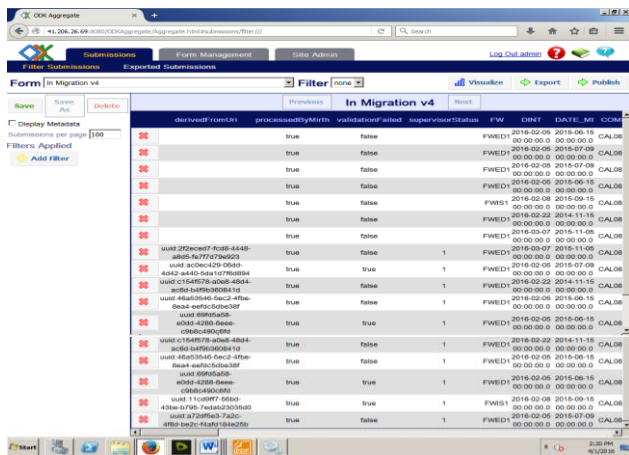


Figure 9. ODK Aggregate showing records with failed and passed validation checks



Figure 10. OpenHDS Mobile Helper showing downloaded form with error

In Figure 11, the Mobile Interop server shows the forms that were transmitted by Form Submission channel to the Interop server. The form group with the label *Ongoing with one submission* indicates that the form has never been previewed by the Data quality supervisor, while the group with label *Ongoing with more than one submission* shows that the form has been previewed by the supervisor but must have failed validation test. Such forms must be previewed again. Finally, the forms with Finished label indicate that the form passed all validation checks and was sent successfully onto the OpenHDS database through the Sender channel.



Figure 11. Mobile Interop showing form status

Figure 12 shows Mirth Connect dashboard with summary of data submissions as well as details of all the channels and forms processed. These include date and time of last deployed channels, the number of records for each form received by the respective channels, number of records queued, sent, filtered and errored. It is interesting to note that, all the records could now drop onto the OpenHDS database after all initial validation failures were traced and resolved; hence the reason for the '0' results along the fields *Queued* and *Errored* on the Mirth Connect dashboard.

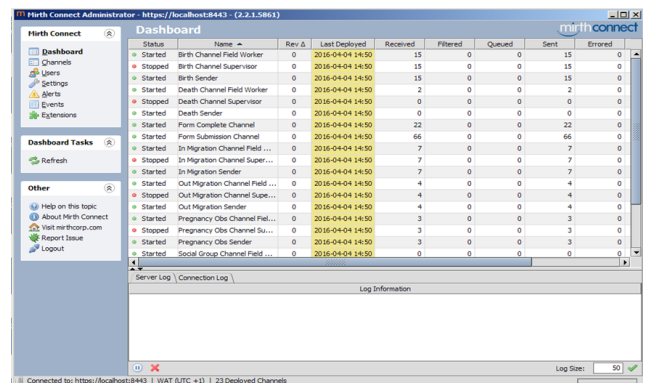


Figure 12. Mirth connect showing forms successfully sent to OpenHDS database

IV. CONCLUSION

This study has demonstrated the use of Mirth Connect (*now NextGen Connect*), an open source ETL in the building and maintaining of channels that interface between two applications (ODK and OpenHDS), transforming the data into standard format suitable for the destination application. Data were filtered, validated, and routed based on user-defined criteria using Mirth Connect as demonstrated in this paper. The software implementation in this study was at the Cross River HDSS, University of Calabar, Nigeria. Twenty-six channels were developed for import of data from ODK to the OpenHDS repository running on MySQL server. With contextual modification, this tool can be extended to other

health and demographic surveillance systems (HDSSs) under the INDEPTH Network (<http://www.indepth-network.org/>).

V. FUTURE SCOPE

The focus of this study had been on the implementation of longitudinal data import from ODK to OpenHDS. The latter system being an emerging data system recommended by the INDEPTH Network for the over 42 health and demographic surveillance system (HDSS) sites across Africa, Asia and the Oceania. Future research will focus on integrating community-based data from OpenHDS with DHIS2, a health facility-based aggregate data platform adopted by many developing countries for the purpose of monitoring health interventions in these settings. This will facilitate comparisons between facility-based data and community-based data, for better health intervention outcomes.

ACKNOWLEDGMENTS

The authors duly appreciate the support of Professor Martin Meremikwu, Principal Investigator and other Research Team colleagues as well as Fieldworkers at the Cross River Health and Demographic Surveillance System (Cross River HDSS), University of Calabar, Nigeria for providing the platform for this research.

REFERENCES

- [1] Ye, Y., Wamukoya, M., Ezeh, A., Emina, J., & Sankoh, O. (2012). Health and demographic surveillance systems: a step towards full civil registration and vital statistics system in sub-Saharan Africa? *BMC Public Health*, 12(741). doi:10.1186/1471-2458-12-741
- [2] Homan, T., Di Pasquale, A., Kiche, I., Onoka, K., Hiscox, A., Mweresa, C., Mukabana, W., Takken, W. & Maire, N. (2015). Innovative tools and OpenHDS for health and demographic surveillance on Rusinga Island, Kenya. *BMB Research Notes*, (2015)8:397. DOI: 10.1186/s13104-015-1373-8
- [3] Arikpo, I., Mbotto, I., Okoro, A., James, U., Aquaisua, E., Osonwa, K., Ushie, M., Enang, E., Asangansi, I., Meremikwu, M. (2013). Optimum hardware, software and personnel requirements for a paperless health and demographic surveillance system: a case study of Cross River HDSS, Nigeria. *Information and Knowledge Management*. 3(3) 77 – 83.
- [4] Shu, N. C., Housel, B. C., Taylor, R., W., Ghosh, S. P. & Lum, V. Y. (1977). EXPRESS: A data extraction, processing, and restructuring system. *ACM Transactions on Database Systems* 2(2):134-174. DOI: 10.1145/320544.320549
- [5] Roth, M. T. and Schwarz, P. (1997). "Don't Scrap It, Wrap It! A Wrapper Architecture for Legacy Data Sources," *Proceedings of the Conference on Very Large Data Bases (VLDB)*, Athens, Greece, August 1997, ACM, New York.
- [6] Karopka, T., Schmuhl, H., Marcelo, A., Molin, J., & Wright, G. (2011). Towards open collaborative health informatics—The role of free/libre open source principles. *Yearbook of Medical Informatics*, 6(1), 63-72.
- [7] Bortis, G. (2008). Experiences with Mirth: an open source health care integration engine. *Proceedings of the 30th international conference on Software engineering*. ACM, 649-652.
- [8] Lampikoski, T. (2012). Green, Innovative, and Profitable: A Case Study of Managerial Capabilities at Interface Inc. *Technology Innovation Management Review*.
- [9] Shang, Y., Wu, H. & Jiang, K. (2019). The Development of an Aggregated Electronic Health Record in Compliance with Consolidated Clinical Document Architecture. *ICMHI 2019: Proceedings of the third International Conference on Medical and Health Informatics (May 2019)* 71 – 77. DOI 10.1145/3340037.3340039
- [10] Marceglia, S., D'Antrassi, P., Prenassi, M., Rossi, L. & Barbieri, S. (2017). Point of Care Research: Integrating patient-generated data into electronic health records for clinical trials. *AMIA Annu Symp Proc*, (2017) 1262 – 1271.
- [11] Lang, Chris & Bortis Gerald (2007). Webreach mirth 1.5 webinar. Retrieved from http://www.fusionspec.com/webinar/webinar_0002_mp4_MPEG4_large_dl.mp4
- [12] Marketwire (2013). *HealthConnections RHIO Transitioning HIE Platform to Mirth Corporation*. Retrieved from <http://www.marketwired.com/press-release/healthconnections-rhio-transitioning-hie-platform-to-mirth-corporation-1753371.htm>
- [13] Parchariyanon, S. (2011). Evaluation of open-source versus commercial healthcare interoperability tools. *Scholar Archive*, 620.
- [14] Castro, J., Domingo, A., Colomé, J. & Estevez, S. (2011). HL7 in Personal Health System component's integration for Mental Health Treatment. *Journal of Health Informatics*. 3(2011): Especial 2011: 73 – 76.
- [15] National Population Commission. (2014). *Nigeria Demographic and Health Survey*. Abuja, Nigeria: ICF International.
- [16] Arikpo, I., Okoro, A., Esu, E., Aquaisua, E., Mbotto, I., Meremikwu, M. (2019). Differences in Population Dynamics and Uptake of Reproductive Health Services in the Urban and Rural Cohorts of Cross River Health and Demographic Surveillance System of Southern Nigeria. *Developing Country Studies* 9(5) 64 – 71.
- [17] Hartsock, B., MacLeod, B., Roberge, D. & Asangansi, I. (2011). Software Extensibility Strategies for Health and Demographic Systems in Low-Income Countries. *Proceedings 2011 IEEE Global Humanitarian Technology Conference*, Seattle, WA, USA. DOI: 10.1109/GHTC.2011.56
- [18] OpenHDS Wiki (2016). *OpenHDS Modulare Architecture*. Retrieved from <http://wiki.openhds.googlecode.com/git/images/openhdsModularArchitecture.png>
- [19] Clear, J. (2020). *How Innovative Ideas Arise*. Retrieved from <https://jamesclear.com/dont-start-from-scratch>.