2018-12-12

A Framework for Timely and More Informative Epidemic Diseases Surveillance: The Case of Tanzania

Rutatola, Edger P.

Journal of Health Informatics in Developing Countries

Journal of Health Informatics in Developing Countries

http://www.jhidc.org/
Vol. 12 No. 2, 2018

Submitted: July 29th, 2017 Accepted: October10th, 2018

A Framework for Timely and More Informative Epidemic Diseases Surveillance: The Case of Tanzania

Edger P. Rutatola[1], Zaipuna O. Yonah[1], Devotha G. Nyambo[1], Geofrey J. Mchau[1], Albogast K. Musabila[2]

[1]The Nelson Mandela African Institution of Science and Technology, P.O Box 447 Arusha, Tanzania
[2]Mzumbe University, P.O Box 1 Mzumbe, Morogoro, Tanzania

Abstract:
Background: A number of health facilities in the United Republic of Tanzania use different Hospital Management Information Systems (HoMISs) for capturing and managing clinical and administrative information for report generation. Despite the potentials of the data in the systems for use in epidemic diseases surveillance, timely extraction of the data for integrated data mining and analysis to produce more informative reports is still a challenge. This paper identifies the candidate data attributes for epidemic diseases surveillance to be extracted and analyzed from the Government of Tanzania Hospital Management Information System (GoT-HoMIS). It also examines the current reporting setup for epidemic diseases surveillance in Tanzania from the health facilities to the district, regional, and national levels.
Methods: The study was conducted at the Ministry of Health, Community Development, Gender, Elderly, and Children (MoHCDGEC), Tumbi Designated Regional Referral Hospital (TDRRH), Muhimbili University of Health and Allied Sciences (MUHAS), and Mzumbe Health Centre, all in the United Republic of Tanzania. A total of 10 key informants (medical doctors, epidemiologists, and focal persons for various health information systems in Tanzania) were interviewed to obtain primary data. Data entry process in the GoT-HoMIS was also observed. Documents were reviewed to broaden understanding on several aspects.
Results: All the respondents (100%) suggested patients' gender, age, and residence as suitable attributes for epidemic diseases surveillance. Other suggested attributes were occupation (85.71%), diagnosis (57.14%), catchment area population (57.14%), vital status (57.14%), date of onset (57.14%), tribe (42.86%), marital status (42.86%), and religion (14.29%). Timeliness, insufficient immediate particulars on an epidemic-prone case(s), aggregated data limiting extensive analytics, missing community data and ways to analyze rumors, and poor data quality were also identified as challenges in the current reporting setup.
Conclusion: A framework is proposed to guide researchers in integrating data from health facilities with those from social media and other sources for enhanced epidemic disease surveillance. Data entrants in the systems should also be informed on the essence and applications of data they feed, as quality data are the roots of quality reports.

Keywords: District Health Information System; Integrated Disease Surveillance and Response; Epidemic Diseases Tanzania; Data Integration; Health Datasets; Health Management Information System; Data Mining, Data Warehouse.

[1] Edger P. RUTATOLA; The Nelson Mandela African Institution of Science and Technology, P.O BOX 447 Arusha, Tanzania; Tel: +255 658 526 256; Email: rutatolae@nm-aist.ac.tz; eprutatola@mzumbe.ac.tz; edgerrutatola@gmail.com

1. Introduction

Increase in number of local software developers and availability of open source software has led to improved development and management of various Hospital Management Information Systems (HoMISs) in developing countries [1, 2]. In the United Republic of Tanzania (URT), a number of Hospitals (public and private) operate different Information Systems for storage and manipulation of clinical and administrative information [2, 3, 4]. Adoption of the Information systems is made possible due to the support and emphasis from the Government on development and operationalization of such systems. Notable advantages have been gained from the systems, including but not limited to generation of reports that help in hospital administration as well as monitoring clinical operations. Moreover, successive systems such as the District Health Information System (DHIS2) have been installed to generate more comprehensive and integrated reports on different aspects including epidemics surveillance from districts to national level [5]. Ideally, these systems are expected to generate quality and vast datasets for diversified analytical purposes, as quality data is a success factor for generation of useful reports [6]. Good quality data can be used to find demographic and clinical patterns in disease cases from multiple hospitals and provide more insights to epidemiologists as well as detailed and reliable statistical based reports such as the Tanzania Malaria Indicator Survey, 2017 [7, 8].

One among the Hospital Management Information Systems installed in Tanzanian health facilities is the Government of Tanzania – Hospital Management Information System (GoT-HoMIS). It is a system developed by local experts and currently installed in more than one hundred and seventy (170) health facilities [2]. While such a wide adoption of the system could be advantageous for data mining and analytics, the GoT-HoMIS nodes are currently not centralized and hence analysis of collected data is only limited to the host facility. At present, aggregated clinical data from all health facilities in Tanzania are integrated and stored in the DHIS2 on a monthly basis, where analytics are performed, and reports are generated [9]. From the aggregated data, only limited analytics and patterns can be observed. Epidemics surveillance needs to be timely and comprehensive so that actions towards an epidemic are effective [10].

This paper explores the potentials of the patients' clinical and demographic data collected by the GoT-HoMIS for epidemics surveillance. It also examines existing systems and setup for epidemic diseases surveillance in Tanzania, from where a case is confirmed at the health facility to when reports are shared at the district, regional, and national epidemiological teams. Furthermore, a framework for integrating health datasets generated from different health facilities, community data gathered from rumors, and other environmental data such as meteorological data is proposed. With comprehensive integrated dataset from the diverse sources machine learning algorithms can be applied for enhanced analytics to aid epidemics surveillance and informed decision making. The proposed framework is expected to guide researchers on enabling the application of data mining tools on the resulting datasets to promote generation and dissemination of more comprehensive and timely reports for epidemics surveillance and control.

2. Methods

2.1 Study population

The study population comprised of medical doctors, epidemiologists, health information systems focal persons, and health information systems administrators. These were either based at the Ministry of Health, Community Development, Gender, Elderly, and Children (MoHCDGEC), or among three health facilities located in Morogoro, Dar es Salaam, and Pwani Regions in the United Republic of Tanzania. The chosen health facilities were Tumbi Designated Regional Referral Hospital (TDRRH) (Pwani Region - Tanzania), Mzumbe Health Centre (Morogoro Region - Tanzania), and the Muhimbili University of Health and Allied Sciences (MUHAS) (Dar es Salaam Region - Tanzania). These were purposively selected based on their roles and position in the Tanzanian health sector (MoHCDGEC); experience in operating HoMISs, specifically the GoT-HoMIS (TDRRH and Mzumbe Health Centre); and having a nominated HoMIS pioneer (MUHAS).

A total of ten (10) representative respondents from the mentioned facilities were involved in the study. These were obtained through snowball sampling, based on their key roles and knowledge on epidemiology, and experience in working with the GoT-HoMIS or other Health Management Information systems and innovations in the country. Seven (7) of the ten respondents were medical doctors and/or epidemiologists, who were further divided into two categories. The first category is of those who are stationed in hospitals. This category had four (4) medical doctors; one being stationed at Mzumbe Health Centre, two at TDRRH, and one from MUHAS. These were chosen following their experience in working with HoMISs, particularly the GoT-HoMIS. The second category (Team) comprised of doctors and epidemiologists stationed at the MoHCDGEC in Dar es Salaam, Tanzania. This category had three (3) doctors/epidemiologists; one being the national IDSR focal person, another one being the Public Health Emergency Operations Centre (PHEOC) manager, and lastly a full-time epidemiologist. This category was selected based on their key knowledge on epidemiology and their roles in the current systems and innovations in support of the same.

The remaining three (3) respondents were not of the medicine or epidemiological background. One was a statistician at the MoHCDGEC and a national DHIS2 focal person. The other two were systems administrators having more than four years of experience working with the GoT-HoMIS.

2.2 Data Collection

Semi-structured interviews were conducted to obtain primary data. The interviews intended to identify key attributes to be extracted from the GoT-HoMIS nodes in the health facilities that can help in facilitating timely monitoring and analysis of epidemic cases. The other part of the interviews, specifically dedicated to respondents from the MoHCDGEC, aimed to derive information on the current setup for reporting on disease cases, especially epidemics, from when the diagnosis is made at the health facility to the district, regional, and national responsible teams.

Observation method was also employed on the patients' registration processes at the Tumbi Designated Regional Referral Hospital. Firsthand observation was carried out on how patients were registered over time. The objective was to examine the quality of the data fed into the GoT-HoMIS in terms of correctness and completeness.

Moreover, documents were also collected and reviewed to obtain additional knowledge. Review was done on literature and various documents focusing on health information systems and the setup for reporting and monitoring in the Tanzania's health sector. Some of the documents were obtained online, whereas some were provided by focal persons at the MoHCDGEC.

3. Findings

3.1 Respondents' Profiles

Seven (7) of the ten respondents (medical doctors and/or epidemiologists) had academic qualifications and working experiences as presented in the charts of Figs. 1 and 2:

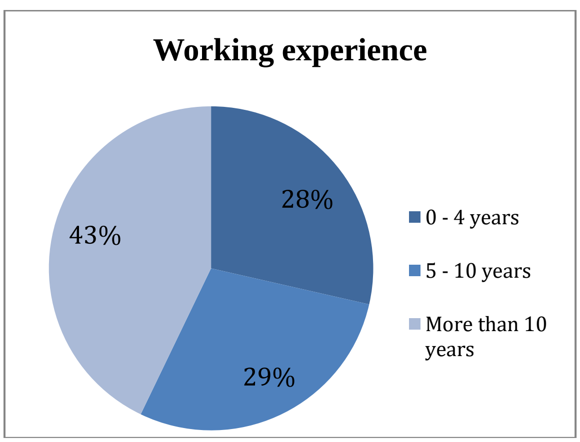

Figure 1: Respondents' working experience.

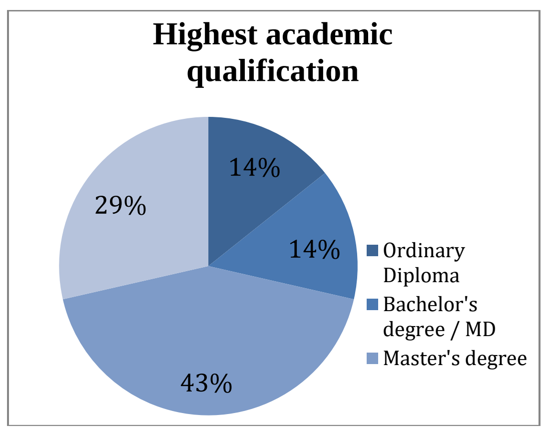

Figure 2: Respondents' highest academic qualification.

3.2 Needed Patients' Particulars for Enhanced Epidemiology

During interviews, respondents (with exception of the system administrators) were provided with a list of patients' particulars collected by the GoT-HoMIS when a patient is registered at the Hospital for the first time. The particulars included first name, middle name, surname, gender, date of birth, age, mobile number, tribe, residence, marital status, occupation, country, and next of kin (name, residence, relationship, phone number). The respondents were asked to select particulars out of the list and suggest any other particulars (demographic and clinical) that can be used to enhance epidemiology and epidemics surveillance. Out of the list, 100% (n = 7) of the respondents recommended the inclusion of gender, age, and residence as parameters in epidemics surveillance. Furthermore, 85.71% (n = 6) of the respondents voted for occupation, 57.14% (n = 4) for diagnosis, and 42.86% (n = 3) for both tribe and marital status. In parallel, 14.29% (n = 1) suggested religion despite it not being in the list. The ones in support of tribe and religion associated them with some cultural practices that may lead to transmission of diseases. Figure 3 is a visualization of the suggested attributes along with the number of respondents in support of each:

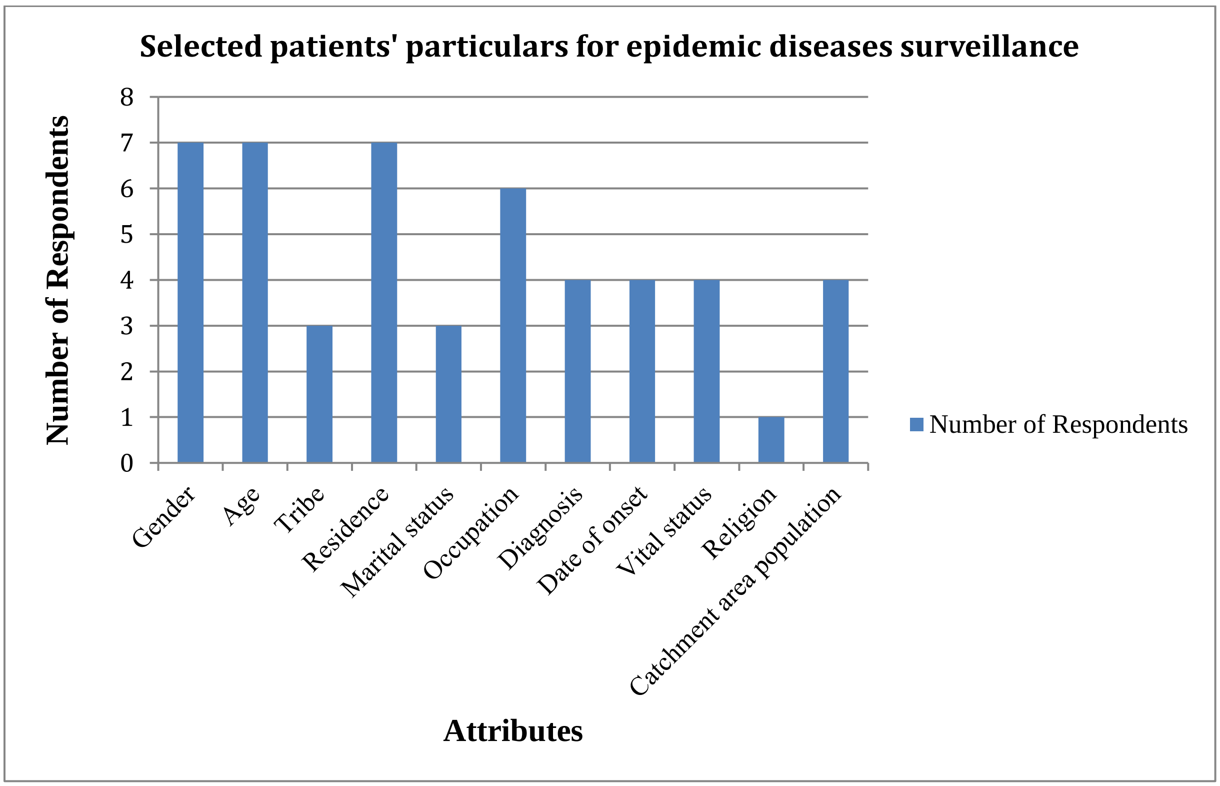

Figure 3: Selected particulars from the GoT-HoMIS for epidemics surveillance.

Catchment area population and vital status, though not included in the list provided to the respondents, were each proposed by four of the respondents (57.14%). They further explained that the two attributes would be useful to provide insights on the severity of the case and enhance informed decision making on mitigation strategies. The catchment area population and the vital status would give insights to the attack rate (AT) and case fatality rate (CFR), respectively. For example, reporting fifty (50) cases of a disease in a population of two hundred (200) people is quite different from fifty (50) cases in a population of five thousand (5000) people. Moreover, five cases of a disease being confirmed at a hospital and all five-people dying within a short time is different from five cases of a disease where all or a significant number of the affected are still alive past the first week. Vital status was also closely associated with date of onset, which also was proposed by 57.14% (n = 4) of the respondents.

3.3 The Current Reporting Setup at the MoHCDGEC

The respondents team at the MoHCDGEC were further interviewed on the reporting process and diseases surveillance. The target was to understand the current setup and the role of the existing systems and innovations in facilitating timeliness, completeness, and enhancement of the reports. The current model derived from interviews responses is as presented in Fig. 4:

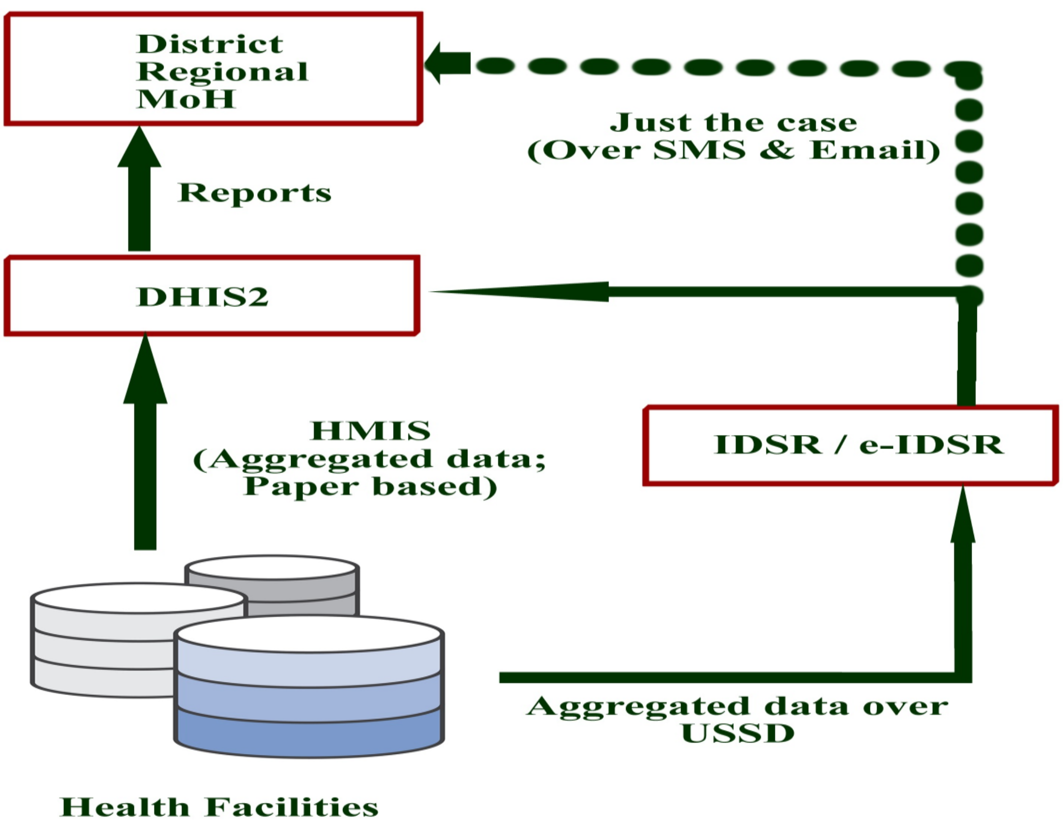

Figure 4: Current setup for disease cases reporting.

It was revealed in the interviews, as presented in Fig. 4, that the health facilities are the primary sources of information. These capture information about patients and their visits at the facilities. Different hospitals operate different Hospital Management Information Systems such as the GoT-HoMIS, Jeeva and Care2x; while, some are still paper based. Harmonization of the collected data is done through the Health Management Information System (HMIS/MTUHA), paper-based registers that contain aggregated reports on different aspects regarding clinical matters. There are a total of 16 registers, each having its own distinctive purpose and information. The patients' clinical data from the health facilities through HMIS reports are inserted in the District Health Information System (DHIS2). This operation is done on a monthly basis. Moreover, the data inserted in the DHIS2 is already in aggregated manner. The DHIS2 generates various reports that are circulated to the district, regional, and national health teams. Despite the quality of reports generated by the DHIS2, nature of the primary data entry system (from hard copy registers) and being on monthly basis results in challenges of timeliness, correctness, and completeness of reports. Some disease cases require immediate alert to epidemiologists and time-to-time follow-up, which is impossible with the DHIS2. In addition to the shortcomings, it was reported that there are often mismatches between the aggregated data presented by the DHIS2 and the data present in the health facilities.

The respondents further explained that in response to the need for timeliness of reporting and alerts for some diseases, especially outbreaks, the MoHCDGEC developed the Integrated Disease Surveillance and Response (IDSR) and its more enhanced innovation e-IDSR. These were developed to simultaneously feed data to the DHIS2 and alert district, regional, and national responsible personnel in case of an outbreak. The e-IDSR utilizes Unstructured Supplementary Service Data (USSD) to pass information to the DHIS2 as well as targeted epidemiologists. Selected people at the health facilities provide information to the system, and alerts are immediately forwarded to the epidemiologists in forms of emails and text messages. The alerts however, contain just the case (diagnosed disease), the name of the health facility where the diagnosis has been confirmed, and phone number of the contact person at the health facility for the sake of follow-up. Figure 5 shows a sample text message sent to epidemiologists.

Figure 5: e-IDSR framework [11].

The epidemiologists also pointed out that this has proven to be insufficient information to work with in terms of analysis and looking for patterns in the outbreak. On a few selected diseases (epidemic prone diseases), weekly follow-up is done by feeding data into the IDSR (in aggregated manner), which consequently feeds them into the DHIS2. Reports can then be generated from the DHIS2 grouping the aggregated number of cases gender and age wise as seen in the sample report format in Fig. 6.

FORM 3 C: WEEKLY REPORTED NEW CASES / DEATHS DURING AN EPIDEMIC AT REGION LEVEL

Region: Week beginning: Week ending:

S/N	DISEASES	< 5				> 5			
		C		D		C		D	
		M	F	M	F	M	F	M	F
1	AFP								
2	Anthrax								
3	Blood Diarrhea								
4	Cholera								
5	CSM								
6	Human Influenza/SARI								
7	Keratoconjuctivitis								
8	Measles								

Figure 6: Template for an aggregated report generated by the DHIS2 [11].

One of the challenges facing e-IDSR usage is the USSD not being efficient in times of outbreaks and in large hospitals where there is a large number of disease cases. USSD are also subject to timeouts when there is delay in the feeding process. Moreover, it was documented that the MoHCDGEC must pay a monthly fee for the aggregator that collects and aggregates data from the USSDs and feeds them to the DHIS2 [11]. The e-IDSR is currently operational in 15 regions in the URT. Furthermore, three (3) respondents pointed out that data and reports generated from the health facilities only reflect a fraction of health issues and disease cases present in the country. Disease cases in the community may not be captured unless the sufferer reports to a health facility, which is not always the case. The respondents further requested for a way that community data (disease cases) can be captured and integrated in the reports for better understanding and planning of diseases mitigation strategies. Moreover, the epidemiologists mentioned rumors as one of the main sources of their information of outbreaks. Due to the advancement of technology, internet connectivity, and smartphones ownership, most rumors are aired on social media platforms. It was further stated that a significant number of the received rumors prove to be valid, and they reach the epidemiologists through their connections quicker that reports from hospitals.

3.4 Proposed Data Integration Framework

The findings of the survey pointed out a need for a data integration framework that would guarantee inclusion of data from various sources (identified patients' particulars from the GoT-HoMIS and related data from other HoMIS in the country, social media, meteorological data). A framework has been envisioned to guide the integration of data from the diversified sources, joint analytics, and timely as well as detailed presentation of reports to enhance epidemic diseases surveillance and informed decision-making. The framework targets prompt analytics and reporting to support the Public Health Emergency Control Centers (PHOEC) as emphasized by the World Health Organization [12]. Figure 7 illustrates the proposed Data Integration Framework.

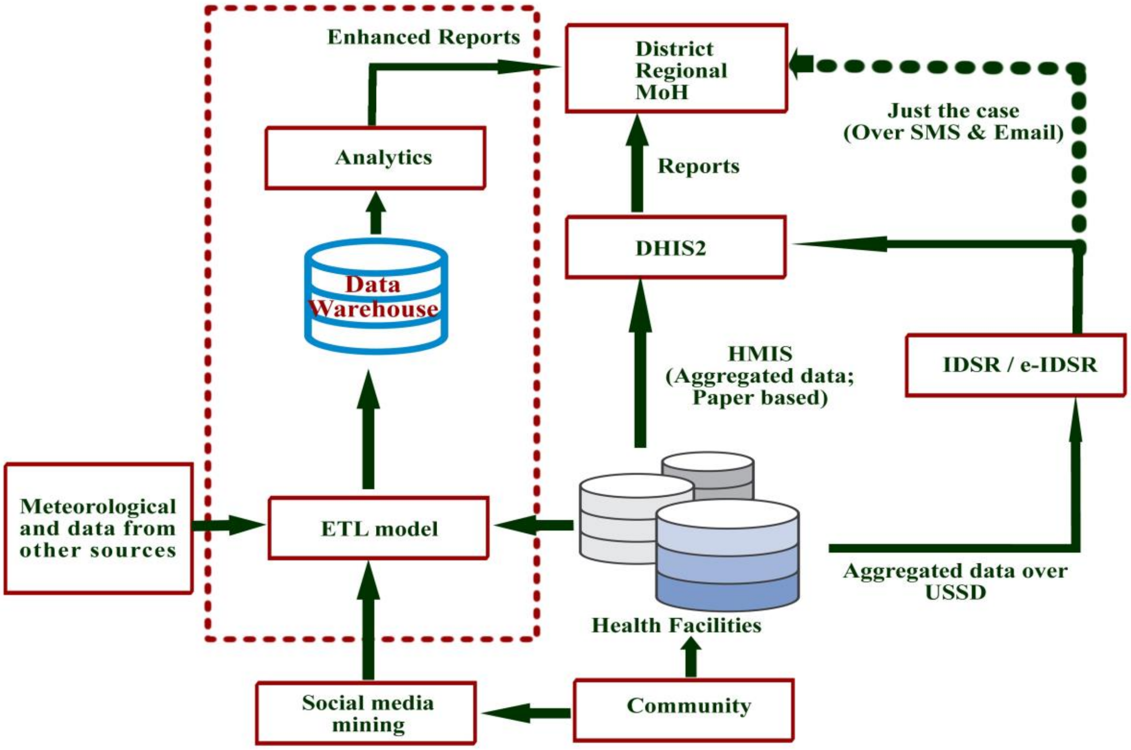

Figure 7: Proposed Data Integration Framework.

The data integration framework complements the existing reporting setup illustrated in Fig. 4. The extraction, transformation, and loading (ETL) model is targeted to fetch data from the HoMISs in the health facilities (GoT-HoMIS, Jeeva, Care2x), transform them accordingly and load them into a data warehouse for joint analytics. To enhance analytics and epidemic diseases surveillance, the ETL model will also extract and transform data from social media and load it to the data warehouse. To some extent, inclusion of means to jointly analyze rumors and epidemic diseases related posts from social media will mitigate the problem of missing analysis of community data for sufferers who do not visit health facilities.

Prompt alerts to the epidemiological teams based on timely extraction of rumors from social media will also increase effectiveness in decision making and controlling the case(s). Addition of meteorological data for integrated analysis will enable observation of patterns of disease cases in relation to the weather or climate changes.

All the aforementioned joint analytics are missing in the existing setup at the MoHCDGEC. The diversification of data sources will increase the possibility of patterns observation and extensive analytics, and consequently enhance epidemic diseases surveillance. On the resulting dataset, superior to the existing, data mining and machine learning algorithms can be applied to produce comprehensive reports and prediction of outbreaks for effective control of the epidemics. The framework does not obliterate the existing setup and innovations; the e-IDSR should still send alerts to the epidemiological teams at the district, regional, and national levels. The alerts, however, should be linked to the associated comprehensive reports that are generated from analysis on the integrated dataset.

4. Discussion

Two objectives were targeted in this research paper. The first objective was to explore the suitability of the GoT-HoMIS for epidemic diseases surveillance through the use of stored patients' particulars; and the second objective was to evaluate the current setup for epidemic diseases surveillance at the MoHCDGEC in Tanzania. From the first objective, the particulars essential for the epidemic diseases surveillance were found to be the patients' gender, age, tribe, residence, marital status, occupation, diagnosis, date of onset, vital status, religion, and catchment area population. It was established that 81.82% of the identified particulars are currently being captured and stored in the GoT-HoMIS. The exceptions are religion and catchment area population. From the identified importance of the patients' religion with ties to disease-prone practices, just as culture, the GoT-HoMIS developers should be instructed to add the field in the patients' registration form. The catchment area population can be fetched from the national census data and updated accordingly. The GoT-HoMIS can therefore be judged as a potential data source for necessary analytics in support of epidemic diseases surveillance and control.

The second objective was also successfully achieved as the current reporting setup at the MoHCDGEC has been analysed. From the health facilities, the HMIS has been identified as the bedrock for all disease cases analysis and is the primary data source to the DHIS2. While it has national-wide coverage and varieties of aggregated facility-level data, it is error-prone and still struggles with timeliness and completeness. Moreover, its frequency is not efficient for the targeted level of analysis for real-time surveillance. The MoHCDGEC's innovation for the adoption of the e-IDSR is commendable. It has successfully solved the gap of timely alerts to response teams at district, regional and national levels in case of outbreaks. Through the utilization of USSD technology it also manages to serve in hospitals with no electronic HoMIS or with poor Information and Communication Technology (ICT) infrastructure. However, reporting just the case as shown in Fig. 5 has proved ineffective to the epidemiologists. More parameters as identified in the Fig. 3 and enhanced analytics should accompany the alerts to facilitate the surveillance and informed decision-making. Furthermore, the aggregation of data before being inserted into the DHIS2 makes it less effective for epidemic diseases surveillance. Only limited analytics can be extracted from the hosted data.

The course of integrating community data in health-related analytics as recommended by some respondents is genuine and paramount. Reporting trends of a disease case based solely on the number of people who visit health facilities is partial; there are other cases in the community affecting people who cannot or will not visit health facilities. Social media can be viewed as a potential data source for the community data, whilst working on further innovations. Fortunately, the number of Tanzanian's accessing the Internet through their mobile phones from the year 2012 has increased more than threefold, from about six (6) million people in 2012 to more than nineteen (19) million in December 2017 [13]. This increase has a ripple effect on social media subscription and usage. The posts on social media can be mined to obtain epidemic diseases related posts as well as those of epidemic-prone environments [14]. A good example of a case where the spread of an epidemic was widely posted on social media is the case of Ebola, in which countless posts were being shared on social media [15, 16]. Once the posts that contain targeted phrases or mere mention of epidemics have been extracted, they can be forwarded to epidemiologists timely, and be utilized for epidemic diseases surveillance. If well designed, social media mining can be a great contribution to epidemic diseases surveillance as well as other health related issues [15, 17, 18]. Moreover, people may post about health endangering issues such as spill of chemicals or leakage of sewers, and these can be captured and shared to epidemiologists for action. The reports from the extracted information can be analyzed along with those from hospital records and enhanced reports made available to epidemiologists.

A Data Integration Framework with an enhanced data warehouse as the nucleus for integrated analytics to support epidemic diseases surveillance is proposed (Fig. 7). Data needs to be captured from all the identified sources, that is HoMISs, social media and additional ones such as meteorological data as well as catchment area population (see Fig. 7). The data warehouse should be designed to host all the data and formulate a unified dataset that can be used for analytics and epidemics surveillance. The data loaded in the proposed warehouse should be transformed (on individual patient-wise) to allow extreme data mining operations. Aggregation will be done within the data warehouse as part of analysis should the need arise, and not before. Extraction of raw data from the sources will also reduce the mismatch in reports, contrary to what is currently experienced by the DHIS2 from the reports present on the health facilities. For the hospitals running electronic HoMIS, a new module needs to be introduced for extracting the identified parameters in Fig. 3, and loading them to the staging area, where they will be transformed and loaded into the proposed data warehouse. In case of an outbreak, the epidemiologists can still receive a notification as they do now from the e-IDSR, but through the proposed framework they can find timely system-generated analysis and determined patterns among new and historic cases.

The elements of quality data can be argued to include availability, cleanliness, completeness, correctness, and timeliness of the data. From observation done on patients' registration process, the data entrants in the systems lack knowledge on the final applications of the data they feed. Consequently, every now and then a patient is registered with incomplete particulars such as occupation and marital status, which have been identified to be crucial in the monitoring of spread of epidemic diseases. The findings support the recommendation by WHO [12], and Kimaro and Twaakyondo [19] for the need to educate all the health workers who feed patients' demographic and clinical data in the HoMISs on the essence of quality data for analysis and reports generation.

5. Conclusion and Recommendation

The study has identified the parameters to be extracted from Hospital Management Information Systems needed for epidemic diseases surveillance. The GoT-HoMIS has the potential to realize the targeted level of surveillance as it records 81.82% of the needed parameters, has large market share, and is government-owned. While existing systems at the MoHCDGEC for epidemic diseases surveillance have prospective strengths, several shortcomings have been identified to include insufficient analytics to support timely epidemic diseases surveillance, limited data mining, poor quality of data fed in the system, and missing mechanisms to capture community data. To counter the shortcomings, this paper has proposed a data integration framework to guide research and development of interventions to enable applications of machine learning and data mining in epidemic diseases surveillance with a case study of the United Republic of Tanzania's health sector. The paper also recommends educating the personnel involved in feeding data in the systems on the essence of quality data and the applications of data for analysis, as this may help improve the quality of data in the systems.

6. Conflict of Interest

The authors declare no conflict of interest.

7. Acknowledgement

Sincere appreciation is extended to the Project 2 (P2) of the VLIR-UOS GRE@T programme at Mzumbe University for funding this research. Furthermore, we wish to thank all the respondents that participated in the study together with the institutions with which they are affiliated. Lastly, gratitude is registered to the Nelson Mandela African Institution of Science and Technology for the support throughout the work.

8. References

1. Karuri J, Waiganjo P, Daniel OR, Manya A. DHIS2: the tool to improve health data demand and use in Kenya. *Journal of Health Informatics in Developing Countrie*s. 2014 Mar 18;8(1).
2. President's Office, Regional Administration and Local Government (PO-RALG). Government of Tanzania – Hospital Management Information System (GoT-HoMIS) [PowerPoint Slides]. 2017 [Cited 6th May 2018]. Available from: https://www.healthdatacollaborative.org/fileadmin/uploads/hdc/Documents/Country_documents/Tanzania_GOT-HOMIS_presentation_12Sept2017.pdf
3. Nyasulu P, Kasubi M, Boniface R, Murray J. Understanding Laboratory Methods and Their Impact on Antimicrobial Resistance Surveillance, at Muhimbili National Hospital, Dar es Salaam, Tanzania. *Advances in Microbiology*. 2014; 4:33-8.
4. Wambura W, Machuve D, Nykänen P. Development of Discharge Letter Module onto a Hospital Information System. *Journal of Health Informatics in Developing Countries*. 2017 Dec 10;11(2).
5. Sheikh YH, Nyella E. Exploring the Power of Technology in the Institutionalization of Health Information Systems: An Actor-Network Analysis of Information Systems Integration. *Journal of Health Informatics in Africa*. 2018 Jan 21;4(2).
6. Parvathi I, Rautaray S. Survey on data mining techniques for the diagnosis of diseases in medical domain. *International Journal of Computer Science and Information Technologies*. 2014;5(1):838-46.
7. Molloy GJ, Stamatakis E, Randall G, Hamer M. Marital status, gender and cardiovascular mortality: behavioural, psychological distress and metabolic explanations. *Social science & medicine*. 2009 Jul 1;69(2):223-8.
8. National Bureau of Statistics (NBS), Office of the Chief Government Statistician (OCGS, Zanzibar), Ministry of Health, Community, Development, Gender, Elderly and Children (MoHCDGEC), Ministry of Health (MoH) Zanzibar. Tanzania Malaria Indicator Survey (TMIS) Key Indicators 2017. 2018 [cited 5th May 2018]. Available from: https://nbs.go.tz/nbs/takwimu/TMIS2017/TMIS2017_KeyIndicatorEng.pdf
9. Darcy NM, Somi G, Matee M, Wengaa D, Perera S. Analysis of Data Dissemination and Use Practices in the Health Sec-tor in Tanzania: Results of desk review and interviews with key stakeholders. *Journal of Health Informatics in Africa*. 2017 Nov 5;4(1).
10. Health Metrics Network (HMN). Framework and Standards for Country Health Information Systems. Geneva, World Health Organisation 2008.
11. Ministry of Health, Community Development, Gender, Elderly and Children (MoHCDGEC). Tanzania e-IDSR System [PowerPoint Slides]. 2018.

12. World Health Organization (WHO). WHO Tanzania Annual Report 2016. 2016 [cited 8th May, 2018]. Available from: http://www.afro.who.int/publications/who-country-office-annual-report-2016
13. Tanzania Communications Regulatory Authority (TCRA). Quarterly Communications Statistics Report October- December 2017 Operator Returns. 2018 [cited 5th May, 2018]. Available from: https://www.tcra.go.tz/images/documents/telecommunication/TelCom_Statistics_Dec_2017.pdf
14. Corley CD, Cook DJ, Mikler AR, Singh KP. Text and structural data mining of influenza mentions in web and social media. *International journal of environmental research and public health*. 2010 Feb 22;7(2):596-615.
15. Fernández-Luque L, Bau T. Health and social media: perfect storm of information. *Healthcare informatics research.* 2015 Apr 1;21(2):67-73.
16. Househ M. Communicating Ebola through social media and electronic news media outlets: A cross-sectional study. *Health informatics journal*. 2016 Sep;22(3):470-8.
17. Nikfarjam A, Sarker A, O'connor K, Ginn R, Gonzalez G. Pharmacovigilance from social media: mining adverse drug reaction mentions using sequence labeling with word embedding cluster features. Journal of the *American Medical Informatics Association.* 2015 Mar 9;22(3):671-81.
18. Paul MJ, Sarker A, Brownstein JS, Nikfarjam A, Scotch M, Smith KL, Gonzalez G. Social media mining for public health monitoring and surveillance. *InBiocomputing* 2016: Proceedings of the Pacific Symposium 2016 (pp. 468-479).
19. Kimaro HC, Twaakyondo HM. Analysing the hindrance to the use of information and technology for improving efficiency of health care delivery system in Tanzania. Tanzania *Journal of Health Research.* 2005;7(3):189-97.