

Developing cloud-based digital data management mechanism for health-related data

Sandhya S. Khadse, Vrushali V. Kulkarni*, Sundaram Kartikeyan, Vishal V. Kulkarni and Bharat Gadegaokar

PSM Department, Rajiv Gandhi Medical College, Kalwa, Thane (W), India

Abstract

Objective: To build an easy-to-use, secure and efficient data collection platform to compile health-related data.

Materials and Methods: An android device with Google App Engine and a digital data collection form was used to collect the data of health camp beneficiaries.

Results: A total of 4747 participants were screened in 10 health camps in this study. Using 10 android devices at a time, a new mechanism was developed to collect systematic data in a short period of time. By combining reliable android device, mobile data connection, a free open-source data collection frame, a scalable cloud-based server and with the team of android device operators, a low-cost, greatly reliable platform was developed, which provided real-time data collection with analysis.

Conclusion: A reliable Android device was combined with a free open-source data collection framework, a powerful device security tool, a scalable cloud-based server, to build a low-cost, highly reliable, maintenance-free, user-friendly data collection platform that can deliver well-formatted results to study organizers immediately on completion of data collection. This study successfully demonstrated the value of a user-friendly digital data collection platform for collecting health-related data.

Keywords: Digital data collection, CBS mechanism, digital data, Google Sheets.

*Correspondence Info:

Dr. Vrushali Vishal Kulakrni,
Assistant Professor,
PSM Department,
Rajiv Gandhi Medical College, Kalwa,
Thane (W), India

*Article History:

Received: 03/06/2019

Revised: 25/07/2019

Accepted: 29/07/2019

DOI: <https://doi.org/10.7439/ijbr.v10i7.5198>

QR Code



How to cite: Khadse S, Kulkarni V, Kartikeyan S, Kulkarni V, Gadegaokar B. Developing cloud-based digital data management mechanism for health-related data. *International Journal of Biomedical Research* 2019; 10(07): e5198. DOI: 10.7439/ijbr.v10i7.5198 Available from: <https://ssjournals.com/index.php/ijbr/article/view/5198>

Copyright (c) 2019 International Journal of Biomedical Research. This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/)

1. Introduction

Major advances have been made in the last decade in low-cost, real-time technologies to assess disease, behaviour and social interactions. These advances are due to increased computational superiority and digital technologies expertise [1].

To receive and transmit electronic data, a medium known as “digital health” is used. This technology facilitates improvement of individual health data, coordinates health care services and offers new opportunities to improve people’s health [2]. Cloud computing is a self-service web-based infrastructure that enables the user to access computing resources anytime from anywhere [3, 4]. A strong system of data collection, consolidation and compilation techniques enables achievement of these objectives. In order to keep up with

the enhancements in this field, the data processing mechanism required a relative level of rebuild.

Traditional paper-based surveys, the most common method for monitoring and evaluation, are relatively expensive, time consuming and largely liable to human error. The research life cycle (research design, data-collection, data validation and statistical analysis) may extend over months or years [5]. Many researchers opt for hard copy registers or Excel-based systems to process digital data. These traditional methods are imperilled by numerous risks, such as, prolonged time, requirement of excess dedicated man power, lack of data security, duplicate entries, selection of wrong attributes, and incorrect data.

Data collection with digital technologies is increasingly used in settings with resource limitations [5].

To get evidence-based results, organizations need systematically informed and improved data collection. Research form, an integral part of the toolkit of research work, is applicable in many intervention situations [6]. This research paper delves into the transformation of a data management system from an Excel-based system to an advanced cloud-based scalable and secure mechanism (CBS mechanism).

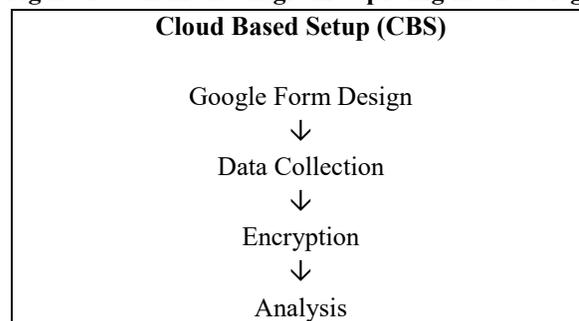
2. Material and methods

Multiple options were explored to overcome the barriers in data management system. Options were like to buy new software from market, to deploy an Enterprise Resourcing Platform (ERP) system, to opt for latest technology or to continue with the traditional method of data management. With this bunch of options, authors have studied advantages and disadvantages of various methods for collection and management of data.

To develop a secure, efficient and user-friendly data collection tool, author developed Cloud Based Secure (CBS) mechanism using open source data tool named Google Sheets (Google LLC, Menlo Park, CA, USA). It allows data encryption and security requirements right at the server level itself. CBS mechanism was evaluated for digital data collection but opting for dedicated network setup and secured storage space entailed high financial costs. Other globally available options were reviewed for free provision of cloud environment to build the setup and

Google became the first choice owing to several reasons that included: free 15 GB storage space per user (additional storage is available at reasonable cost ~2000 Indian Rupees per year for 100 GB), global accessibility with maximum data security, data validation, data accuracy, data availability and scalability. The schematic demonstration of the flow of data through cloud-based setup is depicted in Figure 1.

Figure 1: Schematic diagram depicting methodology

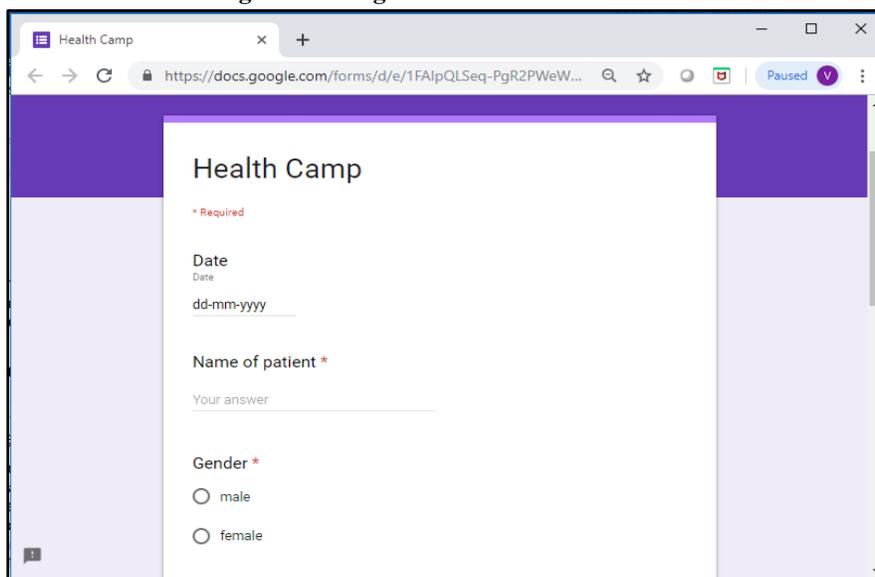


A Google Form connected to a Google spreadsheet helped to address the requirement to the core level. The form allowed liberty to make the data entry user friendly. The customizations associated to the fields enabled maintaining the required format for various attributes.

2.1 Data collection form:

With the help of Google account, Google form was used to define the questionnaire while collecting the data of health camp. The authors created online Google form to collect the data of camp beneficiaries. (Figure-2)

Figure 2: Google form for data collection



As per the requirement of health camp, the requisite variables were defined for data collection. Based on syntactically modified form of World Wide Web of Google form, the form allowed building questions with a variety of answer formats to be displayed, such as, multiple choices, checkbox, text entry, any date-time questions and many more. Some questions were of multiple-choice

question (MCQ) type where options were made available to choose from whereas few questions, such as, postal address of patients and diagnosis of patient’s problem were retained as open text, which allowed descriptive answers. Also, if answer for a particular question was expected in a particular form, such as, number entry or a specific notation, instructions were displayed after substituting the answer to

get it checked by the user. This form enabled read, display and logical rules applicability to control progression. CBS mechanism allowed the definition of custom constraints for various questions, ensuring that all required sections were completed with answers within acceptable ranges. Only after filling-up the required answers for mandatory questions, the user could submit this data on the Web.

2.2 Development

CBS mechanism in terms of Google form can function on any android device. Hence, mobile phones,

computers and tablets were used while compiling the data. The platform software of this cloud-based setup was free-of-cost, with no dependency on other software. It was configured using the ready-to-deploy server software in conjunction with Google App Engine. This App Engine offers unlimited scalability and does not require server maintenance. On completion, forms automatically get transmitted to a secure data worksheet. These worksheets further can be exported to an Excel or MS Access format when required.

Figure 3: Data compilation

Timestamp	Date	Name of patient	Gender	Age	Disease type
11/13/2018 22:45:26	11/14/2018	Test	female		13 skin
11/13/2018 22:45:32	10/18/2018	Test1	male		25 skin
11/13/2018 22:46:00	10/20/2018	Test2	female		26 Paediatrics
11/13/2018 22:47:01	11/8/2018	Test3	male		35 General
11/13/2018 22:47:18	11/14/2018	Vktesting	female		13 Medicine
11/13/2018 22:47:51	11/7/2018	Test4	male		45 Ophthalmology
11/14/2018 13:02:38	7/28/2018	Sunanda Adekar	female		68 Ophthalmology
11/14/2018 13:02:51	7/28/2018	Babaji Suryaji Naik	male		62 Ophthalmology
11/14/2018 13:03:02	7/29/2018	Yashwant pawar	male		70 Medicine
11/14/2018 13:03:10	7/29/2018	Deepak Sawant	male		49 Ophthalmology
11/14/2018 13:03:12	7/28/2018	Maruti Amame	male		56 Dental
11/14/2018 13:03:17	7/28/2018	Shalini Pitale	female		70 Ophthalmology
11/14/2018 13:03:26	7/28/2018	Rahiben Ravai	female		40 Ophthalmology
11/14/2018 13:03:49	7/28/2018	Suryakant Pathe	male		35 Ophthalmology
11/14/2018 13:03:50	7/28/2018	Balaji Naik	male		69 Dental
11/14/2018 13:03:56	7/28/2018	Sachin Vaidya	male		42 Ophthalmology
11/14/2018 13:03:56	7/28/2018	Arun Patil	male		50 Urinary
11/14/2018 13:04:16	7/28/2018	Sachin Vaidya	male		42 Dental
11/14/2018 13:04:25	7/28/2018	Dattatray Dhotre	male		53 Ophthalmology
11/14/2018 13:04:26	7/28/2018	Ulhas Salvi	male		58 Ophthalmology
11/14/2018 13:04:27	7/29/2018	Chandrakant Gode	male		64 Ophthalmology
11/14/2018 13:04:43	7/28/2018	Kanchan Shirkar	female		72 Dental
11/14/2018 13:04:46	7/29/2018	Sashikant karde	male		38 Ortho
11/14/2018 13:05:06	7/28/2018	Walter Pais	male		62 Ophthalmology
11/14/2018 13:05:09	7/28/2018	Jayashree Jadhav	female		60 Ophthalmology
11/14/2018 13:05:09	7/28/2018	Bhimrao Jadhav	male		62 Dental
11/14/2018 13:05:10	7/29/2018	Ashok Shigvan	male		53 Ophthalmology
11/14/2018 13:05:37	7/28/2018	Reshma Dalvi	female		52 Dental

2.3 Implementation

With the traditional method of data entry using MS-Excel (Microsoft Corporation, Redmond, WA, USA) data of 1947 participants of six health camps (conducted between April 2018 and June 2018) were entered by a single operator on single device. The data included patients' names, age, gender, mobile phone number, address, their illness and their respective diagnoses.

Due to zero cost, fast speed, secured single database, globally accessible cloud-based storage technique, Google forms were preferred. The associated database was provided by Google Sheets making the data analysis and reporting simpler. The setup was fast and simple. The tests conducted while the form was being built enabled customizing the Form for optimum user experience. A simultaneous trial-run from multiple devices helped demonstrate the ease of data entry and data compilation when compared with traditional system. Flexibility in device usages for data entry in Google form was seen here. Hence dependency on particular machine or server is less. Single database has been generated on cloud/web. It can be accessed from anywhere and storage

devices are not required here. On the cloud-based server, submission of each entry was individually secure, each using a unique key. However, data remained encrypted until manually downloaded and decrypted which is known only to the study organizers as server access was restricted to study organizers. The new system, thus deployed, had its own advantages over and above its default features like internet connectivity and device capability, etc.

In proposed method of Google Sheets, Google Form was generated in the English language for registration. Medical students, volunteers and clerical staff participated as data collectors. An online training session of ten-minute duration was conducted for data collectors, wherein, information was provided about the data collection system and online video on the digital data collection system was shown. With the help of hotspot technology, a common internet data connection and hyperlink of Google form was provided to all data collectors. The information of 2800 patients of various seven health camps (conducted between July 2018 and December 2018) were filled by using the hyperlink of Google form in a very short time. The information was regarding patients name, age, gender,

mobile number, address, their illness and respective diagnosis. At a time, multiple users were substituted this information. Database was generated for 2800 beneficiaries in which, data were uploaded in different time slots by multiple users.

3. Observations and results

Two different methods were deployed to enter the data of camps. Using CBS mechanism, a greater number of data entries could be generated in lesser time. (Table-1)

Table 1: Number of camp data entries with respect to time and methods

Camp Period	Number of camps	Digital Data Entry Method	Duration for data entry	Number of data sets
April to June 2018	06	Traditional (Excel-based)	2 months	1947
July to Dec 2018	07	Latest (CBS Mechanism)	15-20 hours #	2800

Including real-time reporting

4. Discussion

Latest technology of CBS mechanism has been adopted in this study. Since the health personnel had sufficient clinical knowledge and experience, combining the digital data collection tool with health-related data was a successful approach [7]. In the present study, authors have used CBS mechanism and health camp beneficiaries data which gave faster and real time data collection. A digital platform has been developed to gauge the prevalence of sepsis in Wales [8]. In another study [9], a cloud-based system was deployed for monitoring blood pressure of individuals. Open Data Kit (ODK) has been found to be resilient and dependable in developing countries [10, 11]. In Kenya, ODK was found to be a cost-effective data collection method during household visits and it has been used to develop a clinical decision support system [12]. Its use in South Africa resulted in enhanced data quality, reliability, and ease of data export for analysis [13]. In the present study, CBS mechanism was found to be a trustworthy, independent and cost-free tool to collect and compile health data. Features of cloud-based setups include the ability to deploy own setup, server, data collection, analysis and exert control over encryption keys without third-party involvement and thus it offers substantial advantage over other commercially available frameworks. The transition from paper-based surveys to digital data collection using handheld devices has been discussed in another research paper [5]. In our study, migration of data management system from traditional Excel-based system to an advanced CBS mechanism has been successful. This data collection platform can benefit other health care researchers and medical students in streamlining their data collection projects.

5. Conclusion

This study successfully demonstrated the value of a digital data collection platform for collecting data in community-based health camps. We have effectively combined a reliable Android device, a free open-source data collection framework, a powerful device security tool, a scalable cloud-based server, to build a low-cost, highly reliable data collection platform that requires little training or maintenance and can deliver well-formatted results to

study organizers immediately on completion of data collection.

References

- [1]. Christian J, Dasgupta N, Jordan M, Juneja M, Nilsen W, Reites J. 21st Century Patient Registries. EBook addendum to Registries for Evaluating Patient Outcomes: A User's Guide. Chapter 3. AHRQ Publication No. 17(18)-EHC013-EF. Bethesda, MD, USA: National Center for Biotechnology Information; 3rd Edition. 2018: 41-63.
- [2]. WHO releases first guideline on digital health interventions, April 2019. Available from: <https://www.who.int/news-room/detail/17-04-2019-who-releases-first-guideline-on-digital-health-interventions>. Accessed on 30th May 2019.
- [3]. Mell P, Grance T. The NIST Definition of Cloud Computing. *Communications of the ACM* 2011; 53(6): 50.
- [4]. Vaquero LM, Rodero-Merino L, Caceres J, Lindner M. A Break in the Clouds: Towards a Cloud Definition. *ACM SIGCOMM Computer Communication Review* 2009; 39(1):50-55.
- [5]. Fitzgerald G, Fitz Gibbon M. A Comparative Analysis of Traditional and Digital Data Collection Methods in Social Research in LDCs - Case Studies Exploring Implications for participation, empowerment, and (mis)understandings. Preprints of the 19th World Congress. The International Federation of Automatic Control, Cape Town, South Africa. 2014: 24-29.
- [6]. Laws S, Harper C, Marcus R. Research for Development: A Practical Guide. London: Sage; 2003.
- [7]. Szakmany T, Ellis G, Lundin RM, Pignatelli I, Sharif B, Joshi S, et al. Sepsis in Wales on the general wards: results of a feasibility pilot. *Br J Anaesth* 2015; 114(6):1000-1001.
- [8]. Sharif B, Lundin RM, Morgan P, Hall JE, Dhadda A, Mann C, et al. Developing a digital data collection platform to measure the prevalence of sepsis in Wales. *J Am Med Inform Assoc* 2016; 23(6):1185-1189.
- [9]. Kuo M-H. Implementation of a Cloud-based Blood Pressure Data Management System. In: Cornet R,

- Stoicu-Tivadar L, Hörbst A, Parra Calderón CL, Andersen SK, Hercigonja-Szekeres M. (Eds). *Digital Healthcare Empowering Europeans*. Amsterdam: IOS Press; 2015: 882-886.
- [10]. Brunette W, Sundt M, Dell N, Chaudhri R, Breit N, Borriello G. Open Data Kit 2.0: expanding and refining information services for developing regions. Proceedings of the 14th Workshop on Mobile Computing Systems and Applications, Jekyll Island, GA, USA. February 2013. <http://dx.doi.org/10.1145/2444776.2444790> Accessed on 30th May 2019.
- [11]. Raja A, Tridane A, Gaffar A, Lindquist T, Pribadi K. Android and ODK based data collection framework to aid in epidemiological analysis. *Online J Public Health Inform* 2014; 5(3):228.
- [12]. Anokwa Y, Ribeka N, Parikh T, Borriello G, Were MC. Design of a phone-based clinical decision support system for resource-limited settings. Proceedings of the 5th International Conference on Information and Communication Technologies and Development. Atlanta, GA, USA. March 2012. <http://dx.doi.org/10.1145/2160673.2160676> Accessed on 30th May 2019.
- [13]. Kabuya C, Wright G, Odama A, O'Mahony D. Routine data for disease surveillance in the undeveloped region of the OR Tambo district of the Eastern Cape Province. *Stud Health Technol Inform* 2014; 197:103-107.