

Prospectus

Health Data Infrastructure Development in Low and Middle Income Countries
(Technical Topic)

An Evaluation of the Socio-Political Interactions of Health Data Infrastructure to Inform Ethical
Data Governance
(STS Topic)

By

Luke McPhillips

10/30/19

Technical Project Team Members: Angela Yi, Bhavana Channavajjala, Nathan Ohene, Rex
Focht, Sarah-Winston Nathan and Victoria Rho

On my honor as a University student, I have neither given nor received unauthorized aid on this
assignment as defined by the Honor Guidelines for Thesis-Related Assignments.

Signed: *Luke McPhillips*

Approved: _____ Date _____

Rider Foley, Department of Engineering and Society

Approved: _____ Date _____

Donald Brown, Department of Engineering Systems and the Environment

Introduction

Examining big data security and privacy issues with the collection of health data, Patil and Seshadri (2014) cite a study from the Ponemon Institute, an organization conducting independent research on data privacy, that concludes 94% of hospitals have one or more security breaches in the past year. As the need for patient medical data to compete in the economic marketplace of healthcare provision increases, there is urgency for the implementation of robust structures of data governance. Governments and healthcare systems around the world are increasingly seeking to explore public policy options to protect patient privacy but support the usage of data for improved care. A clear example of this in the United States is the Health Insurance Portability and Accountability Act.

Data-driven healthcare solutions are not a trend just for the high-income world. Stakeholders in low- and middle-income countries (LMIC) also seek to build systems for collecting, storing, and analyzing data for improved patient outcomes. But the difficulties are steeper, Tuti et al. (2015) provides a non-exhaustive list of challenges faced in clinical data management in Kenya:

1. Resource context: There is shortage of staffing, limited funds, unreliable power sources.
2. IT infrastructure: Non-comprehensive electronic health record systems, poor internet connectivity, complex data storage and access procedures.
3. Pre-existing reporting structures: existing national health information system requirements on data capture and storage, designed to address stakeholder needs.
4. Data collection tools: Non-standardized collection protocols, usage of paper records.

(p. 185)

My capstone team is taking on a small portion of this challenge. Under our advisor, Doctor Donald Brown, we will develop an algorithm deploying image capture and optical character recognition techniques to scan perioperative (surgical data) flowsheets from medical centers in Rwanda. This data will be captured and transferred securely to the University of Virginia for storage and further research, with remote access provided to physicians in Rwanda. Our goal is a system to facilitate the growth of a stronger, independent data infrastructure built out within Rwanda. The client is a UVA anesthesiologist who travels through Doctors Without Borders each year to the University Teaching Hospital of Kigali, which will be the partner site for this project.

Complications arise with the existence and storage of private, personal health data. Wyber et al. (2015) cite work from a study completed at the London School of Economics, writing:

The risk of accidental or intentional breaches of data security may be particularly high in settings with high levels of illiteracy and corruption that are undergoing rapid technological transition. In many such settings, legislation supporting the privacy and security of information services is frequently underdeveloped and rarely enforced. (p. 205)

The authors call attention to the complexity of technological advancement, and the difficulty of maintaining a power equilibrium that shines light on the array of stakeholders that comprise data systems. I will argue that the construction of health data systems must recognize their inherent political nature and the human impact on evolution of their applications.

Technical Topic

The technical project can be broken into three segments: image classification, data storage and data protection and access. Surgical data at the University Teaching Hospital of

Kigali is collected in handwritten flowsheets. Our team's goal is to develop an end-to-end system for the scanning, upload, and analysis of these flowsheets. Similar data collection work has been done electronically in Kenya through REDCap, a proprietary Vanderbilt clinical software. Sileshi et al. (2017) implemented a simple electronic data collection system to digitize a 132-field form of surgical data and conduct statistical tests on surgical outcomes. This collection was done in real-time using tablets and did not include a means for digitizing data stored in older paper records. Our project will take a two-pronged approach, providing a means for the automated digitizing of both current and older surgical records.

Scanned flowsheet images require the construction of deep learning algorithms to extract data that can be stored in a structured manner. The images in Appendix A display the pre/perioperative flowsheet (Durieux, 2019). The record contains an array of information types that can be categorized into three analytical bins:

1. Graphical: Graphical data will require building an algorithm that recognizes the spatial relationships of points on the graph to assign values to locations. The data will be stored as a time series.
2. Structured handwritten text: Includes fields on the form with finite options for entries. Examples include types of surgical procedures, medications, dosages. Our team will utilize a classification algorithm to match handwritten words against a set library of potential terms.
3. Checkboxes: We will use optical mark recognition techniques analyzing light patterns to predict whether a box is empty or filled.

Our team is leveraging open source code to perform these tasks.

The data extracted from the medical record will be stored in a relational database at the University of Virginia for further academic research both in the United States and Rwanda, and our team will look to integrate our results with the data systems that do already exist in Rwanda. Verbeke et al. (2012) discuss this in their use of OpenClinic, an opensource software, at The University Teaching Hospital of Kigali. In a 2010 survey, they found that nearly 70% of patients could be identified by electronic record (p. 11). Our team will map the data stored at UVA through our project back to these records using patient identifiers. This will best support further analytics and extraction of information.

The final component of the project is the definition of database accessibility guidelines. We will address this towards the start of next semester, contingent on successful development and testing of our technical system. Our team will seek to answer the following questions:

- Who should have access to this data?
- What are potential use cases for how the data will be used?
- What legal requirements surround the international transfer of patient data?
- How do we anticipate the needs of this infrastructure to evolve?

These questions will be the foundation for our team's exploration of open source software's impact. While the example discussed at the beginning of this section utilized proprietary software (REDCAP), Kanter et al. (2012) discuss what social and technical advantages have led to the download of opensource medical data platforms in over 50 countries for efficient and ethical data storage:

Open source platforms were preferred over proprietary products...(They) require no license fees, can be tailored to fit the needs of the project, and have support from the development community contributing to and using the particular product...This includes

cost, flexibility, ownership, and the benefit of a large passionate, and involved developer and implementer community. (p. 193-194)

These benefits have been clear to our team, it is what has supported all our work so far and enabled us to build code we could not understand ourselves. The definition of accessibility guidelines will mark a shift in thinking, where our team will examine the system we build and try to identify risks that come with using open source software with patient medical data.

Our problem approach will result in a client deliverable that can ideally be used in a transitory manner as the University Teaching Hospital's health data architecture scales. We also aim for our solution to be based in principles that will support usage in further medical facilities in a wider range of LMIC. Allen et al. (2007) created OpenMRS, an opensource clinical data platform. Having seen success in implementation with over 10 million data entries uploaded about 48,000+ patients, the software has been adopted in Kenya, Rwanda, South Africa, and Lesotho with further expansion underway (p. 386). Our team is already in contact with doctoral students in South Africa who want to use our system, and view this as evidence that information technology systems can be built for common usage in differently structured societies.

STS Topic

Delving deeper into the data accessibility and protection component of my capstone project, I will leverage two frameworks towards analysis of the socio-technical development of data applications and governance in LMIC. Large-scale health data systems are inherently political, the risks and opportunities associated with the storage of private data naturally associate with centralized political systems that leverage strong institutional structures to protect individual liberties. Winner (1980) references Plato as evidence for his claim of inherent techno-politics:

Plato goes on to suggest that governing a state is rather like being captain of a ship or like practicing medicine as a physician. Much the same conditions that require central rule and decisive action in organized technical activity also create this need in government. (p. 129)

The reference towards practice of medicine holds incredible relevance, implying the political connotations of medical care provision have prevailed for 1500+ years. Medical environments continue to lend themselves to the same centralized power structure and hierarchical order of command for the support of decisive, safe actions. Tiffin, George and Lefevre (2019) draw on their personal experiences operating with health data in South Africa and India to support this, calling for the establishment of strong data governance procedures to mitigate potential power imbalances. Two pillars of data governance they define are ethics/informed consent and legal frameworks. Without regulation, data systems can instead be used as a tool to perpetuate the inequities that health systems aim to reverse through improved care. It is the opportunity for misuse that therefore makes data political, and inherently requires transparency and centralized management.

This framework for analysis is not complete without the complementary consideration of the rapidly evolving nature of health innovation. While I will argue the above point that technical health systems maintain inherent politics, the procedures, methods and basis of care is evolving according to the needs of diverse social groups. Consider the framework illustrated by Pinch and Bijker (1987, p. 119-120). The needs of social groups such as patients, providers, and other vested system interests are constantly being re-defined by expanding data capabilities.

Reciprocally, the functionalities of medical systems are simultaneously evolving in response to those needs. Our capstone team's client (Dr. Marcel Durieux) briefly discussed a program in Rwanda where blood is flown by drone to regional medical centers. In this example, the needs of

patients and rural medical providers are defining the construction of this technical system and in response the needs of those individuals and organizations are evolving to higher orders of complexity. Providers may soon need to use data to predict required blood types and quantities.

The capstone project itself supports similar interpretation. Hospitals in LMIC are seeking to use health data to provide more customized patient journeys and improve doctor training programs. As they develop their capabilities, the needs of patients evolve to now include data rights. The technology must now respond and this need is the genesis of our work. The necessity for using a two-pronged framework of analysis is driven by the increasingly complex profiles of stakeholder groups. In his article on techno-politics, Winner (1980) references Engels' examples of railroads and shipping as inherently political technologies:

Similar lessons are adduced in Engels's analysis of the necessary operating conditions for railways and ships at sea. Both require the subordination of workers to an "imperious authority" that sees to it that things run according to plan. (p. 129)

These traditional systems are either publicly owned by shareholders or privately controlled. Ownership of capital is clear, and there is a clear organizational structure from those who conduct business strategy to those who perform field work. In contrast, debate exists over who actually owns data and whether that is the right question to truly frame the conversation of societal power. Rosner (2014) argues:

However, there is a great deal of evidence for citizen concern with control of information about them. Perhaps, then, ownership is not the appropriate framework from which to address people's anxieties and the broader rights of data protection and privacy. Control regimes and rights are better perspectives from which to consider the state of personal data. (p. 627)

It is this notion of centralized control versus private ownership which requires the usage of two frameworks. In the technical dimension of this paper I discuss the privacy of HIV data, a clear example of how human identity is tied to personal health data. Drilling deeper into [what topic], I will argue that the play of this notion of self-ownership against the larger notion of “control” is what must drive a human-centric development of the societal framework of values within which the requisite political hierarchies and governance systems of data operate.

Research Question and Methods

How do the system requirements of health data collection, maintenance and analytics interact with institutional power and inequities of LMIC? Data is an irreplaceable component of the human future, but considering the barrage of data breaches in the past decade, it is clear that a culture of safe and ethical usage has not yet permeated any corner of the globe. Underprivileged populations are at most risk. Beck, Gill & Lay (2016) conducted focus groups across LMIC with medical field stakeholders and found that, “Of the 49 countries claiming to have developed (data) privacy laws, 55% reported that they had not developed any guidelines for the implementation of such laws.” The concerns raised by this finding are the basis for my research question and the examination of the nuances of data governance in LMIC, and how it is shaped by institutions, people, and data systems themselves.

I will use a case-study approach to provide a review of the cultural and political complexities that can inform the development of cohesive health data systems. Examining 4-5 LMIC across the globe (including Rwanda), I will leverage a multi-directional set of techniques to provide just support for the bilateral framework outlined above. Policy analysis will drive the evaluation of how data systems have altered the public space and pervaded political dialogue. I

will use the techniques outlined by Mutimukwe, Kolkowsa and Gronlund (2019) to examine the difference between existence and implementation, and how to use government literature as a lens for understanding the local politics of data.

An evaluation of health data systems in the selected countries will ground the conversation for how the needs of individuals, medical groups, international NGOs and the like are shaping the construction of data systems. The referencing of the drone system and my capstone project above are two examples. Binagwaho and Scott discuss crucial characteristics of health data application, and call attention to the intertwining of politics with socially defined health goals:

Each country should promote policies that foster synergies across sectors. The health sector cannot achieve health-related goals as an isolated, stand-alone system. For example, it needs the involvement of local leaders who govern by proximity and ensure the implementation of social policies... (p. 204)

This evaluation will include a standardized review of medical data usage in each country's health system, I will define a set of criteria against which to score each country's progress in improving its digital health systems. Sample criteria include patient data accessibility, private sector funding for public health initiatives and the construction of data-sharing network across medical facilities. The goal of this portion of the analysis is to understand the success stakeholders have had in advocating for their interests, it will give further insight into how health systems are transforming and what the actual results of stakeholder advocacy and government legislation are.

Conclusion

The culmination of this capstone project will be a data architecture for the automatic scanning of surgical flowsheets into a secure database, ideally to be based in Rwanda in future

iterations. Access will be provided back to doctors in Rwanda for performance feedback and further training. I will analyze the ethical construction of data governance procedures in LMIC, focusing on how techno-politics and social construction of technology frameworks can explain the interaction of data control versus ownership in light of the political inequities of LMIC. These deliverables could potentially provide collective direction towards the development of a technical solution to the data necessities of LMIC through methods of responsible implementation. The existence of data collection, storage, and analytics is a public good only if utilized in an ethical and intentional manner reflective of the value that healthcare is a human right. I will seek to conduct my analysis in the first two months of 2019, draft my thesis by spring break and finish editing by the beginning of April.

References

- Patil, H. K., & Seshadri, R. (2014). Big Data Security and Privacy Issues in Healthcare. 2014 IEEE International Congress on Big Data. doi: 10.1109/bigdata.congress.2014.112
- Tuti, T., Bitok, M., Paton, C., Makone, B., Malla, L., Muinga, N., ... English, M. (2015). Innovating to enhance clinical data management using non-commercial and open source solutions across a multi-center network supporting inpatient pediatric care and research in Kenya. *Journal of the American Medical Informatics Association*, 23(1), 184–192. doi: 10.1093/jamia/ocv028
- Wyber, R., Vaillancourt, S., Perry, W., Mannava, P., Folaranmi, T., & Celi, L. A. (2015). Big data in global health: improving health in low- and middle-income countries. *Bulletin of the World Health Organization*, 93(3), 203–208. doi: 10.2471/blt.14.139022
- Sileshi, B., Newton, M. W., Kiptanui, J., Shotwell, M. S., Wanderer, J. P., Mungai, M., ... McEvoy, M. D. (2017). Monitoring Anesthesia Care Delivery and Perioperative Mortality in Kenya Utilizing a Provider-driven Novel Data Collection Tool. *Anesthesiology*, 127(2), 250–271. doi:10.1097/ALN.0000000000001713
- Durieux, M. (2019). *Anesthesia Record. University Teaching Hospital of Kigali*.
- Verbeke, F., Karara, G., Van Bastelaere, S., & Nyssen, M. (2012). Patient Identification and Hospital Information Management Systems in Sub-Saharan Africa: A Prospective Study in Rwanda And Burundi. *Rwanda Medical Journal*, 69, 7–12. Retrieved from www.rwandamedicaljournal.org/previous-issues
- Kanter, A. S., Borland, R., Barasa, M., Iiams-Hauser, C., Velez, O., Kaonga, N. N., & Berg, M. (2012). The Importance of Using Open Source Technologies and Common Standards for Interoperability within eHealth: Perspectives from the Millennium Villages Project. *Advances in Health Care Management Health Information Technology in the International Context*, 189–204. doi: 10.1108/s1474-8231(2012)0000012013
- Allen, Christian & Jazayeri, Darius & Miranda, Justin & Biondich, Paul & Mamlin, Burke & Wolfe, Ben & Seebregts, Christopher & Lesh, Neal & Tierney, William & Fraser, Hamish. (2007). Experience in Implementing the OpenMRS Medical Record System to Support HIV Treatment in Rwanda. *Studies in health technology and informatics*, 382–386. doi: 129. 382-6. 10.3233/978-1-58603-774-1-382.
- Winner, L. (1980). Do Artifacts Have Politics? *Daedalus*, 109, 121–136.
- Bijker, W. E., Hughes, T. P., & Pinch, T. (1987). The Social construction of technological systems: new directions in the sociology and history of technology.
- Rosner, G. (2014). Who owns your data? Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing Adjunct Publication - UbiComp 14 Adjunct. doi: 10.1145/2638728.2641679

- Beck, E. J., Gill, W., & Lay, P. R. D. (2016). Protecting the confidentiality and security of personal health information in low- and middle-income countries in the era of SDGs and Big Data. *Global Health Action*, 9(1), 32089. doi: 10.3402/gha.v9.32089
- Mutimukwe, C., Kolkowska, E., & Grönlund, Å. (2019). Information privacy practices in e-government in an African least developing country, Rwanda. *The Electronic Journal of Information Systems in Developing Countries*, 85(2). doi: 10.1002/isd2.12074
- Binagwaho, A., & Scott, K. W. (2015). Improving the world's health through the post-2015 development agenda: perspectives from Rwanda. *International Journal of Health Policy and Management*, 4(4), 203–205. doi:10.15171/ijhpm.2015.46

