PUBLIC HEALTH INFORMATION SYSTEMS: FROM DATA TO KNOWLEDGE

A PREPRINT

Bell R Eapen
Information Systems
McMaster University
Hamilton, ON, Canada
eapenbp@mcmaster.ca

Norm Archer
Information Systems
McMaster University
Hamilton, ON, Canada
archer@mcmaster.ca

Kamran Sartipi
Information Systems
McMaster University
Hamilton, ON, Canada
sartipi@mcmaster.ca

January 3, 2020

ABSTRACT

Public Health Information Systems (PHIS) in the new era of Big Data, Machine Learning and Predictive Analytics, should be based on effective knowledge management (KM) theories, strategies and frameworks. We espouse a maturity model for PHIS based on the health impact pyramid metaphor to represent the transition from data to a knowledge-based system. We apply the knowledge-based view to describe how PHIS is different from other health information systems. We also propose pragmatic solutions to the challenges of knowledge management in public health and propose guidelines for evaluating software artifacts in PHIS.

Keywords Public Health Information Systems · Knowledge Management · Health Data Analytics

1 Introduction

Public Health Information Systems (PHIS) are entering a new era of Big Data, Machine Learning (ML), Artificial Intelligence (AI) and Predictive Analytics. Democratization of these technologies over the years led to wider availability of high-performance computing and trained workforce. Emerging techniques in Big Data, ML and AI have various applications such as epidemic surveillance [1] and health policy evaluation to name a few [2]. The importance of preventive health in checking the rising healthcare expenditure is widely recognized. However, the PHIS remain behind other health information systems (HIS) in adopting these emerging paradigms.

Public health is defined as “the art and science of preventing disease, prolonging life and promoting health through the organized efforts of society” [3]. The fundamental difference between clinical medicine and public health is that the former emphasizes the care of individual patients, while the latter focuses on the health of the population with an emphasis on prevention. The customers of public health are not always *patients* who require medical care, but clients who seek services some of which have a societal impact. In this article we use the term ’patients’ to refer to both.

HIS are knowledge management systems "dealing with processing data, information, and knowledge in health care environments" [4]. HIS are socio-technical systems that include people and processes in addition to technology. HIS can be patient-centric (electronic health records), decision-centric (clinical decision support systems) or population-centric in which case they are called PHIS. In addition to individual and group health, PHIS collect and analyze data related to environmental health such as food and water safety, toxicology, animal health, and social determinants of health. Though the health of the community and the health of an individual are inextricably linked, both require different approaches.

Information systems that create, store, retrieve, transmit and apply knowledge are called Knowledge Management Systems (KMS) [5]. Knowledge Management (KM) theories, strategies and frameworks are mostly studied in the context of the firm [6]. PHIS can benefit hugely from the application of KM principles and techniques. Public health organizations require KM at the societal level that is distinct from KM at the corporate level used to gain competitive advantage [7]. The sharing of knowledge assets related to public health can benefit all stakeholders, thereby creating an advantage for society as a whole [8].

Globally public health organizations strive to become digitalized with the adoption of PHIS. Studies show that PHIS have an uncertain impact on population health[9]. PHIS like most HIS, are designed and developed by IT experts with little knowledge of healthcare workflow, while the clinicians who evaluate the systems are unaware of the intricacies of system development [10]. This knowledge gap has led to a significant technology adoption gap in public health where paper-based KM is not uncommon [11]. The existing Information System (IS) theories on technology adoption and most existing evaluation frameworks thus have failed to capture the complexities of the healthcare domain [12]. This is stifling the inevitable transition from data-driven decision making to knowledge-driven evidence-based decision making [13].

The remainder of this article is organized as follows: First, we adapt the health impact pyramid, a popular metaphor in public health to PHIS, to explain the need for a transition from a data-centric view to a knowledge-based view. Next, we propose a maturity model for PHIS followed by a brief description of a few popular PHIS that are in widespread use. Theoretical solutions are not enough for public health. We therefore offer some pragmatic solutions to public health processes such as data collection and data warehousing for collaborative analytics and visualization. Finally, we propose guidelines for evaluating new PHIS artifacts for their potential impact.

2 Health Impact Pyramid and PHIS

The Health Impact Pyramid is a popular public health metaphor used to describe the impact of public health interventions [14]. It is a 5-tier pyramid with individual-specific interventions at the top and societal determinants such as socioeconomic factors at the bottom. It is based on the idea that actions represented by the base of the pyramid require less effort while having the greatest population impact.

The health impact pyramid aligns with the informatics pyramid that outlines the relationship between data, information and knowledge. In public health, data are the *facts* collected using various data collection forms. The analytics team converts data into indicators that can be easily understood by decision makers. There is a growing need for integrating information gathered at various levels of the pyramid into knowledge that can be applied for long-term sustainable healthcare.

We have integrated the informatics pyramid into a health impact pyramid in the context of PHIS. In our model, (Figure: 1) the apex of the pyramid is formed by activities that support the tacit knowledge of individuals, such as training, followed by actions that externalize knowledge in an explicit way. (e.g. Documentation). Design of data collection forms and databases form the middle layer followed by the data to information conversion process; health data analytics.

Knowledge dissemination and utilization depend on data visualization and the level of health information exchange between systems. PHIS are typically implemented with minimal exchange of information between systems. Information systems that enhance information flow across the continuum of care is vital for health system integration[15]. Hence, data visualization and interoperability among PHIS form the base of the pyramid operationalized through dashboards and health information exchange. It often happens that public health organizations stop at the data collection stage and fail to pass it down the chain with minimal additional effort. Next, we propose a maturity model for PHIS based on the knowledge management value chain [16] and other existing HIS maturity models.

2.1 The PHIS maturity model

Healthcare organizations and their information systems progress through various stages of growth and development towards an advanced maturity level. There are several maturity models for assessing hospital information systems such as the HIMSS maturity model for electronic health records [17] and Quintegra Maturity Model for electronic healthcare[18]. As IT infrastructure, PHIS adoption, and Interoperability are the priority areas in public health, we identified and integrated existing maturity models in these functional domains such as the NEHTA interoperability maturity model (IMM), the Healthcare analytics adoption model (HAAM) and the NHS infrastructure maturity model (NIMM) [19]. Our recommendation is based on the common themes identified in these existing models, contextualizing it to public health.

At *stage 0*, data collection is fragmented, with analysts using paper instruments to collect data. The analysis is done manually using tools such as spreadsheet software. In the next stage (*stage 1*), some form of health information system is available for computerized data collection. However, these disparate data collection software systems do not interoperate and the analysis is done manually. In *stage 2* PHIS is supplemented with an electronic health record

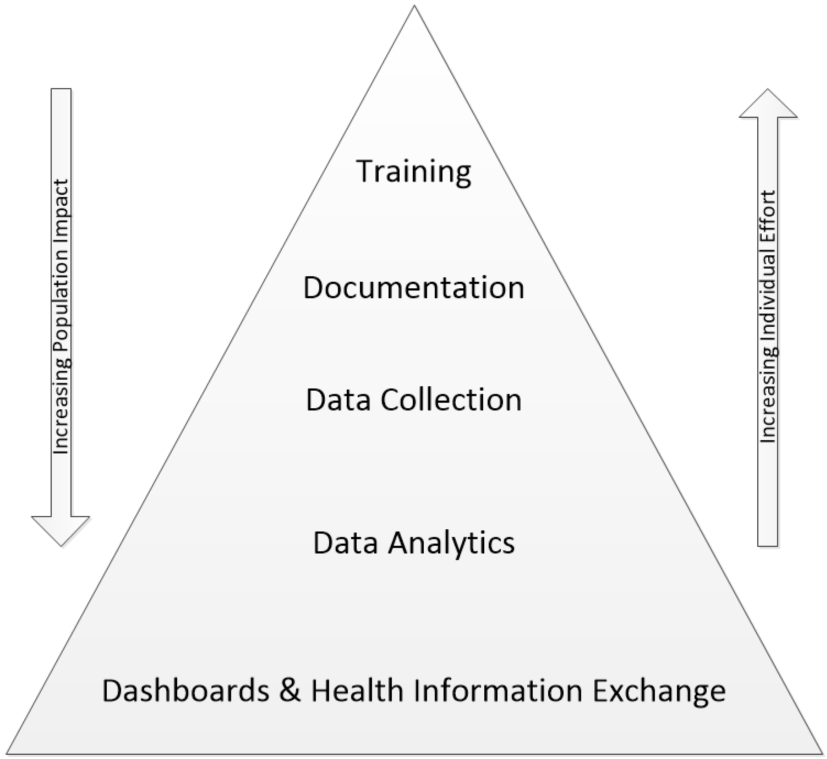

Figure 1: The PHIS Pyramid model.

(EHR) for collecting longitudinal patient data. Interoperability is not present, but there is usually a progression towards standardizing vocabularies. Reporting is still manual at this stage.

At *stage 3* standards and vocabulary-based interoperability exist, leading to efficient data collection. Reporting is still manual, but complex reports and visualizations driven by various stakeholder requests become possible. *Stage 4* is characterized by integrated geographic information systems (GIS) and data visualization. IT infrastructure now becomes stable with adequate support for privacy and security. Reporting, with emphasis on continuous quality improvement is possible, with team level data sharing and collaboration.

Stage 5 is characterized by the appearance of a data warehouse and real-time public-facing data visualization dashboards. Reporting becomes semi-automated with enterprise level knowledge sharing and collaboration. At *stage 6* integration with external HIS becomes possible with fully automated reporting. HIS interoperability becomes well established and operationalized, leading to regional data sharing and collaboration networks. Predictive analytics is implemented at the highest level (*stage 7*) leading to automated surveillance alerts based on decision support algorithms and a variety of integrated data sources.

The proposed PHIS maturity model is summarized in Table 1. This model connects people, process and technology illustrating the transition from data-driven decision making to knowledge-driven decision making as systems become increasingly mature.

2.2 PHIS software

In this section, we present some popular PHIS artifacts in widespread use. Our aim is not to provide an exhaustive list, but to present some salient features of PHIS.

The District Health Information Software (DHIS) is an open source PHIS developed by the Health Information Systems Programme (HISP)[11] that is used in more than 40 countries around the world. The latest version DHIS2, supports aggregation of data, organizational hierarchy and data visualization dashboards. The event-level data can be used to monitor epidemics. It provides APIs for integration with other HIS.

Epi Info [20] is a suite of software programs to support public health workflow developed by the Center for Disease Control and Prevention (CDC). Epi Info can be used for questionnaire design, data entry and analysis, mapping,

Table 1: PHIS Maturity Model

Stages	Features
STAGE 7	1. Automated public health surveillance alerts based on decision support algorithms and a variety of data sources. 2. Clinical Risk Intervention & Predictive Analytics. 3. Optimized eHealth interoperability.
STAGE 6	1. Integration with external data sources and automated reporting. 2. Established regional data sharing and collaboration networks with well-defined eHealth interoperability.
STAGE 5	1. Data Warehouse implementation with anonymized data. 2. Data marts and real-time public facing data visualization dashboards. 3. Semi-automated reporting. 4. Enterprise level of knowledge sharing and collaboration.
STAGE 4	1. Integrated system with Geographical Information Systems and dashboards for data visualization. 2. Manual reporting with emphasis on continuous service improvement. 3. Team level knowledge sharing and collaboration.
STAGE 3	1. Integrated PHIS and EHR with automation of data collection, and analytics. 2. Adoption of eHealth standards for interoperability. 3. Manual request-driven reporting.
STAGE 2	1. Use of PHIS and EHR with minimal or no interoperability. 2. Semi-automated data collation and analytics. 3. Manual reporting with standardized vocabularies 4. Problem-driven approach.
STAGE 1	1. A principal health information system, with several auxiliary systems for data collection. 2. Manual collation and analysis of data. 3. No eHealth interoperability. 4. Reactive and ad-hoc services with a focus on avoiding downtime.
STAGE 0	1. Fragmented data collection. 2. Manual collation and analysis of data.

graphing, and reporting. It can also be used for outbreak investigations and disease surveillance. It has a standalone desktop version, and a cloud-enabled version that can be accessed on any web-browser.

Panorama [21] is a PHIS developed by IBM, mainly for supporting communicable disease surveillance and management across Canada. The system provides public health agencies with tools for monitoring, managing, and reporting on public health. The Panorama project was implemented to enable the sharing of data on immunization and public health outbreaks across Canada, following the SARS outbreak in 2003 [22]. However, Canadian provinces have other disparate systems in place for communicable diseases surveillance, such as the iPHIS system in the Province of Ontario [23].

The existence of PHIS in itself does not guarantee success. Next, we discuss how technology can strengthen processes such as efficient data collection and analytics. We recommend some pragmatic solution based on our experience and previous work.

2.3 Semantic data capture

Template based knowledge capture is vital for the success of any PHIS. Validated data capture instruments such as InterRAI [24], Nurse-Family Partnership (NFP) [25] and Ontario Common Assessment of Need (OCAN) [26] are in use for various patient groups. Vendors build information systems that convert these instruments into electronic forms for data capture. It is common for healthcare organizations to realize that expensive and resource intensive PHIS do not adequately support end-to-end management and maintenance of such data collection forms (hereafter *E-Forms*).

An ideal E-Form framework should have a centralized repository, semantic interoperability, modular architecture, consistent and accurate rendering with embedded procedural logic [27]. Such a framework would enable the submission of the collected data to other systems that adopt the standard, making multi-level data collection and integration efficient. We adopted a pragmatic approach by leveraging the existing FHIR Questionnaire specification to operationalize such an E-Form framework. We have integrated various open-source tools into a framework and called it *FHIRForm* [27]. We hope that this open-source initiative will be useful to public health agencies in developing an effective data management strategy, especially in resource-deprived areas.

2.4 Data warehousing and reporting

Data warehousing is the process of integrating various operational data sources, and transforming and loading them into a separate database in a manner structured for reporting, analysis and data mining. A public health Data Warehouse (DW) should be simple, flexible, stable, and timely with good data quality [28]. Public health DWs face common challenges arising from the nature of epidemiological data and the temporal relationship between clinical events that can lead to modelling issues. Data privacy and security are of concern. The success of a data warehouse depends on factors such as support from higher management, availability of resources, and user engagement[29]. Linking and integrating various public health data sources allow performance monitoring, and improves the quality of patient care and patient safety [30]. Without the adoption of information systems that have the appropriate capabilities, analysis and reporting of data trapped in PHIS may be difficult and resource intensive. Though we have no panacea to offer, we recommend some architectural best practices that may be helpful in dealing with these challenges.

The vital piece in data warehousing and reporting from public health databases is a common data model (CDM) to ensure interoperability and collaboration. The Observational Medical Outcomes Partnership (OMOP) project developed one such CDM [31] which was subsequently taken up by the Observational Health Data Sciences and Informatics (OHDSI) collaborative. The OMOP CDM has an associated vocabulary and the relationships among concepts are explicitly specified. [32]. OHDSI provides several software tools to enable analyses to be executed rapidly and efficiently. General ETL tools for source systems are also available from the open-source community [33]. CDMs are generally patient-centric data models that extract data from electronic health records (EHR). The subtle yet important difference between patient-centric longitudinal data collection in hospital-based health systems (EHR) and group-centric PHIS is not immediately apparent. We propose a taxonomy based on principles of knowledge management systems (KMS) to make this distinction clear, and pose the need for a data model suitable for public health.

3 Knowledge-based taxonomy of HIS

A knowledge need is a situation where the provision of timely and appropriate knowledge improve performance. In the healthcare context, knowledge need can be broadly divided into the following:

1. Patient-specific knowledge (Ex: Blood sugar level of the patient)
2. Clinical domain knowledge (Ex: Differential diagnosis of patients with high blood sugar)
3. Public health knowledge (Ex: Prevalence of High Blood Sugar in the community)

These three knowledge needs are satisfied by different classes of health information systems. Patient-specific knowledge is provided by electronic health records (EHR), clinical domain knowledge is provided by clinical decision support systems (CDSS) and public health knowledge is provided by public health information systems (PHIS).

Public health knowledge creation requires data integration at various levels but requires less effort to contextualize and reuse. It is usually created by the community or several (often publically funded) collaborating healthcare organizations. A knowledge source may "augment" the existing knowledge of users in a particular domain or "substitute" a non-experts knowledge with that of an expert. The quality and lifespan of knowledge are moderate. The sharing of knowledge assets can benefit all collaborating organizations, thereby creating an advantage for society as a whole. The knowledge captured in PHIS is semi-structured. These PHIS characteristics are typical of a knowledge community [16].

PHIS should support reporting and analysis at various levels, utilizing visualizations and GIS functionality for decision support. PHIS needs to gather group level data in a privacy-preserving manner, but at the same time may need to follow up high-risk individuals longitudinally. Patient-centric observational data models such as OMOP may be useful in some cases, but may not be sufficient to assess policy impacts at the group level. Also, public health practitioners need to respond promptly to new threats such as epidemics [34].

CDM is inadequate for representing data such as environmental factors. A composite person-location based Public Health Data Model (PHDM) may be more suitable for a public health DW. In a CDM geographic or organizational location is an attribute of a person, while in a PHDM, this is a dimension over which other data can be summarized. DHIS2 defines geographic units for aggregating data. In some cases, the longitudinal health history of an individual becomes important too [35]. Pragmatically this can be achieved by making PHIS exchange data with EHRs. OpenMRS is an open-source EHR specifically designed for resource-deprived areas with a modular architecture for extending functionality [36]. OpenMRS has a module available for making data exchange possible with DHIS2 supporting individual and group data management. DHIS2 supports industry standard messaging services such as Apache Kafka [37] and RabbitMQ [38] for efficient messaging that can be used for communicating with EHR systems.

PHIS informs policy making, measure policy adoption and assess the policy impact [39]. Inappropriate use and presentation of public health indicators may lead to negative consequences for public health [40]. There are several open-source tools available for public health, but public health policymakers often find it difficult to choose the right solution for their needs. Next, we present some guidelines for PHIS evaluators to assess new PHIS artifacts.

4 Guidelines for evaluating PHIS software

Evaluating a PHIS artifact for potential impact is not different from evaluating other HIS. First, we recommend an assessment of the artifact's design style. It is crucial to include clinical domain experts to evaluate the style and the plausibility of the artifact's proposed use. If the artifact has a knowledgebase, it must be reliable and valid. The credibility of the knowledge base is important as many PHIS knowledge bases degrade with time and lose relevance [41]. Hence it is imperative to have an adequate strategy for maintaining and monitoring the knowledge base.

System usage is an important indicator of system success. Usage intentions and actual PHIS system use do not always match. This paradox has not been adequately explored and maybe related to factors beyond technology such as workflow problems and end-user training [42]. PHIS are socio-technical systems that cannot be wholly assessed in purely technical terms. The potential instrumental, economic and humanistic outcomes of the PHIS [43] must be assessed. Health is a global problem, and the humanistic outcomes of the PHIS should be assessed with a pragmatic and emancipatory worldview.

Open-source software has had a significant impact on public health, especially in lower income countries. When properly implemented open-source systems may play a significant role in reducing health inequalities [44]. However, the sustainability of open-source software should be thoroughly evaluated before expensive implementations. Standard compliance of PHIS is also vital in long-term sustainability. Commercial software vendors may have proprietary data standards that may lead to vendor lock-in.

Finally, it is essential to evaluate the impact of the PHIS on health outcomes as in other HIS. This step is often ignored because of the practical difficulties involved. However a few recent studies that showed a negative effect of HIS on patient outcomes and patient safety have kindled interest in the assessment of HIS impact on patient outcomes [45].

5 Discussion

The use of PHIS is vital in monitoring lifestyle-related disorders, communicable disease surveillance, child and adolescent health and sexual health. PHIS informs policy making, measure policy adoption and assess the policy impact [39]. PHIS is essential for assessing Continuous Quality Improvement (CQI) in Public Health [46].

Public health organizations collect knowledge at the societal level that is important for all stakeholders irrespective of their group affiliations. In KM parlance, everybody has an interest in the societal semantic memory [5]. The structured data collection instruments that we espouse [27] aid the conversion of societal episodic memory (such as epidemics) into societal semantic memory. PHIS and public health experts are grappling with emerging paradigms such as big data, machine learning and artificial intelligence (AI) applications in healthcare [47]. Clinical medicine traditionally has a positivist worldview and relies on randomized controlled trials to seek objective reality. The new generation of deep learning and AI applications that are growing in popularity are less amenable to positivist evaluations though many have numerical algorithms at their core. This is especially true for 'black box' solutions that aid decision making [48].

It is important to assess the implementation, adoption, and impact of various PHIS implementation projects, considering the amount of public funding involved [49]. Societal benefits in the context of PHIS are hard to define, quantify and compare. Benefit evaluation exercises often start post hoc with the intention of demonstrating potential benefits to the funder. Benefit evaluations usually start with a biased research question; *How do we measure the benefits?* An evaluator should gather credible evidence, by an appropriate evaluation design that engages all stakeholders, and the conclusions must be justifiable and actionable. The presence of various stakeholders and their differences in priorities make PHIS benefit evaluation, a complex undertaking that requires expertise and knowledge of clinical outcomes, IT design and implementation processes and the cognitive limits of the potential users [50].

6 Conclusion

Effective and efficient PHIS is required for evaluating the impact of public health interventions and measuring public health outcomes. In addition to collection and analysis of data, PHIS should support its transformation to knowledge for effective reuse [51]. Such a knowledge-based view would help evaluators gain insights into the socio-technical foundations of PHIS software artifacts.

References

[1] Simon I Hay, Dylan B George, Catherine L Moyes, and John S Brownstein. Big data opportunities for global infectious disease surveillance. *PLoS medicine*, 10(4):e1001413, 2013.

[2] Travis B Murdoch and Allan S Detsky. The inevitable application of big data to health care. *Jama*, 309(13):1351–1352, 2013.

[3] Donald Acheson. Public Health in England. *London: Department of Health*, 1988.

[4] Alfred Winter, Reinhold Haux, Elske Ammenwerth, Birgit Brigl, Nils Hellrung, and Franziska Jahn. Health information systems. In *Health Information Systems*, pages 33–42. Springer, 2010.

[5] Maryam Alavi and Dorothy E Leidner. Knowledge management and knowledge management systems: Conceptual foundations and research issues. *MIS Quarterly*, pages 107–136, 2001.

[6] Michael A Hitt, R Duane Ireland, and Ho-uk Lee. Technological learning, knowledge management, firm growth and performance: an introductory essay. *Journal of Engineering and Technology management*, 17(3-4):231–246, 2000.

[7] Karl M Wiig. Knowledge management in public administration. *Journal of Knowledge Management*, 6(3):224–239, 2002.

[8] Yung-Lung Lai, Maw-Shin Hsu, Feng-Jyh Lin, Yi-Min Chen, and Yi-Hsin Lin. The effects of industry cluster knowledge management on innovation performance. *Journal of Business Research*, 67(5):734–739, 2014.

[9] Carla AbouZahr and Ties Boerma. Health information systems: the foundations of public health. *Bulletin of the World Health Organization*, 83(8):578–583, 2005.

[10] David C Mohr, Stephen M Schueller, Enid Montague, Michelle Nicole Burns, and Parisa Rashidi. The behavioral intervention technology model: an integrated conceptual and technological framework for ehealth and mhealth interventions. *Journal of Medical Internet Research*, 16(6), 2014.

[11] Jørn Braa, Eric Monteiro, and Sundeep Sahay. Networks of action: sustainable health information systems across developing countries. *MIS Quarterly*, pages 337–362, 2004.

[12] Mike W Chiasson and Elizabeth Davidson. Pushing the contextual envelope: developing and diffusing is theory for health information systems research. *Information and Organization*, 14(3):155–188, 2004.

[13] Syed Sibte Raza Abidi. Knowledge management in healthcare: towards 'knowledge-driven'decision-support services. *International Journal of Medical Informatics*, 63(1-2):5–18, 2001.

[14] Thomas R Frieden. A framework for public health action: the health impact pyramid. *American Journal of Public Health*, 100(4):590–5, apr 2010.

[15] Esther Suter, Nelly D Oelke, Carol E Adair, and Gail D Armitage. Ten key principles for successful health systems integration. *Healthcare Quarterly (Toronto, Ont.)*, 13(Spec No):16, 2009.

[16] Alan R Dennis and Iris Vessey. Three knowledge management strategies: Knowledge hierarchies, knowledge markets, and knowledge communities. *MIS Quarterly Executive*, 4(4):399–412, 2005.

[17] HIMSS. The EMR Adoption Model. *HIMSS Analytics: Innovative Research/Informed Decisions*, 2008.

[18] Balaji Sharma. Electronic Healthcare Maturity Model (eHMM), 2008.

[19] João Vidal Carvalho, Álvaro Rocha, and António Abreu. Maturity Models of Healthcare Information Systems and Technologies: a Literature Review. *Journal of Medical Systems*, 40(6):131, jun 2016.

[20] A G Dean, J A Dean, A H Burton, and R C Dicker. Epi Info: a general-purpose microcomputer program for public health information systems. *American journal of Preventive Medicine*, 7(3):178–82, 1991.

[21] Paul Webster. National electronic disease surveillance: a dream delayed. *Canadian Medical Association. Journal*, 185(9):E365, 2013.

[22] Lawrence E Frisch, Elizabeth M Borycki, Alyse Capron, Abla Mawudeku, and Ronald St John. Public health informatics in canada. In *Public Health Informatics and Information Systems*, pages 603–618. Springer, 2014.

[23] Colin D'Cunha. Sars: Lessons learned from a provincial perspective. *Can J Public Health*, 95(1):25–26, 2004.

[24] John P Hirdes, Gunnar Ljunggren, John N Morris, Dinnus HM Frijters, Harriet Finne Soveri, Len Gray, Magnus Björkgren, and Reudi Gilgen. Reliability of the interrai suite of assessment instruments: a 12-country study of an integrated health information system. *BMC Health Services Research*, 8(1):277, 2008.

[25] David L Olds. The nurse–family partnership: An evidence-based preventive intervention. *Infant Mental Health Journal*, 27(1):5–25, 2006.

[26] Mike Slade. An evidence-based approach to routine outcome assessment: Commentary on… use of health of the nation outcome scales in psychiatry. *Advances in Psychiatric Treatment*, 18(3):180–182, 2012.

[27] BR Eapen, A Costa, N Archer, and K Sartipi. Fhirform: An open-source framework for the management of electronic forms in healthcare. *Studies in Health Technology and Informatics*, 257:80–85, 2019.

[28] JAMA. Publication of Updated Guidelines for Evaluating Public Health Surveillance Systems. *JAMA*, 286(12):1446, sep 2001.

[29] Barbara H. Wixom and Hugh J. Watson. An Empirical Investigation of the Factors Affecting Data Warehousing Success. *MIS Quarterly*, 25(1):17, mar 2001.

[30] Seyed Ali Mussavi Rizi and Abdul Roudsari. Development of a public health reporting data warehouse: lessons learned. *Studies in Health Technology and Informatics*, 192:861–5, 2013.

[31] J Marc Overhage, Patrick B Ryan, Christian G Reich, Abraham G Hartzema, and Paul E Stang. Validation of a common data model for active safety surveillance research. *Journal of the American Medical Informatics Association*, 19(1):54–60, jan 2012.

[32] George Hripcsak, Jon D Duke, Nigam H Shah, Christian G Reich, Vojtech Huser, Martijn J Schuemie, Marc A Suchard, Rae Woong Park, Ian Chi Kei Wong, Peter R Rijnbeek, Johan van der Lei, Nicole Pratt, G Niklas Norén, Yu-Chuan Li, Paul E Stang, David Madigan, and Patrick B Ryan. Observational Health Data Sciences and Informatics (OHDSI): Opportunities for Observational Researchers. *Studies in Health Technology and Informatics*, 216:574–8, 2015.

[33] Bell Raj Eapen and contributors. Hephaestus - data warehouse and ETL tools for open source EMRs. `https://github.com/dermatologist/hephaestus`, 2018.

[34] CJE Metcalf, WJ Edmunds, and J Lessler. Six challenges in modelling for public health policy. *Epidemics*, 10:93–96, 2015.

[35] Elaheh Pourabbas and Arie Shoshani. The composite data model: A unified approach for combining and querying multiple data models. *IEEE Transactions on Knowledge and Data Engineering*, 27(5):1424–1437, 2015.

[36] Burke W Mamlin, Paul G Biondich, Ben A Wolfe, Hamish Fraser, Darius Jazayeri, Christian Allen, Justin Miranda, and William M Tierney. Cooking up an open source EMR for developing countries: OpenMRS–a recipe for successful collaboration. In *AMIA Annual Symposium Proceedings*, volume 2006, page 529. American Medical Informatics Association, 2006.

[37] KMM Thein. Apache kafka: Next generation distributed messaging system. *International Journal of Scientific Engineering and Technology Research*, 3(47):9478–9483, 2014.

[38] Maciej Rostanski, Krzysztof Grochla, and Aleksander Seman. Evaluation of highly available and fault-tolerant middleware clustered architectures using rabbitmq. In *Computer Science and Information Systems (FedCSIS), 2014 Federated Conference on*, pages 879–884. IEEE, 2014.

[39] Ross C Brownson, Jamie F Chriqui, and Katherine A Stamatakis. Understanding evidence-based public health policy. *American journal of public health*, 99(9):1576–1583, 2009.

[40] Julian Flowers, Pamela Hall, and David Pencheon. Mini-symposium—public health observatories: public health indicators. *Public Health*, 119(4):239–245, 2005.

[41] John Fox, David Glasspool, Vivek Patkar, Mark Austin, Liz Black, Matthew South, Dave Robertson, and Charles Vincent. Delivering clinical decision support services: there is nothing as practical as a good theory. *Journal of Biomedical Informatics*, 43(5):831–843, 2010.

[42] Joshua R Vest, L Michele Issel, and Sean Lee. Experience of using information systems in public health practice: findings from a qualitative study. *Online Journal of Public Health Informatics*, 5(3):227, 2014.

[43] Ahmed Abbasi, Suprateek Sarker, and Roger HL Chiang. Big data research in information systems: Toward an inclusive research agenda. *Journal of the Association for Information Systems*, 17(2), 2016.

[44] Ernesto TA Marques Jr, Romulo Maciel Filho, and Paul Nordstrom August. Overcoming health inequity: potential benefits of a patient-centered open-source public health infostructure. *Cadernos de saude publica*, 24:547–557, 2008.

[45] Michael I Harrison, Ross Koppel, and Shirly Bar-Lev. Unintended consequences of information technologies in health care—an interactive sociotechnical analysis. *Journal of the American medical informatics Association*, 14(5):542–549, 2007.

[46] Alex Price, Robert Schwartz, Joanna Cohen, Heather Manson, and Fran Scott. Assessing continuous quality improvement in public health: Adapting lessons from healthcare. *Healthcare Policy*, 12(3):34, 2017.

[47] Fei Jiang, Yong Jiang, Hui Zhi, Yi Dong, Hao Li, Sufeng Ma, Yilong Wang, Qiang Dong, Haipeng Shen, and Yongjun Wang. Artificial intelligence in healthcare: past, present and future. *Stroke and Vascular Neurology*, 2(4):230–243, 2017.

[48] Judith E Dayhoff and James M DeLeo. Artificial neural networks: opening the black box. *Cancer: Interdisciplinary International Journal of the American Cancer Society*, 91(S8):1615–1635, 2001.

[49] Peter Littlejohns, Jeremy C Wyatt, and Linda Garvican. Evaluating computerised health information systems: hard lessons still to be learnt. *BMJ*, 326(7394):860–863, 2003.

[50] Elske Ammenwerth, Stefan Gräber, Gabriele Herrmann, Thomas Bürkle, and Jochem König. Evaluation of health information systems—problems and challenges. *International journal of medical informatics*, 71(2-3):125–135, 2003.

[51] Hong Chen, David Hailey, Ning Wang, and Ping Yu. A review of data quality assessment methods for public health information systems. *International Journal of Environmental Research and Public Health*, 11(5):5170–5207, 2014.