MEDINFO 2017: Precision Healthcare through Informatics
A.V. Gundlapalli et al. (Eds.)

doi:10.3233/978-1-61499-830-3-361

An Integrated Surveillance System to Examine Testing, Services, and Outcomes for Sexually Transmitted Diseases

Brian E. Dixon[a,b], Guoyu Tao[c], Jane Wang[b], Wanzhu Tu[b,d], Sarah Hoover[b], Zuoyi Zhang[b], Teresa A. Batteiger[d], Janet N. Arno[d]

[a] *Richard M. Fairbanks School of Public Health, Indiana University, Indianapolis, Indiana, USA,*
[b] *Center for Biomedical Informatics, Regenstrief Institute, Indianapolis, Indiana, USA*
[c] *Division of STD Prevention, Centers for Disease Control and Prevention, Atlanta, Georgia, USA*
[d] *Indiana University School of Medicine, Indianapolis, Indiana, USA*

Abstract

Despite laws that require reporting of sexually transmitted diseases (STDs) to governmental health agencies, integrated surveillance of STDs remains challenging. Data and information about testing are fragmented from information on treatment and outcomes. To overcome this fragmentation, data from multiple electronic systems spanning clinical and public health environments were integrated to create an STD surveillance registry. Electronic health records, disease case records, and birth registry records were linked and then stored in a de-identified, secure server for use by health officials and researchers. The registry contains nearly 6 million tests for 628,138 individuals over a 12-year period. The registry supports efforts to understand the epidemiology of STDs as well as health services and outcomes for those diagnosed with STDs. Specialized disease registries hold promise for collaboration across clinical and public health domains to improve surveillance efforts, reduce health disparities, and increase prevention efforts at the local level.

Keywords:

Sexually Transmitted Diseases; Registries; Public Health Informatics

Introduction

Sexually Transmitted Diseases (STDs)

Undiagnosed and untreated sexually transmitted disease (STD) is associated with adverse outcomes such as infertility, increased HIV transmission and acquisition, and adverse pregnancy outcomes. Several STD health services are recommended by the Centers for Disease Control and Prevention (CDC) to protect the reproductive and sexual health of young men and women. Recommendations include: annual chlamydia and gonorrhea screening of sexually active women ≤24 years, pregnant women, and older at-risk women; chlamydia and gonorrhea screening of anatomic sites of exposure (urethral, rectal, or pharyngeal) of men who have sex with men (MSM); retesting of all infected persons after treatment for chlamydia or gonorrhea; and syphilis testing of pregnant women as well as sexually active MSM [1].

Surveillance of STDs and STD Services

A core function of public health is the assessment of disease prevalence and burden as well as the utilization of health care services, also referred to as public health surveillance [2]. Ministries around the globe seek to perform surveillance on STDs as well as the utilization of STD health services. They further seek to monitor the quality of health services received by at-risk groups, assess adherence to recommendations for chlamydia and gonorrhea testing and retesting, syphilis testing, test results, patient and partner treatment, and the incidence of adverse outcomes related to STDs.

Assessing STD prevalence, burden and utilization of health services is challenging, because available data sources are limited by small sample sizes, incomplete demographic information, cross-sectional design, insufficient periods of follow-up time, and incomplete information about the services provided [3]. Access to a longitudinal data source with complete demographic and clinical information for individual patients is challenging for public health agencies and researchers, especially in the United States, given the fragmented delivery of care in public and private settings. Furthermore, there are even fewer data sources that capture an entire geographic community as opposed to a population defined by a single institution that provides care or insurance (such as a managed care population). While data sources such as population health surveys provide partial information, none have been able to provide all the information required to assess community access, utilization and quality of services, and the incidence of adverse outcomes following an STD.

Specialized Disease Registries

Centralized data registries have become important informatics tools for surveillance and research in a variety of public health contexts, including cancer treatment [4], immunization programmes [5], and injury prevention [6]. In fact, expanded health policies in the United States, referred to as "meaningful use" criteria for electronic health record (EHR) systems, include disease registries as a 'public health' criterion for the years 2013-2018 [7]. These policies encourage providers to submit patient-level information to specialized registries.

Once populated, disease registries can be reused for a variety of purposes, including clinical performance improvement, surveillance of disease incidence, and research on the utilization of health services [8; 9]. In essence, disease registries serve as integrated surveillance systems that support a wide range of clinical and public health functions.

Research Objective

Given the need for better community-level surveillance of STDs and STD health services as well as the past success of

other population disease registries, we sought to develop a longitudinal, comprehensive patient-centric registry to examine STDs and STD health services in a large metropolitan area. We hypothesized that the registry would support analysis of STDs and STD services as well as ongoing surveillance practice among public health agencies in that community.

Methods

We created a registry for all individuals tested for one of three STDs (chlamydia, gonorrhea, syphilis) between January 1, 2003, and December 31, 2014, by healthcare providers in the Indianapolis MSA (metropolitan statistical area). To create the registry, we gathered data from clinical and public health sources, linked individual patient records, and created a secure environment to facilitate collaborative access for surveillance and research. Our work occurred in partnership with local, state, and federal public health partners and was approved by the Institutional Review Board (IRB) at Indiana University.

Geography and Population Information

According to the 2010 census, Indiana ranked 15th among the states by population with just under 6.5 million residents. Consistent with national data, STDs are over-represented in racial and ethnic minorities (cases per 100,000 population). For example, the 2015 rate of gonorrhea among African Americans was 836 compared to the rate among Caucasian 87.7 and Hispanic individuals 85.0. The rates for chlamydia were 2234 for African Americans, 319 for Caucasians, and 545 for Hispanics for primary and secondary syphilis (26.8, 6.6 and 16.6, respectively).

The Indiana State Health Department (ISDH) STD Control Program divides the state's 92 counties into ten districts for morbidity reporting and disease intervention purposes. These district offices are the recipients of contracts with the STD Program for the state's approximately 30 disease intervention specialists. The Marion County Public Health Department (MCPHD) STD Control Program has responsibility for STD reporting in District 5, which includes Marion County (Indianapolis) and the seven surrounding counties: Boone, Hamilton, Hancock, Hendricks, Johnson, Morgan, and Shelby. This district makes up the majority of the Indianapolis MSA. District 5 (population of 1.7 million) and Marion County (population of 903,393) account for the largest share of Indiana's STD morbidity. In 2015, District 5 accounted for 39% of the state's chlamydia morbidity, 47% of the state's gonorrhea, and 60% of the state's primary and secondary syphilis. This partially reflects the district's racial health disparities, which is substantially more diverse than the state.

According to the CDC's 2015 STD Surveillance Report, Indiana reported a total of 28,886 cases of chlamydia and ranked 27th among states in rate (437.9/100,000), while Marion County ranked 25th among United States counties and independent cities at 949.3 cases/100,000 population. Indiana ranked 23th among states for gonorrhea with a case rate of 118.9/100,000 population, while Marion County ranked 16th among United States counties and independent cities in the rate of gonorrhea cases with 344.1 cases/100,000 population.

Residents of District 5 receive STD diagnostic and treatment services through the Bell Flower Clinic, the STD Control program of MCPHD, which also houses the District 5 reporting site. The program is operated by the Health and Hospital Corporation, which also operates MCPHD and safety net hospital for the county. The Bell Flower Clinic, therefore, serves those at highest risk. Of the unique patients at the Bell Flower Clinic, 57% were African-American, 33% were Caucasian, and 7% were other. Seven percent were Hispanic, mostly of Mexican descent.

Data Sources

Data for the registry came from three distinct sources:

1. The Indiana Network for Patient Care (INPC), a regional health information exchange (HIE) network that contains longitudinal EHRs for patients who received care in the Indianapolis MSA.
2. MCPHD Bell Flower Clinic, the STD Control Program which houses an information system where disease investigators enter details about STD cases reported to public health for the Indianapolis MSA.
3. MCPHD Birth Registry, a vital records information system used at MCPHD to capture data on all births in Marion County, in which Indianapolis is located.

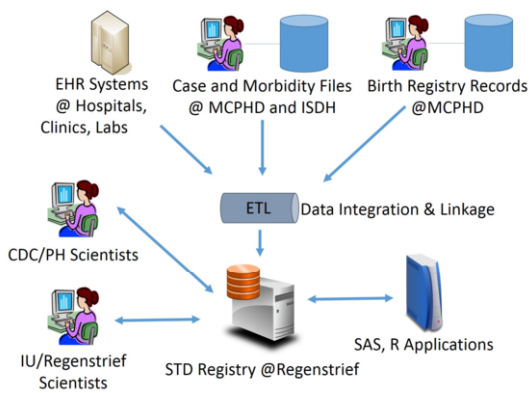

Figure 1 – Diagram depicting data sources, how data are integrated, and how data are accessed for surveillance

Indiana Network for Patient Care

The INPC is one of the largest community-based HIE networks in the United States [10; 11]. The INPC connects over 90 healthcare facilities, including hospitals, physicians' practices, pharmacy networks, long-term post-acute care facilities, laboratories, and radiology centers. The INPC maintains over five billion structured observations for over 12 million individuals; nearly one million electronic healthcare transactions are processed every day.

From the INPC, we extracted demographic data (e.g., gender, age, race, county of residence), STD laboratory testing data (e.g., lab test, date of test, result), co-morbidity data (e.g., pregnancy status, HIV status, ICD diagnoses) at time of STD test, encounter data (e.g., visit date, visit type), and medication history (e.g., drug name, drug class, date of dispense).

MCPHD Bell Flower Clinic and ISDH Morbidity Data

From the files at the Bell Flower Clinic, we extracted demographic data (e.g., gender, age, race), STD laboratory testing data (e.g., lab test, date of test, result), co-morbidity data (e.g., pregnancy status, HIV status) at time of the STD test, and medication information (e.g., drug name, drug class, date of dispense).

Because STD treatment may not be fully captured by the INPC, we extracted treatment of STD morbidity information

from the ISDH reporting database, SWIMSS (Statewide Investigating, Monitoring and Surveillance System).

MCPHD Birth Registry

The MCPHD Birth Registry contains records on all live births in hospitals and birthing centers in Marion County as reported by birth registrars. From the birth registry, we extracted pregnancy outcomes (e.g., date of delivery, infant weight, gestational age), STD laboratory testing data (e.g., lab test, date of test, result), and co-morbidity data (e.g., HIV status) at time of delivery.

Record Linkage, Integration, and Preparation

Data were integrated from the three distinct sources using a two-step process (Figure 1). First, individuals were linked across datasets. Next, data from each source was extracted and combined into a single, patient-centric data registry.

The INPC employs an advanced, probabilistic matching algorithm that matches patient identities using first name, last name, social security number (when available), date of birth, phone number (when available), and gender [12]. The two MCPHD datasets were independently linked to the INPC using the enterprise master person index (eMPI), based on that algorithm. Individuals in the MCPHD STD Case and Morbidity Files who did not match to an INPC individual were imported into the registry as new clients. Only data for individuals in the MCPHD Birth Registry who matched an INPC individual were imported from the vital records system.

Once patient identities were linked, longitudinal data for each unique individual were extracted, transformed, and loaded from the three sources into the registry. Each unique individual was given a de-identified or pseudonymised "client ID" that did not resemble his/her medical record number or any identifiers in the MCPHD datasets. Birth dates were transformed to ages and other identifiable information was removed. The ability to re-identify individuals exists to enable capture of new and updated information using a key between the medical record number and the client ID. Only the data manager at Regenstrief can perform data updates. Registry users cannot access such details to ensure confidentiality of records.

Data Management

The deidentified, linked registry datasets are hosted on a secure, virtual server at the Regenstrief Institute (Figure 1). The encrypted server is password-protected and managed by the technical services division at Indiana University (IU). The datasets are stored as a collection of interoperable data files, enabling them to be interpreted by all major analytical software tools. The data files require 20GB disk space.

Authorized users include public health scientists at the CDC and MCPHD as well as scientists working at Regenstrief and IU. The virtual environment affords users the opportunity to leverage a wide range of analytical tools, including SAS, R, and SPSS. Analyses of the data can be performed within the IU high-performance computing environment without necessitating download of the data onto local computers or drives. Output from the analyses, such as tables, charts, and graphs, can be downloaded from the servers to support in public health or academic reports.

Results

The registry contained 5,093,863 STD tests for 628,138 unique individuals collected over a 12-year period. In Table 1, we present the demographics of the individuals who were tested and those who tested positive for an STD in comparison to the overall demographics for the Indianapolis MSA. Although the area was well-balanced with respect to gender, a greater proportion of females were tested for STDs. This is likely due to clinical guidelines that recommend screening pregnant women and young, sexually active women for STDs. African American individuals were proportionately tested more and had a greater proportion of disease, than other races. These data highlight both a racial disparity in disease burden as well as the fact that providers are more regularly screening this population.

Table 1 – Demographics for individuals in the registry

Demo-graphic	Individuals Tested for an STD N=628,138	Individuals Positive for an STD N=119,751	Population of the MSA N=1,988,817
Gender			
Male	17.5%	25.3%	49.3%
Female	82.4%	74.6%	50.7%
Race			
African-American	25.0%	61.9%	15.3%
Caucasian	48.2%	25.1%	79.2%
Asian	0.4%	0.2%	2.9%
Hispanic	3.4%	2.7%	6.5%
Age			
5-17		10.4%	18.3%
18-24		40.0%	8.8%
25-44		25.8%	27.6%

In Figure 2, we summarize test results and positivity over time across all three STDs. Positivity is defined as the number of positive laboratory tests that confirm the presence of disease divided by the total number of lab tests analyzed. The overall volume of tests captured by the registry increased through 2011 then plateaued as result of the growth in the contributing data sources to the INPC from providers joining the HIE network to comply with 'meaningful use' incentives. Growth in the volume of data captured by the INPC resulted in decreased positivity; the number of negative tests grew by a factor of 3, while the number of positive tests increased by a factor of 2.5 (from 10,044 in 2003 to 25,606 in 2014). While the total number of positive STD cases grew dramatically, this growth is attributed to increased electronic lab reporting rather than an outbreak of disease.

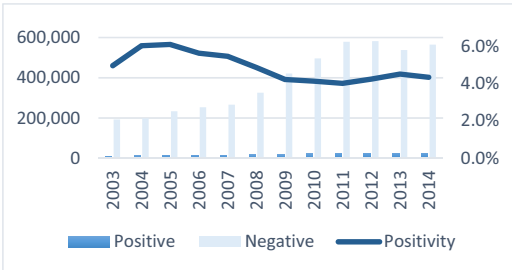

Figure 2 – Longitudinal test results and positivity for chlamydia, gonorrhea and syphilis. Test results are depicted as bars; positivity is depicted as a trend line

Discussion

By integrating three disparate sources of routinely collected clinical and public health data, we have created a novel registry containing longitudinal data on individuals tested for STDs in a large metropolitan area. The STD registry is an important public health informatics resource as it affords surveillance and research on STD testing, services, and outcomes. While each data source may exist independently in states and nations around the world, very rarely are EHR systems, vital records, and STD morbidity files linked and used to examine those tested or treated for STDs.

Most often, health departments only have access to positive laboratory results, which are required by law to be reported to health authorities [13]. While electronic laboratory reporting of positive test results improves the completeness and timeliness of public health reporting [14; 15], the lack of negative test reporting prevents health departments from examining whether individuals at risk for STDs are receiving recommended screening. Moreover, while health departments maintain both STD morbidity and vital record information systems, many health departments fail to link these data to examine outcomes for pregnant women or populations at risk for poor birth outcomes. Therefore, integrating clinical and public health datasets allows for an expanded evaluation of preventative services, clinical guidelines, and outcomes experienced by those with STDs.

Our work demonstrates the feasibility of creating a specialized STD registry for conducting surveillance and research. Building the registry further highlights three lessons for the biomedical and public health informatics communities. First, specialized registries that cross clinical and public health boundaries can be created in a way that preserves privacy and confidentiality. Second, record linkage is a crucial aspect of creating a registry. Third, health IT policies affect the breadth and depth of specialized registries.

Maintaining Privacy and Confidentiality of Health Data

Individuals and health organizations can be fearful of centralized, monolithic databases that contain protected health information [16]. Therefore, healthcare providers may be wary of releasing identifiable information to public health authorities, except when required by law to do so.

To create our registry, we leveraged the Regenstrief Institute, a neutral third party with experience in protecting health data. Regenstrief is a business associate with healthcare providers, public health authorities, and the INPC [11]. As a convening, trusted partner, the Institute was able to bring clinical and public health organizations together to exchange identifiable data that could be linked and then de-identified for storage in a secure, common environment that affords surveillance and research by multiple users. The role of neutral third parties is supported by prior HIE research [17]; therefore, public health authorities should look to HIE networks or other third parties to support creating and maintaining specialized registries.

The Critical Role of Data Linkage

One of the most important and challenging aspects of creating the registry was record linkage. Linkage is important because uniquely identifying individuals is critical to pulling fragmented EHRs together for tracking an individual's STD testing and services longitudinally.

Because the original data sources independently maintain distinct, unique identifiers for individuals and the United States lacks a universal health identifier, there was no easy method to link individuals at the start of the project. While the probabilistic algorithm used by the INPC's eMPI is excellent, it is not flawless. Therefore, each public health source had to be independently matched to the INPC, and then the two matched sets had to be linked using a third round of matching. Due to the lack of universal identifiers, each round of matching involved some degree of manual review and a decision threshold for determining correct matches had to be established. The necessity of manual review prohibits automation and scaling of specialized registry creation.

One potential solution for others is a client registry (CR) [18]. A CR adjudicates identities across EHR and other health data systems, like vital records and morbidity information systems, producing a centralized MPI to link identities across data sources. CRs have been demonstrated in HIE networks emerging in several countries, including Rwanda [19]. The CR should be further studied and applied to specialized disease registries.

Robust Policies Facilitate Specialized Registries

The STD registry is but one example of a specialized health data registry. While a wide range of registries for injuries, vaccines, and diabetes existed before the HITECH (Health Information Technology for Economic and Clinical Health) Act of 2009, the "meaningful use" program's incentive for clinical providers to contribute data to a specialized registry encourages clinical-public health data exchange. Local health authorities struggle to receive data that are currently not covered under existing public health laws. While new laws can be written to require data exchange, health authorities have an opportunity to leverage existing policies, like HITECH, to work with clinical providers to create and sustain population health surveillance through specialized registries.

When creating registries, health authorities should consider the unique health needs of their jurisdiction. Community health assessments, an activity involving the gathering of input from a wide array of stakeholders on the important issues facing a community, are another opportunity to work with healthcare providers to identify key health issues that might benefit from a specialized disease registry. Diabetes may be a top priority in one nation, while hypertension might be a top priority in another jurisdiction. Working with healthcare providers to identify the health priorities of a community may lead to better participation in the registry as well as progress in "moving the needle" towards higher quality of care and outcomes for at-risk populations.

Future Directions for the STD Registry

The STD registry allows our team to explore many important questions relevant to public health practice and research. Our team is currently conducting the following analyses and plans to disseminate results in the coming year:

- Utilization of STD Services: Understanding where individuals present for STD services is critical for appropriately allocating available resources. Using the data available within the registry, we are examining testing locations and positivity rates of individuals to determine where individuals present for STD care and whether a positive result increases the likelihood of presenting to a specific location.
- Testing and Outcomes for Pregnant Women: Women should be screened and treated for STDs while pregnant. Using the available testing data for women who either tested positive for pregnancy or delivered a baby, we are examining the proportion who received an STD test; of those, which women were positive and the birth outcomes for women who tested positive.

In addition, we seek to expand the capacity for the registry to support other research and surveillance of STD testing, services, and outcomes. In the next year, we plan to link the registry to other unique public health datasets, including the Immigrant Tuberculosis and All Refugee Application (ITARA) database. This database includes information on Indiana state immigrant medical exams. Incorporating these data will facilitate an analysis of newly immigrated citizens for incidence as well as risk factors associated with STDs. The registry will continue to be hosted at Regenstrief for use by public health researchers as well as epidemiologists in local, state, and federal agencies.

Conclusion

Using multiple data sources, we successfully linked and integrated data relevant to the testing, treatment, and outcomes for individuals with STDs to create a specialized STD registry. Registries like this one are increasingly feasible to build using informatics approaches. Specialized disease registries are critical to understanding the epidemiology of disease and enable collaboration across clinical and public health domains to improve surveillance, reduce health disparities, improve health services for individuals with disease, and increase prevention efforts at the local level.

Acknowledgements

The research reported in this publication was supported by the Centers for Disease Control and Prevention (CDC), United States Department of Health and Human Services (HHS), under Contract Number 200-2011-42027. The content is solely the responsibility of the authors and does not necessarily represent the official views of CDC or HHS.

References

[1] Centers for Disease Control and Prevention, 2015 Sexually Transmitted Diseases Treatment Guidelines, in, U.S. Department of Health and Human Services, Atlanta, GA, 2015.

[2] L.M. Lee and S.B. Thacker, The cornerstone of public health practice: public health surveillance, 1961--2011, *MMWR. Surveillance summaries: Morbidity and mortality weekly report. Surveillance summaries / CDC* **60 Suppl 4** (2011), 15-21.

[3] M.B. Ivankovich, J.S. Leichliter, and J.M. Douglas, Measurement of Sexual Health in the U.S.: An Inventory of Nationally Representative Surveys and Surveillance Systems, *Public Health Reports* **128** (2013), 62-72.

[4] S. Mehra, R.M. Tuttle, M. Milas, L. Orloff, D. Bergman, V. Bernet, E. Brett, R. Cobin, G. Doherty, B.L. Judson, J. Klopper, S. Lee, M. Lupo, J. Machac, J.I. Mechanick, G. Randolph, D.S. Ross, R. Smallridge, D. Terris, R. Tufano, E. Alon, J. Clain, L. DosReis, S. Scherl, and M.L. Urken, Database and Registry Research in Thyroid Cancer: Striving for a New and Improved National Thyroid Cancer Database, *Thyroid* **25** (2014), 157-168.

[5] Progress in immunization information systems - United States, 2012, *MMWR Morb Mortal Wkly Rep* **62** (2013), 1005-1008.

[6] R.J. Mitchell, C.M. Cameron, and M.R. Bambach, Data linkage for injury surveillance and research in Australia: perils, pitfalls and potential, *Aust N Z J Public Health* **38** (2014), 275-280.

[7] Medicare and Medicaid Programs; Electronic Health Record Incentive Program--Stage 3 and Modifications to Meaningful Use in 2015 Through 2017. Final rules with comment period, *Fed Regist* **80** (2015), 62761-62955.

[8] G.W. Hruby, J. McKiernan, S. Bakken, and C. Weng, A centralized research data repository enhances retrospective outcomes research capacity: a case report, *J Am Med Inform Assoc* **20** (2013), 563-567.

[9] B.E. Dixon, E.C. Whipple, J.M. Lajiness, and M.D. Murray, Utilizing an integrated infrastructure for outcomes research: a systematic review, *Health Info Libr J* **33** (2016), 7-32.

[10] P.G. Biondich and S.J. Grannis, The Indiana network for patient care: an integrated clinical information system informed by over thirty years of experience, *J Public Health Manag Pract* **Suppl** (2004), S81-86.

[11] J.M. Overhage, The Indiana Health Information Exchange, in: *Health Information Exchange: Navigating and Managing a Network of Health Information Systems,* B.E. Dixon, ed., Academic Press, Waltham, MA, 2016, pp. 267-279.

[12] S.J. Grannis, J.M. Overhage, S. Hui, and C.J. McDonald, Analysis of a probabilistic record linkage technique without human review, *AMIA Annu Symp Proc* (2003), 259-263.

[13] T.J. Doyle, M.K. Glynn, and S.L. Groseclose, Completeness of notifiable infectious disease reporting in the United States: an analytical literature review, *Am J Epidemiol* **155** (2002), 866-874.

[14] J.M. Overhage, S. Grannis, and C.J. McDonald, A comparison of the completeness and timeliness of automated electronic laboratory reporting and spontaneous reporting of notifiable conditions, *Am J Public Health* **98** (2008), 344-350.

[15] B.E. Dixon, J.J. McGowan, and S.J. Grannis, Electronic laboratory data quality and the value of a health information exchange to support public health reporting processes, *AMIA Annu Symp Proc* **2011** (2011), 322-330.

[16] R.V. Dhopeshwarkar, L.M. Kern, H.C. O'Donnell, A.M. Edwards, and R. Kaushal, Health care consumers' preferences around health information exchange, *Ann Fam Med* **10** (2012), 428-434.

[17] N.M. Lorenzi, Strategies for Creating Successful Local Health Information Infrastructure Initiatives in: U.S. Department of Health and Human Services, ed., Assistant Secretary for Policy and Evaluation, Washington, DC, 2003.

[18] T.D. McFarlane, B.E. Dixon, and S.J. Grannis, Client Registries: Identifying and Linking Patients, in: *Health Information Exchange: Navigating and Managing a Network of Health Information Systems,* B.E. Dixon, ed., Academic Press, Waltham, MA, 2016, pp. 163-182.

[19] Jembi Health Systems, Rwanda Health Enterprise Architecture (RHEA), in, 2012.

Address for correspondence

Brian E. Dixon, PhD
Regenstrief Institute, Inc., 1101 W. 10th St., RF 336
Indianapolis, IN 46202 USA
+1 (317) 278-3072
bedixon@regenstrief.org