Distribution Agreement

In presenting this report as a partial fulfillment of the requirements for an advance degree from Emory University, I agree that the School of Public Health shall make it available for inspection and circulation in accordance with its regulations governing materials of this type. I agree that permission to copy from, or to publish this report may be granted by the professor under whose direction it was written, or in his/her absence, by the Director of Career Education, when such copying or publication is solely for scholarly purposes and does not involve potential financial gain. It is understood that any copying from, or publication of this report, which involves potential financial gain will not be allowed without written permission.

_____

Shella Farooki, M.D. Date

AUTOMATED RADIOLOGY DATA AND INFORMATION TRANSFER (ARDIT): A PILOT STUDY AT EMORY HEALTHCARE IN CONJUNCTION WITH THE AMERICAN COLLEGE OF RADIOLOGY (ACR)

BY

Shella Farooki, M.D.

Degree to be awarded: M.P.H.
Career M.P.H.

Mark Conde, Committee Chair	Date
Mythreyi Chatfield, Ph.D., Field Advisor	Date
Kimberly Applegate, M.D., M.S., Committee Member	Date
Sheri Chernetsky Tejedor, M.D., Committee Member	Date
Melissa Alperin, MPH, CHES Chair, Career MPH Program	Date

AUTOMATED RADIOLOGY DATA AND INFORMATION TRANSFER (ARDIT): A PILOT STUDY AT EMORY HEALTHCARE IN CONJUNCTION WITH THE AMERICAN COLLEGE OF RADIOLOGY (ACR)

BY
Shella Farooki, M.D.
M.P.H., Emory University, 2013
M.D., Albert Einstein College of Medicine, 1994
B.S., Massachusetts Institute of Technology, 1990

Thesis Committee Chair: Mark Conde

An abstract of
A Thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements of the degree of
Master of Public Health in the Career MPH program
2013

Abstract

AUTOMATED RADIOLOGY DATA AND INFORMATION TRANSFER (ARDIT): A PILOT STUDY AT EMORY HEALTHCARE IN CONJUNCTION WITH THE AMERICAN COLLEGE OF RADIOLOGY (ACR)

BY
Shella Farooki, M.D.

Background: Data exchange between clinical healthcare and public health is vital to improving outcomes. As electronic medical records (EMRs) become more prevalent, this vast and valuable data source will be increasingly leveraged in order to gain more insight into population health.

Purpose: The purpose of this study was to create an automated method of radiology data transfer from Emory Healthcare to the General Radiology Improvement Database (GRID) registry for quality improvement at the American College of Radiology (ACR). Automated radiology data and information transfer (ARDIT) would allow Emory Healthcare's participation in GRID without requiring numerous, repetitive, manual, monthly web-based data manipulations.

Methods: Workflow analysis of Emory's current data extraction procedure was performed. De-identified turnaround time (TAT) data in hours by modality was chosen as the metric of interest and extracted from RadNet, Emory's radiology information system (RIS). TAT was defined as the time from radiology exam completion to the time of radiologist final signature on the report. Data transformation utilizing Excel files from Emory's structured query language (SQL) output, GRID's data measures and data dictionary for TAT was performed. We designed a new database query using SQL to extract TAT data from the external database sourced from RIS. The output was formatted as an Excel file allowing import into GRID.

Results: An ARDIT model for TAT data sharing was created using an external database server that transformed Emory's current TAT data into a format that can be transferred directly to GRID via Secure File Transfer Protocol (SFTP). ARDIT has functional capacity to share TAT data with the ACR and can be expanded to incorporate other GRID measures and metrics for seamless data transfer.

Conclusions: EMRs are an important source of data for public health informatics platforms such as registries. The ARDIT model created has functional capacity for sharing TAT data between

Emory Healthcare and the ACR. Additional SQL code will be required to share the remaining GRID measures. The technology for facile health information exchange between EMRs and public health platforms does exist, but rapid implementation is limited by resources and human factors.

AUTOMATED RADIOLOGY DATA AND INFORMATION TRANSFER (ARDIT): A PILOT STUDY AT EMORY HEALTHCARE IN CONJUNCTION WITH THE AMERICAN COLLEGE OF RADIOLOGY (ACR)

BY
Shella Farooki, M.D.
M.D., Albert Einstein College of Medicine, 1994
B.S., Massachusetts Institute of Technology, 1990

Thesis Committee Chair: Mark Conde

A Thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements of the degree of
Master of Public health in the Career MPH Program

2013

ACKNOWLEDGEMENTS

I would like to thank the faculty at the Rollins School of Public Health at Emory University for their support and dedication to the field as well as formal and informal input on this work. Thank you, in particular, to a few amazing teachers and mentors: Frederic Grant, Ph.D., M.P.H., M.B.A., Jon Lipsky, M.B.A., and Marc Overcash. Thanks to my committee members, Sheri Chernetsky Tejedor, M.D., and Kimberly Applegate, M.D., M.S.. I am thankful for the kind help of Jessie Knighton. I am greatly indebted to Mark Conde, my Thesis Advisor and Committee Chair, for his humor, wisdom, and mentorship and to my outstanding field advisor, Mythreyi Chatfield, Ph.D., Senior Director for Data Registries at the American College of Radiology.

I would also like to thank the talented students in my cohort who provided valuable feedback and problem solving throughout the process. Special thanks to Viju Sebastian for opening up my world to SQL, Melanie McLeod, M.A. for challenging me and keeping me on task, Madhavi Golakoti, M.C.A. for instilling just the right amount of confidence in me when I needed it most, and Cameron Tran for helping me navigate the Emory infrastructure. And thanks to students in other tracks of the cMPH program, David Gahn, M.D., M.P.H. and Maggie Hassan, B.D.S., who also helped me in this endeavor.

No words can express the gratitude I feel for my wonderful parents and family. I am so fortunate to have amazing family and friends. I am the third physician in my family, following in my late Grandfather's and my cousin's footsteps, to obtain an MPH, so this is an honor for me.

I also want to dedicate this work to my mentors who won't see me graduate, the late David J. Sartoris, M.D., M.P.H. and the late Francis Dono, DO, who inspired me and showed me the bridge between Medicine and Public Health.

Alyssa and Sasha, this is dedicated to you. As I watch with fascination the evolution of technology with Big Data, advanced analytics, and information exchange, I cannot even imagine what the future will hold for you. You will look back on this paper in amazement that Mommy worked in the stone ages, when it wasn't easy for hospitals to share information, when smart phones weren't even smart, and when humans were the obstacles in the application of their own technology.

Table of Contents

Chapter 1: Introduction

Examination of Context and Background

In its groundbreaking report, *To Err is Human: Building a Safer Health System*, the Institute of Medicine (IOM) recommended over a decade ago the creation of voluntary, national mandatory reporting systems to collect and analyze standardized clinical information in an effort to better understand and reduce medical errors. These reporting systems are a key feature of high-reliability organizations and are a prerequisite for quality improvement (1). Reporting systems also function to identify adverse events, to compare metrics, to establish benchmarks and areas for improvement, and to increase safety and reliability. Error reporting systems allow for performance accountability and improved safety through the analysis of important information. Furthermore the IOM report states, "good reporting systems are a tool for gathering sufficient information about errors from multiple reporters to try to understand the factors that contribute to them and subsequently prevent their recurrence" (1).

Industries outside of healthcare and public health, such as aviation, have utilized reporting systems to improve safety and reliability in a high-risk milieu, and military aviation in particular has contributed significantly to modern safety systems (2). In healthcare, anesthesia was the first medical specialty to use reporting systems for quality improvement, observing a 10-20% decrease in mortality and morbidity (3). There are numerous benefits of reporting systems including decreased "morbidity and mortality, improved patient and referring physician satisfaction, reduced health expenses and medical liability costs" (4). When physicians participate in national or international registries for example, patient care is improved, costs are lowered, and physician learning is enhanced (5). Despite these benefits, a major challenge in

implementation is stakeholder buy-in for sufficient participation in order to create an adequate response system (1).

Following "*To Err is Human*," the IOM issued "*Crossing the Quality Chasm*" in 2001. In this follow up report, information technology was recognized as a valuable tool to potentially improve the safety, effectiveness, timeliness, efficiency, and equity of healthcare and to make healthcare more patient-centered. The IOM recommended that lawmakers and government commit to building a national healthcare infrastructure given the body of evidence supporting the value and utility of automated reminder systems, online support groups, clinical decision support systems, telemedicine, and improved communications stemming from various informatics solutions. Additionally, the IOM supported involving the patients in their healthcare information via the Internet and educating patients about the benefits of automation of clinical data (6). This philosophical statement changed the paradigm of clinical and public health towards a more technology-focused model. In 2006, the Royal Australian and New Zealand College of Radiologists developed a voluntary incident reporting system in an effort to improve quality and patient safety (7). This was the first national registry specific for Diagnostic Imaging, and since November 2009, over 800 incidents have been recorded in the system (7). The creation of this registry was important as it forged the connection between Radiology and public health.

In response to the IOM call to collect standardized information, the American College of Radiology (ACR) developed the National Radiology Data Registry (NRDR). NRDR is a group of quality databases, one of which is the General Radiology Improvement Database (GRID) which collects and aggregates various performance indicators such as turnaround times, patient wait times, patient satisfaction, and other metrics for the purpose of quality improvement (8).

The guiding principle underlying registries is cyclical quality improvement that empowers imaging facilities to drive patient satisfaction and clinical quality.

The field of radiology has been scrutinized in particular due to the large number of radiological procedures performed and the massive increase in radiation delivered to patients from increased utilization of computed tomography (CT) examinations (9). Thus, there was a need to better understand if all these additional CT examinations were really necessary and if imaging facilities were optimizing their radiation doses. The Dose Index Registry (DIR) collects data on dose indices and provides comparisons by exam to help participating facilities target protocols for analysis and optimization. NRDR registries, including GRID were developed so that radiologists nationwide could make more objective and evidenced-based decisions in their practices and be able to compare their quality metrics to other facilities regionally and nationally. GRID was able to provide "accurate and objective national and regional benchmarks" for radiology practice process measures, outcomes and incidents (8). A national quality database system such as GRID is most effective when it is part of a comprehensive quality improvement strategy (4).

Currently, GRID is one of six registries under the umbrella of the National Radiology Data Registry, or NRDR, existing at the ACR since 2008. The compendium of registries serves to enhance the specialty of radiology and patient care by providing a national perspective on how diagnostic radiology is practiced in the United States (10). The mission of NRDR is to "aid medical imaging facilities with their quality improvement programs and efforts to improve patient care by comparing facilities' data to those of their region and the nation...and to do so easily and correctly" (11).

GRID works through a network of facilities in the United States that sign an NRDR participation agreement to submit data to the registry. Facilities register online and pay a one-time $500 registration fee along with an annual fee based on the number of radiologists that practice at that facility. Facilities are responsible for entering data into NRDR and assuring the data entered meets quality standards. ACR registry data anonymize any limited Protected Health Information (PHI) that may be collected; PHI is not disclosed. GRID collects aggregated data and does not contain any patient level information. NRDR uses a Limited Data Set (LDS) for quality improvement research purposes only. GRID registered facilities may choose between Green and Gold levels of participation; gold level participation tracks more quality metrics than the Green level and provides outcome measures. Once participation agreements are executed, centers may enter data manually on web-based forms or upload the data as flat files in a specified format. Once centers are set up and have sent their data, they receive semi-annual reports on their performance as compared to other participating facilities (8). In terms of demographics, GRID has the highest participation from metropolitan community facilities followed by academic facilities with the highest participation rates in the Northeast and Midwest (12).

Major limitations to the voluntary participation in GRID are similar to any of the challenges related to benchmarking: limited standardization, lack of ability to automate data entry, and inadequate actionable informative feedback. GRID is difficult to use, and it is cumbersome to populate manually the 149 total data fields that must be submitted for each month although facilities can collect data and submit it every six months (12). Adding to this impediment is the challenge for facilities to find the required data. Since participation in GRID is voluntary, underreporting is an issue when attempting to analyze population data to establish national benchmarks. GRID's formal enrollment at the time of this paper is 54 fully registered

facilities with 22 active facilities that contribute data (12). Poor participation in GRID is thought to be multifactorial: "suboptimal data collection tools, lack of anonymity due to limited enrollees, and lack of awareness of existence of ACR registries." (4). By contrast, the ACR's Dose Index Registry (DIR) has over 750 facilities enrolled, and over 450 facilities currently contributing data. With DIR, the radiation dose information from the CT scanner is automatically sent to a Digital Imaging and Communication in Medicine (DICOM) node with Triad, software developed for facility interoperability that allows DICOM images to be downloaded and viewed (13). The DICOM node is a computer terminal at the facility that is loaded with the software and is part of a network of computers. The automation and "hands off" approach removes the manual labor barrier and invites increased participation.

In other clinical specialties such as cardiology, registries such as the American College of Cardiology's (ACC) PINNACLE do not charge a fee and provide EMR system integration (SI) mapping solutions to allow accurate data capture and automatic population of the registry data element (14). Because of the resource intensive nature of registry participation, third party organizations have developed that specifically provide IT solutions for registry participation (15). Additionally, EMR vendors are becoming more cognizant of the public health need to share data and may become more proactive in facilitating this process. Since there are no interface solutions like the ACC's at the ACR, this thesis examines the process of building a bridge towards more automation and less burden on IT personnel. The current ACR registry paradigm could change in the future particularly if registry based solutions become more prevalent.

Problem Statement, Purpose, and Research Question

Data collection is integral to the establishment of nationwide registries and surveillance systems and is challenged by our current vertical silo landscape for healthcare where individual healthcare systems are able to integrate and communicate vertically but not as well or at all horizontally across networks. In some instances, data collection from within a vertical silo or from one standalone facility can be challenging. One needs to know where the data lives, in what format does it exist, what it looks like, and how to obtain it. Determining these steps is not always easy. At the 2012 Annual Imaging Informatics Summit and NRDR User Group Conference, a speaker representing Rockingham Memorial Hospital/Sentara Healthcare in the User Group Forum, emphasized the difficulty that facilities have in data collection (16). This sentiment is expressed by other current GRID participants who also share in that same struggle (12). The frustration and manpower requirement related to data collection is a potential deterrent from participating in GRID, and ideally, the ACR would like to help alleviate the cumbersome and labor intensive steps in data transfer in order to remove obstacles from participation (12).

The purpose of this study is to develop a semi-automated or automated method of data collection and transfer for Emory Healthcare in order to facilitate their participation in GRID. Accomplishing this goal enhances and improves public health by bringing Radiology as a subspecialty from an individual health delivery perspective to one that is populational. This study also aims to better understand the availability and barriers to obtaining health care data for extraction purposes so that future automated transfer of the designated data from the healthcare facility level to a national registry can be accomplished using a defined message structure. This thesis is specifically focused on the processes of finding the data source, access, transformation, and transfer. Although GRID encompasses numerous process and outcome measures, structural

measures, protocols and safety procedures, for the purposes of this study the metric of radiology report turnaround time (TAT) is considered. The radiology report turnaround time metric is the time from the completion of the radiology examination to the time the radiologist electronically signs the final report. The research question to be answered here is how can this data be sent automatically from Emory Healthcare to GRID? Emory's data is useful to GRID since the more facilities participate, the more representative the data is within GRID. It is also important to identify and understand the barriers that the facility faces when having to collect, aggregate, and send data to GRID. It is postulated that a method to streamline data entry for Emory Healthcare is feasible and that the new process will be more efficient, automated, and accurate than manual monthly entry.

This thesis is focused on improving the data collection process for GRID, specifically the turnaround time metric. The ACR would like to transform the data collection process from manual entry to a more automated electronic process for all participating facilities (12). However, even with new automatic upload capabilities, many radiology practices are burdened with collecting information from multiple outpatient centers and/or hospitals (4). Determining where the data is and how to get it is a major problem for many imaging facilities particularly when dealing with disparate information systems (12). Standards are only one part of the larger complex problem of integration. Another obstacle is that the metrics and outcome measures in GRID may differ from a practice's internal metrics, so facilities may not be measuring those metrics. Use of Health Level 7 (HL7) International listeners may be helpful in the future in streamlining the process from both interoperability and interfacing standpoints, but the ACR does not yet have HL7 capacity for GRID and interoperability issues cannot be solved with HL7 alone (4, 12). The ACR has a vested interest to develop methods to aid in the automated

collection of data for GRID including potential collaboration with Radiology Information Systems (RIS) and workflow vendors and interested facilities such as Emory Healthcare (12).

Theoretical Framework

The ACR currently utilizes a C-store listener to collect dose-related data for CT scans; the CT scanner from a participating facility automatically sends the appropriate DICOM structured reporting (SR) object to a personal computer (PC) located at the imaging facility. TRIAD software is used to de-identify and transmit the data to the Dose Index Registry (DIR). The ACR does have an HL7 listener that can be potentially integrated with TRIAD so that facilities could send HL7 feeds to the note and IP address where TRIAD is installed. Ideally, this model can be used to extrapolate a solution to import data to GRID.

At the 2012 Annual Imaging Informatics Summit and NRDR User Group Conference, radiologists discussed how they are able to share their data with the ACR via registries (17). At Rockingham Memorial Hospital (RMH)/Sentara Healthcare, a community hospital in Harrisonburg, VA, GRID participation was used to aid in decreasing patient wait times by 10% in 2011 and to decrease turnaround time in all modalities (CT, Magnetic Resonance Imaging, Nuclear Medicine, etc.) by providing feedback and actionable data to the facilities. The speaker emphasized, however, that the greatest challenge was data collection. Additionally, their participation was limited, and an automated method of sharing did not exist. For 2013, RMH will be measuring current metrics, adding wrong exam/patient/site metrics and adding digital radiology repeat rates to their reports (16). For RIS, the challenge lies in local codes and vocabulary that must evolve to a single standard if interoperability and data sharing is ever to occur. However, this is a universal challenge for all reporting needs. Additionally, RIS are

vendor specific and therefore, an informatics solution that automates data extraction at one facility will not necessarily work at another facility. Thus, a solution for data sharing at the facility level must be customized to address the uniqueness of that particular facility in terms of RIS, local codes/vocabulary, and information technology infrastructure. The facility specific or silo nature of EMRs are vendor driven. Vendors have a disincentive to collaborate and support interoperability because they need to preserve their market share and do not want to reveal trade secrets to competitors. This is particularly true of the big vendors, but this could change in the future as free and/or open source EMRs become more prevalent in the marketplace.

Significance

In creating an informatics solution for Emory Healthcare to transmit data to GRID, a model to extract the correct data in an efficient way is being created and establishes a framework for other institutions. If the pilot at Emory is successful, more data points from Emory can be incorporated so that participation in GRID is more seamless and comprehensive. Additionally, a successful pilot at Emory may be translatable elsewhere. The study implications could improve participation in GRID, allowing for more accurate national data and benchmarks. The impediments encountered at Emory and the methods in which they were handled can prove useful for other facilities as well. Lastly, and most importantly, data sharing amongst institutions and between clinical medicine and public health is crucial for improving patient outcomes.

Method and Rationale for Target Journal Selection

The target journal for publication of this work is the Journal of the American College of Radiology (JACR). This decision is based on the ACR's role as a stakeholder and this work being part of a larger effort aiding healthcare networks in collecting and submitting data, and the

role of JACR as the most-read journal on topics related to practice issues in radiology. The model developed in this thesis will be incorporated as part of a joint paper with other collaborators.

Chapter 2: Review of the Literature

The PubMed search terms utilized for the literature review were: registry, informatics, radiology, electronic health data exchange, electronic medical record, health quality, HL7 messaging, data transfer, and interoperability, and combinations of these terms.

EMRs are rich and vast sources of patient information for large populations and therefore are desired systems from both research and public health standpoints. Additionally, they provide useful data and can strengthen the collaboration and coordination between clinical health and public health (18). EMRs have also been shown to enhance registries by providing unique or updated clinical information (19). Given the trend towards establishing a national medical record and national EMR adoption, it makes sense to leverage EMRs to fulfill public health needs, research, and data warehousing (20) .

Unfortunately, EMRs are challenging sources of data for clinical and public health use for several reasons. First, EMR data is structured and stored for healthcare transactions (usually as a relational database) related to patient care and billing, and the desired data may not be readily available for searching and retrieving (21, 22). Data retrieval can also be burdensome on a system and result in slower performance. The radiology, laboratory, pharmacy, and clinical documentation aspects of an EMR often reside in separate areas within the EMR and contain different data, sometimes with different standards and structure (23). For example, radiology

uses Digital Imaging and Communications in Medicine (DICOM) standards but laboratory uses Logical Observation Identifiers Names and Codes (LOINC). Additionally, when hospitals utilize external radiology or laboratory facilities, the data from these facilities are not necessarily transferred to the hospital's EMR, thus rendering an incomplete picture of the patient's care (23). Even within an integrated healthcare system, an EMR is a transactional database that may not provide all of the required data. Data that is not important to clinical care but important for research or public health may be absent from the patient's medical record (24). In addition, clinical providers decide where in the EMR to place information and also may use different terminology for the same disease state (24). Data extraction from EMRs is challenging since the data can be unstructured, such as clinician notes, or incomplete or difficult to locate, requiring mapping exercises (25). The design and development of specific EMRs for medical subspecialties such as oncology may make outcomes research more facile (26). Searching an enterprise data warehouse, if available, for the required attributes can also be difficult even though more information is stored in the data warehouse than in the EMR (21).

The quality of data within the EMR affects research study results, and one must take into account the idiosyncrasies of EMRs when utilizing them as a data source (24). Free-text within the EMR, particularly clinician/provider or nursing notes, can also make searching for specific research terms difficult and can lead to interpretation bias, but coding alone may not be granular enough for some research purposes (22, 23). For example, in a Norwegian study of 14 practices using Winmed EMR, data extraction related to diagnosis code alone was found to be variable depending on how staff members utilize their EMR. Therefore, how users and physician practices utilize their EMR is an important factor when performing data extraction for research purposes (27). The clinical, business, and research workflow all impact data quality, accuracy,

and reliability. Vocabulary variability can also present challenges when abbreviations and different names for the same disease exist. For example, "type 1 diabetes mellitus" is often abbreviated as "DM1" (24). Other potential problems of utilizing EMRs is lack of socio-economic data, missing data from use of both paper and electronic records, and reliance on billing codes to identify diseases (24).

Tradititionally, EMRs are vendor specific and competition in the EMR market inhibits the desire and motivation for developing interoperability between EMRs. However, in the past year, EMR vendors have approached the ACR in attempts to develop mechanisms for interoperability with registries(28). There is a paucity of literature on HL7 messaging and integration of EMRs with public health information systems. However, Emory's EMR, Millennium, does have an automated datafeed to the Georgia Immunization Registry (GRITS). Data linkage issues as well as ethics, privacy and confidentiality of electronic health information are also important considerations (29). Despite these challenges, methods for data extraction from EMRs have been developed to further research or public health efforts (20). The simplest method is to perform a query using the EMR software but one is restricted to simple searches. Queries can also be created to generate reports by start and end date or by diagnosis. On a more complex level for example, Boolean logic can be used to generate reports by diagnosis and allow selection of a second filter such as medication or co-morbidity. Other complex data extraction method requires execution of structured query language (SQL) commands through an interface with the EMR or use of database tools to conduct complex searches (24).

In certain instances, the collaboration between the clinical and public health sectors occurs through the creation of a health information exchange which can result in improved individual and population outcomes (30). The paradigm of a health information exchange (HIE)

allows for facile exchange of information between the public health realm and the clinical sector, and how that information exchange occurs is variable. According to Balog, there was a perfect storm for immunization data exchange with the passage of the Health Information Technology for Economic and Clinical Health (HITECH) Act and Meaningful Use (MU) in the sense that a provider's EMR must connect with public health via immunization data transfer to registries using HL7 V2 standards (31). The challenge remains as to how to capture provider EHR data and transfer it to Immunization Information Systems (IIS) while maintaining high quality of data and representativeness. Electronic record exchange between EHR vendors and Immunization registries can occur as real time exports, batch exports, or by query. To achieve successful exchange, Balog suggested a three step approach; Step One was Investigation and Discovery and entails research on the vendors, providers, stakeholders, and business processes. Step Two was testing and evaluating the HL7 messages and system to assure accuracy, completeness, and functionality. Step Three was implementation including confirmation of successful data transmission and determination of frequency of data transfer. This paper was important in establishing a methodology and framework for EHR data exchange (31).

In Louisiana, an innovative, secure bidirectional electronic health information exchange was created for HIV/AIDS and linked patient level EMR data with statewide public health surveillance data. Multiple disparate Louisiana hospitals were connected with the Office of Public Health (OPH) over a designated wide area network (WAN) connection via a secure point to point tunnel, and an HL7 interface allowed for data exchange to the OPH's Louisiana Public Health Information Exchange (LaPHIE). The purpose of the exchange was to identify patients with HIV/AIDS who had not received HIV care for longer than 12 months so that providers could be notified. After LaPHIE confirmed the patient's demographic information, an HL7

standard alert message (Patient Problem message or PPR) was sent to the clinician interface. Due to privacy and confidentiality issues, only physicians and nurses could view these alerts. After a clinical action was performed, a standard HL7 patient problem response (PPR) message was sent to LaPHIE and logic was utilized to determine if the patient should remain in the system for follow up or if the issue had been resolved and the patient name removed. Some major challenges with this implementation were determination of data ownership, ethical issues regarding HIV diagnosis and sharing information related to that diagnosis, and legal and ethical issues related to individual rights versus public health protection and promotion. LaPHIE continues to be important because it demonstrated real-time, bidirectional HIE between multiple EMR systems and a public health surveillance system with alert feedback to clinicians (32). Within a two year period, there were demonstrable significant improvements in HIV- related utilization of health services and outcomes after implementation of LaPHIE (30).

Hernandez et al were successful at extracting patient data from hospital EMRs with a cancer registry, the Florida Cancer Data System (FCDS). They utilized three years of data from 2007 to 2010 from a large hospital system in Florida to identify 12,804 unique breast tumors. ICD-9 codes were used as the trigger. The researchers found that registry information can be enhanced by adding data obtained from the EMR regarding surgery, radiation, and chemotherapeutic regimen. Even the most comprehensive clinical registry can be improved by the addition of information provided by EMRs, thus contributing to patient centered outcomes research (19). Some challenges for these researchers included the lack of standards for data collection, variations in coding and transmission of data, and exclusion of certain data sets from various hospitals (19). This study had important implications since cancer registries often lack information on chemotherapeutic regimens or radiation therapy information (33).

In 2011, the Lucile Packard Children's Hospital (LPCH) devised an integrated tool using HL7 messaging to connect their EMR with the California Immunization Registry (CAIR) (34). Although there are many institutions that use unidirectional information exchange from their EMR to an immunization registry, only a few use bidirectional information exchange to extract information from the registry to populate the EMR (34, 35). This was the first visual integration with bidirectional exchange between an EMR and an immunization registry, and the advantage of having this information flow back to the provider is that immunizations administered elsewhere become easily accessible within the provider's workflow. Hospital immunization data including historical data was transferred with HL7 code to the registry. A "smart link", or web based icon was then created so that it would appear in the "Patient Summary" area in the chart thus alerting physicians and nurses. Clicking on the icon would direct the healthcare worker to the CAIR database and with an institutional log on, the worker could access the patient's immunization history (34). A post implementation survey several months after the rollout revealed that the majority of providers thought the link improved efficiency and increased the probability that their patient's record was up to date (34). The web-based visual integration with unidirectional exchange via HL7 is unique in this circumstance and on this scale. The only other similar solutions were accomplished on smaller, internal EMR systems (35).

Using the nation's largest HIE, the Indiana Health Information Exchange (IHIE), a network encompassing 70 hospitals and 18,000 physicians, researchers linked clinicians to public health by connecting practitioners to an immunization information system (IIS) via the Child Health Improvement through Computer Automation (CHICA) system (36). CHICA was built on an OpenMRS medical database (OpenMRS, Indianapolis, IN) and uses an HL7 message processor that communicates with its unique EMR (37). CHICA also provided decision support

for its primary care providers in pediatrics, and its medical logic module (MLM) interpreted and extracted data from rules encoded in Arden Syntax (Health Level Seven International, Ann Arbor, MI) (38). CHICA had a unique user interface which consisted of forms called Adaptive Turn Around Documents (ATAD) which were clinical forms that were printed and scanned. These forms were completed by physicians, nurses, or patients and then scanned in a standard document scanner. A Teleform Reader interpreted the handwriting and checkboxes for the immunization history. For physicians without access to EMR, faxes were utilized for data exchange. The CHICA Immunization Assistant (CHIA) received the fax from the pediatrician's office, and then the coded fields on the form were extracted and read by the Teleform Reader using optical character recognition. CHIA then generated an HL7 VXQ, or vaccination record query, that was sent to the Children and Hoosier Immunization Registry Project (CHIRP), Indiana's statewide IIS. The HL7 messages were sent to CHIRP via the virtual private network (VPN) of the Indiana Network for Patient Care (INPC). CHIRP then matched the VXQ against records in the registry. When a patient match occurred, CHIRP faxed a form to the clinician with a barcode for unique identification and vaccination data with recommendation for what vaccines should be administered and the dates. While at the pilot stage in 2011, the large scale deployment of such an exchange was complex and sophisticated.

This asynchronous method of data transfer worked for Indiana physicians without EHRs and did not affect busy clinical workflow, however, manual data exchange was not nearly as rapid as synchronous communication methods (36). It was vitally important that information flowed from the clinical to the public health realm, and these investigators understood that "a major challenge for public health informatics is facilitating the exchange of information" between the two sectors (36). This team astutely noted that often public health data arrived on

paper or forms that were filled out manually and transmitted via postal mail or fax. Even when reporting was electronic, the initial data entry was often manual. This resulted in underreporting for surveillance systems and registries. If data could flow easily between the two environments, more accurate and rapid assessment of health and disease could take place. With the Health Information Technology for Economic and Clinical Health (HITECH) Act, MU criteria place pressure on providers to link Electronic Health Records (EHR) to immunization registries in some fashion. In this capacity, the path for communication between EHRs and registries is paved. Although faxing information over phone lines was not ideal, it was important that information transfer was occurring in some manner at least until real-time EHR connections were established. A major limitation was the quality of the fax machine which sometimes precluded automated recognition and necessitated manual intervention (36).

The IHIE is closely associated with the Regenstrief Institute, an entity which has designed and implemented two clinical messaging systems both called DOCS4DOCS. The system received HL7 messages from data sources and then delivered the reports/results to physicians in a variety of manners: web, fax, or HL7 to an EMR. ATADs were converted to portable document format (PDF) files so that the system could handle the information and have it cross domains and still abide by security policies. Some important lessons learned were that hospitals sometimes changed vendors and EMRs were often upgraded, resulting in sudden loss of functionality of messaging. New coding systems and HL7 changes from hospitals also presented challenges for data transfer (39).

The US Department of Veterans Affairs (VA) created their own method of autopopulating a registry using EMR data (40). The VA developed local Clinical Case Registries (CCR) which utilized their already established EMR to allow physicians to create

customizable reports. These reports were then aggregated and selected clinical and/or demographic data from the EMR was captured, structured, and placed in a relational database system for statistical analysis. Previously, the VA had limited local reports being transferred to the Immunology Case Registry (ICR) for human immunodeficiency virus (HIV) and the Hepatitis C Case Registry (HCCR). The new CCR used International Classification of Diseases -9th Revision (ICD-9) codes and Logical Observation Identifiers Names and Codes (LOINC) to identify desired patients for the ICR and HCCR registries. LOINC was used to identify positive test results for HIV and Hepatitis C. A CCR coordinator then manually and periodically reviewed the list to assure accuracy and confirm the diagnosis prior to transmission of data to the national registry. This framework was not without challenges, namely local data security, determination of who had access to CCR, and the need for continual software updates as EMRs updated. A major accomplishment, however, was the ability to create a local registry that tied into a national registry allowing national data aggregation from all VA facilities down to the unique patient level. HL7 extraction of data was performed on a nightly basis. Since patient data security was a top priority, access to the local CCR was restricted. A limitation of this design was that the national CCR may not have been consistent with the local registry due to nightly addition of data and some sites missing data. Strengths of this study include the robust EMR utilized by the VA that allows for population management tools, and the CCR has been useful beyond its original scope. For example, the CCR helped to identify patient safety issues when the Food and Drug Administration (FDA) issued alerts about HIV medications by generating facility lists of patients on those medications. Additionally, since the VA is a national system, data from the CCR can be utilized for cost modeling and to assist federal agencies. In the VA experience, computer generated lists for patients who are potential cases along with a

manual confirmation of the case resulted in accurate registry lists. Since local software is evolving constantly, the CCR data extraction process must constantly be assessed so that critical data is not being overlooked (40). These lessons learned are translatable to other systems and provide value to the approach of registry design and population.

Registries can be created even in low income, underdeveloped nations which much less available resource than in the United States (41-44). For example, in Pakistan, a locally developed, customized Karachi Trauma Registry (KITR) was created and populated using existing medical records, trauma diagnosis based on ICD-9 codes, and open source software (41). The motivation for development of this software was the recognition that 90% of trauma deaths occur in low and middle income countries, and these countries often cannot afford expensive commercial off the shelf (COTS) products. Additionally, information gained from data collection could improve the process of treating these patients and ultimately clinical outcome. The authors developed an electronic trauma registry using Windows-XP TM based software on a PC requiring Pentium III or higher and storage on SQL Server 2005 R and SQL Server express R. For data collection, the authors used drop-down menus and checklists and limited the amount of free text that could be utilized. The development took 23 months and cost approximately $9600. The registry was able to analyze mechanism of injury, severity of injury, length of stay, and survival probability. To increase efficacy, the authors suggest providing provider based data collection methods or making a standardized data collection tool (41).

Paper based records can also be used to populate data in a registry as demonstrated by the Kampala Trauma Registry in Uganda. A lesson learned in Uganda was to keep the data set minimal since providers have resource constraints. The registry contained 5,210 records from an urban 1,500 bed tertiary hospital and a 100 bed district hospital (44). In Haiti, paper based

records were also used, but data entry and analysis was performed with Epi Info (Centers for Disease Control and Prevention, Atlanta, GA) (42). Data exchange in low resource settings must be designed and adapted to the unique circumstances of the environment and its available resources. An overarching theme for all of these systems is the need for accurate patient identification and registries, however rudimentary, are important for improving outcomes (41-44).

Electronic medical record support for public health (ESP) systems have been spawned to harness EMR data and analyze it for potential use in public health surveillance (45, 46). Developed by the Harvard Center of Excellence in Public Health Informatics, ESP code is open source and is able to capture data for infectious diseases, diabetes and notifiable diseases as well as provide syndromic surveillance (45). Open source code can aid low resource nations in more efficient health information exchange (41). ESP is an automated surveillance system in Massachusetts and Ohio and includes over 1 million patients (45). Limitations of using EMR data for surveillance, as other investigators have noted, are billing and coding practices which may limit the identification of true cases, and using a single component of the medical record is not as sensitive or specific than using multiple components of the medical record (45). Another platform of ESP is RiskScape which is a graphic display for public health surveillance that can search by zip code, disease, age, ethnicity, and other parameters. Other investigators have used EHR systems in combination with other data sources to provide medical subspecialty data. For example in ophthalmology, investigators used data from Kaiser Permanente EMR, Veterans Health Administration EMR, and Centers for Medicare and Medicate (CMS) to create a comprehensive eye and vision health surveillance system (47).

Another electronic medical records retrieval system (ERS) was created to improve the efficiency of clinical research. In this study, 800,000 clinical cases in an EMR were extracted and evaluated to determine which patients fit the eligibility criteria for research on osteoporosis medications. Researchers manually converted the narrative data from the EMR to codes or parameters. A data model was manually populated so that narrative criteria were converted into entity level criteria. This aided in extracting patients who met the criteria and excluding those that did not. To identify target patients, logical queries were defined in the ERS, and then the ERS could automatically generate the SQL necessary to extract the data based on the logical queries. Executing the logical queries generated the targeted patient list. Certain items are flagged so that the investigators could confirm data accuracy. A total of 7,062 patients were on the target list, and data extraction took approximately three months with an additional four months for investigator confirmations and statistical analyses. An important lesson learned was that conversion of narrative data to computable criteria within a data model was efficient and was independent of EMR database structure. The authors suggest possibly organizing the EMR extracted data before using it for research, and using ICD-10 codes alone were insensitive for the specific diagnosis yielding only 35 of the 72 cases, or 48.6% (21).

It is critical to remember that the method of data extraction affects the quality of the data in the registry particularly in our current landscape towards greater information exchange. How data is transferred to a registry or the method by which data is received can affect attributes of a registry. A large study assessing 757,476 de-identified demographic records and 2,634,101 vaccine records in 2010 demonstrated that the completeness of data was best for records arriving as flat files. Completeness was not as good for manually-entered data or HL7 records. Completeness proportions were defined as the number of demographic fields completed divided

by the total number of demographic fields, and the number of complete immunization fields divided by the total number of immunization fields. For demographics, the fields that had to be completed included name (first, last, middle), address, social security number (SSN), birthdate, medical organization. Investigators also looked at the minimum/mandatory data items such as name (first, last), birthdate, provider organization, vaccine type, and service date. The mean completeness for all immunization records was high at 99.28%, and batch flat files had the highest completeness record for demographics of 90.76% while HL7 records had highest mean completeness for vaccination records at 99.5%. Additionally, manually-entered data and HL7 immunization records demonstrated greater timeliness than data that was imported as flat files. Timeliness was defined as the number of days between vaccine administration and submission of data to the IIS. Thus, provider-immunization information system exchanges impact data quality in different ways. This study was limited, however, because it only considered timeliness and completeness for data quality metrics and did not examine accuracy (48). Immunization systems (IIS) are mature and HL7 messages have been a part of the landscape since 2001, so it is logical to use IIS as a model for registries and surveillance systems in public health informatics.

Linking datasets is another method by which to enhance the quality of data and information within a registry. The 2005-2009 Nebraska Cancer Registry Data was linked to Nebraska hospital discharge data captured an additional 5% of potentially missed cases and is a useful strategy to finding new cases, updating treatment regimens, and performing treatment surveillance. The linked dataset also found 12 percent more treatment cases for colorectal cancer patients and 14 percent more treatment cases for breast cancer patients (33). Linking could potentially address some of the previous shortcomings of EMRs.

This chapter summarizes a number of methods to extract EMR data for public health surveillance or registry use. EMR data extraction for public health and research use can be challenging, and the models presented here were reviewed so that a method could be developed for Emory Healthcare. According to McDonald, an inexpensive solution of patient health data exchange lies in informatics standards (23). Global acceptance of standards allows for easier interoperability. A common theme in the models presented is that each institution or groups had to deal with the problem of interoperability and variations in data exchange standards. Until standards become more uniform in the future, interoperability and exchange will require customization. Data linkage, investment in technological infrastructure, and collaboration between public and private sectors are all needed in order to improve health services research (49). Researchers may in the future want to extract and link data within a certain EMR component, but linking EMR data between health care entities is challenging. Moving forward, establishing and developing data standards, removing barriers such as inability of states to link CMS data with cancer registries, and collaboration will improve the public health informatics landscape (49). This study addresses the gap of radiology data flow between Emory Healthcare and the American College of Radiology's GRID registry. Utilizing models from previous work, this study addresses stakeholder requirements and past lessons learned to design a method of data exchange between the two entities.

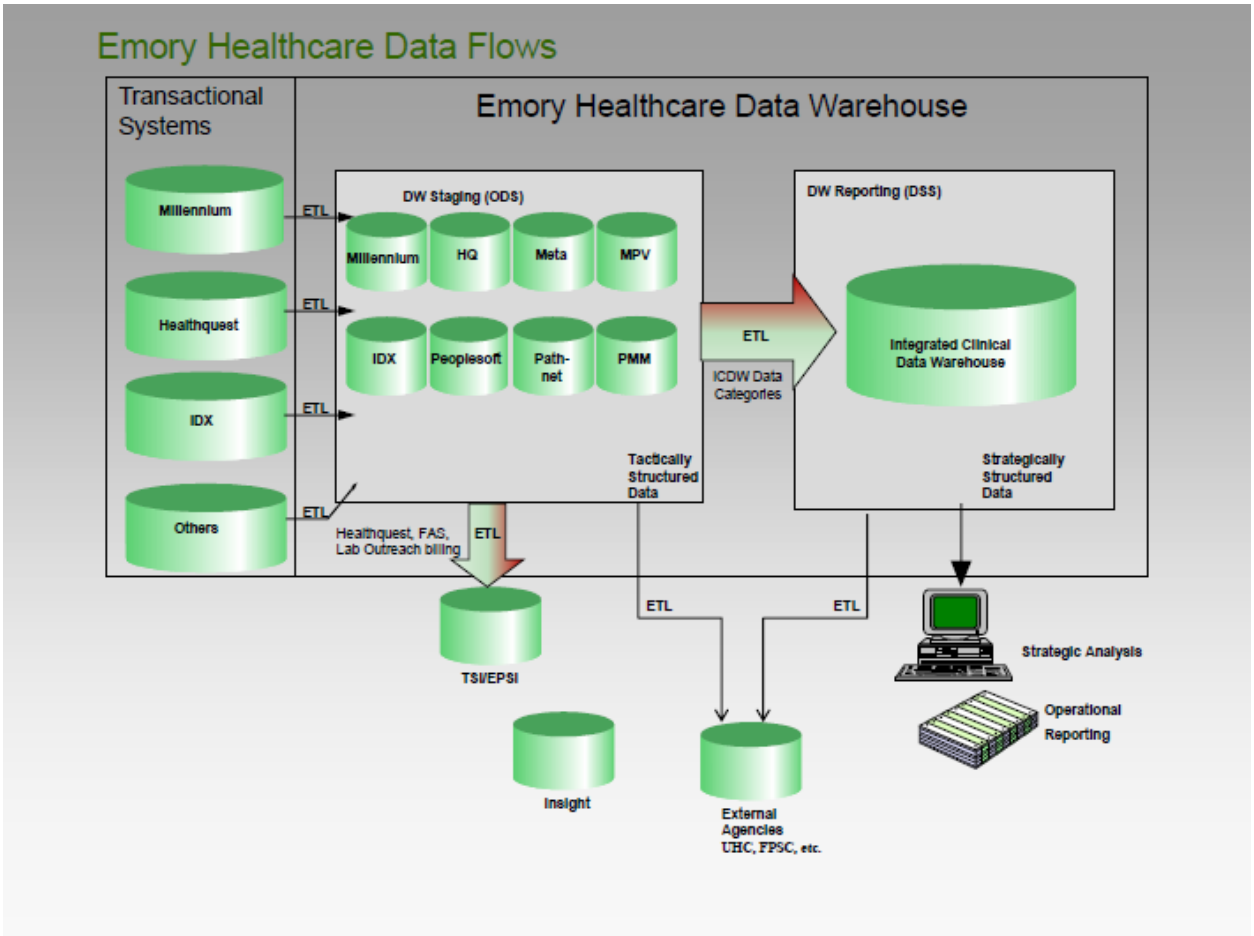
Chapter 3: Methodology, Approach, and Solution

Methodology and Approach

A workflow analysis of the current Radiology Quality and Safety data extraction at Emory Healthcare was performed to identify opportunities for how and where a solution for data

extraction could be implemented without interruption of workflow or additional expense. The Emory Healthcare dataflow diagram depicting the technical infrastructure of the entire Healthcare system is shown in Figure 1. The database systems that manage the relevant data for this study are RadNet which is a RIS which exists within Millennium (Cerner, Kansas City, MO), Emory Healthcare's EMR (Powerchart, Cerner, Kansas City, MO) and the Emory Healthcare Clinical Data Warehouse (CDW) (50).

Figure 1 Data Flow Diagram of Emory Healthcare depicting transactional systems and their relationship with the Clinical Data Warehouse. The Radiology data is located within the RadNet application of Millennium.



Several informal telephone interviews were conducted in 2012 and 2013 with a data analyst, a radiology decision support analyst, and a radiologist at Emory Healthcare who served

as the Vice Chair for Quality and Safety in order to define, map, diagram, and confirm the workflow process (51). On a weekly basis, a data analyst dedicated to radiology applications sends a pre-programmed SQL query to RadNet. The analyst receives flat files that are copied and pasted into an Excel spreadsheet (Microsoft Office, Seattle, WA). The decision support analyst then de-identifies the data. The Excel files reside on a shared drive with raw data, and the spreadsheet is manipulated by the analyst numerous times before it is emailed to the radiology staff as part of their Quality and Safety program. Because of the row limitations of Excel, this process was performed weekly as opposed to monthly, and the monthly report was a compilation of the weekly data. A review of the monthly report revealed that Emory collected almost all of the required data for GRID. This was an important discovery since most institutions struggle with finding this type of data (4, 16, 28).

The next step was stakeholder identification. In this study, the stakeholders were the Vice Chair for Quality and Safety at Emory Healthcare, the radiologists, Decision Support Analysts at Emory Healthcare for Radiology, and the Registry Director at the ACR. The stakeholders represented different aspects of the study, for example, the Decision Support Analysts represented an operational aspect whereas the Vice Chair for Quality and Safety had a managerial and strategic role. Stakeholders were interviewed to identify and understand stakeholder needs and requirements in particular functional, operation, and interface requirements. Equally important was identification of constraints related to revenue, design, and process.

The ACR's requirement for GRID was that any flat file could be accepted as a monthly report or Emory could perform a manual monthly web based form entry via GRID's user interface (UI). The ACR did not have an HL7 feed capacity (28) . Emory's requirements were

that the solution should not require additional revenue, significant IT resources, new or additional software, or significantly disrupt workflow. The solution also had to allow Emory to transfer the data in an efficient manner, avoiding the situation of requiring an employee to manually type in the required fields for GRID on a monthly basis. An automated push of these data from RadNet was not evaluated due to resource constraints. Interviews were a critical step in determining which department and person(s) had stewardship of the data and when/if data or departmental permissions or agreements were required. A challenge of this and many other informatics projects was navigating the institutional informatics red tape, understanding the institutional infrastructure in order to find the correct and appropriate personnel, navigating through the necessary steps for allocating resources, and obtaining permission to view/share data.

This study was considered non-Human Subject Research (non-HSR). The de-identified data was already collected by the data steward, and there was no risk of re-identification. Institutional Review Board (IRB) exemption was not required as per the Data Disclosure Decision Tree Request by Emory Covered Entity Source guidelines dated 9/16/2011. A Radiology Data Analyst acted as the honest broker, overseeing the data and processes utilized for the study. At no point was the data used for this study re-identifiable for determination of protected health information (PHI). A data use agreement was signed, and data steward approval was granted by Dr. Carolyn Meltzer. The Department of Radiology approved data sharing on January 8, 2013.

Teleconferencing with the Vice Chair for Quality and Safety in Radiology, Radiology Decision Support Analysts, and the ACR representative occurred on multiple occasions in 2012 and 2013 in order to understand the best steps in the workflow process for data extraction. There was minimal impact upon workflow when the Excel spreadsheet that was already being

generated from the de-identified and manipulated data was utilized for transformation. Thus, the data transformation for GRID occurred downstream from the data analyst's workflow process. A single radiology metric, namely radiology report turnaround time (TAT) was the quality measure that was studied. TAT is defined as the time between radiology exam completion and final signature on the report by the radiologist. TAT is reported for each radiologist by specialty or division, and Emory has 11 divisions. The TAT was not organized by modality in the current workflow. The target percentage for reports signed within 24 hours was 85%. One limitation of this process was that the completed time stamps for exams were dependent on technologists completing the exam manually (a required click in the RIS) and was not an automated or machine driven event.

The data attributes for TAT were reviewed. The Excel file contained columns for examination order complete date, order complete time, order procedure (or modality), order start date and time, and the final date and time. The TAT as defined by both Emory Healthcare Department of Radiology and the ACR was the time it took in hours from completion of the radiology examination to the final date and time which is when the radiologist finalized the report. Thus, TAT is equivalent to difference between the final date and time of the final report and the complete date and time for the exam. This metric was chosen because it was simple to measure, clinically important, and is a common service metric for many healthcare facilities.

Data transformation was performed with SQL Server Integration Service (SSIS) (Microsoft, Redmond, WA). This step was necessary in order to calculate TAT, stratify it into the categories specified by GRID, and convert it to a format that could be transferred.

In the approach to establish interoperability between Emory Healthcare and the ACR's GRID registry, a few important questions were raised. How much data can be pushed versus pulled by query? Is there an interoperable linkage that can be created? Since a query already exists that extracts the data we need, how can the registry be autopopulated via HL7 messaging? On the registry side, is it feasible to pull the data with periodic queries as previously discussed? ACC's PINNACLE has a solution that automatically pulls the data from the EMR and formats it for direct population into the registry (14). Autopopulation could present a problem as most institutions would want to review the data for quality control prior to submission to assure there are no typographical errors or outliers. Is utilizing a third party vendor such as Fig MD a viable option(15)? In answering some of these questions, the evaluation of other industries for ideas was undertaken. For example, how are financial or business transactions taking place between institutions and will those models translate to healthcare?

Our approach also had to consider impediments such as data permission requirements, and process of approval within a large academic medical center, and policy constraints related to data sharing.

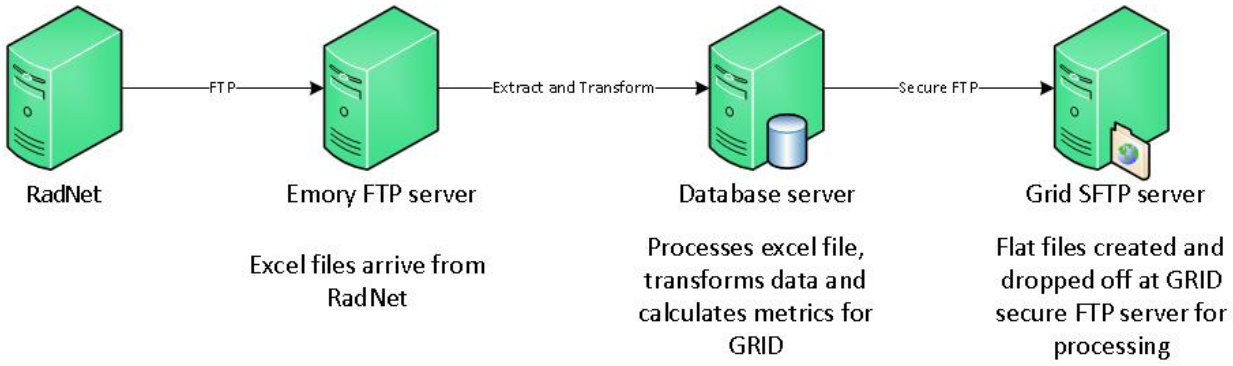
Solutions

Two possible solutions to the data transfer problem were explored. In the first solution, the de-identified data in the Excel file was opened and a second sheet behind the original data was created. A template consisting of formula embedded rows and columns in the second sheet was created to calculate TAT based on the difference between the exam complete date and time and the final date and time. However, in order for this calculation to occur, a person must copy the correctly labeled columns in the first sheet and paste them correctly into the second sheet.

Although this rapid and simple solution worked, it was decided that the potential for human error and the additional steps required did not warrant further development of this solution. Eventually, all of the GRID metrics, not just TAT will be sent from Emory to the ACR, therefore, the programming of potentially dozens of formulae and copying and pasting numerous large columns are not feasible.

The second model leveraged the features and versatility of RadNet and the existing business intelligence tool (52). A preprogrammed SQL query was already being utilized to generate a table of radiology data. Therefore, it was possible to create and run a different SQL query to extract and transform only the data elements that pertained to GRID. The flat file that was generated from that SQL query could then be uploaded to GRID via file transfer protocol (FTP). Due to resource constraints, the Excel file that was already extracted from RadNet was transformed using SSIS within a transform database external to Emory Healthcare. The author collaborated with a subject matter expert (SME), an experienced business intelligence data architect, with over 15 years of experience (53). The author utilized the measures and data dictionary provided by the ACR to communicate GRID requirements to the SME (54, 55). The extract, transform, and load (ETL) process is shown in Figure 2.

Figure 2 Extract, Transform, and Load (ETL) Method. RadNet stores the clinical data. Via file transfer protocol (FTP), Excel files arrive on the Emory server. Data is extracted from the server to the transform database server. The database server processes the excel file, transforms the data and calculates metrics for GRID. In the future, flat files can be created and sent to GRID.



The relevant columns for TAT in the Excel spreadsheet were renamed, and the SME wrote SQL code to calculate TAT in hours, average TAT by modality, and stratification of TAT into less than 12 hours, 12-24 hours, or greater than 24 hours (Table 1). In the transform database, the query was executed on a sample data set, and two separate outputs were obtained (Tables 2-3).

Table 1. SQL code that converted the dates and times of radiology exams listed in the data analyst's Excel spreadsheet to TAT by modality.

```
SELECT

        /* concatenate the exam complete  date and the exam complete time from the spreadsheet to create a new field
        CompleteDatetime */
        DATEADD( mi, DATEPART(mi,ExamCompleteTime),  DATEADD( hh,  DATEPART(hh,ExamCompleteTime) ,
        ExamCompleteDate ) )  as CompleteDateTime  ,

        /* Concatenate the date and time part from the columns FinalTime and FinalDate */
        DATEADD( mi, DATEPART(mi,FinalTime),  DATEADD( hh,  DATEPART(hh,FinalTime) , FinalDate ) )  as
FinalDateTime,

        /* Calcualte the time elapsed between the CompletionDatetime and the FinalDateTime */
```

```
        ROUND ( (DATEDIFF ( mi, DATEADD( mi, DATEPART(mi,ExamCompleteTime),   DATEADD( hh,
        DATEPART(hh,ExamCompleteTime)  , ExamCompleteDate ) ) , DATEADD( mi, DATEPART(mi,FinalTime),   DATEADD(
        hh,  DATEPART(hh,FinalTime)  , FinalDate ) )  )) /60.00, 2 )   as Difference,

        OrderProcedure

        /* Load the data into a temporary table */
        INTO #Temptable

FROM RAD


 /* Now calculate the average turn around time and select the data from the temporary table */

SELECT
        AVG ( difference ) TAT, OrderProcedure
FROM
        #Temptable
GROUP BY
        OrderProcedure
ORDER BY
        OrderProcedure
```

Table 2. The output of sample data set obtained from selected fields of the Excel worksheet. The average TAT in hours is displayed by different order procedures.

Average TAT	OrderProcedure
5.47	CT Abdomen + Pelvis w/ + w/o IV Contrast
37.97	IR Cath PICC Replacement
114.97	IR Cath Plcmnt Tunneled
1.40	MG Diagnostic Mammo Digital Bilat w/ CAD
101.73	MRI Abdomen w/ + w/o Contrast
1.91	NM Brain Scan SPECT
17.71	PET CT Abdomen+ Pelvis w/contrast
3.30	US Abdomen Complete
13.79	XR Abdomen 1 View (KUB)

Table 3. Post Transformation Data Table. This is an example of an output of transformed data categorized by modality and the stratified TAT data required by GRID.

Modality	Month of exam	Year of exam	Month-Year	Modality month year	TAT in Hours	Less than 12 hours	12-24 hours	More than 24
CT	10	2012	201210	CT201210	46.65	0	0	1
CT	10	2012	201210	CT201210	3.53	1	0	0
CT	11	2012	201211	CT201211	6.17	1	0	0

This solution is designed so that a flat file can be transferred to GRID in the future via secure file transfer protocol (SFTP). Since GRID collects data on a monthly basis, monthly

batch push on Emory's side could be performed for data transfer. Conversion from extensible markup language (XML) to HL7 could also be accomplished easily should the ACR adopt HL7 for GRID in the future. On the GRID receiving end, there are data cleaning features which reject inconsistent data such as biopsy reports that are non-diagnostic. In terms of TAT, if the "number of exams completed this month" = 10, then the subsets (number of reports signed < 12 hours later, number of reports signed between 12 and 24 hours, and number of reports signed greater than 24 hours) must be equal to or less than 10. If a record has a duplicate key, for example, a year or year/month, it overwrites the previous record with that key. If duplicate data is accidentally submitted on two different records with different keys, then both records will be added to the database assuming both keys and data are otherwise valid (56). In creating data sharing solutions, understanding the capabilities, limitations, and requirements on both sides is essential.

Chapter 4: Discussion

As the nation trends towards greater EMR adoption, data within EMRs will become more prevalent. This valuable data source will be increasingly desired for utilization by public health agencies. The ability to aggregate and analyze EMR data will improve especially in today's landscape as big data advanced analytics (BDAA) plays a greater role in public health. New technologies such as BDAA can be leveraged to analyze both in real time and retrospectively the patient exhaust data in EMRs in order to discover correlations or causations not previously recognized. Therefore, data exchange between clinical healthcare and public health is vital to recognizing associated factors and behaviors related to disease as well as improving outcomes.

Our present landscape of disparate EMRs and health systems challenges us from multiple standpoints including semantics, vocabulary, local codes, and standards.

If a healthcare facility wishes to exchange data, then exploration of all of the potential or existing data sources must be undertaken in order to extract the desired data. Once the data is extracted, one must determine if it is in the appropriate format and transformation is required. An HL7 listener that can automatically extract the pertinent data could be created to streamline this process, but this is not always possible (57). As discussed in Chapter 2, there are numerous methods of data extraction and transfer from an EMR system to a public health informatics platform or registry: real time bidirectional HIE, point to point data transfer via HL7, faxing papers, manual entry of paper forms, creation of external database as intermediate or staging database, scanning documents, batched flat files, open source software, creation of a local registry to communicate with national one, data entry via Epi Info, and manual extraction with SQL. Emory's EMR, for example, has an automated data feed to GRITS, and the CDW has a unidirectional automated data feed to Syndromic Surveillance at the Centers for Disease Control and Prevention (CDC) (58). EMR data extraction for public health and research use can be challenging, and the aforementioned models presented demonstrate creative problem solving methods as well as a spectrum of obstacles that required tackling. A common theme in all of these scenarios was that each institution or groups had to deal with their unique problem of interoperability and variations in data exchange standards. Until standards become more uniform in the future, interoperability and exchange will require customizing and tailoring to the institutional needs, capabilities, policies, and procedures. Data linkage, investment in technological infrastructure, and collaboration between public and private sectors are all needed in order to improve health services research (49). In the future, researchers may want to extract

and link data within a certain EMR component, and linking EMR data between health care entities is challenging in today's world. Moving forward, establishing and developing data standards, removing barriers such as inability of states to link CMS data with public health data, and collaboration will improve the public health informatics landscape (49). Most likely, data sharing will become more seamless in the future. Another important consideration related to data sharing is the recognition that the method of data sharing can have direct impact on the timeliness and completeness of data in the receiving platform and may affect representativeness and other attributes in an unknown fashion. For example, batched flat files have been shown to be more complete than manually entered or HL7 feeds (48). Data extraction also can affect the efficiency and speed of the EMR, and this is an important factor when considering automated feeds. Additionally, the set up and maintenance of automated feeds are not "free" but in fact require time, money, and resources. These factors must be accounted for when an institution is deciding whether or how to participate in a registry(58).

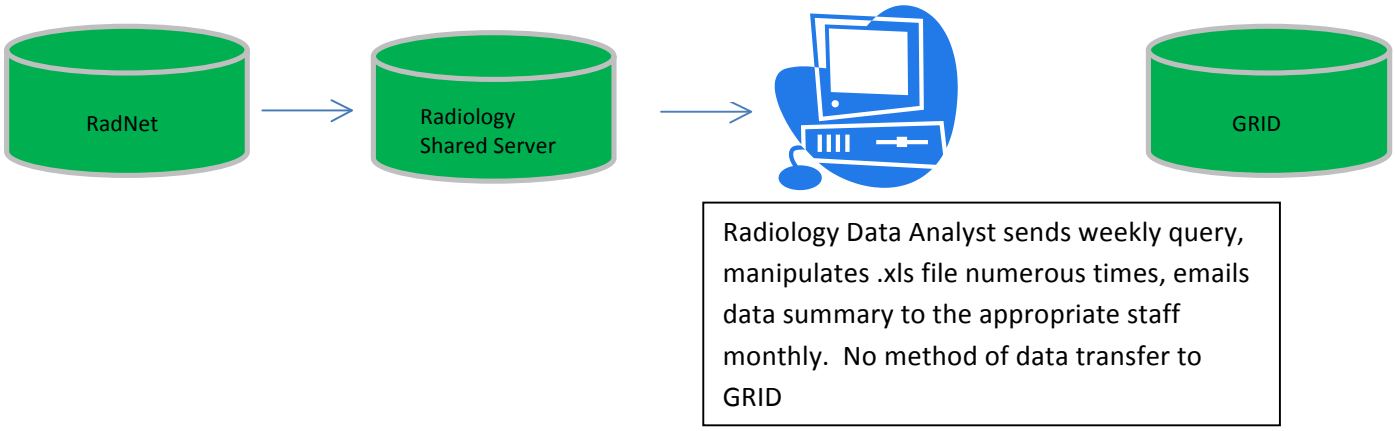
ARDIT is the automated method of radiology data transfer from Emory Healthcare to the General Radiology Improvement Database (GRID) registry at the American College of Radiology (ACR). ARDIT allows Emory Healthcare's participation in GRID without requiring numerous, manual, monthly web-based data field entries. Figure 3 demonstrates both Emory's current process and the proposed solution. The current process does not connect to GRID, and the proposed solution does connect to GRID and allows more versatile data transfer to other platforms and/or agencies should Emory desire that. Collection of clinical data from EMRs such as Emory's to registries and surveillance systems will improve the representativeness of the system and has widespread potential in establishing more accurate national benchmarks, guiding practice and policy decisions, and improving clinical outcomes and patient safety. As

demonstrated in the VA CCR study, a system may prove useful beyond its scope. This phenomenon could also become true for GRID. For example, as participating facilities increase and GRID becomes more robust, its data may be valuable and desirable to other public health informatics platforms such as the Health Center Patient Satisfaction Survey administered by the Health Resources and Services Administration (HRSA) and the U.S. Department of Health and Human Services (HHS)(59).

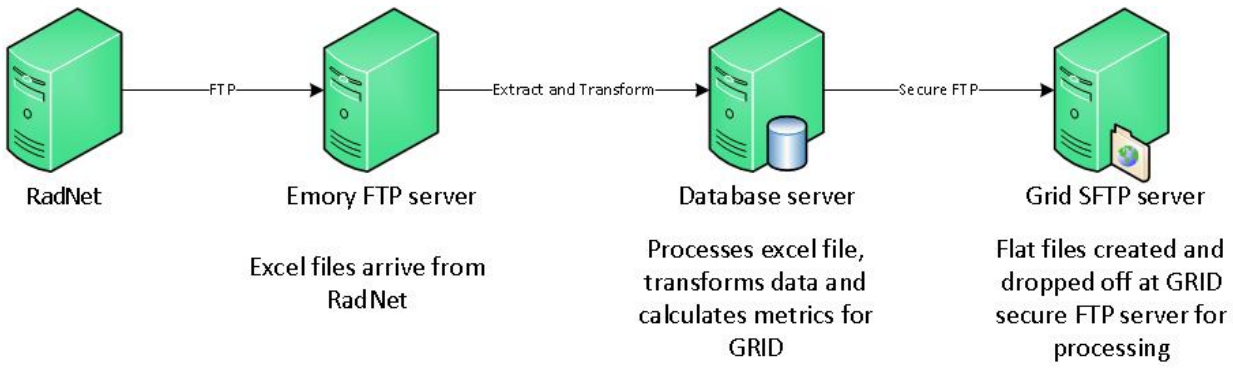
Workflow analysis of Emory's current data extraction procedure was performed to determine the least obtrusive method of data extraction. De-identified TAT data in hours by modality was chosen as the metric of interest because it was a simple yet important metric. Data transformation via creation of a transform database was performed. This model lays the groundwork for future metrics to be used for submission to GRID for Emory and also for outside institutions with Cerner products. ARDIT lives in development mode on the Emory server, and when resources are available in the future, the database can be expanded and formalized into Emory's infrastructure. Additional metrics can be added and sent. Lessons learned from this thesis can also be helpful for the ACR who may have inquiries from other healthcare facilities related to data submission.

Figure 3 Existing Process versus Proposed Solution. The existing process does not communicate with GRID and requires manual manipulation of Excel (.xls) files by the Data Analyst for sharing amongst department members. The proposed solution removes the manual manipulation of the .xls files and transforms and has the functional capacity to transfer the TAT data to GRID. Data from the transform database can also be sent to other agencies if so desired by Emory.

Existing Process:



Proposed Solution:



The barriers that were encountered during the development process include the time and effort in learning the Emory infrastructure (who, what, where, when) in terms of the data flow process, workflow process, and stakeholders. Data permissions and IRB exemption were temporary administrative obstacles that were fairly easy to overcome. However, institutional policy regarding alteration of a current information technology process and additional IT resources required paperwork and approval which is still pending at the writing of this thesis. It may take several months or longer to obtain approval for the official use of the model SQL query and eventual use of the transform database. Automated data feeds require resources to design, implement, and maintain and thus require time and money. Additionally, most institutions would want to review their raw data before automatically sending it to a registry or other destination. The reason for this is to correct missing data, outliers, or typographical/nonsense errors and to assure data quality. Thus, automated data feeds are not 100% automated since a human needs to oversee or review the data prior to submission (58). Although protective and important, in general, policy structure is constrictive and places even greater restraint on the methodology of integrating systems. This is true for any organization, and IT professionals employed by that organization will need to manage ETL procedures that touch databases with ePHI (58). Time and procedure kept the scope to ARDIT narrowed to an external database transformation which was then incorporated to the Emory server to function in development mode. Equally important is determining how to advocate data sharing in order to mitigate risk and extend the business model for years to come.

This study had several limitations. First, the focus was narrow as the ARDIT model was customized as an exact fit for a single institution. The solution is not in of itself translatable to another institution although facilities with Cerner Millennium may be able to implement a

similar if not identical solution. Another limitation of the study is that only a single metric was utilized, however, this is not unusual for a pilot study. Future work is recommended to expand the current development model so that more metrics are transformed. Eventually, the entire submission can be uploaded as a correctly formatted flat file.

Lastly, the technology for data and information exchange is readily available, but the obstacle preventing rapid data sharing or sharing at all, it is the human element. Humans, not technology, are the barriers in our current struggle for interoperability and exchange, and this is due to a number of reasons: finance/cost, resources, security, policies, regulation, legal and ethical issues. Other industries are capable of moving massive amounts of data in real time on a daily basis, and this is evident in the financial sector. Because healthcare data is personal and has potential for social stigmatization, careful handling cannot be understated. Protecting PHI is constraining from a legal standpoint and affects to a certain extent what we can do and how fast we can do it. Protecting PHI and interoperability are like risk and benefit; both must be weighed. We also should recognize that sometimes what we believe is a technology problem is actually a human problem (57).

Nonetheless, it is critical to continue to develop solutions for data sharing between EMRs and registries, particularly in the field of diagnostic radiology where there are no benchmarks or standard industry metrics for TAT, outcomes, or accuracy of interpretations, for example. Thus, GRID's value lies in the development of these much needed metrics for clinical quality and patient safety and assuring that these metrics are accessible, useful, and actionable for practitioners and healthcare facilities. The usefulness and effectiveness of GRID will increase as the number of participating facilities increase, and increased participation could allow GRID to expand as a research tool for radiology outcomes.

Future recommendations would be to develop the transform database for ARDIT in the Emory environment so that test data could be sent to GRID via secure FTP. Additional inquiry is required to determine if ARDIT development into a permanent semi-automated datafeed at Emory is feasible and/or desirable. This decision would be determined at the departmental and institutional levels and project prioritization, IT resources, potential grant money and funding would all play critical roles in determining the future of ARDIT.

Chapter 5 – Journal Article

References

1 Institute of Medicine (U.S.). Committee on Quality of Health Care in America. To Err is Human: Building a Safer Health System. Washington DC: National Academy Press; 2000. p. 9-99.
2 Wiener EL, Nagel DC. Human factors in aviation. San Diego: Academic Press; 1988.
3 Stoelting RK. About APSF: foundation history. 2012 [cited; Available from: www.apsf.org/about_history.php
4 Itri JN, Krishnaraj A. Do we need a national incident reporting system for medical imaging? J Am Coll Radiol. 2012 May;**9**(5):329-35.
5 Larsson S, Lawyer P, Garellick G, Lindahl B, Lundstrom M. Use of 13 disease registries in 5 countries demonstrates the potential to use outcome data to improve health care's value. Health Aff (Millwood). 2012 Jan;**31**(1):220-7.
6 Institute of Medicine (U.S.). Committee on Quality of Health Care in America. Crossing the quality chasm : a new health system for the 21st century. Washington, D.C.: National Academy Press; 2001.
7 Jones DN, Benveniste KA, Schultz TJ, Mandel CJ, Runciman WB. Establishing national medical imaging incident reporting systems: issues and challenges. J Am Coll Radiol. 2010 Aug;**7**(8):582-92.
8 ACR. 2013 [cited; Available from: http://www.acr.org/Quality-Safety/National-Radiology-Data-Registry/General-Radiology-Improvement-DB
9 Schauer DA, Linton OW. National Council on Radiation Protection and Measurements report shows substantial medical exposure increase. Radiology. 2009 Nov;**253**(2):293-6.
10 Patti JA. The national radiology data registry: a necessary component of quality health care. J Am Coll Radiol. 2011 Jul;**8**(7):453.
11 Weinreb J. NRDR User Group Conference. Annual Imaging Informatics Summit and NRDR User Group Conference; 2012 October 11-12; Washington, DC; 2012.
12 Chatfield M. Annual Imaging Informatics Summit and NRDR User Group Conference. 2012 October 11-12; Washington, DC; 2012.
13 ACR. Welcome to Triad OA. [cited; Available from: https://triad.acr.org/index.htm
14 American College of Cardiology Foundation. PINNACLE Registry 2013 [cited 2013; Available from: http://www.cardiosource.org/News-Media/Publications/Cardiology-Magazine/PINNACLE-Registry.aspx
15 Fig MD. Fig MD. 2013.
16 Lough J. Annaul Imaging Informatics Summit and NRDR User Group Conference; 2012 October 11-12; Washington, DC; 2012.
17 American College of Radiology. Annual Imaging Informatics Summit and NRDR User Group Conference,. Annual Imaging Informatics Summit and NRDR User Group Conference; 2012; Washington DC; 2012.
18 Calman N, Hauser D, Lurio J, Wu WY, Pichardo M. Strengthening public health and primary care collaboration through electronic health records. American journal of public health. 2012 Nov;**102**(11):e13-8.
19 Hernandez MN, Voti L, Feldman JD, et al. Cancer Registry Enrichment via Linkage with Hospital-Based Electronic Medical Records: A Pilot Investigation. Journal of registry management. 2013 Spring;**40**(1):40-7.
20 Prokosch HU, Ganslandt T. Perspectives for medical informatics. Reusing the electronic medical record for clinical research. Methods of information in medicine. 2009;**48**(1):38-44.

21 Yamamoto K, Sumi E, Yamazaki T, et al. A pragmatic method for electronic medical record-based observational studies: developing an electronic medical records retrieval system for clinical research. BMJ Open. 2012;**2**(6).
22 Baus A, Hendryx M, Pollard C. Identifying patients with hypertension: a case for auditing electronic health record data. Perspectives in health information management / AHIMA, American Health Information Management Association. 2012;**9**:1e.
23 McDonald CJ. The barriers to electronic medical record systems and how to overcome them. Journal of the American Medical Informatics Association : JAMIA. 1997 May-Jun;**4**(3):213-21.
24 Terry AL, Chevendra V, Thind A, Stewart M, Marshall JN, Cejic S. Using your electronic medical record for research: a primer for avoiding pitfalls. Fam Pract. 2010 Feb;**27**(1):121-6.
25 Kristianson KJ, Ljunggren H, Gustafsson LL. Data extraction from a semi-structured electronic medical record system for outpatients: a model to facilitate the access and use of data for quality control and research. Health Informatics J. 2009 Dec;**15**(4):305-19.
26 Kanas G, Morimoto L, Mowat F, O'Malley C, Fryzek J, Nordyke R. Use of electronic medical records in oncology outcomes research. ClinicoEconomics and outcomes research : CEOR. 2010;**2**:1-14.
27 Treweek S. The potential of electronic medical record systems to support quality improvement work and research in Norwegian general practice. BMC health services research. 2003 Jun 6;**3**(1):10.
28 Chatfield M. Telephone Conversation ed; 2013.
29 Peng C, Kesarinath G, Brinks T, Young J, Groves D. Assuring the privacy and security of transmitting sensitive electronic health information. AMIA Annual Symposium proceedings / AMIA Symposium AMIA Symposium. 2009;**2009**:516-20.
30 Magnus M, Herwehe J, Gruber D, et al. Improved HIV-related outcomes associated with implementation of a novel public health information exchange. Int J Med Inform. 2012 Oct;**81**(10):e30-8.
31 Balog J. A Three-Step Approach for Creating Successful Electronic Immunization Record Exchanges between Clinical Practice and Public Health. Online journal of public health informatics. 2012;**4**(3).
32 Herwehe J, Wilbright W, Abrams A, et al. Implementation of an innovative, integrated electronic medical record (EMR) and public health information exchange for HIV/AIDS. Journal of the American Medical Informatics Association : JAMIA. 2012 May-Jun;**19**(3):448-52.
33 Lin G, Ma J, Zhang L, Qu M. Linking cancer registry and hospital discharge data for treatment surveillance. Health informatics journal. 2013;**19**(2):127-36.
34 Stevens LA, Palma JP, Pandher KK, Longhurst CA. Immunization registries in the EMR Era. Online journal of public health informatics. 2013;**5**(2):211.
35 Kairys SW, Gubernick RS, Millican A, Adams WG. Using a registry to improve immunization delivery. Pediatric annals. 2006 Jul;**35**(7):500-6.
36 Downs SM, Anand V, Sheley M, Grannis SJ. The Last Mile: Using Fax Machines to Exchange Data between Clinicians and Public Health. Online J Public Health Inform. 2011;**3**(3).
37 OpenMRS. 2013 [cited; Available from: http://openmrs.org/
38 Health Level Seven International. Arden Syntax. 2013 [cited 2013 8/13/13]; Available from: http://www.hl7.org/Special/Committees/arden/index.cfm
39 Barnes M. Lessons learned from the implementation of clinical messaging systems. AMIA Annual Symposium proceedings / AMIA Symposium AMIA Symposium. 2007:36-40.
40 Backus LI, Gavrilov S, Loomis TP, et al. Clinical Case Registries: simultaneous local and national disease registries for population quality management. Journal of the American Medical Informatics Association : JAMIA. 2009 Nov-Dec;**16**(6):775-83.
41 Mehmood A, Razzak JA, Kabir S, Mackenzie EJ, Hyder AA. Development and pilot implementation of a locally developed Trauma Registry: lessons learnt in a low-income country. BMC emergency medicine. 2013;**13**:4.

42 Schultz CR, Ford HR, Cassidy LD, et al. Development of a hospital-based trauma registry in Haiti: an approach for improving injury surveillance in developing and resource-poor settings. The Journal of trauma. 2007 Nov;**63**(5):1143-54.
43 Nwomeh BC, Lowell W, Kable R, Haley K, Ameh EA. History and development of trauma registry: lessons from developed to developing countries. World J Emerg Surg. 2006;**1**:32.
44 Kobusingye OC, Lett RR. Hospital-based trauma registries in Uganda. The Journal of trauma. 2000 Mar;**48**(3):498-502.
45 Klompas M, Murphy M, Lankiewicz J, et al. Harnessing electronic health records for public health surveillance. Online journal of public health informatics. 2011;**3**(3).
46 Klompas M, McVetta J, Lazarus R, et al. Integrating clinical practice and public health surveillance using electronic medical record systems. American journal of public health. 2012 Jun;**102 Suppl 3**:S325-32.
47 Elliott AF, Davidson A, Lum F, et al. Use of electronic health records and administrative data for public health surveillance of eye health and vision-related conditions in the United States. American journal of ophthalmology. 2012 Dec;**154**(6 Suppl):S63-70.
48 Hills RA, Revere D, Altamore R, Abernethy NF, Lober WB. Timeliness and data element completeness of immunization data in Washington State in 2010: a comparison of data exchange methods. AMIA Annual Symposium proceedings / AMIA Symposium AMIA Symposium. 2012;**2012**:340-9.
49 Bradley CJ, Penberthy L, Devers KJ, Holden DJ. Health services research and data linkages: issues, methods, and directions for the future. Health services research. 2010 Oct;**45**(5 Pt 2):1468-88.
50 Douglas B. Emory Healthcare Clinical Data Warehouse. Atlanta, GA: Emory Quality Academy; 2012.
51 Knighton J, Mulvey T. 2012.
52 Office of Information Technology EU. 2013 [cited 2013; Available from: http://it.emory.edu/catalog/ehc_clinical_data_warehouse/
53 Sebastian V. Collaboration ed; 2013.
54 American College of Radiology. General Radiology Improvement Database (GRID): Measures. Reston, VA; 2009.
55 American College of Radiology. General Radiology Improvement Database (GRID): Data Dictionary. Reston, VA; 2010.
56 Meyer L. Email communication; 2013.
57 Digital Health Summit. Digital Health Summit 2013 - Man + Machine: Big Strides in Big Data. 2013 1/18/13.
58 Chernetsky Tejedor S. Personal Communication. 2013.
59 U.S. Department of Health and Human Services. Health Center Patient Satisfaction Survey. 2013 [cited 2013 September 23]; Available from: