Health Care Information Exchange: Using EDpCloud™ to Automatically and Securely Synchronize Data between Heterogeneous Systems, Remote Geographic Sites, *Private, Public and Hybrid Clouds* (July 2016)

Abderrahman Aba El Haddi, *Senior Member, IEEE (elhaddi@ieee.org)*

***Abstract*—one of the challenges faced by health care providers and payers is the secure sharing and exchange of protected health information (PHI) using effective Health Information Technology (HIT) and Health Information Exchange (HIE) solutions. This paper describes a case study where EDpCloud was used to solve HIE as well as workflow and data synchronization.**

***Index Terms*—Protected Health Information, PHI, Health Information Exchange, HIE, Accountable Care Organizations, ACO, Electronic Health Records, EHR, Health Information Technology, HIT.**

I. INTRODUCTION

THE goal of healthcare providers and payers is to provide better patient outcomes, to increase patient safety, to lower risks and costs while complying with a myriad of necessary regulations to protect patients and taxpayers.

To ensure quality and continuity of care and to lower risks and costs, secure exchange of Protected Health Information (PHI) is critical. Such information includes a long list of data types and meta data that must be exchanged only between authorized people and systems. The exchange is governed by privacy laws, legal frameworks and many regulations [5]. Health Information Exchange(HIE) allows all providers and patients to access the same critical patient medical information required to deliver quality care safely, rapidly when needed [2][5]. Furthermore, HIEs promise to:

- Reduce costs to patients, payers, providers and society
- Reduce risks to patients and providers
- Reduce readmissions
- Reduce errors
- Improve outcomes
- Reduce duplication of labs, x-rays and other testing.

In a town hall event held in November 2015, Gray Plant Mooty (GPM) solicited community input on challenges and barriers to HIE in Minnesota. 95% of participants experienced some problems with sharing PHI and 84% indicated that some Minnesota laws impede the exchange of PHI for patient treatment [1].

72% of Minnesota's hospitals that are not using EPIC Electronic Health Records (EHR) report that their providers do not have necessary clinical information available electronically from outside providers. [6].

A series of interviews of doctors, nurses and health IT staff by EnduraData showed that several factors impede the sharing of PHI. These factors include the need for multiple consents, heterogeneous systems, and vendor use of lock-in as a barrier to entry, non-interoperability, laws and regulations and the availability of resources for the providers and profit goals [3]. Furthermore, many providers have opted for the use of the traditional isolation model such as firewalls, network access control to protect data from emerging threats [5].

The difficulties of exchanging data have several implications for the quality of care, its continuity, and its costs and for patient health outcomes. Many states such as Minnesota, Texas, and Utah have either mandated interoperable EHRs or have issued directives to solve the interoperability problems. A minimum set of requirements and economic incentives have emerged in some states and at the federal level [2]. Health IT, NIH, Human services, CMS are all pushing toward interoperability. Yet, you can visit a hospital for a procedure and the hospital staff may have no way of exchanging information electronically with the doctors who may be providing the services for a procedure but belong to a different organization (Even if they are located within the same facility within a few meters away from each other). These providers cannot exchange information which may save the patient's life and reduce costs.

To illustrate the severity of the problem, imagine the following real scenario:

a. A primary care physician(PCP) refers a patient to a specialty clinic
b. The primary care doctor has no way of exchanging data with the specialty clinic
c. The primary clinic refers the patient to a specialty clinic for a procedure in a hospital that has a joint venture with the specialty clinic. The specialist changed the patient's prescription in the specialty clinic's EHR (before the procedure)
d. The med list for the same patient in the specialty clinic and in the PCP's EHR have now diverged
e. The hospital cannot get the data electronically from the clinics
f. The hospital now has three different med lists for the same patient: two in hard copy and one in the hospital's EHR.

If we add other care givers to the previous workflow list, the risks, costs and duplicate procedures go up exponentially. The potential for errors and for data leaks increases dramatically as well.

II. Data Types and the Need to Know

PHI needs to be exchanged between payers and providers on a regular and need to know basis while complying with various regulations and procedures. Providers may belong to the same organization, but most often they belong to multiple distinct health care and non-health care organizations as illustrated by the previous example.

Figures 1 and 2 illustrate health information exchanges that may need to occur to ensure quality patient care and an economically functioning health system and business model [1], [6].

Employers and payers need to exchange employee health plan enrollment data that is very sensitive from the employee's privacy perspective but also from a competitive, human capital, acquisitions and business processes perspective for both the employer and for the payer. Automating and securing the exchanges will protect all parties, will reduce costs and risks and will also protect the competitive advantages of the payer and of the employer.

Figure 1 shows the importance of exchanging information between other care providers where patients may not be able to provide their medical history. This category of providers may include nursing homes, mental health care institutions, outpatient clinics or other care centers. For these reasons, regulations exist to organize health care intermediaries and healthcare information organizations (HIO) to facilitate such exchanges to protect patients. This is why providers recommend their patients carry their medical lists with them.

A growing number of states are implementing Accountable Care Organization (ACO) models in their medical assistance programs to allow for health organizations to collaborate to raise the quality of care, to reduce costs and to allow the collaborating organizations to share in the savings, the costs as well as the risks. Furthermore, the pay for performance makes it more imperative for ACOs to be able to exchange data in a secure fashion and in a manner that protects patients' information and privacy, yet contribute to improving care for him using a data driven approach. Data must be moved in an automated and secure fashion to reduce costs, risks and errors [3] [4]. Minnesota Health Department reported that providers are struggling to share data [1] [6]. Table 1 shows the difficulty the providers are having with sharing data as of this writing.

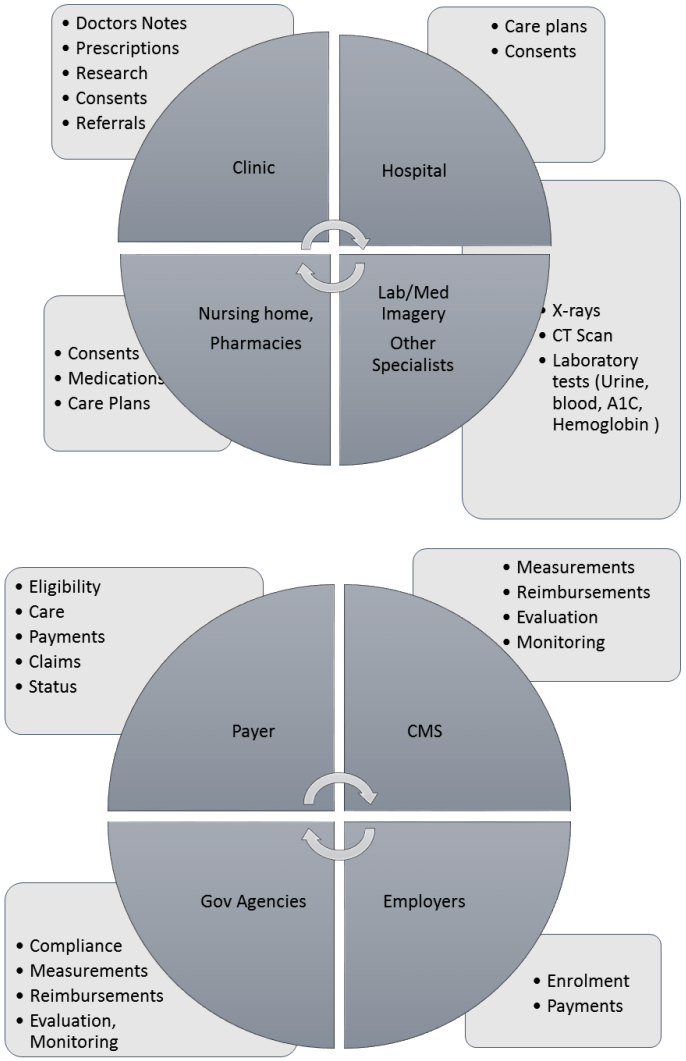

Figure 1. HIE Between many health care organizations. Adopted from GPM [1]

Table 1: Difficulty of Interoperability & Exchanging data (Source: 2016 eHealth profile, MDH, Health IT, MN, USA)

Facilities	Exchanging with unaffiliated providers	Routinely sending summary of care	Integrating summary of care into EHR
Hospitals	72%	74%	19%
Clinics	69%	40%	12%

There are many reasons for these data exchange issues. Chief among them are the laws that govern data practices, vendor lock-in and interoperability issues. The impediments to information exchange impacts both delivery systems and patient care.

Last week, I had a physical at 8:30 am but my doctor was delayed by over 45 minutes. He explained that he waited for another provider to send him an elderly patient's health records. The patient needed immediate attention because he was allergic to some medication (Neither the man nor his wife could remember) and the doctor needed the medical records in order to prescribe pain medication. That "prescription" could also put the patient at risk if he is allergic to it. Not having the records could be fatal but having them increased costs that propagate throughout the system. Fast forward in my journey, a medical technician walks into the room with an EKG machine, takes some measurements and prints them. I asked "What now?" He said he would scan and import them to the EHR. This not only increases the cost but also the risk of error and highlights the need for a glue that will ingest the data automatically.

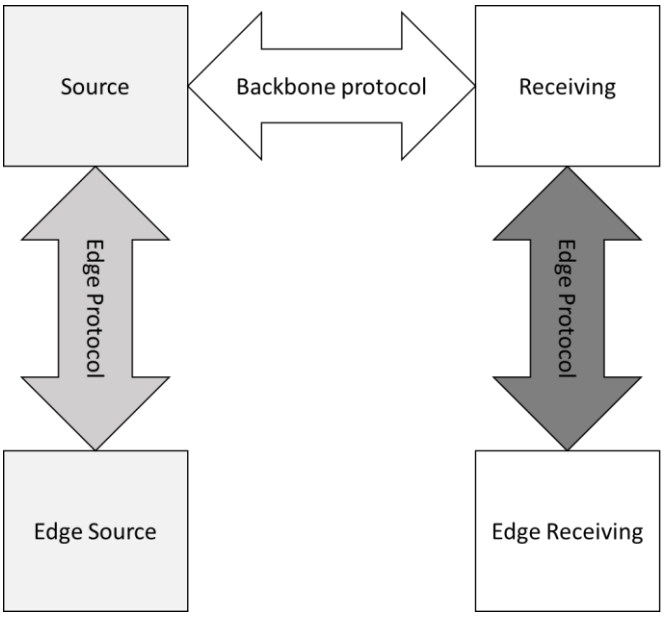

Figure 2: Health Information Exchange between and within two health care entities.

Even with the aforementioned difficulties and struggles, according to Minnesota Department of Health (MDH), hospitals and clinics are using data from their EHRS to make effective health care decisions and to improve care quality. Many of the providers use data repositories and EHRs to improve quality of care, identify high-risk patients and maintain a chronic disease registry. Hence it is critical that the data is available to be leveraged for effective patient care and for improving the health of the population and of members of society.

III. EDpCloud as a Solution

In this section, we will discuss how a large healthcare payer and provider (HPP) with over 15 billion dollars in revenues is leveraging EDpCloud to synchronize data in real time [3], to update and move millions of files each month between systems, workflow processes and cities. The HPP needed to synchronize files from various providers, employers' enrollments, dentists, billing and more. EnduraData helped the HPP synchronize data in real time between Linux servers in multiple cities and integrated with the healthcare payer and providers' operations and processes.

The systems in this case were running Red Hat Linux (However, EDpCloud is also available for Windows, Mac, Solaris and other UNIX Operating Systems on both physical and virtual machines). The environment is very critical for HPP's operations and could not be down for maintenance except for a very brief window of time in the early hours on Sunday mornings once every few months. Half of the systems must be available to ingest incoming data and to queue it automatically for secure transmission and for automatic replication between systems in different sites making it all the more challenging. In addition, our HPP customer needed to replicate the root file systems ("/" to "/") under a Red Hat Linux server farm. Load balancers directed the various users to the available servers. Hence, if a customers' data was not available, operational tasks would fail. HPP prides itself in effective operation, hence the bar was set very high and EnduraData had to perform.

EnduraData's HPP customer is a sophisticated payer and provider, has an infrastructure that is reliable and extensive, has highly skilled and well organized operations and technical staff and has a mission to reduce pain and suffering and to serve its patients and customers. HPP's stakeholders' needs are critical and demanding. EDpCloud was up to the challenge and was installed to retire another competing product. Untangling the first part was very complex and risky. An extensive amount of time was spent on the planning stage and on the verification of the order in which servers were upgraded to the new EnduraData EDpCloud Software.

A. Configurations

The first step was to create a simulation lab inside EnduraData's private and public clouds. Multiple cloud vendors were also used to provide a test bed for EnduraData's implementation team. Nodes were setup in New York, San Francisco, Chicago, Vancouver, Singapore, United Kingdom and Germany to include the worst-case latencies and latencies similar to the customer's environment. The test bed was used to model the customer's environment (except for the latency where worst and best cases were used). Multiple factorial experiments (data types x meta data x file numbers x file sizes x compression rates x I/O streams x I/O types …) were setup to run 24/7 for weeks. EnduraData massive I/O (mio) generation and measurement tool was used to create a mix of operations and dummy data files (which were comparable to HPP's data sets) in order to find the optimal number of communication channels between the systems to take advantage of parallel I/O, and to reduce data synchronization time and increase throughput. No data from HPP was used in this test process at all.

B. What was replicated?

The customer needed to replicate over 30 million files each month. File sizes ranged from a few kilobytes to over 32 GB each for orders, notes, X-rays and over 90 GB for some kernel dumps, virtual machine backups, regular backups and dumps.

The following operations on files where replicated in a bidirectional fashion as they took place in real time:

- Creates
- Writes
- Truncates
- Renames of files and directories
- Deletes
- Symbolic links
- ACLs
- File ownership and group changes
- Permissions changes
- Etc.

One of the biggest challenges in bidirectional file replication is the rename operation across multiple systems. Even on the same local system, this operation can be very expensive for large directories and large files, even when the source and the new name are on different file systems on the same host let alone be on remote systems located hundreds of miles away from each other. Nevertheless EnduraData's technology achieved faster renames (across cities) than renames across file systems on the same localhost.

C. Dealing with bidirectional replication risks using includes and excludes

Bidirectional replication can be risky, but is a necessary operation to mirror file systems. It must be approached with care. If the software or the system administrator is not sophisticated perils may follow. Imagine you replicated "/etc" from one machine to another, or "/boot" to "/boot".

Therefore, it is vital that you pay attention to your include and exclude patterns and to the type of software used. EDpCloud allows you to define inclusion and exclusion patterns. These are regular expression patterns that tell EDpCloud to replicate something if it is in the specified include list and never replicate it if it matches the pattern in the excludes list. Includes and excludes patterns are used to further refine replication policies.

D. Compressing data only when it lends itself to compression

EDpCloud uses adaptive compression. Hence, when files are compressible, EDpCloud compresses them and reduces the bandwidth used during synchronization. X-rays and other medical images do not compress very well. EDpCloud learns that over time and applies that knowledge to adapt its compression levels.

E. Protecting data in transit and at rest with encryption

To ensure data protection, all communications are encrypted using AES 128 by default. Other encryption methods may also be configured. System administrators may also configure EDpCloud to leave files encrypted at rest or to be decrypted if the key is available on the remote system.

F. Access Control

To ensure data privacy and security, each sender and each receiver define the replication policies. Only authorized and authenticated senders can send data or receive data. Multiple access control layers are built into EDpCloud.

G. Extensive transaction history for audits of data changes

Extensive history allows the sending and the receiving parties to keep track of which files were modified and sent, which ones were received, when and how much data was sent and where the files originated from or delivered to.

H. Real time, scheduled or on demand

EDpCloud on Windows and Linux can be configured to operate in real time, on demand or scheduled modes or in a combination of these modes. In real time mode, each time a file changes, the changes are propagated to one or more remote locations allowing the other partners to receive data continuously as it is created or changed. All other platforms can send data on demand or using a schedule. In all modes, only the portions of files changed are sent (deltas). This reduces the amount of bandwidth used, increases the transfer speeds and

reduces windows of vulnerabilities to data loss in case of a disaster or human errors.

I. No proprietary file formats, no need for restore

EDpCloud does not use any proprietary file formats. Files are readable by any other applications immediately. There is no need to restore the data to use it (But you may also restore it to one or more remote machines if desired), allowing data to be connected and ingested to other systems and applications.

J. Post and preprocessing hooks

EDpCloud can be configured to trigger scripts or other applications to further transform data when received or before sending it. This allows you to create a powerful middleware or glue between your data and other applications (databases, reporting, analytics, etc.) and workflows.

K. How does this work?

EDpCloud can be configured to move data internally between systems or externally between networks, systems and sites. EDpCloud watches a file system for file changes and then examines the policies to find out where data needs to be delivered. Figure 3 is an example of a configuration where an internal node synchronizes to another internal node that is allowed to communicate and synchronize data to another partner. In case of network failures, EDpCloud continues to journal file changes and continues to resend file changes to the remote systems until it succeeds, or a configurable threshold of retries has been reached.

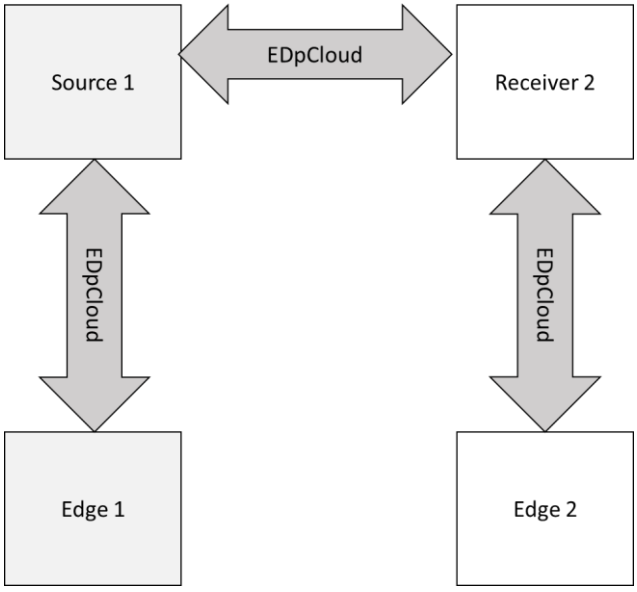

Figure 3: An inter-connect to move data between edge and a source and between the source a receiver.

Figure 4 depicts a complete interconnect where data can be replicated and synchronized between internal edges and external sources. This is suitable for joint ventures, partnerships. In all cases the sender has full control over what to send and when to send it and the receiver has full control on who can send to it, what to receive, etc. In this case, any data that changes on edge1 can be changed on source 1 and on receiver2 as well as on edge2.

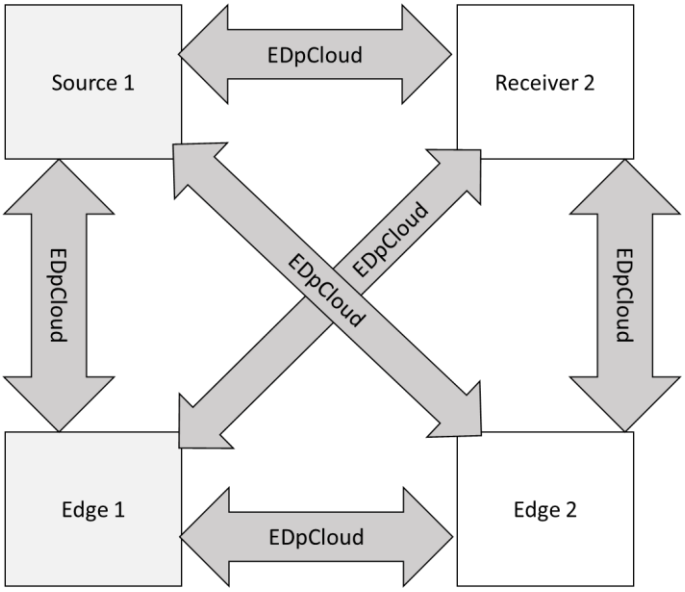

Figure 4: A completely interconnected mesh where file changes in one node are also propagated to other nodes.

IV. CONCLUSION

Helathcare providers and payers rely heavily on data for decision making. Patient care is ensured by multiple providers that usually require the exchange of health information. EDpCloud is being used to exchange healthcare information in real time, securely and automatically which in turn reduces costs, risks and errors and improves quality of care.

A. A El Haddi is the architect of EDpCloud. He is the CTO and founder of EnduraData. El Haddi holds a BS/DAG, an MS/DE in AEng, (from IAV and the University of Minnesota), an MS in computer Science from the University of Minnesota and an executive MBA from the University of St Thomas. More information about El Haddi can be found in his google scholar and linkedin profiles: (https://www.linkedin.com/in/aelhaddi, http://scholar.google.com/citations?user=NWfII8sAAAAJ&hl=en)

References:

[1] J. Berg, T Johnson, J. Marotte Gray Plant Mooty (2016) "Barriers to Sharing Health Information in Minnesota. State Initiatives to Manage Them." Minnesota 12th Annual e-Health Summit, June 7th 2016.
[2]"Health Information Exchange (HIE) Healthit.gov/providers-professionals/health-information-exchange/wht-hie
[3] A. A. El Haddi (2016), EDpCloud cross platform file synchronization and replication: www.enduradata.com
[4] S. Lins, P. Grochol, S Schneider and A. Sunyaev (2016) "Dynamic Certification of Cloud Services". IEEE Computing Edge, June 2014, p 14-19.
[5] D. Kotz, C. Gunter, S. Kumar and J Weiner (2016) "Privacy and Security in Mobile Health: A Research Agenda", IEEE Computer, June 2016:p22-29.
[6] Minnesota Department of Health, Office of Health Information Technology, Minnesota e-Health Profile, 2016, http://www.health.state.mn.us/e-health/assessment.html

Your feedback is important for us, please email,
Call 952-746-4160 or visit www.enduradata.com