INTERNATIONAL JOURNAL OF MEDICAL INFORMATICS 81 (2012) 821–827

journal homepage: www.ijmijournal.com

# Where should electronic records for patients be stored?

*Vijay Lapsia*[a,*], *Kenneth Lamb*[b], *William A. Yasnoff*[c,d,e]

[a] *Department of Medicine, Mount Sinai School of Medicine, New York, United States*
[b] *The Chronic Disease Research Group, Minneapolis, MN, United States*
[c] *NHII Advisors, Arlington, VA, United States*
[d] *Division of Health Sciences Informatics, Johns Hopkins University, Baltimore, MD, United States*
[e] *Institute for Healthcare Informatics, University of Minnesota, Minneapolis, MN, United States*

ARTICLE INFO

*Article history:*
Received 4 May 2012
Received in revised form
10 August 2012
Accepted 22 August 2012

*Keywords:*
National Health Information Infrastructure
Electronic Medical Record
Health Information Exchange
Health Record Bank
Personal Health Record
Patient-Centered Medical Home

ABSTRACT

*Introduction:* The importance of a nationwide health information infrastructure (NHII) is widely recognized. Patient data may be stored where it happens to be created (the distributed or institution-centric model) or in one place for a given patient (the centralized or patient-centric model). Minimal data is available regarding the performance implications of these alternative architectural choices.

*Objective:* To help identify the architecture best suited for efficient and complete nationwide health information exchange based on the large-scale operational characteristics of these architectures.

*Design:* We used simulation to study the impact of health care record (data) fragmentation and probability of encounter on transaction volume and data retrieval failure rate as markers of performance for each of the above architectures.

*Results:* Data fragmentation and the probability of encounter directly correlate with transaction volume and are significantly higher for the distributed model when the number of data nodes >4 ($p < 0.0001$). The number of data retrieval failures increases in proportion to fragmentation and is significantly higher for the distributed model when the number of data nodes $\geq 2$ ($p < 0.0059$).

*Conclusion:* In simulation studies, the distributed model scaled poorly in terms of data availability and integrity with a higher failure rate when compared to the centralized model of data storage. Choice of architecture may have implications on the efficiency, usability, and effectiveness of the NHII at the point of care.



## 1. Introduction

The Health Information Technology for Economic and Clinical Health (HITECH) Act calls for the Office of the National Coordinator for Health Information Technology (ONC) to develop "a nationwide health information technology infrastructure that allows for the electronic use and exchange of health care related information" [1]. Through this legislation, the federal government has committed unprecedented resources on a multiyear incentive program to support the adoption and use of electronic health records (EHRs) [2]. Getting healthcare data in an electronic format is foundational to eventually enabling patient data sharing. Ultimately, the goal of this nationwide health information infrastructure (NHII) is to provide secure access to comprehensive electronic patient

* *Corresponding author at*: 1468 Madison Ave, Box 1243, New York, NY 10029, United States. Tel.: +1 212 241 2264; fax: +1 212 987 0389.
E-mail address: vijay.lapsia@mssm.edu (V. Lapsia).
1386-5056/$ – see front matter © 2012 Elsevier Ireland Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.ijmedinf.2012.08.008

information when and where needed resulting in measurable improvement of the quality of individual care [3]. Undoubtedly, efficient exchange of information across the multiple settings where care and services are delivered is a central and necessary component of coordinated, accountable and patient-centered models of care, all strategic to achieving the grand vision of the HITECH Act. However, multiple major challenges continue to exist in delivering on this ambitious target [4]. Even though exchange of information is the cornerstone of good clinical care and vital to all the objectives of our health care system, the level of health information exchange in the U.S. is extremely low, both in terms of infrastructure and provider participation [5]. A recent survey of health information exchange organizations in the United States found that of the 179 regional health information organizations surveyed, none met the criteria for the comprehensive health information exchange needed to substantially improve care quality and efficiency [6].

Most health information exchange efforts have employed a distributed model, with data passed between providers' systems. In this architecture (also known as the institution-centric model) patient information is stored locally by the provider/institution that generates it. Proponents of the distributed model cite institutional autonomy and better scalability as key benefits [7]. In this model, institutional autonomy is a part of the architectural design primarily to avoid many of the obstacles and concerns related to privacy fears, confidentiality issues, regulation, and proprietary interests. However, several concerns have been raised about the real world performance of the distributed type of architecture, including effectiveness in actual practice, fragmentation of data sources, accuracy and completeness of health care records and real time availability of data at the point of care. While the distributed model may serve institutional interests, it may not be the best architecture to provide accurate, reliable, complete and timely access to the patient's health record. An alternative to this system is the centralized (or patient-centric) model, in which the patient record (or a minimum subset) is also stored centrally in a Health Record Bank (HRB) of the patient's choice [8,9] (See Fig. 1). So far, there is very little direct performance data that can be used to recommend one paradigm over the other. We developed a simulation model to compare performance and gain better insight into the large-scale operational characteristics of the abovementioned architectures.

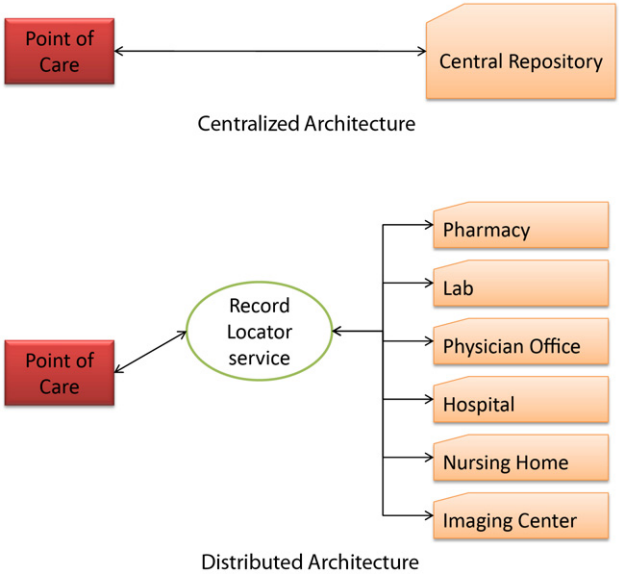


**Fig. 1 – Centralized vs. Distributed architecture.**

## 2. Methods

### 2.1. *The simulation model*

The NHII currently consists of a network of heterogeneous electronic medical record (EMR) systems (each EMR system is also known as a node). These loosely connected nodes are expected to share data electronically using standardized secure communication protocols built on top of the internet [10]. A file in this model (the distributed or institution-centric model) represents one patient's medical record at a specific site of care. Users needing comprehensive information for a given patient must request the patient's files via queries to all the patient's prior sites of care. Therefore, in the distributed model many nodes store patient data, so each patient's medical record is split over multiple nodes. In contrast, in the centralized model each patient's records from all sites of care are also located on a central server.

In the distributed model multiple queries are needed to retrieve the fragmented patient data from the source nodes, whereas only a single query is required to obtain a patient's file in the centralized model. By design, large distributed systems with heterogeneous data sources incur a query performance penalty [11]. Various methodologies and techniques that optimize the query performance and improve scalability and workload adaptability have been proposed and validated [11–15]. An example of such optimization is the use of a 'Record Locator Service' (RLS) or similar index to identify and track the file locations of each patient's records. In the distributed model, a patient's record would be retrieved via queries to the various sites of care documented in the RLS at the time of previous encounters. Query search optimization using a solution such as an RLS dramatically reduces the cost of locating nodes with relevant data in a distributed model [16,17]. *However, unlike the centralized model, the distributed model will still incur the cost of multiple queries to assemble the patient's record, in direct proportion to the extent of fragmentation.* Essentially, the total number of queries required to retrieve a single patient's complete record in the distributed model will at the very least equal the number of nodes across which the record is fragmented.

We selected the emergency room as the simulated point of care since immediate availability of comprehensive patient information in this setting is likely to have a substantial impact on health outcomes. Each provider or entity generating health care related data associated with a patient is represented as a data node in the model. We then identified key factors that impact the performance of the NHII:

**Data fragmentation:** Based on published data and review of care utilization, most patients, even those requiring only an annual physical, are likely to have at least 3 providers