

T E C H N O L O G Y

The New Alphabet Soup: Models of Data Integration, Part 1

Noam H. Arzt, PhD

In November 2004, the National Coordinator for Health Information Technology issued a request for information to solicit public comment on the development and adoption of a National Health Information Network (NHIN).¹

Since then, the alphabet soup has thickened. What do all these new terms, like LHII, NHIN, EHR and RHIO mean to providers and hospitals? How do these institutions fit into the emerging national initiatives, and what role should they play in them?

In a previous column (JHIM, Summer 2005), I defined some barriers and enablers for regional health information organizations (RHIOs). That column defined a framework for integration that focused on two necessary aspects—data integration and application integration. Data integration involves forming valid relationships between data sources. Application integration for data presentation involves

making integrated data available by presenting a unified view of data to a user through a computer application, which can range from a personal computer to a Web browser to a smart card.

This column will focus on models of data integration. A more thorough discussion of these concepts can be found in a response to a request for information for the Office of the National Coordinator for Health Information Technology (ONCHIT) submitted in January 2005.² In this and my next column, I will review five models of data integration that repre-

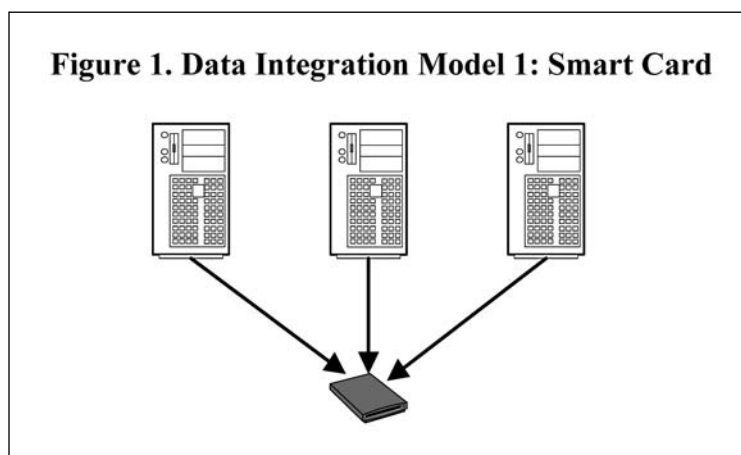
sent points along a continuum from least centralized to most centralized implementation.

The Continuum

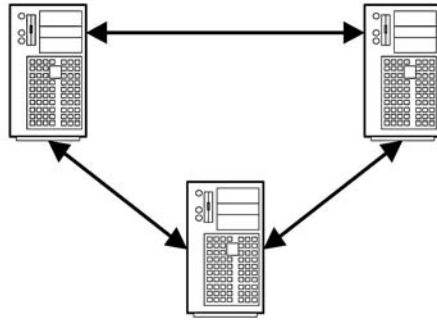
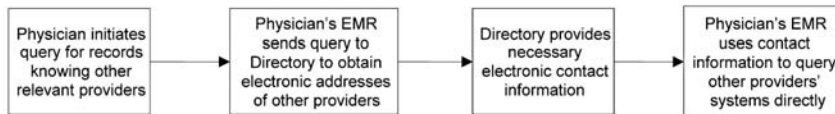
This first model, the smart card, is the extreme in distributed databases: there is no central database at all (see Figure 1). Providers of data store information about a patient directly on the patient's smart card (or other portable data device, like a USB drive), which is carried with the patient from site to site. Authorized users have smart card readers that permit access to patient records

contained on the card as well as the addition of new data. The patient controls access to his or her own data through possession of the card.

This model enables incremental deployment. Data providers and patients can be added to this system over time as quickly or as slowly as a RHIO requires. The technology required to



T E C H N O L O G Y

Figure 2. Data Integration Model 2a: Targeted Peer to Peer**Figure 3. Facilitated Peer to Peer Model Information Flow**

read data from smart cards is relatively inexpensive. There is no requirement for an expensive central database or system or for consolidation of patient records.

On the other hand, the patient must be physically present to access data that is replicated from provider systems onto the smart card. If data changes at its source, there is no ability to update the smart card until the patient returns to the provider. Technology to write data to the smart card is not expensive, but its integration into existing or emerging electronic medical records systems that possess the necessary data may not be simple or inexpensive. In addition, this model does not facilitate community-wide data analysis because there is no central consolidation of data.

The second model, peer-to-peer, can be implemented in a number of variations. In its simplest implementa-

“How do these institutions fit into the emerging national initiatives, and what role should they play in them?”

tion, (see Figure 2), each participating system communicates as needed with its neighboring system. There is no central server. Data is displayed or stored for later display within a local participating system depending on the needs and capabilities of that system. Standards for communication exist for data formats and message types (for example, HL7), vocabulary (for example, SNOMED), and communications techniques (for example, Web Services, SOAP).

This implementation is called

“targeted” because the user must know where to target a request for information for the patient. This implementation can support real-time or batch communications depending on the capabilities of the participating systems.

A variation of this model (broadcast peer-to-peer) introduces the capability of a network broadcast. In this way, a participating system can make an inquiry to participating systems, by mutual agreement, as to whether they have data about a particular patient. An agreed-upon broadcast and response protocol must be developed and implemented for this to be successful. A second variation of this model, facilitated peer-to-peer, introduces a central directory server to help one facility identify participating providers, obviating the need for a network broadcast.

Note that this is not a directory of patients. Standards exist for this type of network-based directory service, such as LDAP. The flow chart in Figure 3 describes a typical information flow for this type of interaction.

This model allows incremental deployment as systems become ready. No replication of data required, although it is possible, and any system can participate, even if they're geographically peripheral, as long as they adopt the standards. There is a lower burden of central coordination because there is no dependence on a central database, than that operated in the facilitated variation.

This model may work well when there's a small number of participants, and it may be less expensive to deploy. However, in some implementations, participants need to know the destination system for the information request, or they need to be patient while “the network” is searched. This model might result in some systems

T E C H N O L O G Y

falling behind and not being able to support inter-system communication because it may be difficult to retrofit some older systems to perform these data queries and transfers. This model will not scale well to many systems, and performance may be slow. Finally, this model does not facilitate community-wide data analysis.

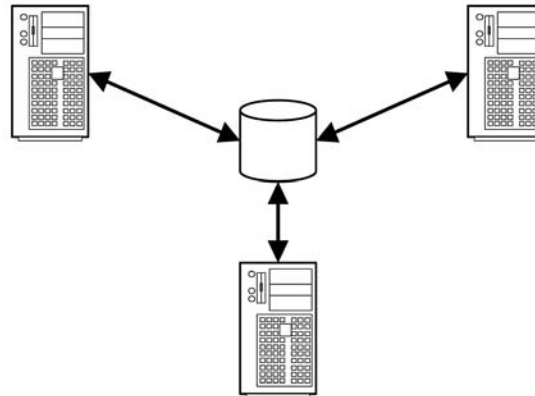
Adding Complexity

The third model, the information broker (some refer to this as the federated model) represents the first real step toward centralized operations (see Figure 4). In this model, a central hub contains a master patient index of all patients contained in all participating systems. The hub does not contain any actual clinical records or other medical data. Within the index, each patient's record is tagged with the participating systems that contain data about that patient.

Ultimately, patient data is retrieved through a two-step process. First, the requestor queries the hub to identify where relevant parts of a patient's record exist, and then the requestor can query those systems to retrieve the relevant data. Standards for communication exist for data formats and message types, vocabulary and communications techniques. This model can support real-time or batch communications.

In this model, the system can discover where relevant records are housed community-wide. There's no replication of clinical data, so data remains close to its source when additions, updates or corrections need to be made. The system as a whole is better protected from inappropriate disclosure because each participating system is still in full control of its patient data. A participating system is free to refuse to respond to a query it might deem inappropriate or insecure.

Figure 4. Data Integration Model 3: Information Broker



This model scales well for many users and systems because only a patient index is located in the central hub, and it can facilitate community-wide data analysis by querying populations or sub-populations from the central hub. It may be easier to integrate existing systems into the network because they need only register their patients with the central hub.

“...the information broker...represents the first real step toward centralized operations...”

On the other hand, for this model to work properly, strong central coordination is required; this is as much an organizational issue as a technical one. There is dependence on the central hub for inter-system communications, so if the hub is unavailable for some reason, system-to-system communications will stop. It is harder for individual systems to

participate because they must be able to communicate with the hub using a standard method.

This model requires two steps, and more time, to get data: first, there's a query to the hub, then a second query to the authoritative system. It may require a large effort to keep demographic records free from duplication, because these records will be collected from numerous disparate sources. After the central hub identifies where the subsequent data queries need to go, other systems may be unavailable at query time. It is more difficult to present a coherent, unified view of the patient because the clinical data continues to reside in distributed systems and not in a single central location.

Where Do We Go From Here?

These first three models offer options from one end of the continuum—the least centralized approach—toward the middle. The information broker model continues to be widely discussed, in part because of its use by the Santa Barbara County Data Exchange,³ a conspicuous RHIO project that has had its ups and downs.

T E C H N O L O G Y

The remaining two models will be discussed in the next column. The partitioned warehouse is a popular model used by the Indiana Health Information Exchange,⁴ one of several current success stories for the RHIO movement fueled by the efforts of the Regenstreif Institute, and embraced by

the collaborative response to the ONCHIT RFI written by the Connecting for Health Collaborative.⁵

The final model, the central warehouse, goes even further down the path of centralizing data resources for ease of consolidation and retrieval.

About the Author

Noam H. Arzt, PhD, is president and founder of HLN Consulting, LLC, San Diego, a consulting firm specializing in healthcare systems integration, especially in public health. He can be reached at arzt@hln.com

References

1. <http://www.hhs.gov/news/press/2005pres/20050603.html>
2. <http://www.hln.com/noam/ONCHIT-RFI-HLNConsulting.pdf>
3. http://www.carescience.com/healthcare_providers/cde/care_data_exchange_santabarbara_cde.shtml
4. <http://www.ihie.com/>
5. http://www.connectingforhealth.org/resources/collaborative_response/toc.php

Congratulations CPHIMS Recipients!



Himss[®]

CPHIMS Certification

Congratulations to the following individuals who have achieved the Certified Professional in Healthcare Information and Management Systems (CPHIMS) credential from August 1 through October 7, 2005. The CPHIMS designation is recognized throughout the healthcare organizations as a national standard of industry knowledge. The designation indicates that an individual has met the eligibility requirements and passed the CPHIMS examination, demonstrating to superiors, colleagues and potential employers their dedication to the industry, dedication to high standards, and pride in their own professionalism.

Randal C. Bablitz—Ottawa, ON, Canada
 Don C. Ellis—Dallas, TX
 Donna R. Eversole—Orange Park, FL
 Melissa A. Foster—Miami, FL
 George A. Gargarella—Ottawa, ON, Canada
 Roger Hertz—Omaha, NE
 Miriam L. Isola—Vernon Hills, IL
 John A. Kupkovits—Tampa, FL
 Lucy Molfetas—Skillman, NJ

Ellen L. Pearce—Walpole, NH
 Catherine A. Sees—Auburn, PA
 Alan Sivik—Coconut Creek, FL
 Susan L. Stefan—Mount Prospect, IL
 Susan C. Sutton—Austin, TX
 Elizabeth A. Trimmell—Baltimore, MD
 Jon L. Walker—Houma, LA

Join this elite group. Go to www.himss.org/getcertified to find out more about the CPHIMS certification program. To see a full listing of certificants go to http://www.himss.org/content/files/cphims_certificants.pdf.