Data Linkage: Ready or Not, It's Here [part 1]

by Lorraine Fernandes, RHIA

In early 2004 Connecting for Health, a public-private collaborative funded by the Markle and Robert Wood Johnson Foundations, engaged a cross-section of experts in security, technology, privacy, consulting, and medical informatics to define how healthcare information could be shared more effectively and broadly. Gail Graham, RHIA, and Lorraine Fernandes, RHIA, represented AHIMA on Connecting for Health's linking work group.

The group believed that making information available on a timely basis will enhance patient safety, increase the quality of care, and reduce the cost of healthcare. The final report, titled "Linking Healthcare Information: Proposed Methods for Improving Care and Protecting Privacy," will be released early this year.

The group's report tackles the issues that surround any initiative to link patient data on a national scale, including privacy, legality, technology infrastructure, and solution architecture. This article describes some highlights of the final report that are particularly relevant to HIM professionals. More information from the final report will be presented in this column in the March issue.

Background and Recommended Approach

Historically, there have been many attempts to link data, such as community health information networks in the 1990s. But more of these endeavors failed than succeeded. Some of the reasons for those failures still exist, while others have since been addressed through standards and new or enhanced technology.

The linking work group strove to define strategies for linking patient data securely, efficiently, and effectively, while balancing the need to protect privacy with the need to deliver health information on a timely basis to facilitate patient care. The underlying principle of this effort is to match patients with their information at various sites of care and make the information available to caregivers when and where it is needed.

The final report advocates a decentralized approach for linking healthcare records. This approach, which was not readily feasible in years past, allows healthcare data to remain under the control of the clinician and the patient until transfer is authorized. Further, it allows for a phased-in implementation that can complement the launch of electronic health records in many organizations.

Stating some of the benefits, the report says, "The only thing centralized is an electronic directory of providers holding patient records and pointers to those files…sharing is peer to peer, and both the decision to link and the decision to share are made locally, where the records are created."[1]

Thus, the patient remains in control of record linkage and data sharing, with the next level, the provider, supporting the patient's decision. An audit trail revealing who taps the records, when, and the location of the clinical data is available to the patient to facilitate a greater understanding and control of his or her healthcare data.

This decentralized approach allows for rapid deployment in incremental steps, thus producing value from inception. Conversely, linking records based on a new national healthcare individual identifier would require waiting until existing technology can be retrofitted, therefore delaying the value and benefits until the "big bang" occurs.

Improving the Linking of Health Records

Linking patient records is essential to improving healthcare. The report asserts that record linkage:

- **Should not require a national health identifier.** One does not exist today, and there is little probability society and the government in the near term would embrace it. If one were seriously proposed, it would likely encounter major opposition from many fronts.
- **Should be designed to take advantage of a health identifier, but not from a top-down approach.** Locally assigned identifiers meet many of the requirements for a health identifier; that is, the issuing facility and its internal number provide a unique, nondisclosing identifier. This system must also support an individual with multiple identifiers. Clearly, Social Security numbers should not be used as such an identifier since they are disclosing identifiers.
- **Should use algorithmic matching of common data elements commonly available today.** These systems work by comparing records from different databases and assigning probability or scores to potential matches.

The Benefits of Probabilistic Matching

The work group concluded that while probabilistic matching may not yield 100 percent accuracy, the advantages far outweigh the disadvantages and record linkage will be greatly enhanced using this method. Probabilistic matching is in use today in many large healthcare organizations with millions of records (such as Sutter Health, RxHub, Regenstrief Institute, and CareGroup, as noted in the report), thus the work and knowledge gained from these and other installations will make it easier to scale this approach to a national level. Even if a national health identifier were in place, probabilistic matching would serve a valuable purpose, as data entry errors, duplicate assignments, and incomplete data would still exist, and probabilistic matching addresses these common data quality issues.

The work group recommends any probabilistic matching solution should:

- Document current practices for data capture, cleanliness, and possible improvements
- Explore incentives for better capture and cleaning of data
- Develop a reference implementation of probabilistic matching
- Develop a pilot project that tests and improves the reference implementation

The value of clean, accurate data at the point of capture should not be underestimated, and this starts with solid practices for capturing, storing, and maintaining data.

Architectural Principles

The architectural vision of the work group revolves around a federated approach to identify patient records that are housed remotely, as occurs when a patient moves between sites of care. No clinical data will be maintained by a national or regional entity (unless the regional entity already has clinical data). The foundation for the architecture is:

- **Decentralized**—the US healthcare system is structured as such, and we do not see change in this model over the next five years.
- **Federated**—since decentralization remains in play, control and release of data remains in the hands of the providers, and standards and agreements bind the participating parties.
- **Built without requiring "rip and replace"**—no wholesale removal or replacement of systems is anticipated. Rather, incremental enhancements to software are expected, along with enhanced standards and practices.
- **Built through decoupled development**—since there is no "all or nothing" package, incremental advancement and its associated value is envisioned. Participation should require minimal investment and return benefits quickly.
- **Built on top of the Internet**—no new wires. Standards for data encryption, security, and auditing will be required since the Internet will serve as the conduit. Patient privacy and data integrity must be preserved while using the new method of communication.

In the next issue, we will discuss the record locator service, security and privacy, and a reference project. In addition, the author and Carol Diamond, chair of the Connecting for Health work group, will be giving a presentation on the topic at the HIMSS meeting, February 14, 2005.

Note

1. Connecting for Health. “Linking Healthcare Information: Proposed Methods for Improving Care and Protecting Privacy.” Forthcoming. To be available online at www.connectingforhealth.org.

Lorraine Fernandes (lfernandes@initiatesystems.com) is senior vice president of healthcare practice at Initiate Systems, Inc.

Article citation:
Fernandes, Lorraine. "Data Linkage: Ready or Not, It's Here [Part 1]." *Journal of AHIMA* 76, no.2 (February 2005): 62-63.