

# Surveying the RHIO Landscape

## A Description of Current RHIO Models, with a Focus on Patient Identification

In an effort to identify best practices in emerging and existing regional health information organizations (RHIOs), AHIMA's e-HIM<sup>2</sup> work group on patient identification in RHIOs offers HIM professionals a marketplace description of existing RHIO models with a focus on patient identification linkage methods. This practice brief does not address the issues related to privacy and security reflected in HIPAA regulations, nor issues such as data quality within the records. It only discusses the quality of linking methods.

Organized cross-jurisdictional healthcare data-sharing organizations are referred to as RHIOs throughout this practice brief. This is just one of several terms applied to such organizations. Others include health information exchange (a more generic term used by the eHealth Initiative), subnetwork organization (the term used by the Collaborative Response to the Office of the National Coordinator for Health Information Technology), and health information network (a term used by several organizations). As this is an evolving field, it is likely that other terminology will come into vogue in the future.

The work group reviewed current RHIO activity in the US, focusing both on RHIOs funded by grants from the Health Resources and Services Administration (HRSA) and the Agency for Healthcare Research and Quality (AHRQ), as well as independently funded RHIOs. Grants have been provided for the establishment of RHIOs by Blue Cross/Blue Shield and other insurers, nonprofit philanthropic organizations, healthcare systems, and HIM and Internet industry vendors.

The following summarizes the work group's findings on the current status of RHIOs and record linkage methodologies based upon detailed assessment of 21 RHIOs.<sup>1</sup> The group reviewed the purposes and funding sources of the RHIOs, the types of data shared, and the methods of identifying patients and linking records. (The latter was the emphasis of the work group.) Included is information on the participants in the RHIOs, pros and cons for each RHIO record linking method used, patient identification linking fields, who defined the patient identification, the time frame within which they receive data, number of records in the database, and what records are shared. The group's summary document also contains a list of definitions, some of which are incorporated into this practice brief. The data that formed the basis of this document are included in the FORE Library: HIM Body of Knowledge, and further details can be obtained from review of our summary documents.

### Purpose and Funding of RHIOs

The work group identified the main purpose for forming a RHIO as facilitating information sharing among enrolled members of the RHIO using common, nonproprietary stan-

dards for data content and exchange over existing networks and the Internet. The main goals in sharing patient-specific data are to:

- ☒ Improve healthcare delivery by providing immediate, secure, confidential exchange of health information between authorized users
- ☒ Enable providers and patients to make decisions based on near real-time access to health information
- ☒ Provide warning and reminders at point of care
- ☒ Reduce medical errors
- ☒ Prevent adverse drug reactions
- ☒ Encourage participation of patients in their own healthcare and chronic disease management
- ☒ Allow patients, payers, and providers to evaluate quality of healthcare and to make informed choices in where and from whom they obtain care

Of the 21 RHIOs, eight receive grants as their primary source of funding, two utilize incentives, six are funded by membership or dues, and five receive private funding.

There are typically two stages of RHIO development, each with a different form of economic support. Start-up funding for project prototyping and implementing is normally provided through grants, government funding, or vendor funding. Most commonly, grants come from AHRQ or HRSA. Only a handful of these RHIOs reported any HIM involvement in the earliest stages of RHIO development.

The second stage of RHIO development involves maintaining and improving systems, requiring some type of sustainable funding. Typically, this comes by way of subscription or transaction fees for all users or incentives from payers. Incentives provide a certain amount of funding by way of new models for reimbursement, pay-for-performance, rewards for innovation, matching grant programs, and tax credits for investors.

### Record Linking Methods

One of the greatest challenges facing RHIOs is accurately linking electronic records across the disparate health information systems of participating members. Because RHIOs have been formed for a number of different reasons, there is a wide variety of patient identification and record linking methods. HIM professionals are the best resource for offering expertise in patient identification and record linkage because of their historical involvement in management and maintenance of master patient indices (MPIs).

HIM professionals are strongly encouraged to find out what RHIOs are being formed or already exist in their geographic region. Find out how you can be involved in recommending

specific patient identification and record linkage methods best suited to the purpose, size, and organizational structure of a particular RHIO. Often, there are opportunities to be involved in technical subcommittees that develop system specifications for patient identification and record linking. In addition, there may be opportunities to be involved in defining membership and dues categories and developing user authentication and security policies.

### Three Levels of Complexity in Record Linking

Record linking methods are used to electronically link MPI records, electronic medical records, or external clinical results to existing electronic medical records. Additionally, they are used to detect duplicate medical records. A variety of methods can be employed to link records from different data sources, each varying in terms of complexity, efficiency, and accuracy. The record linking subgroup of the AHIMA work group has classified the various record linking methods into three categories: basic, intermediate, and advanced.

#### Basic

The basic record linking method compares selected data elements—most frequently name, birth date, Social Security number, or gender—using exact (identical match of data elements) and deterministic (exact or partial match) linking approaches. This method assumes a high degree of confidence that the match is accurate. Multiple patient identifying data elements are required to prevent false positives. False positives, erroneous linking of two records belonging to two different individuals, should not be tolerated, as they cause two different patients' information to be linked into a single record.

Basic linking is appropriate in facilities or organizations with MPIs containing fewer than 150,000 records or in community settings with a small ethnic population. Basic methods are not recommended for organizations wanting to identify and eliminate all duplicate medical records. Linking records based on exact match of a small set of data elements should be approached with caution. Larger databases that use the basic record linking method often experience very high rates of duplication, often as high as 30 to 40 percent.

#### Intermediate

Intermediate record linking provides more advanced techniques for comparing records by enhancing exact match and deterministic tools with additional logic and arbitrary or subjective scoring systems. Subjective weighting, ad-hoc weighting, fuzzy logic, and rules-based algorithms are examples of intermediate matching tools.

Subjective weighting involves scores assigned by people to field match based on significance of match using rules. Ad-hoc weighting applies numeric values that indicate the overall importance of a comparison relative to other comparisons. The discriminating power of each comparison variable—its importance in determining links—is expressed as a weight.

Fuzzy logic in this context involves data massaging, or rules built to emulate common errors made by users. Examples include rules that transpose digits in a Social Security number, use Soundex or NYSIIS encoding for names, swap first name and last name, swap month and day in date of birth, or search for any date of birth within five years.

When scoring systems are used, the score is the sum of the products of all the comparisons with the associated weights. The score is used to evaluate record pairs and determine links and nonlinks. When weights are applied and summed into scores, the scores for record pairs that should be linked are generally higher than scores for record pairs that should not be linked. In other words, the higher the score, the greater the likelihood that the match is indeed valid.

Intermediate methods are recommended in situations where the organization wants to control the matching attributes and weight assignments. Intermediate methods are not the optimal choice for healthcare organizations with a low tolerance for false positives or in organizations whose primary focus is minimizing duplicate record volume. Non-scientific weight assignment to field matches causes arbitrary record match scores, and if data errors occur that are not programmed in the rules, the two records won't be linked.

#### Advanced

The advanced method employs sophisticated mathematical or statistical algorithms such as probabilistic matching, bipartite graph theory, machine learning, and neural networks.

Probabilistic matching is defined as automatic increase or decrease of field weight match based upon frequency of data item within a database. The more frequently an item appears in a database, the less its validity as a unique identifier. An example would be decreasing the field weight match score for last name if the last name being searched for was Smith.

Bipartite graph theory is a mathematical method for determining similarity between strings of data that models human similarity. It uses mathematical graphs as constructs to determine similarity of data between strings. Machine learning is a discipline that involves mathematics, computer science, artificial intelligence, statistics, and pattern recognition to create software that learns and improves with experience. It attempts to model the human decision making process, taking into account the varying significance of different database fields in establishing whether two possibly differing records actually refer to the same person, and adapts automatically when one or more fields are missing or incomplete.

Neural networks employ machine learning. Although there are a number of definitions of neural networks available, the AHIMA *Pocket Glossary* defines neural networks as "nonlinear predictive models that, using a set of data that describe what a person wants to find, detect a pattern to match a particular profile through a training process that involves interactive learning."<sup>2</sup>

The advanced method is appropriate in healthcare organizations with MPI databases containing 250,000 or more re-

cords, in enterprise MPIs (master patient indices that provide access to multiple repositories of information from overlapping patient populations maintained in separate systems and databases), or in complex organizations with ethnic diversity. Advanced methods are less arbitrary and more error tolerant, leading to the identification of duplicates (or the correct linking of electronic records) created by less obvious errors or when multiple discrepancies in a record occur. Organizations without standardized patient access procedures, with 150 or more patient access registrars in a decentralized registration environment, and with a low tolerance for duplicate records and false positives also benefit from this method.

### Comparing Methods

While each method delivers value to healthcare organizations of particular profiles, analysis of databases using the advanced record linking methods indicates they are far more accurate in successfully linking electronic records. The advanced methods are successful in minimizing false negatives, the failure to link two records together when both records belong to the same individual (which create duplicates), and minimizing false positives (which create “overlay” records). Despite the advantages of advanced methods, smaller organizations may find that the financial outlays required prevent the selection of this method.

As patient identifying data are not perfect, some level of human intervention is required in database management. The reality of electronic health record databases today is such that patient identifying data elements are frequently missing, changed, or entered inaccurately. As healthcare databases get larger and as more integration of healthcare databases occurs, the proper oversight of these databases from a record linking perspective is of high importance. Unfortunately, to date, very few RHIOs have involved HIM professionals in discussions on how to successfully link their electronic records.

Among the 21 RHIOs in this study, the decision of patient identification method was made by varying groups. Decisions were made by hospital project teams with vendor assistance, by vendors independent of hospital involvement, by RHIO project teams with vendor assistance, by ad-hoc committees and technical advisory committees formed by the RHIO, and by information technology professionals. Because HIM professionals have particular expertise in this arena, it is important that they make the appropriate decision makers or groups aware of their interest early in the selection process.

### Data Access

Because of their differing sizes and component organizations, the RHIOs surveyed reported a wide variation in the amount and nature of data collected and made available, the time frame within which the data were shared, and the delivery methods.

Among the more mature RHIOs, real-time Web-based data sharing was the norm, with several specifying HL7 messaging standards. Others shared information by telephone, mail,

or fax. One RHIO provided real-time data to physicians and patients maintaining a personal health record and complied with requests from nonphysician health professionals within 48 hours. Some RHIOs acknowledged that on the basis of currently available technology, response times for certain types of data may be weeks.

Among some of the newer organizations, the decision regarding the types of data to be shared is still being made, and several developing RHIOs are also defining what type of data they will share. The type and amount of information shared ranged from simple demographic data to real-time sharing of entire electronic health records. The data shared by the members of the RHIOs in the work group study included:

- ☒ Demographic data only
- ☒ Payer information
- ☒ Communications between information sharers
- ☒ Physician orders
- ☒ Medication records
- ☒ Radiology images
- ☒ Nursing notes
- ☒ Allergic history
- ☒ Laboratory data
- ☒ Childhood immunization records
- ☒ Eligibility and referral information
- ☒ Complete electronic health records

The vast majority of RHIOs studied did not maintain a centralized data repository. They left the responsibility for maintaining the data with the originating organization and instead provided a method for linking and relaying the data among the RHIO members. Only one RHIO surveyed maintained an actual enterprise master patient index; the remainder used the various pointing systems to identify patients.

The largest RHIO in the study provided access to more than 5 million records, while the smallest currently sharing data included 200,000 records. Among the RHIOs not yet actively sharing data, one anticipated that the number of records to be shared would be approximately 500,000, and several others had not yet determined the number of expected records. As the RHIOs evolve and expand the number of stakeholders, the number of records should increase.

### Summary

Across the country, new RHIOs are being formed every day. The 21 RHIOs studied by the work group provide good examples of the variety of purposes, funding, and record linking methods RHIOs may adopt. As this trend continues to evolve and improve, RHIOs may prove to be a valuable stepping stone on the road to a national system in which a patient’s medical data will be available anywhere, anytime.

Accurate patient identification and linking is the foundation of health technology that is implemented in a RHIO or any similar network that shares patient information. Without accurate patient identification, patient safety and quality of care are compromised. When high percentages of duplica-

tion or overlaying of records occurs in electronic health record databases, physician trust in the system is lost. As HIM professionals, we must be involved in addressing the security and confidentiality of RHIO databases and in defining the record linking method appropriate to the RHIO.

As professionals skilled in patient identification methods and possessing significant organizational skills and personnel management experience, HIM professionals should become involved in this process at the earliest opportunity in the RHIO formation. HIM professionals can participate in long-term planning, business plan development, and organizational structure definition. Future articles will address how HIM professionals can become involved, what particular attributes and skills they can bring to the table, and job descriptions appropriate to HIM professionals in the healthcare information sharing industry. The work group urges all HIM professionals to become involved personally in this exciting new field.

### Notes

1. The 21 surveyed RHIOs were Brevard County Health Information Alliance, CareGroup Healthcare System, CareSpark/Tri-Cities TN & VA, Healthcare Access Program, Jackson Health System, MA-SHARE, Michiana Health Information Network, Northwest Florida RHIO, Rhode Island HIT Project, Santa Barbara County Healthcare Community, Santa Cruz RHIO, St. John's Mercy Health System, St. John's Physicians and Clinics, Taconic IPA/MedAllies, Tampa Bay RHIO, the Community Foundation of Central Florida and Good Health Network, Utah Health Information Network, Western North Carolina Healthcare Networks, Whatcom County Health Information Network, Wisconsin Health Information Exchange, and Winona Health.
2. AHIMA. *Pocket Glossary of Health Information Management and Technology*. Chicago, IL: AHIMA, 2006.

### Prepared by

Wanda Bartschat, MA, RHIA  
 Jill Burrington-Brown, MS, RHIA  
 Susan Carey, RHIT  
 Jennie Chen  
 Sally Deming,  
 Stacie Durkin, MBA, RNC, RHIA  
 Lorraine Fernandes, RHIA  
 Kathy Giannangelo, RHIA, CCS  
 Kerry Heinecke, RHIA  
 Susan Helbig, MA, RHIA  
 Rhonda Hisle, RHIT  
 Gwen Hughes, RHIA, CHP  
 Susan Hull, MPH, RHIA, CCS, CCS-P  
 Beth Just, MBA, RHIA  
 Teresa Knox, RHIT  
 Deborah Kohn, RHIA, CHE, CPHIMS  
 Dianne Koval, RHIA, CPEHR  
 Chrisann Lemery, MS, RHIA  
 Eve-Ellen Mandler, RHIA, MS, CCS  
 Natalie Nichols-Banks, JD, RHIA  
 Michele O'Connor, MPA, RHIA  
 Laurie Peters, RHIT, CCS  
 Rebecca Ramsey, RHIA, CCS-P  
 Rebecca Reynolds, MHA, RHIA  
 Renae Spohn, MBA, RHIA, CPHQ  
 Melanie Thomas, RHIT  
 Frank Waterstraat, PhD, MBA, RHIA  
 Vicki Wheatley, MS, RHIA  
 Gail Wood, MS, RHIA

This work was supported in part by Initiate Systems.

**Authors: (1) Is one RHIO missing from note 1?  
 (2) Should addition be 2005 in note 2?**