Ensuring Data Integrity
In Health Information Exchange
Inaccurate health information may adversely affect the quality of an individual’s healthcare, insurance, and employability. As computerization of health information continues and the scope of organizational exchange of health information widens into health information exchanges (HIEs), maintaining the integrity and completeness of health data is paramount.

The overarching goal of HIEs is to allow authorized users to quickly and accurately exchange health information to enhance patient safety and improve efficiency. Achieving this goal is dependent on the ability to link (match) multiple, disparate records relating to a single individual.

A 2008 RAND report, “Identity Crisis: An Examination of the Costs and Benefits of a Unique Patient Identifier for the US Health Care System,” noted that avoiding adverse drug events, which are often the result of incomplete linking information about a patient’s medications or allergies, could save the healthcare system in the US about $4.5 billion per year. This report also points out that on average an 8 percent duplicate record rate existed in the master patient index (MPI) databases studied. The average duplicate record rate increased to 9.4 percent in MPI databases with more than 1 million records. Additionally, the report identified that the duplicate record rates of the enterprise master patient/person index (EMPI) databases studied were as high as 39.1 percent.

High duplicate record rates within EMPI databases are commonly the result of loading unresolved duplicate records from contributing MPI files. EMPI systems that leverage advanced matching algorithms are designed to automatically link records from multiple systems if there is only one existing viable matching record. If the EMPI system identifies two or more viable matching records when loading a patient record, as is the case when the EMPI contains unresolved duplicate record sets, it must create a new patient record and flag it as an unresolved duplicate record set to be manually reviewed and resolved. Therefore, if care is not taken to resolve the existing EMPI duplicate records, the duplicate rate in an EMPI can significantly grow as additional MPI files are added.

Patient identity integrity is the accuracy, quality, and completeness of demographic data attached to or associated with an individual patient. This includes the accuracy and quality of the data as it relates to the individual, as well as the correctness of the linking or matching of all existing records for that individual within and across information systems.
systems. There are tremendous potential benefits and cost savings within the healthcare industry contingent on accurate patient identification and interoperability. Participation in an HIE can also provide increased efficiency in the delivery of healthcare by permitting access to more complete and timely information regarding individual patients.

**HIM Professional Involvement**

Traditionally, the primary responsibility of the HIM professional has been to provide accurate data to support the delivery of quality healthcare. The HIM professional has to balance the patient’s right to privacy with the need to allow access to health data for legitimate uses. Application of HIM skills within the HIE arena will expand opportunities in the areas of information technology, database management, data analysis, and patient identity. HIM professionals possess the knowledge and skills that add value to the effective planning and implementation of an HIE while ensuring compliance with the HIPAA privacy and security rules.

The unique skill set of the HIM professional is increasingly recognized and sought after in today’s electronic healthcare environment. The increased importance of collecting accurate and timely health data across multiple organizations and the integration of new technology in the healthcare delivery system continue to both shape and expand the role of the HIM professional.

The purpose of this report is to provide general concepts and current examples of methodologies for the development of data integrity and quality standards in HIE.

**HIE Models**

HIEs typically have one of three architectures; centralized, federated (or decentralized), and hybrid. The choice of architecture is driven by the organization’s privacy and security practices.

» In a centralized model each organization transmits patient demographic and clinical information to a shared repository. This centralized repository is queried to obtain a patient’s linked results and other information.

» In a federated model the data source organization maintains custodial control over the patient’s medical record and indices. When requested, data is queried from the data source organization.

» A hybrid model is a mixture of the federated and centralized models. For example, some HIEs use a modified hybrid model consisting of independent databases (federated) for each data source on “virtual servers,” replicating all the benefits of federation while leveraging economies of scale from centralized management capabilities. This approach utilizes “virtual servers” and is believed to be more cost effective and scalable, providing additional security that would not be available through a more conventional approach.

HIE organizations are governed by an HIE board of directors and/or executive committee that provides oversight to specialty committees. The specialty committees develop policy and procedures that ensure health data are accurate, complete, relevant, and up to date. These committees are responsible for subject matter expertise and operational aspects including HIM, finance, technology, communication, and clinical priorities. Each HIE must develop policies to ensure high levels of data integrity, including data content standards and definitions to promote submission of quality data for which the HIE and participating organizations are responsible.
The HIE governance structure specifies the distribution of rights and responsibilities of the board, as well as the managers, shareholders, and other stakeholders. It also spells out the rules and procedures for making decisions on corporate affairs. HIE governance is also focused on promoting corporate fairness, transparency, and accountability. According to the Office of the National Coordinator for Health Information Technology, the governance structure also incorporates the organization’s strategic response to risk.²

Some initial tasks traditionally include:

» Establishment of a governance structure that achieves broad-based stakeholder collaboration

» Setting initial goals, objectives, and performance measures for the exchange of health information that reflect agreement among the healthcare stakeholder groups and that accomplish regional or statewide coverage of all providers for HIE requirements related to meaningful use (MU) criteria

» Establishment of mechanisms to provide oversight and accountability of HIE to protect the public interest

As an HIM professional, how can you get involved in your HIE?

One way for HIM professionals to participate and guide HIE activities regarding data integrity is to join HIE leadership—the board of directors or executive committee—or specialty committees—HIM, finance, legal, audit and compliance, clinician, technology/technical advisory, consumer advocacy, community project, or nominating—that develop policy and procedures that ensure health data are accurate, complete, relevant, and up to date.

Patient Information Exchanged within an HIE

The overall goal of HIE is to provide a patient’s requested clinical information in real time and in a format that allow it to be used effectively by the provider currently seeing the patient. To achieve this purpose the HIE may provide the requested information in a variety of ways. For example, accepted formats for exchange documents are varied and include the Continuity of Care Document (CCD) or Continuity of Care Record (CCR), scanned documents in PDF, TIFF or other format, HL7 records, plain text records, and
the like. In most cases the document formats are dependent on the original source formatting. The “gold standard” is a document that can be “parsed” to provide discrete data elements that can be incorporated into the current provider’s electronic health record (EHR) using standard mapping and conversion techniques.

Although many HIEs have the ability to translate documents into other formats, providing such a service is expensive and requires constant maintenance and review. Therefore, the HIE itself often only provides a view of a document provided by another source.

Many HIEs are still developing their business models and working to establish transactional volumes and records to ensure HIE viability and sustainability. In order to maximize record volume they are willing to work with provider systems to accept records in a variety of formats. More fully developed HIEs permit participating providers to retrieve data elements in standardized formats, which may then be incorporated within the provider’s EHR. The actual incorporation of these data elements into the provider systems is the responsibility of the provider. One strategy some HIEs have adopted, and which others are considering adopting, is incorporation of both the NHIN and the Direct Project protocols in order to facilitate electronic information access.

Incorporating the HIE’s information into a participant’s medical record

There are three models with regard to incorporating patient information obtained from the HIE. One model permits cut and paste with attribution; one model permits scanning the report itself into the chart so that the source is visible; and the third model permits summarizing the content of the received information. The policy in each case is determined by the HIE’s board of directors.

**Patient Identification**

*How a patient’s information is maintained, corrected, or updated when new information is sent to the HIE*

A combination of automated and manual processes are in use in various HIEs to maintain the integrity of any EMPI. Typically the HIE uses the same processes as a self-contained EMPI. It is continuously updated by data sources regarding patients who have opted in and out of the exchange. To a greater or lesser extent, historical information is maintained with all updates as patient matching algorithms match new patient demographic information against existing information. Systems that rely on the algorithm usually set a confidence level, so that if a match does not meet a confidence threshold, then a new patient is created. A report is periodically run to identify likely matches subsequently reviewed by the HIE staff. Depending on staffing availability, HIE staff follow up with the submitting facilities as necessary to verify the validity of the demographic data. Corrections are made as necessary and when indicated. Duplicate patients are merged or linked. Many HIEs also receive HL7 merge messages from each participant. These messages are routed into a work queue and patients are electronically and manually merged, linked, or updated.

The processes to correct demographic data depend on the individual HIEs and their agreements with the participating hospitals or providers. At a minimum, the HIE’s policy should clearly state who can initiate a correction, what notifications are required (by whom, and to whom), and within what time frame of the correction.
Clear and concise policies and procedures are required at both the organizational and HIE levels to ensure corrections are handled in an appropriate manner. For example, in Massachusetts, Hospital A requires that all requests for amendments or corrections to health information be submitted to the hospital within five business days. The HIE is not allowed to independently and arbitrarily make patient amendments to health information. As the HIE will most likely not be the source system for either the patient’s demographic or clinical information, corrections in the HIE environment can be risky.

Since it is essential to propagate any change to all copies of the health record across the continuum of care, all participants within the HIE should understand how and when corrections will be made by the HIE and the impact those corrections may have on the patient’s records. HIM professionals can provide leadership and guidance regarding HIPAA privacy rule amendment requirements for both the source organizations and the HIE.

Process for ensuring all information in the record attributed to a patient actually belongs to the patient

A combination of algorithms and programmatic matching, human review of matches that fail programmatic linking, and periodic sampling and review of data quality should all be represented in a well-rounded data integrity program.

Ideally, before sharing data with an HIE each organization will have made a formal commitment to accurate patient identification by establishing guidelines for data stewardship and data governance with quantifiable expectations and performance standards in order to decrease the number of duplicate patients in their healthcare systems. Many HIE organizations will employ advanced matching algorithms capable of identifying up to twice as many potential duplicate record sets than most duplicate record reports found in healthcare information systems. This increased accuracy in duplicate record detection within the HIE has the potential to highlight the need for greater patient identity scrutiny and possibly improved patient identity management technology with the patient access and HIM departments.
Although improved technology is important, inside each participating organization and within the HIE there should be multiple stakeholders who identify problem areas and monitor performance. This group should also define the format for each field and data definitions for each field (to increase consistency) and create processes for complete and accurate data capture for all required fields. HIM and IT staff, in their collaborative roles, should examine each system interface and any affiliated domains with reasonable tests to assure accurate patient identification. The individual healthcare organizations should establish internal review and acceptance criteria and performance metrics, perform baseline and continuous assessments of patient identification practices, and provide performance feedback to affected staff.

**Data Elements Used in Patient Selection**

**Process to ensure accurate selection, by requester, of information belonging to the correct patient**

HIEs have varying minimum criteria for searching for a patient. At a minimum we believe that the patient’s last name, first name, and date of birth should be used to score and match a patient. If the requester does not supply the required elements, no patient data will be accessed. Not only is it necessary to ensure the information in the index associated with that patient’s identifiers actually belongs to the patient in question, but care must also be taken to ensure the requestor selects the correct patient from the available choices. The literature discusses both “pull” and “push” technology in relation to the combining of clinical data from multiple sources into one view for the requester and the various algorithms used to provide the list of possible candidates, but little information is available regarding how to ensure the requestor selects the correct patient.4

**Process to ensure all available patient information maintained in the HIE is displayed and/or accessible to the requesting party**

In order to perform this task, the HIE must use some type of intelligent search mechanism or protocol that can first present a concise list of possible matching patients and records and allows for minor errors such as typos and “sounds-alike” naming. Once the user selects the specific patient, the HIE must then search its databases for records identified as belonging to that specific patient. The identified records may be contained within the HIE (centralized) or held by an affiliated source (federalized) or may be a hybrid. The list of possible matching patients is returned to the requester, who then selects the patient. A request is then sent through the system to find any available documents for the specified patient. From the list of available documents, the provider selects which document(s) he or she wishes to actually see or review. This selection initiates a query for the actual document(s), which is then retrieved and displayed for the provider. For the HIE to successfully present a comprehensive list of records for the specified patient, there must be some type of patient matching process in place.

**Process/method for locating patients outside the jurisdiction of your HIE (an external HIE query)**

Presently, inter-HIE patient inquiries can occur through the National Health Information Network utilizing the Integrating the Healthcare Enterprise Cross-Community Patient Discovery (XCPD) profile.5 The XCPD offers a means to discover mutually known patients and a method to correlate the patient’s identifiers across those communities.
In conjunction with the XCPD, the Cross-Community Access (XCA) profile provides a method to assemble clinical documents (CCD, CCR, radiology [DICOM] images, and the like).

To illustrate, imagine that Dr. Graff has an encounter with his patient, Elizabeth “Beth” Hexom. Currently, Beth lives in Dallas; however, she recently moved there from Detroit. Thus, the majority of Beth’s past medical history is stored in the clinical systems of provider institutions in the Detroit area. Fortunately, Dr. Graff’s HIE has the ability to discover patient data that exists outside of the local, Dallas-based community. Dr. Graff directs a patient inquiry through the North Texas HIE utilizing the Cross-Community Patient Discovery (IHE XCPD profile) protocol supported by the NHIN to the Detroit HIE and finds the relevant patient identifiers from the Detroit community who represent Beth Hexom. With this information, Dr. Graff can subsequently use the Cross Community Access (IHE XCA profile) to look for documents containing Beth’s past medical history held within the Detroit community.

It is important to note that according to the NHIN Patient Discovery Specification Version 2.0 that HIEs provide advanced probabilistic matching for patient inquiries. Specific restraints are also defined in order to mitigate the risk of false negative and false matches. These restraints require that inter-HIE searches can return multiple records from different healthcare organizations, but can only return one record per “assigning authority” or organization-specific MPI (data source). Therefore if a hospital contains a duplicate record for the requested patient, then the HIE is required to return an error message and is not allowed to return any results (from that data source) to the requesting HIE. The HIE is still able to supply the requesting HIE with non-duplicated patient records, but must block any information from the organization that has not resolved a compromised patient record within their own organizational/enterprise MPI. If multiple matches occur, the requester can resubmit the query with additional demographic information or contact the healthcare provider directly to verify the possible existence of clinical records for their patient.
Secure Data Transfer
HIEs employ a combination of approaches depending on their integration architecture. Some examples include an encrypted virtual private network for interfaced clients, secured sockets layer (a protocol for encrypting information over the Internet) or secured portal with applicable HIPAA safeguards (password complexity, timeouts, and so forth) for portal users.

Of the HIE survey respondents, all are committed to providing secure clinical data exchange as defined by ONC and HITSP, including public key encryption and hardware/software tokens. ONC recognizes that secure data transfer is essential from both a security and privacy perspective. A clinical message service that meets industry security standards will be utilized and required for HIE certification. For example, the Commonwealth of Virginia HIE is committed to implementing the HITSP Secured Communication Channel Transaction to provide the mechanisms to ensure the authenticity, integrity, and confidentiality of transmissions, and the mutual trust between communicating parties.\textsuperscript{8,9} Their objectives include providing:
- Mutual node authentication to assure each node of the other’s identity
- Transmission integrity to guard against improper information modification or destruction while in transit
- Transmission confidentiality to ensure information in transit is not disclosed to unauthorized individuals, entities, or processes

Data stewardship of the data maintained by the HIE
Health data stewardship is a responsibility, guided by principles and practices, to ensure the knowledgeable and appropriate use of data derived from individuals’ personal health information. Data stewardship is everyone’s responsibility. This is a particularly sensitive issue in healthcare today because there is no federal law establishing precise ownership rights or responsibilities. Health data stewardship encompasses the responsibilities and accountabilities associated with managing, collecting, viewing, storing, sharing, disclosing, or otherwise making use of personal health information.\textsuperscript{10}

This is further complicated by HIPAA because organizations must consider the relationship between covered entities, organized healthcare arrangements, business associates, and the obligations for disclosure of information to patients. EMPI managers must consider how data should be distributed in the HIE ecosystem. A patient record source can control data distribution by specifying which entities should see patient data either during patient query or during patient identifier cross reference notification. The patient can control which providers should see their data using a patient consent portal by using the portal to grant or revoke the consent given to a provider.

In 1973, a task force was formed at the US Department of Health, Education, and Welfare—now Health and Human Services—to look at the impact of computerization on medical record privacy. The task force developed a code of fair information practices, consisting of five clauses: openness, disclosure, secondary use, correction, and security.\textsuperscript{11} The value of fair information practices not only rests in providing a framework for privacy laws, but in forming the foundation of an individual organization’s privacy policy, regardless of whether it is a private, public, or nonprofit organization.\textsuperscript{12} The principal
components embodied in this framework are those that make sense for any organization adopting enhanced HIT. They include:

» Individual access
» Correction
» Openness and transparency
» Individual choice
» Collection, use, and disclosure limitation
» Data quality and integrity
» Safeguards
» Accountability

Health data stewardship is becoming increasingly important—not only to ensure privacy protection, but also to ensure the data used to make decisions are sound and properly maintained and retained. Within healthcare, the importance of health data stewardship has arisen because of concerns regarding data management—especially in light of automated collection and enhanced data mining tools that potentially make electronic health data more vulnerable to risk—again, not only in terms of privacy protections, which are of utmost importance, but also in the quality of data with which health-related business and clinical decisions are made.

**Summary**

Ensuring data quality is not a trivial task. Ideally, the health data in an electronic record should be accurate, up-to-date and complete; but unfortunately the real world is far from ideal. High-quality data requires us to have a very clear understanding of the meaning, context, and intent of the data—unambiguous and, ideally, standardized computable definitions of data that can form the basis for future safe decision making.

To facilitate data quality, the ultimate goal of any HIE should be accurate identification of the patient. HIE patient identity and administration has three patient identification profiles: (1) the patient identifier cross-reference profile that matches patients by cross-referencing IDs; (2) the patient demographics query profile queries a central patient information server; (3) patient administration management knows where the patient is, was, or is going. In addition the HIE should assign a unique patient/person identifier by using advanced record matching techniques, for example, probabilistic algorithms, and manual processes, as needed.

Nine influences have been identified as industry standards: system interfaces, algorithms, unique identifiers, business processes, data accuracy, data quality, training, and medical devices. Very high MPI duplication rates have been identified in all arenas. Meeting industry standards regarding data quality could therefore produce tremendous benefits both in terms of monetary savings and quality of care.

Quality information is essential to all aspects of today’s healthcare system, so improving the quality of data, information, and knowledge is paramount as we transition from paper to EHRs. Many errors and adverse incidents in healthcare occur as a result of poor data and information. In addition to threatening patient safety, poor data quality increases healthcare costs and inhibits health information exchange, research, and performance measurement initiatives.

HIM professionals play a critical role by leading initiatives related to standards, technologies, education, and research that are vital for capture, use, and maintenance of accurate healthcare data and facilitating healthcare’s electronic evolution.
Notes


5. Ibid.


7. US Department of Health and Human Services. NHIN Patient Discovery Production Specification, Section 3.1.6, Returning Multiple Entries.


13. The HIE would be wise to invest in and implement the following industry standards:
   2. ASTM–E2553, Guide for Implementation of a Voluntary Universal Healthcare Identification System (no standards exist for the data elements);
   3. HITSP TP22, Patient ID Cross Referencing Transaction Package;
   4. HITSP T23, Patient Demographics Query Transaction;
   5. Health Level Seven (HL7) Version 2.5 or higher.
References


