Healthcare Information Assurance: Identification Issues

Ludwig Slusky, California State University, Los Angeles, CA 90032; E-mail: lslusly@calstatela.edu

ABSTRACT

The paper summarizes identification issues pertaining to healthcare information assurance, including National Provider Identifier, identification in HIPAA, Electronic Medical Records, Electronic Data Interchange, and Disease Management. In conclusion, it discusses the need for further research of interlinks and dependencies among various identifiers of healthcare information to support confidentiality, integrity, and availability of a trustworthy national healthcare information system.

1. INTRODUCTION

Healthcare is a trillion-dollar industry in the United States, accounting for 14 percent of the nation's gross domestic product, with about 10 million employees sharing approximately 400 job titles. While the industry grows, it is also undergoing rapid transformation in the area of Web-based activities and increased security needs.

Healthcare providers and insurance companies amplified their Internet presence with Web-based applications such as online doctor-patient interactions, appointment scheduling, patient records administration, electronic claims, and online referral to specialists, computerized physician order-entry, dissemination of healthcare information over wireless networks to laptop computers and other devices, and many other forms of information delivery.

This paper reviews some issues of identification, de-identification, authentication, privacy and security for Healthcare Information Assurance and their application to a trustworthy national healthcare information system.

2. HIPAA

The Health Insurance Portability and Accountability Act (HIPAA), passed by Congress in 1996, recognizing the need for privacy and security protection of healthcare information, set national standards for identifiable Protected Health Information (**PHI).** It requires all health plans, healthcare clearinghouses, and healthcare providers that operate with PHI to comply with HIPAA's Privacy Rule and Security Rule.

2.1 HIPAA Components and Regulations

PHI is characterized as the following:

- It describes past, present, or future physical or mental health, or condition of an individual; or
- It describes a provision of health care to an individual; or
- It describes a payment for the provision of health care to an individual; or
- It identifies or provides a reasonable basis to believe it can be used to identify an individual.

Three main components of HIPAA are:

- Privacy of patients' PHI with national standards (The Privacy Rule),
- Security of electronic transactions with patients' PHI (The Security Rule), and
- Transactions and code set standards (for claims, enrollment, eligibility, payment, coordination of benefits, etc)

The Privacy Rule implemented on April 14, 2003 has privacy provisions applicable "to health information created or maintained by health care providers who engage in certain electronic transactions, health plans, and health care clearinghouses." [HSS 2006] It protects confidentiality of the individual's PHI when this PHI is used or disclosed in any form—paper, oral, electronic. HIPAA privacy and security regulations enforce accountability and apply to healthcare providers and to anyone who provides financial, legal, business, or administrative support to health care providers or health plans.

HIPAA privacy and security regulations enforce accountability and apply to healthcare providers and to anyone who provides financial, legal, business, or administrative support to health care providers or health plans.

HIPAA Transaction and Codes Sets regulations require that transmission of all healthcare data electronically be based on standard transactions, code sets, and identifiers. Thus, for Electronic Data Interchange (EDI), HIPAA has identified ten standard transactions (e.g., claims and encounter information, payment and remittance advice, and claims status and inquiry) and the code sets to be used in those transactions.

The four categories of code sets for claims are: pharmacy code set (from National Council for Prescription Drug Programs) and dental, professional, and healthcare institutional code sets (all three from Washington Publishing Company)

HIPAA regulations are based on the following Key Concepts:

- *Principle-based*. Complying with a series of security best practices and principles.
- *Reasonableness*. Mitigating all reasonably-anticipated risks by balancing resources and business requirements against the risks.
- *Full compliance.* Having workforce members of the covered entity in compliance with the regulations.
- *Documentation*. Having security processes, policies, and procedures approved and documented.
- *Ongoing compliance.* Revising security policies and procedures as needed, providing regular security training, and building awareness of the workforce.

2.2 HIPAA Security Rule

The HIPAA Security Rule established "national standards for the security of electronic health care information… This final rule specifies a series of administrative, technical, and physical security procedures for covered entities to use to assure the confidentiality of electronic protected health information." [Security 2003]

It specifies administrative, technical, and physical security procedures for covered entities to assure the confidentiality of "all electronic protected health information the covered entity creates, receives, maintains, or transmits." [Security 2003] It requires that each covered entity engaged in the electronic maintenance or transmission of identifiable health information pertaining to individuals assesses potential risks and vulnerabilities to electronically maintained or transmitted healthcare information and develop, implement, and maintain appropriate security measures to protect that information. [HSS 2006]

While HIPAA Privacy Rule applies to all PHI, the HIPAA Security Rule applies only to the **electronically maintained** or **transmitted** subset of PHI (ePHI).

Three basic categories of security mechanisms are: administrative procedures, physical safeguards, and technical security mechanisms. In addition, the security solutions must provide protection against the following:

- any reasonably anticipated threats or hazards to the security or integrity of such information;
- any reasonably anticipated uses or disclosures of such information that are not permitted or required by the Privacy Rule

Other important components of Healthcare information Systems that need protection are Electronic Medical Records EMR (also called electronic Health Records) and National Provider Identifiers (NPI).

Identification, de-identification, and authentication are common security concerns for various healthcare applications. They are addressed within the scope of the HIPAA Security Rule.

2.3 Implementation of HIPAA Privacy and Security Rules

A Privacy and Security Office in charge of consistent HIPAA compliance is responsible for Risk Analysis and Evaluation, as well as reporting of suspected security incidents and incident handling, including documentation, determination of notification requirements, remediation, and reporting to management. Overall, the HIPAA security standard requires comprehensiveness in terms of all aspects of security; scalability; and technological neutrality. Therefore, HIPAA Security Rule implementation must follow the guiding principles listed below:

- *Scalability.* All sizes of covered entities must comply with the rule, from the one-person doctor office to the insurance company with thousands of employees.
- *Comprehensiveness.* Principle of "defense in depth" as a unified security approach
- *Technology neutral.* No requirement for specific security technology (firewall or IDS) making selection a provider's choice.
- *Internal and external security threats* protection. Must protect ePHI against both internal and external threats.
- *Risk analysis*. Must regularly conduct thorough and accurate risk analysis.

Security Rule protecting identifiable PHI makes distinction between ''required'' and ''addressable'' specifications. Required implementation specifications are mandatory and must be met. Addressable specifications, depending on the specifics of the covered entity environment (size, capability, risk), are implemented as follows: if the covered entity determines that a given addressable specification is a reasonable and appropriate safeguard in its environment, it must implement the specification; otherwise, the covered entity may implement another equivalent measure or choose to not implement or substitute it at all if the standard can be met in some other way but to the same end result. [Security 2003]

Required implementation specifications are listed in Table 1.

Examples of addressable implementation specifications include Workforce Security with Clearance Procedure; Facility Access Control with Access Control and Validation Procedures; Access Control with Encryption and Decryption.

Table 1. HIPAA security rule "required" specifications

Standards	***Implementation of Required Specifications***
Administrative Safeguards	
Security Management Process	Risk Analysis
	Risk Management
	Sanction Policy
	Information System Activity Review
Information Access Management	Isolating Health Care Clearinghouse Functions
Security Incident Procedures	Response and Reporting
Contingency Plan	Data Backup Plan
	Disaster Recovery Plan
	Emergency Mode Operation Plan
Business Associate Contracts and Other Arrangements	Written Contract or Other Arrangement
Physical Safeguards	
Device and Media Controls	Disposal
	Media Re-use
Technical Safeguards	
Access Control	Unique User Identification
	Emergency Access Procedure

3. ELECTRONIC MEDICAL RECORDS (EMR)

"The EMR is the core component around which the totality of clinical care IT progress will necessarily revolve." [Hagland 2006]

Identification and authentication of Medical Records Information, i.e., assuring a reliable match of EMRs with patients, involves complex technical and social issues for potentially over 300 million EMRs accessible through a national health information network. It should be based on a technology-assisted authentication of the patient and an EMR compiled from multiple sources (and locations). However, such technological solution is likely to lead to a number of false-positive record authentications for a particular patient and, consequently, may result in significant health harm. Another problematic situation is authentication under accident or severe disability constraints: the patient may be unconscious or unable to communicate. Whatever the solution is it must strengthen portable EMRs and measures of their safety.

3.1 Identification of EMR

EMR Identity Management with the network-wide patient identifier helps to alleviate these problems and to integrate and exchange the patient's clinical and administrative information dynamically within a Healthcare Information Network (HIN), which may include hospitals, primary care physician, specialists, ambulatory care centers, etc.

The scope of the EMR is characterized by vastly different dimensions, such as the following:

- Continuum of care such as preventive, acute, post-acute, sub-acute chronic, long-term care, community and home care.
- Local needs and practice patterns that vary significantly in the information and functionality across the entire country.
- Clinical specialties such as emergency care, laboratory, radiology, pharmacy, intensive care, medical and surgical services, etc.
- Industry sectors such as hospitals, physicians and other clinicians, payers, community and social service organizations, government/public health, and even life science companies.

This inconsistency in data categories is further complicated by a patient's history, which may span many **legacy identifiers**. Thus, information can be sourced from multiple databases that were established over a long period of time. It may also include SSN, driver's license, individual hospital account and medical record numbers, lab system identifiers, pharmacy system identifiers, that had been changed or not valid any longer.

The solution to such disparity of identifying data is a cross-reference repository of identifiers across all systems, such as EMPI - Enterprise Master Person Index, which evolved from the previous hospital master patient index.

An Enterprise Master Patient Index (EMPI) can be described as a database that contains a unique identifier for every patient in the enterprise (including medical centers, clinics, practice offices, etc.) and provides a cross-reference for all systems, data records and applications throughout the healthcare information network. All registration systems would look to the EMPI to obtain patient information based upon several defined identifiers.

An EMPI will have either the deterministic indexing where one can search based on an exact match of the identifying aggregate data (e.g., combination of name, SSN, date of birth, sex) or the probabilistic searching mechanism based on truncated search data (e.g., truncated last name). A widely known example of it is a search using Soundex formula, which indexes names by their sound when pronounced in English so that matching can occur despite minor differences in spelling. Algorithm for Soundex formula are implemented in various computer languages, including Visual Basic. [Gillham 2001]

Implementation of an EMPI depends on the system architecture and may require conversion of the IT systems or a merge of the medical records. There are two types of patient/record identification: passive identification based on the existing patient's ID document (one factor identification and authentication) and active identification using the EMPI data and algorithms.

Development of EMR is progressing in the USA and in other countries. In the USA, there are plans to develop portable EMRs for every individual. However, a maze of EMR ownership and frequent incompatibility of technical, procedural and clinical requirements are hindering integration of EMRs. Thus, EMR systems are often lacking interoperability, offered on a variety of hardware and software platforms. Their approach and design are not consistent and they have no uniformity in interfaces, vocabularies, coding systems.

The key integration issues of EMRs lay in three areas: identification, authentication, and access control within the entire healthcare delivery chain; network security; and stakeholder commitment.

3.2 De-identification of EMR

One important approach for sharing patient publicly while protecting its privacy is de-identifying healthcare records. De-identified healthcare information (e.g., aggregate statistical data or data stripped of individual identifiers) requires no individual privacy protections and is not covered by the Privacy Rule.

New approaches for data de-identification have emerged to improve quality of research while protecting privacy. Among them, the following two methods of de-identification of PHI are most common:

- Statistical de-identification is performed by a qualified statistician who using accepted analytic techniques concludes that the risk of identification is substantially limited, i.e. that the information used alone or in combination with other reasonably available information is unlikely to identify the subject of the information.
- "Safe-harbor" de-identification method allows a covered entity or its business associate to de-identify information by removing specific PHI identifiers.

Researchers indicate that only 30-60% of all personally-identifying information can be found using the straightforward approach of global searching by the patient's name and replacing all occurrences with a pseudo name. Identifying information is often hidden in other correlated data and in the written free-form notes and letters exchanged among doctors. Other techniques yield to much better results to minimize risk to patient confidentiality. [Sweeney 1996]

For example, Scrub system uses multiple detection algorithms executed in parallel to label contiguous characters of text. Each detection algorithm is designed to recognize only one specific entity (e.g., name or address or date, etc.)

These detection algorithms are employed in a way similar to speech recognition: they use local knowledge sources (e.g., area codes, first names, medical terms) to determine whether searched words "sound" like identifiers (e.g.., medical terms, names). Then, the algorithm with the highest precedence and the greatest certainty above a minimal threshold prevails and its results may be made available for future use. [Sweeney 1996]

The accuracy of this technique is relatively high:

- 100% for well-defined references in the upper-lower case counterparts or numerical codes (such as names, addresses, organizations, cities, states, zip codes and phone numbers); however, this accuracy drops down to 94% when such reference is presented in all upper case letter configuration.
- 99% for more obscure references (such as nick names, abbreviations, ID numbers)
- 95% for references not distinguished by upper-lower case (95%)

HIPAA "Safe Harbor" de-identification of EMR requires that each of the 18 identifiers of the individual or relatives, employers, or household members of the individual must be removed from medical record information in order for the records to be considered de-identified. Examples of such identifiers include names, address, all elements of dates (except year), phone, email, Account numbers, biometric identifiers, and others.

"Safe harbor" de-identification may include the assignment of re-identification codes to the de-identified healthcare record information. These re-identification codes must be securely managed to prevent unauthorized access to information linking these codes with corresponding PHI.

4. NATIONAL PROVIDER IDENTIFIER (NPI)

In addition to the EMR identification, HIPAA requires standard unique identifiers for health care providers, as well as for health plans. National Provider Identifier (NPI), due for compliance by May 23, 2007 for large health plans and a year later for small plans, is the standard unique health identifier assigned to health care providers and an important component of Healthcare Information Assurance. NPI's 10-position number has a 9-position unique identifier plus one position for the check-digit data validation. It is intelligence-free, i.e., does not itself convey information about the provider, and it is compatible with health insurance card issuer standard.

All health care providers are eligible to receive an NPI, but only entities covered by HIPAA are required to use the NPI when submitting and processing electronic transactions. (For example, x-ray technicians and dental hygienists may apply for an NPI but are not required to have an NPI.)

Provider types affected by NPI requirements are legal entities characterized as entity type 1 or entity type 2:

- entity type 1 - individuals (e.g., physicians, dentists, nurses, pharmacists, and physical therapists),
- entity type 2 - organization health care providers and suppliers (e.g., hospitals, ambulatory care facilities, laboratories, HMOs, group practices, suppliers of durable medical equipment, pharmacies, etc.

Additionally, an organization may designate subparts of a covered organization healthcare (like departments, divisions), which are not legal entities themselves but need to be uniquely identified in standard transactions with their own NPIs (that does not apply to individuals). [NPI 2006]

The NPI is to replace all "legacy" identifiers that are currently used, such as Provider Identification Numbers (PINs), National Supplier Clearinghouse (NSC) numbers, Unique Physician Identification Numbers (UPINs), etc. It is also permitted to be used for other lawful non-HIPAA transactions and identification.

The Centers for Medicare & Medicaid Services (CMS) developed the National Plan and Provider Enumeration System (NPPES) to assign NPIs. The NPPES is designed to accept health care provider data (including those who do not participate in Medicare) for unique identification and assigning an NPI. The NPPES performs three required functions: (a) assign a single, unique NPI to health care provider; (b) collect/maintain information about health care providers; and (c) reactivate or deactivate NPIs. It also disseminates NPPES information. [NPPES 2006]

5. ELECTRONIC DATA INTERCHANGE (EDI) AND X12 STANDARD

Electronic Data Interchange (EDI), a fundamental component of the healthcare information network, is the computer-to-computer exchange of business data in standard formats (for example, Purchase Orders, Invoices, Shipment Notices, Health Care Claims). The data is structured by patient's identifier, transaction type, and code sets.

The X12 standard commonly used in healthcare networks defines data structure for electronically exchanged documents using 315 or more EDI transaction sets. The documents are organized as data separated by "delimiter" characters (not as fixed length fields) and include: Transaction Sets consisting of delimited data; Functional Groups consisting of related Transaction Sets; and Interchange wrapping Functional Groups. X12 standard does not define a transmission type.

A new Context Inspired Component Architecture (CICA) standard - the XML equivalent of the current X12 standard - enables XML-built business documents in a cross industry setting. The large-scale structure of this architecture has seven discrete levels of granularity - each level builds on the levels below it: from DOCUMENT level to PRIMITIVE level. CICA helps to facilitate a common reusable vocabulary across multiple industries and creates an environment for convergence with other standards of organizations, industry associations or data content committees. [CICA 2002]

6. IDENTIFICATION IN DISEASE MANAGEMENT (DM)

Disease Management (DM) is another area of healthcare information where patient identification is required.

DM has its own specificity in implementation of three tenets of information security: confidentiality, integrity, and availability. The major application processes of DM systems that require security protection are analytic systems, predictive modeling, stratification algorithms. They are at the core of DM and require full confidentiality, integrity, and availability protection. *Analytic systems* integrate data from several sources and provide early identification of members needing services. *Predictive modeling* of medical and pharmacy data (with automated health risk assessment tools and electronic clinical laboratory results) helps to identify high risk members. *Stratification algorithms* are designed to provide informational guidance to a nurse about health status and an acuity level of a patient. [Johnson 2004]

Three components of technological platform for healthcare systems deserve special attention: Web-based applications and data banks, email, and online biometric devices. Web-based healthcare information systems with online data banks for diagnostics, care progress tracking tools, and health care alerts delivered to a portable device are utilized by clinical practitioners more than ever. This information is often integrated into email or other communication systems that need to be protected. It should be noted that biometric devices coupled with communication mechanisms, which provide consistent collection of clinical information, automated transmission of results, and tracking of findings, have inherent information security vulnerabilities. [Johnson 2004]

7. TRUSTWORTHY HEALTHCARE INFORMATION SYSTEM (THIN)

The ultimate goal of IT development at healthcare organizations is building a trustworthy nationwide healthcare information system. The identifiers used for identification, de-identification, and authentication need to be interlinked to support Access Control with confidentiality, integrity, and availability of healthcare information.

The ultimate goal of IT development at healthcare organizations is building trustworthy healthcare information systems to meet the requirements of confidentiality, integrity, and availability specific to healthcare industry.

The information and processes to be protected are diverse in context, workforce preparedness, applications, geography, and technological platforms. Identification, de-identification, and authentication of protected patient information (PHI), medical records (EMR), and providers (NPI) in all kinds of healthcare systems are important for assuring *confidentiality* of information and are addressed in HIPAA's Privacy and Security Rules.

Integrity, i.e. protection from intentional or accidental unauthorized changes, and *availability* are obviously vital requirements for healthcare information and processes. Among two threats to availability of the healthcare information and processes – (a) human actions or natural disasters and (b) network intrusions like denial-of-service – the first threat is more likely to occur than the second one. All three types of control mechanisms – administrative (e.g., access control policies, operating procedures, contingency planning), physical (e.g., off-site backup storage), and technical (e.g., fault-tolerance mechanisms) – play important roles in assuring availability.

The priorities for implementation of confidentiality, integrity and availability (and, consequently, the use of identifiers) may differ from one application to another. For example, in EMR systems confidentiality is clearly a priority, while in DM systems integrity and availability could be of the same or higher concern as confidentiality.

Identification plays a particularly important role in DM applications where risks are characterized by higher single loss expectancy and interlinks to affected members of population are vital (particularly where mass health disaster may occur).

Database vulnerabilities and threats can sufficiently be enforced with common DBMS controls. Although discretionary access controls are prevalent in current healthcare systems, for large systems the preference should be given to the mandatory (policy based) or role-based access controls (with established sensitivity of data and appropriate protection mechanisms).

Use of portable devices, removable media, and email is a common practice in healthcare networks and as such may also have many vulnerabilities. Identification of these devices may also be required to support a high level confidentiality.

EDI exchanging information among various entities is a centerpiece of many healthcare systems. The standard transactions for healthcare systems include identifiers and are covered by HIPAA Transaction and Codes Sets regulations.

8. FUTURE WORK

Integration of various healthcare systems will further complicate issues of identification, de-identification and authentication. Future research will address interlinks and dependencies among various identifiers of healthcare information for modeling a trustworthy nationwide healthcare information system.

9. CONCLUSION

Identification and de-identification play an important role in authentication and security of healthcare information. Recently enacted regulations, while satisfying the needs of relevant systems, do not go far enough in viewing and defining various types of healthcare identifiers in an integrated manner. Model for trustworthy nationwide integrated healthcare systems will have to incorporate critical identifiers and the relationships among them to satisfy requirements of integrated healthcare information assurance.

REFERENCES

HSS (2006). *Medical Privacy - National Standards to Protect the Privacy of Personal Health Information*. Office for Civil Rights – HIPAA. US Department of HSS. http://www.hhs.gov/ocr/hipaa/

Security (2003). *45 CFR Parts 160, 162, and 164. Health Insurance Reform: Security Standards; Final Rule*. Part II. Department of Health and Human Services. Office of the Secretary. Federal Register. Thursday, February 20, 2003

Sweeney, L. (1996). *Replacing Personally-Identifying Information in Medical Records, the Scrub System*. In: Cimino, JJ, ed. Proceedings, Journal of the American Medical Informatics Assoc. Washington, DC: Hanley & Belfus 1996:333-337.

NPI (2006). NPI – Medicare Policy on Subpart Designation. Medical Learning Network. CMS. http://www.cms.hhs.gov/MLNMattersArticles/downloads/SE0608.pdf

NPPES (2006). National Plan and Provider Enumeration System (NPPES). https://nppes.cms.hhs.gov/NPPES/Welcome.do

CICA (2002). ASC X12 Reference Model: Context Inspired Component Architecture (CICA). ASC X12C Communications and Controls Subcommittee. Technical Report Type II. October 2002.

Gillham, Brian (2001). Function for the Soundex Formula. VBCity.com. VB5, VB6. 4/18/2001 http://www.vbcity.com/pubs/article.asp?alias=soundex

Johnson, Alison et al (2004). Disease Management: Changes and Challenges. HCT Project Volume 2. July 2004. http://www.hctproject.com/documents.asp?d_id=2768

Hagland, Mark (2006). Nine Tech trends: Electronic Medical Records. Healthcare Informatics Online. February 2006. http://healthcare-informatics.com/issues