Journal of Medical Systems, Vol. 29, No. 4, August 2005 (© 2005)
DOI: 10.1007/s10916-005-5890-1

Public Health Data Collection and Sharing Using HIPAA Messages

Min Wu,[1,4] Tian Zhao,[2] and Changshan Wu[3]

Public health information has significant value for doctors, public health officials, epidemiological researchers, the general public, and government agencies. Unfortunately, these data are difficult to obtain and are typically collected on as-needed basis and maintained locally. This localized process unavoidably limits the access to important public health data by its users. Moreover, the diversity of data transmission standards and collection techniques make the collected data less usable. This paper proposes a new standardized public health information system based on the HIPAA (Health Insurance Portability and Accountability Act) messages, which are the standard transactions between hospitals and insurance companies. In particular, this paper explores the applicability of HIPAA messages as a data source and transmission standard, and proposes a prototype design of a new system to collect and share public health data using HIPAA messages.

KEY WORDS: HIPAA; public health; data collection; system design.

INTRODUCTION

Public health information has significant value for various users, including (1) public health officials and doctors, who make decisions through analyzing historical disease data and associated demographic information, (2) epidemiological researchers, who conduct extensive studies in analyzing causative agents, vectors, hosts, population at risk, and their relationships to geographical environments, (3) general public, who are willing to access the data and be involved in public participatory decision-making, and (4) government agencies, such as the Center of Disease Control and prevention (CDC), who are responsible for national emergency, especially with the concern of biological warfare.(1) Because of the

[1] Department of Health Sciences, P.O. Box 413, University of Wisconsin-Milwaukee, Milwaukee, Wisconsin 53201.
[2] Department of Computer Science, University of Wisconsin-Milwaukee.
[3] Department of Geography, University of Wisconsin-Milwaukee.
[4] To whom correspondence should be addressed; e-mail: wu@uwm.edu.

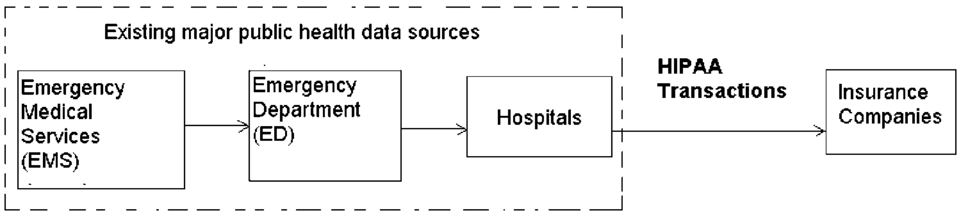

Existing major public health data sources
Emergency Medical Services (EMS)
Emergency Department (ED)
Hospitals
HIPAA Transactions
Insurance Companies

Fig. 1. Existing major public health data sources and HIPAA data.

importance of public health data, in recent years, a number of public health information and surveillance systems have been developed to collect public health data, such as those for the early detection of bio-terrorism events. According to a comprehensive summary[2] of the existing public health systems based on data sources, transmission standards, and collection techniques, most of public health data are collected from emergency medical service (EMS), emergency department (ED), and hospitals (see Fig. 1). ED is the most popular data source and public health information from ED has been collected by most public health systems, such as the Indianapolis Network for Patient Care (INPC)[3] and the Syndromic Surveillance Information Collection (SSIC) system.[4] In addition to ED, EMS and hospitals, as well as others (e.g. physician groups and primary care units), have been adopted as major public health data sources by a few public health information systems.[1] Associated with data sources, different data transmission standards and collection techniques have been developed. Major data transmission standards include Text, eXtensible Markup Language (XML), Simple Mail Transfer Protocol (SMTP), Open Database Connectivity Standard (ODBC), and Health Level 7 (HL7) standard.[1] Moreover, data collection techniques, such as fax, e-mail, file transfer, reports, and data mining have been developed for public health information transmission. As a summary, the existing public health information systems have adopted a variety of data sources, transmission standards, and collection techniques.

To date, the existing public health information systems have not provided adequate support to the government agencies and public health researchers due to some unresolved issues. One major issue is related to data sources. Most of public health data are typically collected on as needed basis and maintained locally. This localized process unavoidably limits the access to important public health data by health care providers and the public. Moreover, it creates problems for data analysis. For example, local data may not include enough incidents with demographic diversity to account for the differences between functional groups, ages, as well as other factors. In addition to the problems associated with data sources, the diversity of data transmission standards and collection techniques makes the collected data less accessible to public health researchers. In particular, most local healthcare data are paper-based and may not be reported or transferred in a timely fashion. In addition, the major electronic data transmission standards of the existing public health systems, such as *Text*, *XML*, or *e-mail*, are not standardized means for data sharing.[1] Therefore, standardized electronic transmission will be more valuable for supporting public health data reporting, especially with the emerging reporting code management tools and automatic reporting techniques, such as text templates. To reach this goal, the CDC and eHealth Initiative Public Private Collaboration have developed implementation guidelines for public health reporting of chief complaint information using version 2.3.1 of HL7 standard protocol.[5] Subsequently, HL7 messages have been utilized in the Real-time Outbreak and Disease Surveillance (RODS) system as the data transmission standard.[6] However, HL7 is not mandatory by federal laws and electronic medical records with HL7 have not been utilized sufficiently by public health practitioners. Therefore, there is a need to explore the possibility of developing a new national data collection system with standardized data sources, transmission standard, and data collection techniques.

This paper proposes a new standardized public health information system based on the Health Insurance Portability and Accountability Act (HIPAA) messages—standard transactions between hospitals and insurance companies (see Fig. 1). HIPAA not only defines security and privacy regulations, but also provides standard transactions and code sets. It has become a national electronic communication standard and has been widely complied by healthcare providers and supported by information technology companies. HIPAA Transaction Sets are standardized federal regulations for all health care providers and payers, so the HIPAA message transmissions should be deployed to a wide geographic area. In spite of its profound impact on overall healthcare industry, electronic communications and transactions, the existing public health systems are inadequate to support data collection based on HIPAA messages, and little is known about the underlying relationships between HIPAA and public health data. This paper explores the applicability of HIPAA messages as a new data source and transmission standard, and proposes a prototype design of a new system to collect and share public health data using HIPAA messages.

The rest of paper is organized as follows: Section 2 gives an overview of HIPAA, with emphasis on HIPAA message content, HIPAA de-identification regulations, and HIPAA message transmission techniques; Section 3 explores the applicability of HIPAA messages as a standardized public health data source and transmission standard and discusses applicable techniques to extract public health information from HIPAA messages. In Section 4, we describe the prototype design of a pubic health information system based on HIPAA messages; and in Section 5, we conclude and discuss future research directions.

OVERVIEW OF HIPAA

HIPAA Message Content

In HIPAA, Congress requires health plans, health care clearinghouses, and health care providers who conduct certain financial and administrative transactions electronically (such as eligibility, referral authorizations, and claims) to comply with each set of the final standards. HIPAA has three major components, including (1) Transactions and Code Sets, (2) Security Regulations, and (3) Privacy Regulations. An example of basic HIPAA transactions is shown in Fig. 2.

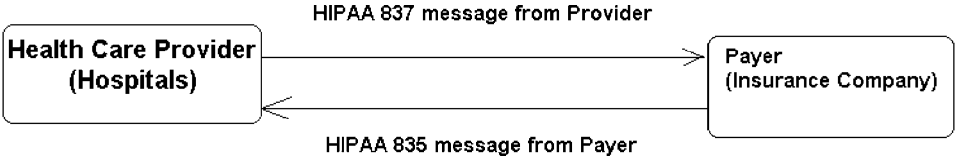

HIPAA 837 message from Provider
Health Care Provider (Hospitals)
Payer (Insurance Company)
HIPAA 835 message from Payer

Fig. 2. Basic HIPAA transactions.

In Fig. 2, a healthcare provider (such as a hospital) originates a transaction and sends claim information (HIPAA 837 message) to a payer (such as an insurance company). The HIPAA 837 message contains claim information about a person who holds a policy with the payer. The payer adjudicates the claim and sends an electronic remittance advice (RA) transaction (835 messages) back to the provider. Because the HIPAA 837 message contains a large amount of valuable health care information, it is virtually a gold mine for anyone who is interested in analyzing but has difficulty locating healthcare data.

HIPAA De-Identification Regulations

In the HIPAA and its new DHHS privacy provisions (De-Identification of Protected Health Information), it is required that public health information not be released when it is possible to be used to identify an individual.[7] HIPAA privacy regulations[8] define 18 individual personnel identifiers. While it is beneficial to release information to researchers and general public, confidentiality is an important issue and needs to be carefully addressed.[9] In particular, the data should be preprocessed to protect unauthorized disclosure of any individual's identity.[10] Such techniques include data aggregation, smoothing, and others. HIPPA privacy rules, however, do contain extensive exemptions for treatment, payment, research, and national priority activities. In particular, confidential public health information may be released to universities and drug companies for research purposes.

HIPAA Message Transmission Techniques

HIPAA standards have received wide technology support. Software industry leaders such as IBM and Microsoft as well as thousands of small software vendors have provided or begun to provide technical support for receiving, sending, parsing, validating, and transforming HIPAA transactions. Major software vendors such as IBM, Microsoft, and GE Medical Systems have their own integrated e-commerce systems that support a suite of functionalities ranging from electronic data interchange to data management systems. Some of the e-commerce systems are for hospital information management with support for billing transactions. To comply with HIPAA regulations, these systems have to be modified to process billing transactions in HIPAA format. For example, IBM's WebSphere[11] Data Interchange supports ASC (Accredited Standards Committee) X12 standards of HIPAA as base function, which enables customers to send and receive HIPAA transactions. It validates the inbound or outbound transmission and then transforms the transaction from X12 to XML. The support for HIPAA includes collaborations for HIPAA transactions that enable integration with customers' existing back-end data processing systems using the IBM WebSphere Interchange Server and the IBM WebSphere Business Integration Collaborations for HIPAA Transactions. Smaller software vendors provide HIPAA compliance component that can be integrated in existing e-commerce systems. For example, iWay corporation's HIPAA eBusiness Adapter is fully compatible with IBM WebSphere Data Interchange Server and can be used as a plugin for translating HIPAA transactions. Other vendors such as HIPAA Accelerator Inc. provide alternative medical management systems that address the needs of HIPAA compliance in the areas of medical record management; privacy and security protection; and information dissemination.

THE APPLICABILITY OF HIPAA MESSAGES IN PUBLIC HEALTH DATA COLLECTION

With knowledge about HIPAA message content, HIPAA de-identification regulation, and HIPAA message transmission techniques, three issues need to be addressed for developing a standardized public health information system with HIPAA messages. The first one involves whether HIPAA messages contain sufficient public health information. This is a prerequisite of utilizing HIPAA messages as a standardized data source for collecting public health information. The second issue relates to HIPAA de-identification regulations. The developed system should not violate the strict HIPAA privacy regulations. And the final one is the implementation issue, which involves whether public health information can be efficiently extracted from HIPAA messages. The following part of this section is devoted to address these three issues and explores the applicability of HIPAA messages in developing a standardized public health information system.

Public Health Information in HIPAA Messages

As a proof of concept, a study has been conducted to explore whether HIPAA messages contain sufficient public health information. Important public health variables, such as age, gender, diagnosis coded with the Ninth Revision of International Classification of Diseases (ICD-9), spatial location, etc. were identified through reviewing the literature in public health research areas, with an emphasis on the practical guide(12) for implementing syndromic surveillance. A number of identified public health variables are listed in Table I. Public health variables embedded in HIPAA messages were discovered through studying the transaction sets of HIPAA, in particular the HIPAA 837 messages sent from hospitals to payers. It shows that almost all the important public health variables can be found in HIPAA transaction segments (see Table I). For example, patient's gender information can be found in the HIPAA DMG (Demographics) segment. Moreover, patient's diagnosis, such as West Nile virus (coded as 066.4 in ICD-9) can be found in the HIPAA HI (high) segment. Overall, this study indicates that HIPAA messages contain most of public

Table I. Preliminary Study of HIPAA Messages for Public Health Variables

Public health variables	HIPAA transaction segments	Notes
Patient birth date	DMG	Available
Gender	DMG	Available
Marital Status	DMG	Optional
Race	DMG	Optional
ICD-9-CM coded diagnosis/procedure	HI	Available
Spatial location (zip code)	N4	Available
Date of Admission	DTP	Available
Others	—	—

health information, which may be extracted for developing a public health information system.

HIPAA De-Identification Regulations

With the concern of the possibility of violating HIPAA privacy regulations when reporting patients' health information, we explored de-identification methods to comply with these regulations. While HIPAA regulations allow healthcare providers to disclose data to public health officials, it requires that 18 individual personnel identifiers be removed from the HIPAA messages for general public. We collected some "de-identify" rules to remove portions of the personnel identifiers, as shown in Table II. More research needs to be conducted to develop more efficient de-identification methods. This preliminary study, however, shows that it is applicable to keep valuable public health information while complying with HIPAA privacy regulations.

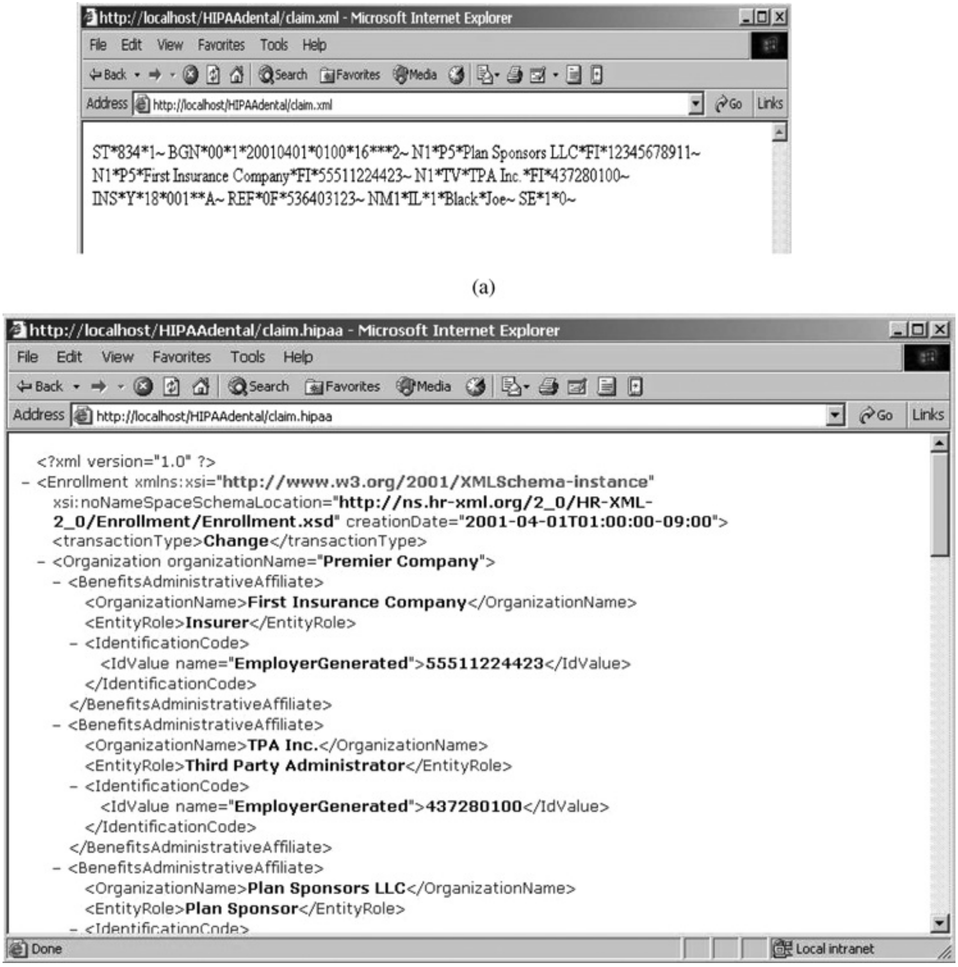
Extracting Public Health Information from HIPAA Messages

To test the ability to automatically extract public health information from HIPAA messages, we developed a system that transforms HIPAA messages to XML documents (and vice versa), and validates the HIPAA messages based on the published implementation guidelines. Its current implementation is limited to dental claims to provide a simple, yet comprehensive enough, application of the conversion to/from HIPAA messages. However, the general design guidelines of this system can be used as a reference in processing other types of HIPAA messages.

Table II. Preliminary Studies of Implemental Rules for "De-Identify" HIPAA Messages

Individual personnel identifiers	HIPAA transaction segments	Implemental rules
Patient name	NM	Remove (re-assignment)
Date of birth	DMG	Year level (aggregated)
Patient state/Zip code	N4	State level (initial three letters of Zip code)
Medical record numbers	REF	Remove (re-assignment)
Patient address (city)	N3	Remove
Date of admission	DTP	Year level (aggregated)
Others	—	—

The process involves the transformation of the incoming HIPAA messages by applying all the mapping rules that have been defined at design time. While the document is being transformed to XML, validation checks are performed. A "Rules file" (provided for the specifics of dental insurance claims) is utilized. Converting from XML to HIPAA is the reverse process, which utilizes the XSLT style sheet file to define the rules of mapping between the two formats. For example, the HIPAA message shown in Fig. 3a can be translated into an XML document shown in the screen shot in Fig. 3b. After transforming HIPAA messages to XML documents, we can easily apply some data manipulation operations on these documents such as erasing certain contents from the documents and retrieving public health information.


http://localhost/HIPAAdental/claim.xml - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites Media
Address http://localhost/HIPAAdental/claim.xml
Go Links

```
ST*834*1~ BGN*00*1*20010401*0100*16***2~ N1*P5*Plan Sponsors LLC*FI*12345678911~
N1*P5*First Insurance Company*FI*55511224423~ N1*TV*TPA Inc.*FI*437280100~
INS*Y*18*001**A~ REF*0F*536403123~ NM1*IL*1*Black*Joe~ SE*1*0~
```

(a)

http://localhost/HIPAAdental/claim.hipaa - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites Media
Address http://localhost/HIPAAdental/claim.hipaa
Go Links
Done
Local intranet

```
  <?xml version="1.0" ?>
- <Enrollment xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
    xsi:noNameSpaceSchemaLocation="http://ns.hr-xml.org/2_0/HR-XML-
    2_0/Enrollment/Enrollment.xsd" creationDate="2001-04-01T01:00:00-09:00">
    <transactionType>Change</transactionType>
  - <Organization organizationName="Premier Company">
    - <BenefitsAdministrativeAffiliate>
        <OrganizationName>First Insurance Company</OrganizationName>
        <EntityRole>Insurer</EntityRole>
      - <IdentificationCode>
          <IdValue name="EmployerGenerated">55511224423</IdValue>
        </IdentificationCode>
      </BenefitsAdministrativeAffiliate>
    - <BenefitsAdministrativeAffiliate>
        <OrganizationName>TPA Inc.</OrganizationName>
        <EntityRole>Third Party Administrator</EntityRole>
      - <IdentificationCode>
          <IdValue name="EmployerGenerated">437280100</IdValue>
        </IdentificationCode>
      </BenefitsAdministrativeAffiliate>
    - <BenefitsAdministrativeAffiliate>
        <OrganizationName>Plan Sponsors LLC</OrganizationName>
        <EntityRole>Plan Sponsor</EntityRole>
      - <IdentificationCode>
```

Fig. 3. Translating HIPAA messages to XML documents for public health information extraction. (a) A HIPAA message for insurance claim. (b) Insurance claim information in XML format.

The application also has the ability to perform *de-identification*. That is, the application can strip off the identification portion of the claim based on certain criteria provided by the user. The identification portion may be related to the insurance subscriber, the insurance provider, or the healthcare institution. Removing the identification section lets the system look at the claim records anonymously. This feature might be quite useful for history tracking purposes or for generating summary reports when the identity of the record is of no importance.

Filtering (querying records according to user-specified criteria) is another available feature of this application. Different portions of the HIPAA message content (identification, diagnostics, illnesses, etc.) can be used as the basis for filtering. When XML files are stored in a database (after transformed from an HIPAA message file), the database management system might be able to provide more extensive querying capabilities. The filtering feature of this application provides the capability to specify simple queries in case the user does not have access to such a database system. In summary, this study indicates that it is applicable to automatically extract public health information from HIPAA messages through transforming HIPAA messages to XML documents (and vice versa) and extracting valuable public health information from the XML documents.

SYSTEM DESIGN

Following the arguments that HIPAA messages are applicable in developing a standardized public health information system, this paper proposes a prototype system to collect public health information by intercepting and filtering HIPAA transaction messages. This system includes three components: (1) HIPAA message filter, (2) HIPAA database, and (3) Geographic Information Systems (GIS) (see Fig. 4). The HIPAA message filter exams the HIPAA 837 message originated from a hospital to insurance companies, collects useful public health information in that message based on predefined filter rules, and sends the result to the HIPAA database. Based on the data sent from HIPAA message filters, the HIPAA database provides functions for information input, storage, management, query, and supports GIS functions. The GIS component, including a map server, web server, and web browser provides functions for interactive display, GIS query, and GIS data dissemination. This prototype system provides a web browser as user interface supported by HIPAA message filters, HIPAA database, and GIS functionalities.

HIPAA Message Filter

The HIPAA message filter is a crucial component in the system and it consists of two parts: a message-converter, which converts HIPAA message to XML files, and an XML-filter, which processes the XML files to retrieve information useful for public health research (see Fig. 5). To retrieve information flexibly, we use a set of formatted filter rules to determine which parts of the XML information we

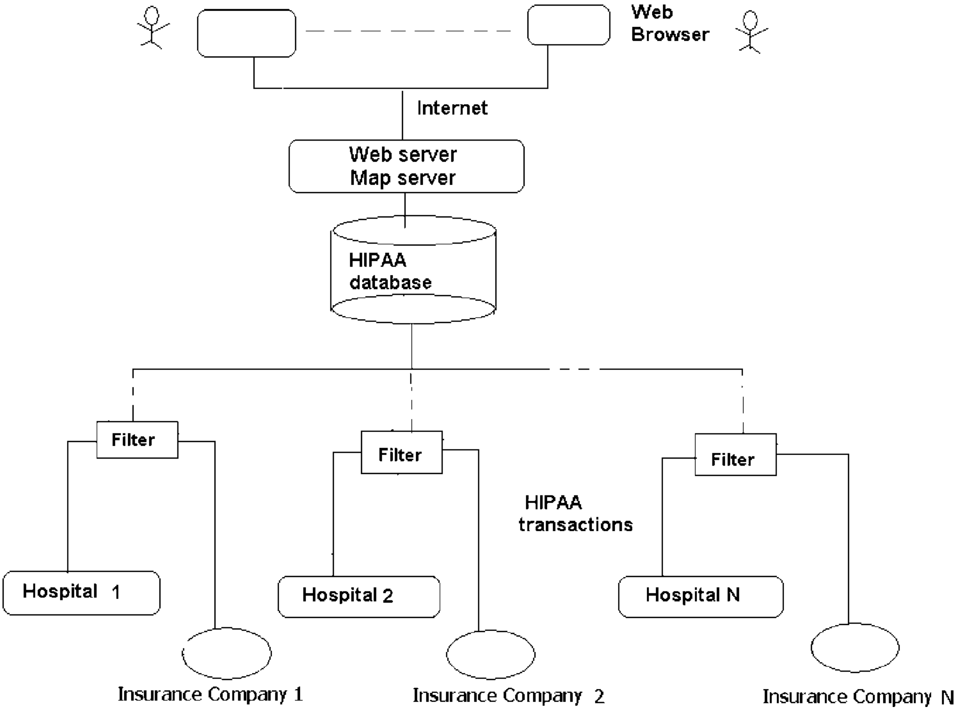


Fig. 4. The architecture of the prototype system.

want to keep. The filter is a client application installed in the host machine of the healthcare providers or payers. Thus, the XML-filter maintains the set of filter rules using either the local file system or a database of the host machine. Prior to the installation of the filter component, we will configure a default set of filter rules. Later, when requirement changes, we will be able to update the filter rules remotely via standard communication protocols (TCP/IP).

The design of the filter rules determines the overall structure of the XML-filter component. We could make the filter rules coarse so that we only filter out non-related data entries of the HIPAA messages. For example, the entries related to financial transactions of a particular patient's visit are removed. We choose to keep all the diagnostic information and some of the patient's geographic data. Such

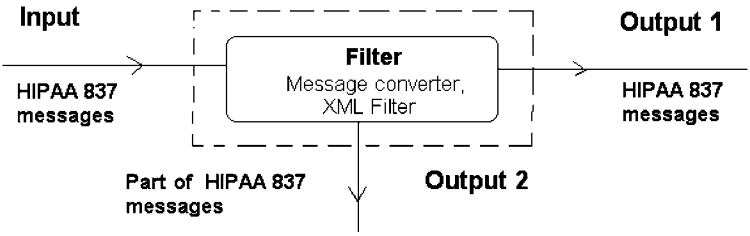


Fig. 5. The filter component in the system.

filtering policy can be viewed as simple transaction logging. Coarse filter rules are simple to implement and efficient for processing. However, the retrieved data may not be precise for our purpose of health demographic research. For instance, if we are interested in a particular kind of disease in one geographic region, then we have to search the whole data set logged by our filter component. Also, since the retrieved data is imprecise, we may have to store a large amount of data in our database system. Another potential drawback is that we are not able to observe the development of certain research experiments in real-time. For instance, we may be interested in detecting cases of fast-spreading contagious-diseases such as SARS as soon as the patients with such disease are discovered. With coarse filter rules, we can only first store raw data in a centralized database system and then mine useful information from the database. Therefore, we also explore the possibility of fine-tuned filter rules that can specify precisely what kind of health information we would like to retrieve and also we need to establish a mechanism to update such filter rules in real-time without stopping the filtering process. Designing fine-tuned filter rules is difficult not only because of the number of cases such as catalogue of diseases we have to consider but because we need to accommodate future requirement changes.

In order to conform the security and privacy requirements of HIPAA, our filter has a verifiable security mechanism and some privacy preserving properties. Since the filter is attached to the host machine of the health providers or payers, we need to provide sufficient security guarantee that the filter does not comprise the security of the software system in the host machine. Also, we follow a strict privacy standard so that the data retrieved by our filter does not reveal privacy information of the parties involved in the HIPAA messages. From the perspective of the host machine, the filter is an application from un-trusted sources. Therefore, limited system resource may be made available to the filter. For instance, the filter may only be able to use file system of specified directories and access network by connecting to a limited number of IP addresses via a specified communication protocol. Specifically, we implement our filter as a software plug-in written in platform independent language such as Java. Most computing platforms support Java Virtual Machine, which provides a security runtime environment for Java-based applications. The host machine could specify a security policy for running the filter programs and the security properties specified in the policy are enforced automatically by the virtual machine. The filter should not comprise the security of host machine either directly or indirectly by containing security loopholes that could be exploited by malicious entities. The main vulnerability of the filter is the communication channel. HIPAA security regulations require methods of protecting the data in transport, such as data encryption, secure sockets, secure shell tunneling, or the use of a virtual private network. We consider using some standard encryption techniques such as public-key cryptography to authenticate the identities of the central database and also the filters in the host machines. Also, using Secure Sockets Layer (SSL) Protocol can safeguard the data communication channel between the filters and the central database. The SSL protocol is supported by a number of language systems such as the Java Secure Socket Extension (JSSE) implemented in Java 2 standard development kit.

HIPAA Database

Connected with the HIPAA message filter, the HIPAA database component should have a clear scope and boundary of this HIPAA database application, which is mainly supporting for public health data in HIPAA messages. For example, we may limit the scope of the major users of the database to public health researchers. The logical of database design should use the relational data model for storing HIPAA transactions and utilize the hierarchical data model for data migration. The HIPAA relational database model includes numbers of tables, such as *Claim*, *Referral*, *Claim-Target*, *Clinical-Info*, *Message-header*, *Claim-Medical-Info*, *Drug-Info*, *Contact-Method*, etc. The data model for 837 transactions is more complex than the model for 835 transactions. The logical and physical design for the HIPAA database should be further optimized based on the feedback in the future. A more systematic database performance tuning is also required in the implementation of the system. If the database design for HIPAA transactions can be successfully validated and published, a new line of data mining research about HIPAA information could be possible.

Geographic Information Systems (GIS)

With the support of HIPAA database, the GIS component has three major functions (see Fig. 6), including (1) GIS interactive display, (2) GIS query, and (3) GIS data dissemination. In particular, GIS interactive display function comprises zoom in, zoom out, zoom last, zoom to full map extent, pan, identify, redraw, etc. These functions provide the flexibilities for researchers with emphasis in an area of their interests. For example, although these public health data are collected nationally, a researcher may be only interested in West Nile Virus in Wisconsin. In this case, he may use these functions to highlight the State of Wisconsin only. In addition to the display functions, GIS query provides necessary information with users' requests. GIS query is divided into two categories: attribute data query and spatial data query. Attribute data query involves searching by attribute information in the HIPAA database. For example, a user may query the geographic distribution of West Nile Virus by specifying the disease name. Spatial query indicates the searches according to spatial locations. For instance, a researcher may be interested

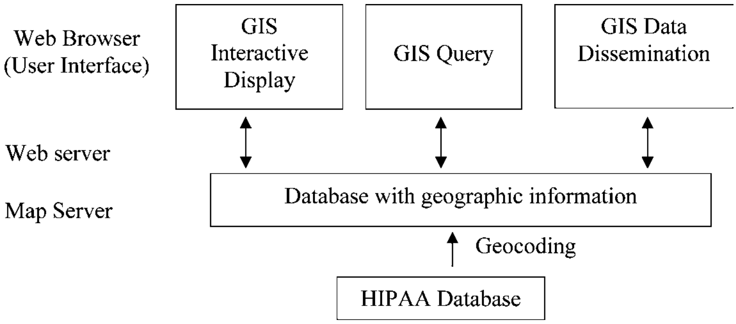


Fig. 6. System functionality and implementation process of GIS component.

in knowing how many accidents took place within 1 mile buffer of a high school. In this case, the researcher can query the information through the location and buffer of the high school using spatial query tools. The other important function of this system is the ability to disseminate public health data. Researchers can request public health data for further analysis. In particular, two formats of data will be provided. The first is the table format for researchers with little or no experiences in GIS and the second is the GIS database format for researchers with more GIS knowledge.

This system can be implemented using Visual Basic language based on the MapObjects libraries developed by Environmental Systems Research Institute (ESRI). In particular, the map display (e.g. zoom in, zoom out, etc.), attribute and spatial query (e.g. search for a specific disease), and data dissemination functions will be realized using the capabilities of the MapObjects. The ESRI Internet Map Server (ArcIMS) will be utilized to fulfill the applications on the Internet. The proposed system could follow typical three-tier architecture (see Fig. 6). The first tier, web browser, provides an interface to users to access and query public health data. The second tier, web server, transmits information between web browser and application server. The third tier, application server, including map server and HIPAA database, performs requests transmitted from the web server and returns the results. This three-tier online mapping system architecture ensures satisfactory public health data interactive display, query, and dissemination functions.

CONCLUSION AND FUTURE RESEARCH

Conclusion

Existing public health information systems adopt a variety of data sources, transmission standards, and collection techniques. Due to problems associated with localized and non-standardized data sources and transmission standards, these systems cannot provide sufficient support to the government agencies and public health researchers. This paper introduces a new data source—HIPAA message, and argues its applicability to the development of a standardized public health information system. As a further step, this paper proposes a prototype system design for collecting public health data with HIPAA messages. In particular, three conclusions may be obtained as follows: First, HIPAA messages may serve as standardized data source, since their formats are strictly defined by federal guidlines and their processing tools are widely supported by technology companies. Moreover, federal regulations mandate that all health care providers and payer use HIPAA messages to conduct financial transactions electronically. Second, HIPAA messages are applicable as a data source for collecting public health information. We have shown that HIPAA messages can provide sufficient amount of public health information while complying with privacy regulations. Moreover, our prototype implementation indicates that it is feasible to extract public health information from HIPAA messages. Finally, a standardized public health information system with HIPAA messages as data resource may be realized. This paper proposed a public health information system design with three major components: (1) HIPAA message filter, (2) HIPAA database, and (3) Geographic Information Systems.

Potential Problems in System Implementation and Future Research

Although the proposed research is in the stage of system design, some potential problems in system implementation can be foreseen and discussed as follows:

1. Minimizing data redundancy is an important aspect of the filter design. Because of the transactional nature of HIPAA messages, the content of many messages may be repetitive so that simple filtering will result in data redundancy that not only overwhelms the data storage capacity of our centralized database system but also makes data mining difficult and results in imprecision in the subsequent analysis of health demographics. The system should have the ability to identify repetitive information exchanged with HIPAA messages. To have such ability, we must record some identification information of health records. Also, we need to avoid recording contents from follow-up messages. For example, when a hospital initializes a billing transaction to an insurance company, the insurance company may send some follow-up messages to correct errors or clarify ambiguities. We need to be able to recognize the repeated information contained in the messages and only record a single instance for each interesting case. The filter rules need to handle the data redundancy problem and storage issues should be optimized basing on the feedbacks in the future.
2. Integration with existing systems in the host machine of the health care providers and payers is one of the most important aspect of system implementation. The filter system needs to intercept the HIPAA messages sent and received by the software system in the host machine. Unfortunately, different health care providers and payers may use different data management software to process the HIPAA messages and it is difficult to build a filter system that can interact with all these products. Therefore, it is desirable to have a standard way to access HIPAA messages that is independent of the data management software. This actually is quite challenging since it requires the cooperation of the software vendors that provide HIPAA transaction support. One possibility is to identify a widely used HIPAA-XML adaptor (which converts HIPAA messages to XML documents for a variety of information systems) and then uses its output as our data source. The alternative is to build customized connections for different data management software.

REFERENCES

1. Cromley, E. K., GIS and disease. *Ann. Rev. Public Health* 24:7–24, 2003.
2. Lober, W. B., Karras, B. T., Wagner, M. M., Overhage, J. M., Davidson, A. J., Fraser, H., Trigg, L. J., Mandl, K. D., Espino, J. U., and Tsui, F. C., Roundtable on bioterrorism detection: Information System–based Surveillance. *J. Am. Med. Inform. Assoc.* 9(2):105–115, 2002.

3. Overhage, J. M., McDonald, C. M., and Tierney, W. M., Design and implementation of the Indianapolis network for patient care and research. *Bull. Med. Libr. Assoc.* 83(1):48–56, 1995.
4. Duchin, J. S., Karras, B. T., Trigg, L. J. *et al.*, Syndromic surveillance for bioterrorism using computerized discharge diagnosis databases. In *Proceedings of AMIA Annual Symposium,* January 05, p. 897, 2001.
5. eHealth Initiative, Available at: http://www.ehealthinitiative.org. Accessed January 30, 2003.
6. Tsui, F. C., Espino, J. U., Dato, V. M., Gesteland, P. H., Hutman, J., and Wagner, M. M., Technical description of RODS: A real-time public health surveillance system. *J. Am. Med. Inform. Assoc.* 10(5):399–408, 2003 Sep–Oct. Epub 2003 June 04.
7. Office of Civil Rights, Medical privacy—national standards to protect the privacy of personal health information: final modifications to the privacy rule, *Federal Register*, August 14, 2002, Available at: http://www.hhs.gov/ocr/hipaa/privrulepd.pdf. Accessed May 7, 2004.
8. Centers for Disease Control and Prevention, HIPAA privacy rule and public health. Guidance from CDC and the US Department of Health and Human Services. *MMWR* 52(Suppl):1–17, 19–20, 2003.
9. Gostin, L. O., Turek-Brezina, J., Powers, M., Kozloff, R., Faden, R., and Steinauer, D. D., Privacy and security of personal information in a health care system, *J. Am. Med. Assoc.* 270:2487–2493, 1993.
10. Croner, C. M., Public health, GIS, and the Internet. *Ann. Rev. Public Health* 24: 57–82, 2003.
11. IBM Websphere software platform, http://www-306.ibm.com/software/info1/websphere/index.jsp?tab=highlights. Accessed May 10, 2004.
12. Mandl, K. D., Overhage, J. M., Wagner, M. M., Lober, W. B., Sebastiani, P., Mostashari, F., Pavlin, J. A., Gesteland, P. H., Treadwell, T., Koski, E., Hutwagner, L., Buckeridge, D. L., Aller, R. D., and Grannis, S., Implementing syndromic surveillance: A practical Guide informed by the early experience. *J. Am. Med. Inform. Assoc.* 11(2):141–150, 2004.