White Paper ■

Advancing the Framework: Use of Health Data—A Report of a Working Conference of the American Medical Informatics Association

MERYL BLOOMROSEN, MBA, DON DETMER, MD, MA

Abstract The fields of health informatics and biomedical research increasingly depend on the availability of aggregated health data. Yet, despite over fifteen years of policy work on health data issues, the United States (U.S.) lacks coherent policy to guide users striving to navigate the ethical, political, technical, and economic challenges associated with health data use. In 2007, building on more than a decade of previous work, the American Medical Informatics Association (AMIA) convened a panel of experts to stimulate discussion about and action on a national framework for health data use. This initiative is being carried out in the context of rapidly accelerating advances in the fields of health informatics and biomedical research, many of which are dependent on the availability of aggregated health data. Use of these data poses complex challenges that must be addressed by public policy. This paper highlights the results of the meeting, presents data stewardship as a key building block in the national framework, and outlines stewardship principles for the management of health information. The authors also introduce a taxonomy developed to focus definitions and terminology in the evolving field of health data applications. Finally, they identify areas for further policy analysis and recommend that public and private sector organizations elevate consideration of a national framework on the uses of health data to a top priority.

■ **J Am Med Inform Assoc.** 2008;15:715–722. DOI 10.1197/jamia.M2905.

Introduction

The availability of complete, accurate health data can improve healthcare experiences for individuals, expand collective knowledge about diseases and appropriate treatments, strengthen insights into the effectiveness and efficiency of healthcare systems, support public health and security goals, and help businesses to address their customers' needs. Aggregation of health information into very large data sets and repositories offers extremely valuable opportunities and benefits—despite limited understanding of these by the general public. Health data can serve as a bridge to achieving many of the goals of the U.S. health system. Large amounts of patient data, available in electronic form, support current-day clinical care and decision-making, and foster public health activities such as surveillance, measurement of outcomes and performance, research, and public policy. Yet despite the growing importance of health data use, the U.S. lacks a coordinated public policy framework to guide users on appropriate health data use practices.

Through this paper, the American Medical Informatics Association (AMIA) seeks to stimulate discussion about and action on a national framework on the uses of health data. We define the concept of data stewardship and offer a rationale for its central role in appropriate use of health data. We urge stakeholders to deepen their understanding of data stewardship principles and encourage them to debate the viability and efficacy of the data stewardship approach. While this paper reflects the discussions at AMIA conferences, the views expressed are the responsibility of the authors and of AMIA and its Board of Directors.

Affiliations of the authors: American Medical Informatics Association (MB, DD), Bethesda, MD.

The American Medical Informatics Association (AMIA) thanks the many participants and presenters for their input on the 2007 AMIA meeting on Health Data Reuse on which this article is based. Charles Safran chaired the meeting steering committee. Doug Barton, Robert Browne, Bart Harmon, Steve Labkoff, Michael Lieberman, Susanne Markel-Fox, Sam Scott, and William Tierney served as members of the Meeting Steering Committee. They were actively involved in and provided valuable input to all aspects of the planning processes. AMIA also thanks the meeting participants who served on the working groups. This includes Christopher Chute, Co-Chairperson Taxonomy Working Group; Stanley Huff, Co-Chairperson, Taxonomy Working Group; James J. Cimino; Robert Dolin; Kent A. Spackman; Judy G. Ozbolt; David W. Bates; Suzanne Bakken; Charles E. Barr; Melissa M. Goldstein, Chairperson Data Stewardship Working Group; Paul C. Tang; Allison Rein; Allen Briskin; Susan McAndrew; John P. Houston; Marcie Wilder; Kenneth D. Mandl; Chuck Friedman; Douglas Barton, Chairperson, De-Identification Working Group; Michael J. Becich; COL William Lang; Bradley Malin; and Peter Szolovits. The AMIA thanks Senator Sheldon Whitehouse for his willingness to make remarks during one of the meeting sessions. The authors also thank the following individuals who reviewed earlier versions of the summary of the 2007 AMIA meeting on Health Data Reuse and provided insight into the topic: Margo Edmunds, Miriam Halimi, Kraig Kinchen, Steve Labkoff, Michael Lieberman, Charles Safran, Elaine Steen, and Freda Temple. The AMIA thanks the organizations that generously supported the 2007 meeting: Deloitte, GE Healthcare, GlaxoSmithKline, IBM, IMS Health, Lewin Group, Lockheed Martin, Eli Lilly and Company, Northrop Grumman, Pfizer, and Thomson Health.

Correspondence: Meryl Bloomrosen, Associate Vice President, American Medical Informatics Association, 4519 St Elmo Avenue Suite 401, Bethesda, Maryland 20814; e-mail: <meryl@amia.org>.

Received for review: 06/19/08; accepted for publication: 08/14/08.

AMIA's 2006 and 2007 Conferences Focus on Health Data Use

In 2006, building upon policy work done over the prior fifteen years, AMIA recognized personal health information (PHI) use as a critical issue for the continued widespread adoption of health information technology. Thus, AMIA convened a panel of experts and stakeholders who represented a variety of backgrounds and work environments to open a dialogue on the topic.[1] The panel reached broad agreement that, while aggregated health data provide value to a wide variety of applications, use of these data pose complex ethical, political, technical, and economic challenges that must be better addressed by public policy. Meeting participants took the first steps toward developing a robust framework for an infrastructure of policies, standards, and best practices to facilitate the collection, storage, aggregation, linkage, and transmission of health data for various uses. This data use framework includes these components:

- Transparent policies and practices.
- Focus on data stewardship and its implications for data control, access, and security, rather than on data ownership.
- Clarification and engagement of an appropriate comprehensive scope of data users and uses (beginning with a taxonomy to define users and uses).
- Formal national leadership and direction.
- Public outreach, awareness, and education regarding the policies and positions of the proposed data use framework.

The 2006 panel also offered recommendations aimed at providing the substance and detail to flesh out these components but much work remained to be done. As a result, in June 2007 AMIA convened an expanded group of experts to further development of the framework (http://www.amia.org/inside/initiatives/healthdata/2007/index.asp). Attendees confirmed that the conversations about health data reuse and the related opportunities and challenges are not simply theoretical issues but must be addressed in the context of a complex and fluid societal environment.

In an effort to learn from other nations grappling with similar health data use issues, the meeting included discussions about the U.K.'s and Switzerland's experience with secondary uses of health data. Two presentations illustrated how collection of electronic health data in primary care settings and submission of the data for selected secondary uses (e.g., health status monitoring, practice performance) are already in place in Britain's increasingly integrated national health service. The presenters stressed that legislation, policy development, and technical security measures are needed to ensure that secondary uses of data are carried out within a safe, ethical framework.[2] A third presentation highlighted relevant aspects of the Swiss healthcare system. In Switzerland, the authorization to use anonymized personal health information for retrospective research is based on an explicit, audited institutional process, with the medical directors of hospitals serving as the data stewards.

During the meeting, participants reviewed work products of the pre-conference Data Stewardship and Taxonomy Working Groups. The work products included a data stewardship definition and data stewardship principles, and a taxonomy depicting dimensions of data use and users. Participants also reviewed a framework tool which partitioned data uses into the four domains of research, quality, public health, and commercial applications.[3] (These products are described later in this paper.) The 2007 meeting resulted in presentations and testimony by AMIA to the American Health Information Community (AHIC), the Office of the National Coordinator on Health Information Technology (ONC), and the National Committee on Vital and Health Statistics (NCVHS).

Refining the Terminology

AMIA's exploration into data uses began with the approach taken by the 1991 Institute of Medicine (IOM) computer-based patient record report, which acknowledged the growth in the types and numbers of uses and users of patient data.[4] While the IOM recognized that patient record uses extend beyond direct patient care, the report focused on certain high-priority uses rather than on all possible functions of the record. The IOM defined primary use as patient care and all other uses as secondary. As a matter of U.S. public policy, HIPAA legislation has de facto broadened the definition of primary use to include business operations and quality of care.

Leveraging the IOM's discussions, AMIA initially made a distinction between primary and secondary data users and uses. *Primary use data* were defined as data collected about and used for the direct care of a patient. *Secondary use data* were defined as non-direct care use of PHI including, but not limited to, analysis, research, quality/safety measurement, public health, payment, provider certification or accreditation, and marketing and other business uses including strictly commercial activities. However, as we reflected on the evolution of the health system over the past two decades and the accompanying growing dependence of successful patient care on uses of PHI, we concluded that a simple division into primary and secondary use had outlived its value. We further concluded that policies and procedures should focus instead on how such data are used, reused, and protected. AMIA asserts that the concept of *data stewardship* should be the choice for expressing this notion. AMIA proposed the following clarification of the primary/secondary data concept in testimony before the NCVHS in August 2007.

> *Reuse of health data occurs when personal health data are used for purposes other than those for which they were originally collected.*

A single category of 'secondary use' did persist among the majority of participants at the 2007 AMIA meeting. This category refers to the use of personal health information, whether individual or collective, when the aim is to generate profits that are outside the bounds of the healthcare system or related health research. Today, the types of uses of data solely for commercial purposes are evolving and attention needs to be paid to the continuum of these uses. The issues surrounding the buying and selling of data are not clear cut since some 'for-profit' companies are integrally involved in essential healthcare operations or research. Different issues come into play depending on whether the purpose of the sale of data is for research and quality or purely for marketing purposes. More discussion is needed to determine the best way to categorize these uses.

Benefits and Challenges of Health Data Use

Biomedical, health services, and policy research, particularly in the areas of population studies and public health, depend heavily on the ready availability of data about patients and populations. Examples of these data include health surveys, clinical trials data, hospital, physician, and laboratory records, state and federal billing and registration data, birth and death records, socio-demographic data, and cancer registry data. Researchers need large volumes of such data to be able to draw meaningful conclusions that are representative of populations.[5] It is problematic if data on certain groups of people are missing, and in some instances, even hypothesis-generating research may be impossible.

There is a natural tension caused by conflicting objectives inherent in many data use and exchange situations. Examples of these tensions include:

- Information needs of the health system versus the wants and needs of individual patients, consumers, and/or health professionals.
- Public safety versus the right to privacy to personal health information.
- Medical-legal, ethical and best practices versus practical demands.
- Technology advancements versus historical approaches to data use practice.[6]

Despite the challenges posed by these tensions, it is generally believed that health data access and sharing are essential for continued progress in promoting quality and continuity of care, patient safety, and research on better treatments.

Developments in the Fields of Data Reuse and Data Stewardship

Discussions about use of health data are not new. The social implications of the growing use of information, computer, and communications technologies are much broader than privacy concerns, encompassing economic, political, legal, and technical issues that cannot be resolved in isolation.[7] Key policymaking groups and stakeholders over the past ten years have issued studies and reports related to health data reuse and data stewardship. For example, the NCVHS has held hearings and workshops on privacy and confidentiality issues including discussion of a stewardship framework for secondary uses of health data. In 2007, the Agency for Healthcare Research and Quality (AHRQ) issued a Request for Information on a National Health Data Stewardship Entity and received 136 responses from stakeholders.

Policy changes, innovations in technology, and other recent events have pushed the discussion about health data use in new directions. For example, state legislation addressing the legality of selling health data[8] and the launch of an array of consumer-targeted products and services illustrate the potential transformative power of emerging trends. Microsoft launched HealthVault (www.healthvault.com), a set of search and personal health record (PHR) tools that enable consumers to control the data that is entered and shared with others. Google is bringing its data storage and organization capacities to the field of medical care and patient records via Google Health (https://www.google.com/health). Intuit (http://quickenhealth.intuit.com) has also entered this field. In 2006, a consortium of companies (Intel, Wal-Mart, Pitney Bowes, British Petroleum America Inc. and Applied Materials) announced a PHR initiative for approximately 2.5 million employees, families, and retirees.[9] Patients and caregivers are creating personal health information systems to help manage chronic illnesses such as "Follow Me" (www.followme.com) and "PatientsLikeMe," www.patientslikeme.com). The potential for profits from these services and products that aim to help consumers manage an increasingly complex and confusing healthcare system often attracts the new players operating in the commercial sphere. The impact of the entry into this field by non-medically-based entrepreneurs is likely to affect the use of data in unforeseen ways.

Reuse of Health Data and Data Stewardship

Data stewardship is pivotal to advancing discussions about the legitimate and appropriate use of health data. Access to personal health data often involves many different data stewards or custodians, such as hospitals, public health clinics, laboratories, physicians' offices, research centers, pharmaceutical companies, third party payers, employers, registries, health information producers, and federal, state, and local government agencies. Variations exist among data stewards and custodians regarding access, control, and data integrity processes that need to be harmonized. Data stewardship encompasses the breadth of activities carried out in varying degrees by all entities that interact with health data, including collection, use, disclosure, management and security of that information. Within each of these aspects there are medical-legal, ethical, and best practice considerations that an individual or organization should consider in the management of health information.[5] It will take the combined efforts of many stakeholders to establish a working environment that promotes acceptance of responsibility, awareness of risk, and establishment of trust for health data use.

The growing volume of health data collected and stored electronically dramatically heightens the importance of data stewardship issues. As described above, these data are at the center of many ongoing scientific, biomedical, and health services research and policy efforts and are critical to quality assessment and improvement projects, which in turn are key to value-based purchasing initiatives. This information is also seen as a critical component of surveillance systems to detect outbreaks of disease, bioterrorism, and adverse events due to drugs or devices.

With an increased interest in reuse of health data, concerns about security and privacy are rising to the forefront. While various state and federal regulations address many of these concerns, they cannot adequately deal with the full range of current or contemplated data reuse scenarios. The biomedical and health informatics community would benefit from the articulation and application of a data stewardship paradigm to enable creditable uses of data while addressing legitimate concerns of privacy and security. Without such a paradigm, it is possible that privacy and security concerns could prohibit or curtail the use of health information, thus inhibiting the benefits this use can bring.

Data stewardship has emerged as a means to balance the rights of individuals to have their personal information protected and their desire for improved health, more effective health services, and a strengthened and sustainable health system. Widespread acceptance of a set of consistent data stewardship principles and application of data handling guidelines for health data collection, storage and exchange would help establish a "chain of trust." In turn, the chain of trust would facilitate transactions and build confidence among consumers that their privacy interests are being given due consideration. Thus, data stewardship is a key building block in the construction of a national framework to guide the reuse of health data. AMIA proposes the following definition of data stewardship:

> *Data stewardship encompasses the responsibilities and accountabilities associated with managing, collecting, viewing, storing, sharing, disclosing, or otherwise making use of personal health information. Principles of data stewardship apply to all the personnel, systems and processes engaging in health information storage and exchange within and across organizations.*

Data stewardship principles comprise:

- *Accountability*, including governance, oversight, and the application of relevant regulations to the appropriate extent and level.
- *Transparency,* including clearly understood policies and procedures regarding data structure, processing, and delivery of data, and business processes and practices.
- *Notice to patients* and other legitimate users.
- *Technical issues*, e.g., data security, and quality, de-identification, and costs of re-identification.
- *Patient consent* of appropriate granularity.
- *Permitted uses and disclosures* including for data aggregation and analyses.
- *Enforcement and remedies.*

The involvement of data stewards and the application of data stewardship principles (including enforcement mechanisms when necessary) would provide sufficient safeguards for legitimate downstream uses of health data. "Trusted data stewards" who adhere to these principles should be able to share data without having to create anew ad hoc data handling guidelines for each transaction. Each partner in the provider/supply chain of health data should accept, both in practice and spirit, an appropriate share of responsibility for maintaining the quality, security, and confidentiality of the data.

The application of sound data analytic principles, as utilized by health services researchers and analysts and clinical researchers (e.g., National Institutes of Health guidelines) is central to the data stewardship paradigm.[10,11] Data must be of a minimum quality (accurate, reproducible, complete, timely, and credible) and data limitations should be acknowledged and described. Organizations and individuals must take an active interest in the accuracy, consistency, and timeliness of their health data.[12] Data stewardship is a multi-faceted function which assigns ultimate responsibility for data quality and integrity to the organization holding the data.[13] Trusted data stewards would be expected to establish, support, and maintain agreed-upon standards of data quality and data management.

Examples of the concept and application of stewardship can be found in the healthcare field as well as in other fields.[14] In 1979, the Department of Health and Human Services Secretary's Advisory Committee on Automated Personal Data Systems analyzed the consequences of using computers to keep records about people[15] and proposed enactment of a federal "Code of Fair Information Practice" for all automated personal data systems. Over the past quarter century, government agencies in the U.S., Canada, and Europe have studied the ways in which entities collect and use personal information, and the safeguards required to assure that these practices are fair and provide adequate privacy protection. The resulting series of reports, guidelines, and model codes represent widely-accepted principles concerning fair information practices.[16–19] For example, the concepts of "circle of trust" and "chain of trust" are central to the data stewardship framework developed by the College of Physicians and Surgeons of British Columbia.[20] The Liberty Alliance, formed in 2001 by 30 organizations to establish open standards, guidelines and best practices for identity management, has promulgated the concept of Circle of Trust as a legal entity and provides guidance on suggested business structures and terminology for a Liberty-enabled technology (http://www.projectliberty.org/liberty/about).[21]

A Taxonomy on Dimensions of Data Use

Discussions about reuse of health data can benefit from ongoing identification of uses and users of PHI. An AMIA working group presented a draft taxonomy at the 2007 meeting. By documenting a comprehensive (albeit not exhaustive) array of data users and users, the working group sought to build consensus around working definitions of these uses and to provide a guide for those developing policy related to health data reuse. The taxonomy is an important tool for sharpening definitions and terminology and helping to inform the greater community about data use terminology; it will require expansion and maintenance to sustain its usefulness. Table 1 presents a selection from this taxonomy.

Framework Tool to Assess the Status of Uses of Health Data

At the 2007 conference, AMIA presented a framework tool which partitions data uses into four domains: research, quality, public health, and commercial. (See Figure 1; domains are indicated in the top, right-hand corner.) In the tool, the seven principles of stewardship (see above) are related to six categories of activities: accountability, transparency, patient consent, cost of re-identification, oversight, and regulation.

The framework tool employs a sliding scale to illustrate a 'scorecard' of the status of activities in each domain as related to each category. Meeting participants adjusted the scale to indicate their opinion of the current status of each activity and the status that they proposed for it in the future (See Figure 1 legend).

Areas for Additional Exploration

The opportunities and challenges discussed in this paper require elaboration in the context of a complex and fluid societal environment. Breakthroughs in medical treatments, advances in biomedical and informatics technologies, new legal challenges and pending legislation, consumer health initiatives, and emerging commercial trends are just some of the factors that will affect the dialogue. AMIA identified several areas deserving further policy analysis and dialogue, based on work done at the 2007 conference. It is worth

Table 1 ■ Selection from Taxonomy on Dimensions of Data Use

Uses of Data

- A. Protect and enhance public health
  - Enable and support biosurveillance
    - Monitor and report vital statistics
    - Monitor and report biometric demographics (e.g. weight, height, blood pressure, normal lab values)
    - Identify, monitor, and report health and illness trends
    - Identify, monitor, and report infectious diseases (e.g. culture, serology, DNA/RNA probe results)
  - Export data to health registries
    - Cancer or rare disease registries
    - Drug and device registries
  - Report toxic exposures (e.g. smoking, Agent Orange)
- B. Develop security and confidentiality algorithms and test de-identification routines
- C. Conduct research
- D. Create and maintain terminology and representation formalisms
- E. Develop and apply decision support for health care providers
  - Develop and test the efficacy of decision support algorithms
  - Develop order sets, rules, and alert
- F. Support quality of patient care
  - Manage quality and outcomes
  - Manage staffing and resources
  - Develop and assess quality indicators
  - Support quality reporting (e.g. HEDIS)
- G. Improve patient safety
  - Conduct pharmacovigilance (post market drug and device surveillance)
    - Detect and analyze adverse and sentinel events
    - Support risk profiling
  - Monitor and survey to prevent patient adverse events
- H. Manage personal health
    - Provide patient-specific feedback and assessments of progress toward health goals
    - Maintain personal health records
    - Provide links to knowledge resources based on personal health information
- I. Educate and credential healthcare providers and assess training activities (e.g. types and outcomes of procedures)
- J. Analyze and Manage Finances
  - Conduct automated billing, claims processing
  - Analyze activity-based charge capture, cost accounting
  - Develop predictive models of costs and accounting
- K. Detect fraud and illicit activity
  - Detect illegal and inappropriate activity (e.g., Medicare upcoding)
  - Report drug screen results to detect illegal drug use
- L. Identify markets and promote sales
  - Conduct market research
  - Target marketing to physicians
  - Target marketing to patients and families

Factors Influencing Authorization for Use of Healthcare Data

- 1. Requirements Imposed on Use of Healthcare Data
- A. Identification Status
  - Patient-identifiable data
  - De-identified data (HIPAA definition)
  - Anonymized data
    - No linkage possible (alteration of PHI, precluding linkage)
    - Relinkable data
    - Linked with protected key (trusted third party)
- B. Consent provided at the time of data collection
  - No consent by the individual
  - Consent by the individual
    - Broad and unspecified
    - Time-limited consent
    - Consented for partial, source specific use (e.g., no psychiatric data)
    - Consented for the particular type of use
- C. Demographic representation
  - Age
  - Race
  - Gender
  - SES
  - Insurance status
- D. Focus on a vulnerable population (e.g. prisoners, pregnant women, undocumented immigrants)
- E. Original collector and aggregator of the data
  - Government
  - Health Plan
  - Other private entity
- F. Proposed user of the data
  - Government agency
  - Academic institution
  - Private, not-for-profit entity
  - Private, for-profit entity
- G. Funding source for use
  - Government agency
  - Academic institution
  - Private, not-for-profit entity
  - Private, for-profit entity
- H. Financial compensation to data collector or data steward for providing data to a second party
  - No compensation
  - Compensation
- I. Beneficiary of use
  - Society
  - Researcher
  - Academic institution/medical center
  - Private, for-profit entity (e.g., financial gain)
- J. Disclosure of use
  - Not disclosed publicly
  - Publicly disclosed
    - Disclosure of results only
    - Disclosure of research methods utilized
    - Disclosure of analytic principles that guide data use
- K. Required level of consent and authorization
  - IRB evaluation not required
  - IRB evaluation required
    - No consent by the individual required
    - Consent by the individual required
- L. Compensation of patients
  - No compensation required
  - Compensation of individual patients required
- 2. Existing and potential sources of data for use
  - A. Public Use Datasets
    - i) Medicare
    - ii) Medicaid
    - iii) CDC surveys (some Primary data use, e.g. NHANES)
  - B. Private Datasets
    - Open-source data
    - Commercial use datasets (at patient level)
    - Pharmacy benefit/claims manager
    - Provider databases
      - Individual providers
      - Aggregated data from provider consortia
    - Consortium databases
      - caBIG
      - CTSA recipients
      - University Health Systems Consortium
    - Aggregated clinical repositories hosted by HIT vendors
    - Personal health records, including patient-entered data
    - Health Information Exchanges (RHIOs, etc)

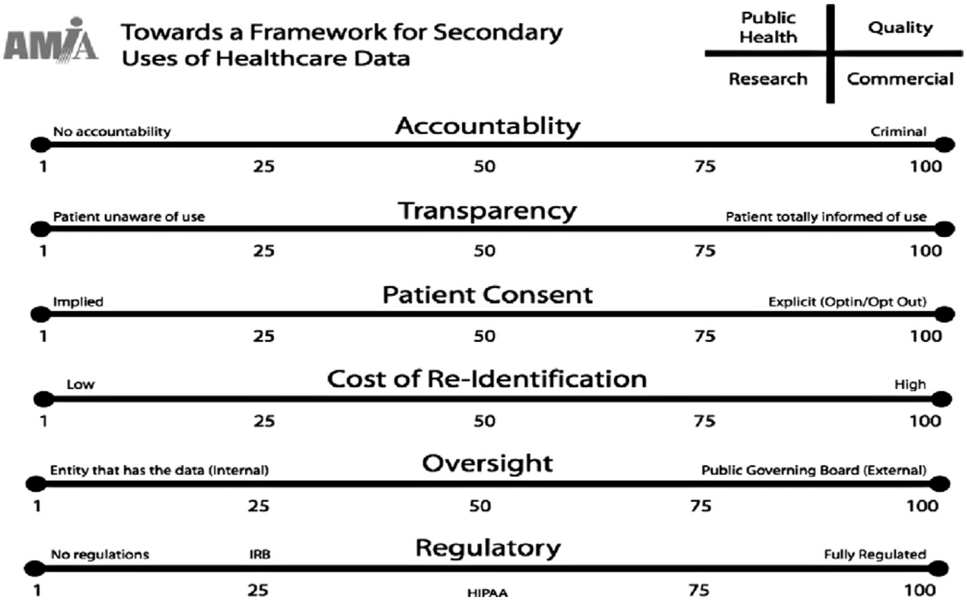

Figure 1. Framework Tool to Assess Status of Health Data Uses. **Accountability**: level of sanctions or penalties for disclosures or inappropriate use of patients' health data. End Points: 1 = No accountability; 100 = Criminal sanctions. **Transparency**: extent to which practices governing use of patients' health data are known and understood by those who disclose or use data and by patients whose data are subject to use. End points: 1 = Patient is completely unaware of uses of health data; 100 = Patient is informed of every use of health data at the time of its occurrence. **Patient consent/notification**: the opportunity offered to patients to allow/permit use of their health data. Notification refers to the mechanism by which patients are informed of their right to consent. End points: 1 = No Choice; 100 = Opt in. **Cost (resources required for) of re-identification**: proxy for the nature, complexity, and extent to which patients can be reidentified in a data base(s). End points: 1 = Low/relatively straightforward; 100 = High/complex and difficult. **Oversight**: extent to which the entity is subject to governance or supervision, including ability to impose remedies for breaches. End points: 1 = Internal, residing with the entity that has the data; 100 = External, residing with a public governing board. **Regulation/Law**: framework of regulations and laws governing health data uses, including penalties and enforcement guidelines. End points: 1 = No regulations or laws; 100 = Fully regulated.

noting that the status in federal or state laws of the concept of health data stewardship, while an important aspect of the overall topic, was outside of the scope of the discussions at the 2007 AMIA meeting and of this paper. Nevertheless, the deliberations of the meeting should be useful to the legal community as they consider the evolving landscape of data stewardship.

Data Stewardship Principles, Policies and/or Guidelines

- Further refine the concept of "trusted data steward" as a means of promoting the legitimate and appropriate exchange of health data across and among entities. Encourage collaboration of public and private sector organizations to refine the data stewardship principles. Questions to be answered include:
  - What factors need to be considered in developing a healthcare data "chain of trust?"
  - How can a robust "chain of trust" be implemented to authenticate an organization's participation?
  - What trust model(s) should be adopted?
- Explore the establishment of a voluntary process for organizational compliance with established health data stewardship principles. Consider the creation of an independent, non-profit, private sector voluntary mechanism, e.g., the Health on the Net Foundation (HON) (http://www.hon.ch/) and the Certification Commission for Healthcare Information Technology (CCHIT) (http://www.cchit.org/about/index.asp).

Ongoing Education and Outreach

- Conduct outreach and education to help improve the public's understanding of societal issues and perspectives around data reuse and to gain the public's trust. Explore opportunities to educate the popular media because they can be the drivers of the public's loss of trust and also provide opportunities to establish and recapture trust.
- Clarify confusion surrounding HIPPA Rules, FDA's human subject protection regulations, and the Common Rule, which may be neither applicable nor adequate to address the complexities of data reuse issues.

Enhancement of Definitions and Terms

- Address terminology issues in the data reuse field. Develop and refine definitions for commonly-used terms; for example, additional granularity of terminology for "commercial uses" needs to be developed, as well as definitions of "circle of trust" and "chain of trust." Definitional issues relating to some frequently used and potentially misunderstood terms still exist: de-identification, re-identification, anonymization, and pseudo-anonymization.[22–24]
- Promote the use of the taxonomy described in this paper to depict data uses. Explore methods of expanding and maintaining the taxonomy.

Summary and Conclusions

AMIA asserts that the proper boundaries for management of health information include its collection, use, access, disclosure, and retention as well as the legal, ethical, business, and fiduciary responsibilities of those entities supplying, maintaining, using, and receiving data. Each dimension is important but historically, some of these have been given careful attention while others have been ignored, relatively speaking, or given lip service. There is an ongoing need to harmonize data use issues such as legislative and regulatory requirements and data access policies, accommodating for differences among and across organizations, institutions and entities, and recognizing variations in the purposes for which health data are used. Recently, there has been a surge in interest in virtually all these dimensions, but policy and practices have typically lagged behind even agreed-upon principles and perspectives in the U.S. With this paper, AMIA seeks to stimulate discussion on these issues by describing various benefits of health data use, refining definitions of relevant terms, presenting an approach to depicting dimensions of health data use and users, proposing an overarching framework for data use, and outlining a definition and principles of data stewardship.

While health data have been used for many purposes beyond direct patient care for decades, the advent and increased deployment of health information technology is complicating the use of such data. One of the key challenges facing the healthcare industry is to reach a workable balance among the value we place on good health care, the effectiveness of healthcare services, and the sustainability of the healthcare system, and the equally compelling value we place on our right to stewardship and confidentiality with respect to our personal health information.[25,26]

AMIA Board of Directors (BOD) Response and Action

By convening the 2007 conference and disseminating this paper, AMIA has further delineated critical issues related to data reuse. The AMIA BOD reviewed the paper and endorsed its findings, conclusions and recommendations. The BOD will continue to encourage other organizations to work collaboratively to continue this important public discourse.

References ▪

1. Safran C, Bloomrosen M, Hammond WE, Labkoff S, Tang PC, Detmer DE. Toward a national framework for the secondary use of health data: an American Medical Informatics Association white paper. Expert Panel. J Am Med Inform Assoc 2007;14:1–9. Epub 2006 Oct 31.
2. Thomas R, Walport M. Data Sharing Review: Data Sharing Review Report, 11 July 2008. Available at: http://www.justice.gov.uk/reviews/datasharing-intro.htm. Accessed Augt 13, 2008.
3. American Medical Informatics Association. 2007 Invitational Conference on Secondary Use of Health Data. Available at: http://www.amia.org/inside/initiatives/healthdata/2007/index.asp. Accessed June 11, 2008.
4. Dick R, Steen EB, Detmer D, (Eds). Committee on Improving the Patient Record, Institute of Medicine. The Computer-Based Patient Record: An Essential Technology for Health Care, Revised Edition. Washington DC: National Academy Press, 1997. Available at: http://www.iom.edu/CMS/3809/22303.aspx. Accessed June 11, 2008.
5. Canadian Institutes of Health Research. Secondary Use of Personal Information in Health Research: Case Studies, November 2002. Available at: http://www.cihr-irsc.gc.ca/e/1475.html. Accessed June 11, 2008.
6. College of Physicians & Surgeons of British Columbia. Committee on Privacy & Data Stewardship. Data Stewardship Framework, Draft Version 2.4, August 22, 2007. Available at: www.cpsbc.ca/cps/general_info/communications/press_releases/2007/09/datastewardship. Accessed June 11, 2008.
7. Brune H. The social implications of information processing. Information & Management 1978;1(3):143–56.
8. United States District Court for the District of New Hampshire. IMS Health Incorporated, et. al. v. Case No. 06-cv-280-PB, Opinion No. 2007, DNH 061.P. Kelly Ayotte, as Attorney General of the State of New Hampshire. Available at: http://epic.org/privacy/imshealth/dist_ct_op.pdf. Accessed Aug 12, 2008.
9. Hoover J. Wal-Mart, Intel, Others to Create Massive Health Records Database. Information Week. 2006 Dec 6. Available at: http://www.informationweek.com/news/management/showArticle.jhtml?articleID=196602073. Accessed Aug 12, 2008.
10. Academy Health. Health Services Research Methods. Available at: http://www.hsrmethods.org/Home.aspx. Accessed Aug 12, 2008. Health Services Research Methodology Core Library Recommendations, 2007. National Library of Medicine. Available at: http://www.nlm.nih.gov/nichsr/corelib/hsrmethods.html. Accessed Aug 12, 2008.
11. Guidelines for Data Quality Assurance in Clinical Trials and Observational Studies. National Heart, Lung, and Blood Institute, National Institutes of Health. Revised April 24, 2001. Available at: http://www.nhlbi.nih.gov/funding/policies/dataqual.htm. Accessed June 11, 2008.
12. Laurent W. The Case for Data Stewardship. DM Review Magazine. 2005 Feb. Available at: http://www.dmreview.com/issues/20050201/1018108-1.html. Accessed June 11, 2008.
13. Geiger J. Data Stewardship Using the Zachman Framework. DM Review Magazine. 1997 Dec. Available at: http://www.dmreview.com/issues/19971201/737-1.html. Accessed June 11, 2008.
14. Smart Card Alliance. Secure Identification Systems: Building a Chain of Trust. 2004 March. Available at: http://www.smartcardalliance.org/pages/publications-secure-id-systems-report. Accessed June 11, 2008.
15. DHHS Secretary's Advisory Committee on Automated Personal Data Systems. Summary and Recommendations. Available at: http://aspe.hhs.gov/datacncl/1973privacy/Summary.htm. Accessed June 11, 2008.
16. Mack J. Pharmaceutical Compliance with Fair Information Practice Principles. Available at: http://www.virsci.com/PharmaPrivacyPolicy_Analysis.pdf. Accessed June 11, 2008.
17. Code of fair information practice. Available at: www.publications.health.sa.gov.au/auinfo/1/. Accessed September 17, 2008.
18. [CAN99] The House of Commons of Canada, 2nd Session, 36th Parliament, 48 Elizabeth II, 1999, Bill C-6. Personal Information Protection and Electronic Documents Act. Available at: www2.parl.gc.ca/HousePublications.aspx?DocId=2330518&Language=e&Mode=1&File=16. Accessed June 11, 2008.
19. Gellman R. Fair Information Practices: A Basic History. Privacy and Information Policy Version 1.4, January 4, 2008. Available at: http://bobgellman.com/rg-docs/rg-FIPshistory.pdf. Accessed June 11, 2008.
20. College of Physicians & Surgeons of British Columbia, Op Cit.
21. Liberty Alliance Project. Schekler V. (ed). Liberty Alliance Contractual Framework Outline for Circles of Trust. March 2007. Available at: http://www.projectliberty.org/liberty/resource_center/papers. Accessed June 11, 2008.

22. Wellner B, Huyck M, Mardis S, et al. Rapid retargetable approaches to de-identification in medical records. J Am Med Inform Assoc 2007; Sep-Oct; 14(5):564–73. Epub 2007 June 28.
23. Szarvas G, Farkas R, Busa-Fekete R. State of the art anonymisation of medical records using an iterative machine learning framework. J Am Med Inform Assoc 2007; Sep-Oct;14(5):574–80.
24. Uzuner O, Luo Y, Szolovitz P. Evaluating the state-of-the-art in automatic de-identification. J Am Med Inform Assoc 2007; Sep-Oct;14(5):550–63. Epub 2007 Jun 28.
25. Detmer D, Steen EB. Learning from Abroad: Lessons and Questions on Personal Health Records for National Policy. March 2006. AARP International. Available at: http://www.aarp.org/research/health/carequality/2006_10_phr_abroad.html. Accessed June 11, 2008.
26. Cylus J, Anderson G. Multinational Comparisons of Health Systems Data, 2006, The Commonwealth Fund, May 2007, Volume 24. Available at: http://www.commonwealthfund.org/publications/publications_show.htm?doc_id=482648. Accessed June 11, 2008.