ORIGINAL ARTICLE

Planning security architecture for health survey data storage and access

M. Rita Thissen and Katherine M. Mason

Research Computing Division, RTI International, Research Triangle Park, NC, USA

ABSTRACT

Sensitive data from health research surveys need to be protected from loss, damage or unwanted release, especially when data include personally identifying information, protected health information or other private material. Researchers and practitioners must ensure privacy and confidentiality in the architecture of data systems and in access to the data. Internal and external risks may be deliberate or accidental, involving unintended loss, modification or exposure. To prevent risk while allowing access requires balancing concerns against providing an environment that does not impede work. The authors' purpose in this paper is to draw attention to basic data security needs for health survey data from the perspective of both the health researcher/practitioner and infrastructure/programming staff to ensure that data are securely and adequately protected. We describe risk classifications and how they affect system architecture, drawing on recent experience with systems for storage of and access to electronic health survey data.

ARTICLE HISTORY

Received 5 September 2018
Accepted 8 March 2019

KEYWORDS

Data security; sensitive data;
health surveys; data
systems; health informatics

1. Introduction

Policy makers at a national, regional, or local level base many of their decisions on data from health research, and frequently the information derives from health surveys. Digital health informatics systems comprise the technology that hosts, protects, receives, and distributes medical or health-related information, underlying nearly all stages of current health survey research. Whether during data collection, storage, quality review, transformation, and analysis, the data are protected by the design and operation of the systems, in the hands of responsible informaticists.

Sometimes those systems are already in place for researchers to use without modification, but often for new data uses and grant-funded or contract research, new infrastructure is put in place to ensure data security and to allow quality processes such as consistency checks, inference of missing data, and deduplication of data records. Researchers doing simulations on health topics may reference survey data to validate or support their findings. Considering the widespread value of health survey data, it is important for health researchers to be fully aware of data security as a basic part of informatics, and to understand how it is assured by the systems they use. Yet a wide gap exists between subject matter experts' understanding of informatics and the knowledge of security experts.

The authors have found that many survey researchers lack a basic understanding of health data security and system architecture, especially those who

are serving as principal investigators or project directors for the first time. Similarly, system architecture, database and programming staff may lack the experience or skill to determine health data sensitivity and may have no experience with health-related privacy laws or best practices for health-data ethics. Each group expects the other to determine appropriate safeguards for the grant or contract's data, and yet neither may have an adequate grasp of data security needs. Our goal in this article is to help bridge the gap. We believe that those knowledgeable in security will learn more about sensitive health data, while health researchers will learn more about security.

Researchers may not always have security on their minds when they consider health surveys as a source of information for their questions in public health, quality of life, health behaviour patterns and trends. Yet the design of a new survey (or plans for analysis of data from a prior one) benefits from considering security in light of the sensitivity of the topic of study and the types of data gathered. Awareness is increasing of the need to protect all forms of data whose release could harm an individual, a vulnerable group or an organisation. The risk of harm, whether accidental or deliberate, rises or falls with many factors: the sensitivity of the subject matter, the presence of direct or indirect identifiers (El Emam, 2008), the extent of planned availability of the data to interested parties, the degree of exposure through electronic access, and the care taken to preserve and present the information. Addressing concerns early through security-

mind system architecture can reduce risk and mitigate the impact of untoward events. While researchers may rely on security experts to recommend or select system components for protecting electronic data, responsibility lies with those in charge of the work.

Data sensitivity depends on context, reflecting the amount of concern for privacy, confidentiality, and security (Lumpkin, 2000). Treatment of sensitive data includes steps taken for protection from loss, prevention of undesired modification, avoidance of unintended release and limitations on who may be authorised for data access. These areas of concern enter the picture in different degrees at different stages of operation. In any given study, health survey data may proceed through one or many processes, such as the following:

- Collection
- Cleaning
- Coding
- Linking or matching
- Analysis
- Dissemination.

Each process handles data “at rest” and data “in motion,” for which security approaches differ, fundamental concepts known to security experts but typically unknown to subject matter or survey experts.

Information is said to be at rest or at an endpoint when it has been stored in a particular location and is used without transfer to another site (Liu & Kuhn, 2010). It is in motion when accessed through an organisation’s private network, web browser or any data transfer mechanism; movement can be as simple as passing from a server to a work station’s screen when a data user views the data or runs reports or analyses on the stored information. Without motion, electronic data cannot be used unless data workers are physically present at a single, physical computer that hosts the values, software, and results, and so the term “access” usually implies data in motion.

Many methods and many types of software and hardware exist for data protection, and the selection and capabilities change rapidly. This article does not attempt to review the plethora of choices, addressing instead the criteria for determining what level of protection is necessary and the information needed by systems experts for achieving that degree of security. By specifying the appropriate level of security without demanding specific approaches, researchers can allow systems experts to select a practical approach to ensuring security from among the many alternatives.

United States (U.S.) federal security standards offer a good starting point for understanding data security needs and the resulting constraints on system

architecture. The National Institute of Standards and Technology (<https://www.nist.gov/>) prepared guidelines for classifying data, released as the Federal Information Processing Systems’ (FIPS) Document 199 (National Institute of Standards and Technology, 2004; Radack, 2004). FIPS 199 defines security classifications of federal information systems. It divides systems into high, moderate and low impact systems based on the potential effect of a breach on individuals and organisations. In March 2006, FIPS Document 200 was released to detail the minimum-security requirements for federal information systems (National Institute of Standards and Technology, 2006). FIPS 200 defines 17 security areas covered under confidentiality, integrity, and availability (CIA) of federal information systems and the information processed, stored, and transmitted by those systems. Information systems employed within the US federal government must follow the categorisation of FIPS 199 and meet the minimum standards of FIPS 200. Other organisations within the country, especially those working with the federal government, are encouraged to adhere to the same standards.

Additional considerations apply when handling or managing health survey data, because of the specific types of content that may be present. Finally, the types of data users and intended types of data usage may influence plans for storage of and access to data. For the newcomer to data security, the multitude of risks, the technical nature of controls and the complexity of assuring a storage and access system that maximises security without impeding work may seem a daunting set of concerns and responsibilities. The authors’ intent is to provide an understanding of security and system vocabulary, to explain common underlying risks for health survey data, and to offer a practical and effective approach for making decisions.

2. Assessing data sensitivity

Much health data are by its nature considered sensitive (German et al., 2001), in that it may lead to harm if known to someone hostile, careless or seeking advantage. Varieties of harm range from hazard to life or well-being as in the case of populations subject to persecution, loss of work opportunities or insurance coverage if medical histories are made known, damage caused by data predators or harassment or unwanted attention of various types. The first step in assessing the risks associated with a particular set of data to be collected, or already collected, is to review the direct and indirect identifiers within it.

Every data set has one or more sources. Items of information may be collected from individuals as

subjects or proxies, from health care or health service providers, or from ancillary sources such as insurance files. Data may be collected through surveys, record abstraction, lab reports, logging by monitoring equipment, or other methods. The file may contain geo-coding for location or indicators of financial liability or payment. Typically, individual subjects or sources are identifiable in the primary data but referenced through anonymised identifiers when information is received from secondary sources. When primary and secondary data items are stored together, all information must be considered traceable and potentially risks harm to the subject or source.

Direct identifiers are those that can be used to connect information to an individual, a small group of individuals, or a vulnerable organisation. The most well-known are names, addresses, telephone numbers, email addresses, website addresses, social media accounts, credit card numbers, driver's license numbers, policy numbers or other data points that are intended for making or maintaining contact. Many direct identifiers comprise a small set of data points, such as the fields of an address, where the apartment number alone provides little identification but, in connection with the street name, town, or zip code, may lead to a single location. Less obvious direct identifiers could include medical appointment dates at known locations, demographics that represent small groups, court or police record identifiers, or laboratory time stamps. Indirect identifiers are those that do not appear to define a small group, such as age or gender, but that may do so under certain circumstances (El Emam, 2008). For example, knowing that a member of a military combat unit is female or knowing that a person's age is greater than 104 may suffice for narrowing an identity search to a very limited number of people.

Researchers often use disclosure analyses (O'Keefe & Rubin, 2015) to assess the presence of direct and indirect identifiers (Kenneally & Claffy, 2010), but such an analysis is typically held in preparation for disseminating data or study findings. Little attention may be given when planning a health survey, setting up data storage and access systems, or processing data during cleaning, coding, and other forms of usage. Yet best practices for protecting privacy recommend limiting collection of direct and indirect identifiers to the minimum necessary for analysis, after which principles of confidentiality and security come into play. Security-minded researchers consider the risks of disclosure before and during data collection and usage, as well as prior to dissemination.

3. Assessing security needs

The necessary level of security for a certain collection of information depends not only on the sensitivity of its contents but also on regulations, requirements, or ethical constraints. Regulations vary by country. This

paper looks only at those in effect within the U.S.; laws in the European Union and other parts of the world differ (Shaffer, 2000). Most research is supported by funding from government, foundations, non-governmental organisations or industry; funding sources and study sponsors often define their own set of requirements for handling and management of data. Perhaps most important are the ethical constraints imposed by researchers themselves and by their organisations; in the U.S., these guidelines and directions generally come from an institutional review board (IRB), whose mission is to protect the rights and welfare of research subjects. IRB panels are set up under the auspices of the U.S. Department of Health and Human Services, whose Office for Human Research Protection has led the country's review of human subjects' rights since the 1970's (Menikoff, Kaneshiro, & Pritchard, 2017).

The best-defined security requirements in the U.S. are those arising from law. The Privacy Act of 1974 (National Archives) defines personally identifying information (PII) as any group of data elements that can identify an individual. This law applies to all data about individuals that is collected, held or used by or for federal agencies, without specifying the technologies of protection, stipulating that agencies, contractors and researchers funded by any branch of the U.S. government are expected to limit the data they collect to the lowest practical level.

The Health Insurance Portability and Accountability Act of 1996 (HIPAA) defines protected health information (PHI) as any individually identifiable health information held by a health care or service provider (Centers for Disease Control, 2003). The law applies to health care and health service providers and all who work with them, whether laboratories, electronic health records companies, data processors, or others. HIPAA rules cover segregation of PHI from PII and many details of data storage and access. For example, patients may sign a HIPAA form at the start of any medical visit in which bio-specimens will sent to laboratories, because the medical care provider must transfer PHI along with the specimen samples. HIPAA's impact on data systems lacks standardisation (Luxton, Kayl, & Mishkind, 2012), and so each grantee, contractor, or independent researcher must oversee legal compliance. Recently it has been suggested that current HIPAA regulations are inadequate in protecting an individual's data from the rapid advances in the use (and misuse) of "big data" (Cohen and Mello, 2018).

Two other laws may apply in some circumstances. The Federal Information Systems Management Act (FISMA) of 2002 places limits on equipment, services and processes for the protection of information held by the U.S. government and its contractors (Hulitt & Vaughn, 2010). Additional requirements may apply to specialised topics of study. Data related to clinical

trials may need to comply with the very complex U.S. Code of Federal Regulations Title 21 Part 11 (often called 21 CFR Part 11), about which many books and articles have been written (See for example, Lopez, 2004). Researchers in the U.S. must be aware of which of these laws apply to their work and to ensure compliance with them, just as researchers in other parts of the world must comply with regulations for their locations.

Standards and guidelines are less binding than law but often form the basis of best practices at well-run organisations. Within the U.S., the FIPS Publication 199 of 2004 classifies the impact of data release as low (limited), moderate (serious) or high (catastrophic). This classification scheme is frequently referenced when determining how to secure data during storage and use, and is commonly called FIPS-low, FIPS-moderate and FIPS-high. Most health data falls into the low or moderate levels of impact, with the high level intended for data affecting national security. FIPS 200 defines minimum standards for system implementation at each of those risk levels. The standards apply to all data used by or for federal agencies, and they may be required by contract, funding terms or policy as a legal requirement for health studies.

IRB guidelines do not carry the force of law nor the consequences of a contractual agreement but reflect the standards and integrity of the researchers' organisations and their sponsors (O'Connor et al, 2008). Typically, an IRB panel reflects carefully on legal and ethical concerns, and their directives may go beyond general practices when they feel that a particularly vulnerable population could be put at risk. On occasion, IRB panels may recommend or require specific electronic safeguards for data at rest, in motion, or both.

The level of security required may depend on who provides the data. HIPAA regulations cover PHI collected, stored or processed in the U.S., unless provided by study participants about themselves or by authorised proxies. When given a choice of data sources, health data analysts may find it less burdensome to receive data directly from participants instead of by medical records abstraction, laboratory systems or medical device services; however, the fidelity and completeness of information may be jeopardised by doing so if the patients forget the precise details of what they were told by the medical team and lack copies of their full medical records (Tisnado et al., 2006).

Security planning must also consider risks inherent in allowing access to the data. Ease of access to information weighs against heightened risk. Depending on the study, data users may be participants who need to enter and review what is being stored about them or the subjects for whom they proxy, providers who may need to enter and review data regarding their patients, data managers or

coders who process data, study managers who oversee operations, and researchers who analyse, summarise and report findings. Each type of data user has a different level of "need to know," ranging from read/write/delete access to a single record through data processors who may read and perhaps edit many of the fields. Though risk levels decline with stronger security measures, some forms of electronic protection may impose such barriers that they affect the quality or quantity of information by deterring data entry, processing, or analysis.

The popular image of data protection centres on hardware and software. However, the human aspects may be equally or even more important. Human threat may be deliberate or accidental, internal or external. Deliberate external threats include hackers, espionage by competitors, or any other planned attempt to damage data from outside the perimeter of the storage environment. Similar attacks could take place internally, such as sabotage by disgruntled co-workers. However, it is more likely that threats to data will occur accidentally by trusted staff or by forces of nature such as floods or fires (Jouini, Rabai, & Aissa, 2014), carelessness with passwords, distraction while editing, neglecting to encrypt files on removable media, casual sharing of data with colleagues, and other inadvertent disclosure. When security precautions are too difficult, the risk of human error increases as users bypass the safeguards, fail to log out when away from the computer or otherwise seek to make their work easier.

Faulty user identification can pose a challenge to data security as well, and so effort must be taken to ensure the identity of system users, whether they are putting data in, modifying it or extracting it. Many technical approaches exist for the purpose of identity management (Todorov, 2007). Difficulties may arise in determining the identity of a participant logging into a health survey website or confirming the identity of the person who answers the phone in a telephone survey. FIPS-moderate security requires two-factor electronic identification, which may mean extra expense by providing smart cards, phone apps or verification messaging. Clearly the need for identification varies by role, with a different level of identity confirmation for participants than for analysts who access the complete data collection. Software or services can confirm and track each login to a website, network, FTP site, or other electronic system.

4. Examples of health data security architecture

This section describes three sensitive health studies that are underway at the authors' organisation on behalf of different sponsors, which the authors present as representative of many research studies that collect and use health survey data. These studies provide insight into what restrictions apply and

what measures were taken to meet them. We feel that these ordinary examples can serve as models for research teams who have not previously defined security systems for health data at varying levels of sensitivity.

One example is a study that was assessed as FIPS-199 low impact, and two are studies assessed at FIPS-199 moderate impact. The authors feel that the moderate category deserves more attention because standards are more stringent, and the number of ways of meeting the standards is greater. We do not present any examples of FIPS-199 high impact categorisation, because we have not encountered any health survey that was assessed at that level, where inadvertent release of data could cause severe or catastrophic damage such as loss of life or danger to national security.

To avoid increasing risk to the studies or their participants, the only information provided here relates to the design of security architecture for each and the reasoning behind the choices that were made. The term “security architecture” refers to descriptions of hardware, software, encryption, user access management, and data isolation practices.

Three studies illustrate different approaches to securing sensitive health survey data. None of them are covered by 21-CFR Part 11, but one is regulated by HIPAA. Two comply with FIPS-moderate classification guidelines, and two have additional sponsor-specific security requirements. One sought and received a waiver for a specific aspect of sponsor-requested security. In each case, careful consideration of the subject matter, presence of PHI or PII, regulations, security guidelines and ethics guided the design of the security architecture to minimise risk without inhibiting data use. These three studies differ in the sources and types of data, the types of security constraints under which they operate, and the architecture and policies they use to protect their health-related data.

4.1. Study 1

This study is a health data collection operation that was classified by its university sponsor at a risk level of FIPS-low with HIPAA compliance. This determination was made because data are collected from multiple sources, including medical providers, medical device services, laboratories, and study subjects through a website. Multiple forms of PII are needed to conduct the study, ranging from contacting data about participants through several types of medical identifiers. Because some of the information comes from medical service providers and device services, HIPAA regulations require storage on servers that host only HIPAA-compliant data systems, with isolation of PII from PHI as well as segregation from non-HIPAA studies.

For this study, data and systems reside on a network of HIPAA-compliant servers maintained by the author’s organisation and certified for FIPS-low storage. In keeping with FIPS-low guidelines, data movement into the website via interactive entry or automated service is protected by Secure Socket Layer (SSL) (Freier, Karlton, & Kocher, 2011), while data deliveries from the FIPS-low system to the sponsor for analysis employ Secure FTP (sFTP) technology (Barrett & Silverman, 2001); both of these encrypt data in motion with SSL handling data at the transmission packet level and sFTP encrypting at the file level. Data files undergo cleaning by staff who are physically located on site with direct network connections, or who use an encrypted virtual private network (VPN) to access the data from offsite. All files are protected by encryption while at rest. The choices of SSL, sFTP and encryption software comply with the FIPS-low classification.

In keeping with best practices for both HIPAA and FIPS-low, as well as health ethics, user access is authorised according to study role: participants see only the data they provide through the website; medical and device service providers deliver data in one-way fashion and lack visibility to any other providers’ data; data managers have full access through the internal network; analysts have full access through files they receive. All data users must log into the network first and the study systems. They are trained in security and sign confidentiality forms specific to the study. These details and more were reviewed and approved by the authors’ organisation’s IRB as well as that of the university.

In short, this study operates under the regulatory controls defined by HIPAA law, architectural and operational policies defined by the FIPS-low classification system, and IRB review of study ethics. Outside of labour, the cost of data security includes SSL licensing, sFTP support and indirect expense to the study from use of the FIPS-low network and HIPAA-compliant servers.¹

4.2. Study 2

This study is also a health data collection operation, classified by its U.S. federal sponsor at FIPS-moderate. The study collects data directly from individuals, which means that PII is present. Information is stored and processed in a certified FIPS-moderate data centre maintained by the authors’ organisation, and then transmitted as files securely to the sponsor. HIPAA law does not apply because medical service providers are not involved, and therefore no data fall under the definition of PHI.

However, the study sponsor requires adherence to or specific waivers from agency-specific policies that augment FIPS-moderate data guidelines. Normally,

the sponsor requires all data systems to be completely isolated from those used for studies conducted by other organisations or agencies, and it directs that PII be stored separately from response data. For data storage, this study occupies a virtual server and dedicated backup system that house only this sponsor's studies, a requirement that greatly increases the cost of data storage but complies with agency policies. PII is maintained in files used for contacting, and other response data is stored separately. The study requested and received a waiver to collect data through a shared software system for interviewing respondents by telephone, because costs were prohibitive for implementing an entire operational data-collection centre for its sole use.

The interviewing system complies fully with FIPS-moderate control. FIPS-moderate security includes all protective measures taken for FIPS-low, plus the addition of two-factor logins, extensive logging of file activity, a higher level of encryption, strict encryption controls for movement of data from the FIPS-moderate environment to any other environment and other safeguards.

All data users are trained not only on general security principles and practices, but also receive training specific to the sponsoring agency. Each data user is given visibility to the data according to the principle of least-access; interviewers see data only as they enter a respondent's answers; data managers have full access; analysts have read-only access. Data files are returned to the sponsor through the agency's own secure file transfer system. All steps of the study and delivery were reviewed and approved by the IRB.

To summarise, this study is not bound by data security and confidentiality law, but it does conform to FIPS-moderate guidelines plus additional measures designated by the sponsor, and the ethical aspects of privacy, confidentiality and security were approved by the IRB. Other than labour, the cost of data security includes operation of an isolated server, fees for isolated data backup services, and indirect expense to the study from use of the FIPS-moderate network.

4.3. Study 3

Study 3 does not collect health data from any primary source. Instead, the sponsoring U.S. federal agency sends medical record abstractions to the project team via an agency-supported sFTP site and receives processed data via the same means. Upon receipt, trained staff review and assign medical codes to several types of textual information that may or may not contain PII. Based on uncertainty of the presence of PII, the sponsor chose to classify this study at FIPS-moderate, taking the precaution of extra security to address the possibility that identifying data might be found. HIPAA regulations do not apply because no medical

providers are involved in the work, and therefore the information is not protected as PHI. The agency, as is common within the U.S. government, added security policies to the contracted work, including restrictions on telework and a requirement for sequestered data storage.

The study complies with FIPS-moderate control and some additional storage and access limitations required by the client. Detailed record-keeping tracks all who have access to the data, and data users must undergo training, read sponsor-provided materials and sign affidavits of confidentiality. Data access is given according to just three roles: coders, who see a single data record at a time and cannot remove data from the web-based coding system; coding managers, who have visibility to all in-progress coding via the same limited-access website; and programming staff, who have full access to load, extract and process data files.

Data at rest are housed on isolated servers within the certified FIPS-moderate environment, per sponsor requirements, and data movement from the website to browser is protected by SSL. Coders and their supervisors may connect to the website through a VPN, but others must be physically present at the study site to work on the data.

In brief, the study is not regulated by law, but does follow the FIPS-moderate classification guidelines and complies with behavioural and storage rules defined by the sponsor. Security costs result from SSL licensing, use of multiple dedicated servers and their backup systems and indirect expense from the use of the FIPS-moderate network.

5. Conclusion

In summary, planning for health survey data security needs to begin early in the project and carefully consider a number of determining factors. For small shops and those who are new to data security, a good starting point is to identify the level of sensitivity of the data overall, carefully reviewing the presence PII or PHI, and determining the minimum "need to know." Review and consideration of each data field may reveal items that are not required for the study's research goals, but which heighten the need for security; possibly those items could be eliminated without harm to the study.

Second, one can assess security needs, starting with a determination of whether laws, policies or other formal systems of control must be applied. Some study sponsors require additional safeguards that must be in place prior to collection or receipt of sensitive data. There is wide recognition in the U.S. of the need to comply with FIPS standards, HIPAA, 21 CFR Part 11, and FISMA for certain studies. Others may not be subject to any regulatory control but may wish to

provide a comparable level of security on the basis of ethical concerns or IRB directives.

Third, researchers should identify, categorise, and monitor all those individuals who will contribute to or use the data in some way. Typically, giving all data users the same level of access no longer provides an adequate level of human control. Training and monitoring may reduce the risk of internal negligence, loss or damage, while system architecture provides security from deliberate outsider threats.

Finally, when the risks and needs are known, architecture alternatives and technological approaches can be evaluated for ease of installation, ease of use or cost, with the intent of providing sufficient security in the most efficient and least burdensome way. This type of careful planning can help ensure that health survey data remains protected and secure at all times while providing researchers with access to the data they need.

Note

1. Because server costs vary enormously with power, storage space, type (physical, virtual), manufacturer and other factors, we are unable to provide information on actual expenses.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

- Barrett, D. J., & Silverman, R. E. (2001). *SSH, the secure shell: The definitive guide*. O'Reilly Media, Inc., Sebastopol, CA, USA.
- Centers for Disease Control and Prevention. (2003). HIPAA privacy rule and public health. Guidance from CDC and the US department of health and human services. *MMWR: Morbidity and Mortality Weekly Report*, 52(Suppl. 1), 1–17.
- Cohen, I. G., & Mello, M. M. (2018). HIPAA and protecting health information in the 21st century. *JAMA*, 320(3), 231–232.
- El Emam, K. (2008). Heuristics for de-identifying health data. *IEEE Security & Privacy*, 6(4) 58–61.
- Freier, A., Karlton, P., & Kocher, P. (2011). *The secure sockets layer (SSL) protocol version 3.0* (No. RFC 6101).
- German, R. R., Lee, L. M., Horan, J., Milstein, R., Pertowski, C., & Waller, M. (2001). Updated guidelines for evaluating public health surveillance systems. *MMWR. Recommendations and Reports*, 50(1), 1–35.
- Hulitt, E., & Vaughn, R. B. (2010). Information system security compliance to FISMA standard: A quantitative measure. *Telecommunication Systems*, 45(2–3), 139–152.
- Jouini, M., Rabai, L. B. A., & Aissa, A. B. (2014). Classification of security threats in information systems. *Procedia Computer Science*, 32, 489–496.
- Kenneally, E. E., & Claffy, K. (2010). Dialing privacy and utility: A proposed data-sharing framework to advance internet research. *IEEE Security & Privacy*, 8(4), 31–39.
- Liu, S., & Kuhn, R. (2010). Data loss prevention. *IT Professional*, 12(2), 10–13.
- López, O. (2004). *21 CFR part 11: Complete guide to international computer validation compliance for the pharmaceutical industry*. CRC Press, Boca Raton, FL, USA.
- Lumpkin, J. R. (2000). Perspective: E-Health, HIPAA, and beyond: The chair of the expert advisory board on health data outlines the most important issues in developing a secure health information system. *Health Affairs*, 19(6), 149–151.
- Luxton, D. D., Kayl, R. A., & Mishkind, M. C. (2012). mHealth data security: The need for HIPAA-compliant standardization. *Telemedicine and e-Health*, 18(4), 284–288.
- Menikoff, J., Kaneshiro, J., & Pritchard, I. (2017). The common rule, updated. *New England Journal of Medicine*, 376(7), 613–615.
- National Archives. The Privacy Act of 1974. Retrieved from <https://www.archives.gov/about/laws/privacy-act-1974.html>
- National Institute of Standards and Technology. (2004). Standards for security categorization of federal information and information systems. Retrieved from <https://csrc.nist.gov/csrc/media/publications/fips/199/final/documents/fips-pub-199-final.pdf>
- National Institute of Standards and Technology. (2006). Minimum security requirements for federal information and information systems. Retrieved from <https://csrc.nist.gov/csrc/media/publications/fips/200/final/documents/fips-pub-200-final.pdf>
- O'Connor, M. K., Netting, F. E., & Thomas, M. L. (2008). Grounded theory: Managing the challenge for those facing institutional review board oversight. *Qualitative Inquiry*, 14(1), 28–45.
- O'Keefe, C. M., & Rubin, D. B. (2015). Individual privacy versus public good: Protecting confidentiality in health research. *Statistics in Medicine*, 34(23), 3081–3103.
- Radack, S. (2004). *Federal Information Processing Standard (FIPS) 199, standards for security*. National Institute of Standards and Technology, Gaithersburg, MD, USA.
- Shaffer, G. (2000). Globalization and social protection: The impact of EU and international rules in the ratcheting up of US privacy standards. *Yale Journal of International Law*, 25, 1.
- Tisnado, D. M., Adams, J. L., Liu, H., Damberg, C. L., Chen, W. P., Hu, F. A., ... Kahn, K. L. (2006). What is the concordance between the medical record and patient self-report as data sources for ambulatory care? *Medical Care*, 44, 132–140.
- Todorov, D. (2007). *Mechanics of user identification and authentication: Fundamentals of identity management*. Auerbach Publications/CRC Press, Boca Raton, FL, USA.