INTERNATIONAL COMPUTATIONAL LAW JOURNAL

CODEX SOURCE ARCHIVE

A HIPAA “Expert Pathway” Method to Achieve More Efficient Access to Protected Healthcare Information

James H. Jose, Rachele Hendricks-Sturrup

Computational Law Report · Release 3.0 · 2021-12-07

Reader edition prepared by The Research University under CC BY 4.0. This cover was added; the original PDF pages follow unchanged. Original authorship, publication identity and source links are preserved. This independent republication is not sponsored or endorsed by Stanford University, Stanford Law School, CodeX or MIT.

Volume 1 · Issue 5 · Evidence, Interpretation and Machine Judgment

Original source and license evidence are recorded on the article landing page.

MIT Computational Law Report •

A HIPAA "Expert Pathway" Method to Achieve More Efficient Access to Protected Healthcare Information

James H. Jose, MD, Rachele Hendricks-Sturrup, DHSc, MSc, MA

Published on: Dec 07, 2021

License: Creative Commons Attribution 4.0 International License (CC-BY 4.0)

ABSTRACT

The growing demand for multi-institutional sharing of electronic healthcare record (EHR) data for research combined with the complex and time-consuming process of generating, negotiating, and executing data use agreements (DUAs), remains a significant frustration for researchers and research administrators. We highlight an administrative method of data sharing among healthcare researchers that leverages the “expert pathway” described in the Health Insurance Portability and Accountability Act (HIPAA) rule, which can allow sharing of protected health information (PHI) with a reduction in burdens related to DUAs and Institutional Review Boards (IRB) reviews. Specifically, we describe how data can be held by an infrastructure custodian or data enclave operating under a single DUA contract (versus multiple, project specific DUAs), allowing, with adequate controls, researchers to review private, aggregated results as computational outputs to database search queries. We discuss how this administrative method can reduce complexity in PHI data sharing, uphold reliable patient privacy, and leverage the HIPAA expert (versus safe harbor) pathway.

Key words: HIPAA Privacy Rule; data sharing; data use agreement; protected health information; de-identification, anonymization, differential privacy, statistical disclosure control; data enclave.

---

1. Introduction

Despite the growing demand for multi-institutional sharing of electronic healthcare record (EHR) data for research purposes, the complex and time-consuming process of generating, negotiating, and executing DUAs remains a significant frustration for researchers and research administrators [1,2]. A survey of 17 research organizations by the Administrative Data Research Facilities Network reported that “We heard from almost every organization that setting up the necessary legal infrastructure is a time-consuming, nebulous, and unpredictable process” [3]. A contributor to complexity is discomfort with data privacy protection levels under HIPAA’s safe harbor de-identification methods that has been characterized as “broken promises” because they do not adequately safeguard data subjects from reidentification [4,5,6,7,8]. Some have argued that even privacy protections offered by sophisticated anonymization methods cannot be effective without administrative controls that can narrow conditions of use and thus allow credible estimation of risk [9,10,11]. A panel of the National Academy

of Medicine proposed that de-identification is not alone adequate to protect privacy [12].

Data enclaves are operational units that embody principles arising from legal, informatics, technology, and ethics sources pursued in design of privacy systems. The operational characteristics of data enclaves have been extensively reviewed and they are widely considered the most secure repository for sensitive information [13,14,15,16]. Many exemplary implementations have been described, particularly for government entities and social sciences research. But data enclaves require unique infrastructure and practices that may conflict with the current methods allowed under the HIPAA “safe harbor” rule. Of concern is that an HHS panel concluded that “Despite this increasingly important advantage, Expert Determination is used less frequently than Safe Harbor” and “The lessons from de-identification research are not informing day-to-day practice” [17]. We believe that privacy goals would be advanced if data enclave methods were applied to nearly all multi-institutional data sharing, but such acceptance will require enclaves to become more appealing to researchers. We argue that a specific use of the HIPAA rule “expert pathway” with enclave methods can enable access to PHI without the need for researchers to seek DUA or IRB approvals.

2. The HIPAA expert pathway can enable access to healthcare data without a DUA requirement

The use of the HIPAA de-identification “expert pathway” can enable analysis of a data set that contains PHI while maintaining the access transaction compliantly de-identified. HIPAA offers two methods of compliant de-identification. The most commonly used is the “safe harbor” method which requires the removal of 18 standard identifiers such as name, address and precise dates. Alternatively, the “expert pathway” allows organizations to share a data set after “a person with appropriate knowledge of and experience with generally accepted statistical and scientific principles and methods for rendering information not individually identifiable” certifies that the risk of re-identification is “very small.” Although this usually involves the use of anonymization methods, the rule does not place restrictions on measures that can be employed to limit the ability of recipients to re-identify individual patients. Updated guidance from HHS in 2012 stated that the expert pathway may “consider the technique of limiting distribution of records through a DUA or restricted access agreement…the specific details of such an agreement are left to the discretion of the expert” [18]. It was also noted that a DUA is not required for researchers to share de-identified data under these administrative controls.

A key step in designing a more efficient sharing process is to recognize the distinction between "access to data" and "possession of data." With appropriate technology enabled, researchers in many cases can address research questions by querying a data set without “physical possession” of the raw data. The data custodian role can be served by a trusted steward who manages the data storage site to ensure effective security and access control while enabling the system to return aggregate results. Researchers can gain remote access to enclave services via virtual machine functionality isolated from users’ computers and using analytic tools that run inside the enclave. Data cannot be downloaded by the researcher’s computer without a DUA that specifies fair use and custody expectations.

Critical to this schema is the ability of the custodian to provide a highly reliable level of privacy protection that can flexibly respond to emerging threats. Enclaves enable this by abandoning sole dependence on de-identification measures and creating a framework with administrative measures in which re-identification risk can be more effectively assessed. These measures include use of statistical control measures that limit access to results with re-identification risk, monitoring of query activity, and policies limiting data access to a population of aligned, trained researchers [10,11,19,20].

While these methods require significant infrastructure and trained staff, they allow for a system in which the primary DUA is no longer the obligation of the researcher but is held by the steward-custodian. The DUA defines the administrative rules the custodian must implement to receive data. Custodians in this schema function as a data keeper extension of the healthcare organization positioned between researchers and data to allow viewing of aggregate results without exposing personally identifiable information. Researchers only have access to aggregate data that is reliably de-identified before they see it. Access is usually restricted to qualified researchers authorized by the custodian organization and its partners. Accessing data by this method eliminates the need for IRB approval and DUAs on the part of the researcher because de-identified data is not subject to HIPAA rules. When the DUA is held by the custodian, access can be administered as “role-based” rather than project or user based. Rather than waiting months for approval of an IRB and negotiation of a multi-institutional DUA, researchers can test a hypothesis at will. Patient privacy protections are in the hands of a data coordinating staff well equipped to maintain and monitor them, a feature that should encourage organizations to share data. One potential schema is detailed in Table 1.

1. **Data is released to a custodian under a Data Use Agreement (DUA). This data may include Protected Health Information (PHI).**
• The custodian under this contract may not directly engage in research, publish findings or analyze data other than for database maintenance. • The custodian must incorporate a substantial suite of analytic tools on the storage site. Offering data curation services and informatics consultation can improve utility.
2. **Authorized researchers have remote access to the data.**
• Researchers cannot download raw line-item data but are able to analyze data with tools on the custodian's web site or cloud service. • Query results with elevated risks of re-identification are minimized by masking results with < 11 subjects and principles of "statistical disclosure controls." • While a DUA is required for custodians, it is not required for researchers.
3. **Criteria for authorized users is determined by the governing board.**
• Researchers should agree to an "Acceptable Use" statement provided at login that stipulates their organization's policy for data privacy and use of enclave data. • Limiting user access to qualified, aligned researchers accountable to an identified organization is an important privacy and security protection.
4. **Linkage to external databases can in selected instances be performed by the custodian using a hashed ID strategy, although data integrity may limit this to instances where substantial infrastructure resources are available.**
5. **Organizations should perform their own risk analysis of the potential to bypass protections with artful methods. The custodian should monitor and audit for proper use as described in the DUA.**

Table 1: Features of a data enclave that can lower DUA/IRB requirements

3. Conclusion

The administrative and computational method discussed leverages the HIPAA expert pathway to enable broader access to PHI by recognizing the role of custodial and administrative methods in privacy protection. While broadening access to PHI, this method can increase the accountability for privacy since many tasks rest with a committed team that can implement and monitor a system designed to offer best protections and engage emerging threats. While data enclaves have been in existence for some time, they are not commonly promoted to compliantly access PHI by altering

the investigator's relationship with IRB and DUA requirements. With the increasing frequency of requirements that investigators submit data sharing plans with research protocol proposals, it is all the more important to find methods which can induce the transition to safer, more efficient methods of sharing patient information.

References

1. Rockhold F, Nisen P, Freeman, A. Data Sharing at a Crossroads. N Engl J Med 2016 375:1115-1117 DOI: 10.1056/NEJMp1608086
2. Sim I, Stebbins M, Bierer B. et al. Time for NIH to lead on data sharing. Science 2020 367:1308-1309. DOI: 10.1126/science.aba4456
3. https://repository.upenn.edu/cgi/viewcontent.cgi?article=1002&context=admindata_reports
4. Ohm P. Broken promises of privacy: responding to the surprising failure of anonymization UCLA Law Review 2010 57:1701. Accessible at SSRN: https://ssrn.com/abstract=1450006
5. Sweeney L, Yoo J, Perovich L, et al. Re-identification risks in HIPAA safe harbor data: a study of data from one environmental health study. Technology Science 2017 2017082801 PMID: 30687852
6. Na L, Yang C, Lo C. Feasibility of reidentifying individuals in large national physical activity data sets from which protected health information has been removed with use of machine learning. JAMA Network Open 2018 doi:10.1001/jamanetworkopen.2018.6040
7. Rocher L, Hendrick J, de Montjoye Y. Estimating the success of re-identifications in incomplete datasets using generative models. Nature Communications 2019 10:3069 Accessible at: https://doi.org/10.1038/s41467-019-10933-3
8. Narayanan A and Shmatikov V. "Robust De-anonymization of Large Sparse Datasets," 2008 IEEE Symposium on Security and Privacy (sp 2008), Oakland, CA, 2008, pp. 111-125, doi: 10.1109/SP.2008.33.
9. Kolata G. Your data were 'anonymized'? These scientists can still identify you. The New York Times July 24, 2019, Section A, Page 8. Accessible at: https://www.nytimes.com/2019/07/23/health/data-privacy-
10. Rubinstein I, Hartzog W. Anonymization and risk. 2015. Washington Law Review 2016 91:703-760. NYU School of Law, Public Law Research Paper No. 15-36. Available at: SSRN: https://ssrn.com/abstract=2646185
11. Lagos l, Polonetsky J. Public versus nonpublic data: the benefits of administrative controls. Stanford Law Review 2013 66:103-109. Accessible at:

https://review.law.stanford.edu/wp-content/uploads/sites/3/2016/08/66_StanLRevOnline_103_LagosPolonetsky.pdf
12. Institute of Medicine 2015. Sharing Clinical Trial Data: Maximizing Benefits, Minimizing Risk. Washington, DC: The National Academies Press. https://doi.org/10.17226/18998.
13. Levenstein M, Lyle J. Data: sharing is caring. Advances in Methods and Practices in Psychological Science 2018 1:95-103. Accessible at: DOI: 10.1177/2515245918758319
14. Platt R, Lieu T. Data enclaves for sharing information derived from clinical and administrative data. JAMA 2018 320:753–754. doi:10.1001/jama.2018.9342
15. Groves R, Harris-Kojetin B. Protecting privacy and confidentiality while providing access to data for research use. In: National Academies of Sciences, Engineering, and Medicine 2017. Innovations in Federal Statistics: Combining Data Sources While Protecting Privacy. Washington, DC: The National Academies Press. Accessible at: https://doi.org/10.17226/24652.
16. Lane J, Schur C. Balancing access to health data and privacy: a review of the issues and approaches for the future. Health Services Research 2010 45:1456-1467. Accessible at: DOI: 10.1111/j.1475-6773.2010.01141.x
17. Stead W, Chair National Committee on Vital and Health Statistics letter to Secretary of HHS Thomas Price. https://www.ncvhs.hhs.gov/wp-content/uploads/2013/12/2017-Ltr-Privacy-DeIdentification-Feb-23-Final-w-sig.pdf
18. “Guidance Regarding Methods for De-identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule” November 26, 2012. Accessible at: https://www.hhs.gov/sites/default/files/ocr/privacy/hipaa/understanding/coveredentities/De-identification/hhs_deid_guidance.pdf
19. Samarati P, Sweeney, L. 1998. Protecting privacy when disclosing information: k-anonymity and its enforcement through generalization and suppression. Tech. Rep. SRI-CSL-98-04, SRI Computer Science Laboratory, Palo Alto, CA. Accessible at: https://epic.org/privacy/reidentification/Samarati_Sweeney_paper.pdf
20. Griffiths E, Greci C, Kotrotsios Y, et al. Handbook on Statistical Disclosure Control for Outputs. The Safe Data Access Professionals Working Group 2019. Accessible at: https://doi.org/10.6084/m9.figshare.9958520

Header image generated with Wombo