HHS Public Access

Author manuscript

Comput Inform Nurs. Author manuscript; available in PMC 2021 January 19.

Published in final edited form as:

Comput Inform Nurs. 2020 September ; 38(9): 427–430. doi:10.1097/CIN.0000000000000677.

Data infrastructure for sensitive data: Nursing's role in the development of a secure research enclave

Lisa C. Lindley, PhD, RN, FPCN, FAAN[1], **Radion Svynarenko, PhD**[1], **Theresa L. Profant, BS**[1]

[1.] University of Tennessee, Knoxville, College of Nursing

Nurse scientists are increasingly using protected, sensitive data in their research. Sensitive data are files that include person-level information that would enable a researcher to identify a study participate or reveal private, personal information. These data can be a valuable resource for building evidence for healthcare practice and health informatics. Increasingly, the level of security needed to conduct sophisticated data analysis for health informatics reaches beyond personal computer equipment or hard drives.[1] Agencies that own data available for nursing research require advancements in data security such as the Centers for Medicare & Medicaid Services (CMS). As a result, this article describes a recent collaboration between the University of Tennessee (UT) College of Nursing and Office of Information Technology in the technology advancement of nursing science and data management where data files can be processed in an environment protected to the highest standards for privacy, confidentiality, and data security in the computing industry.

Types of Data

Data are often classified as protected health information (PHI), personally identifiable information (PII), or sensitive data. PHI is an individual's health information that is created by a healthcare provider and that identifies or could reasonably identify a study participant. Governed by the Health Insurance Portability and Accountability Act (HIPAA), common PHI identifiers include name, social security number, account number, and biometrics.[2] PII data also include information about a study participant that might enable identification, such as address, social security number, driver's license, or biometrics.[3] When drawing from PHI and PII, sensitive data such as social security number, name, medical information, and account information are present and can be used for direct identification.

Databases used in healthcare research frequently include PHI and/or PII variables. For example, Medicaid data contains PHI and PII. These data elements include zip code, ICD-9/10 codes, and date of birth. Medicaid is one of the largest insurance providers to children in the US. These data are critical in pediatric research because it is one of the few national data sources about children's health, especially at end of life.[4] Nurse researchers, such as our team, who investigate issues of access, quality, and cost of care for children in

Correspondence: Lisa C. Lindley, PhD, RN, FPCN, FAAN, College of Nursing, University of Tennessee - Knoxville, 1200 Volunteer Blvd. #147, Knoxville, TN 37996-4180, 865-974-0653, llindley@utk.edu.

Declaration of Conflict of Interest:
The authors declare no conflicts of interest with respect to the authorship and/or publication of this article.

hospice, also need social security numbers to confirm dates of death. Our data are highly sensitive. We have an ethical and contractual requirement (i.e., data use agreement) to ensure our Medicaid data are secure. While most nurse researchers are able to use their organizations' HIPPA-compliant data storage sites, such as Google Drive or OneDrive, these sites many not be sufficient for highly sensitive data. They may not meet the requirements of the data owner. New approaches to data infrastructure are needed and secure enclaves are one approach.[5]

Interprofessional Partnership

The UT College of Nursing and the UT Office of Information Technology (OIT) have a strong history of inter-professional partnership. In the past, nursing researchers have worked with OIT staff to identify and create data protocols (e.g., data destruction). However, in 2018 our research team and OIT found ourselves meeting at the right place and right time regarding secure data storage and analysis. The university was just beginning to create a secure research enclave (Figure 1). The enclave is data infrastructure for sensitive data that operates on its own secure computer server with limited access. It is analogous to data lockdown, which requires multiple steps in the user authentication process.[5] There is limited access into and out of the enclave for researchers and their data. Once in the enclave, a researcher has a remote, secure, and customized computing environment and data storage, but they are not able to access the internet or any other files or systems.

Our encounter with the secure enclave began when our team received a federal award to conduct pediatric hospice research with Medicaid data. Although we had worked with Medicaid data in the past, the rules of engagement with the Centers for Medicare and Medicaid Services (CMS) had changed, especially regarding data security.[6] CMS now requires a more detailed data management plan that includes a secure data infrastructure for highly sensitive data. There are also restrictions on the type of data infrastructure including no cloud-based systems. This negated our organization's traditional data environments, such as OneDrive and Google Drive. One option our team explored was to create a locked-down computer that had no internet access and that only one investigator could use. Given that our team operates at a distance and postdoctoral fellows were joining the team to work on the data, this option was not appropriate. In consultation with the UT Office of Research and Engagement (ORE), OIT, and the Advanced Computing Facility (ACF), we identified a secure enclave as meeting the needs of our data owner (CMS) and research team. However, the secure enclave was only a prototype at that time. The secure enclave was conceptualized with two parts: 1.) secure information processing (SIP) for Big Data analytics using a Linux operating system and 2.) a virtual machine (VM) statistical computational environment using a Windows operating system. From the time of our first meeting, the SIP was operational within four months, followed by the VM six months later. The development of these secure data environments was a work in progress, and our research team was the fourth client in the SIP. Once our research data was operational, however, we migrated to the VM where we were the first client at the university.

Beta Testing

As the first client in the secure enclave VM, our nursing research team became the beta testers for the developers at OIT over a 1-year period. In this role, we assisted in secure enclave process developments such as the data transfer protocol. When the secure enclave was first operational, data had to be manually uploaded onto the secure enclave server and no information could be extracted. We worked with OIT to test a variety of system upgrades for data transfer to and from the enclave. Activities included physically taking the files to the OIT office, using an encrypted external drive, and deploying an encrypted file transfer service (SFTP).

Our team also tested a variety of hardware and software modifications to the secure enclave. We assisted in the development of a secure data transfer node when it was first installed in the enclave. The node is a “window” to the external world that only the enclave administrator has access to for moving data in and out of the enclave. Our team tested adding software to the enclave environment, such as encryption and word processing software.

Other beta testing activities included general enclave operations. We prototyped adding a new user to the enclave and addressing systems problems (e.g., a systems crash). For example, the process of adding new team members has been standardized to include authorization by the study Principal Investigator and confirmation by the Office of Sponsored Programs that the new person has completed compliance training. Our team reviewed best practices for communications with OIT and within our team. At the onset of the enclave, we simply emailed the enclave support staff if we needed assistance. This process evolved to formally submitting an OIT job ticket that is routed to the enclave group. The ticket is assigned a number for tracking and all communications on the job are electronically linked to the ticket for ease of documentation. Our team also created internal project documentation to track which computers were accessing the enclave and monitor the flow of files into and out of the enclave. Our role as beta tester has provided an important service to our organization and has improved the functioning of the secure enclave.

Future Concerns

We anticipate engaging with OIT on future developments in the secure enclave. First, project memory space in the VM is restricted, which may cause workflow problems for researchers. When our VM project space was configured, we requested 1TB of memory. This was substantially more than a standard secure enclave VM project space at our organization, which is 32GB. As our research team continues to develop and save working files, we anticipate that our file space will exceed 1TB. We will need to work with OIT to plan for this event without having the enclave slow down or lock us out because we have exceeded our space allotment.

Second, adding new projects to the VM will present infrastructure challenges. We expect to add new research projects to the enclave. Some of the projects will include new data files, while others will add data to an existing project. Although OIT is planning for a distinctly

new project space for each new research project, there will be teams like ours that need to use existing files. The question will be whether we can duplicate a project file in the enclave or create an internal linkage between project files.

Third, the enclave will continue to evolve and require ongoing beta testing. As the secure enclave matures, there will be new developments. On the immediate horizon, we expect that OIT will develop mechanisms for researcher movement between the Big Data/SIP and VM sides of the enclave. They will also establish procedures for data breach drills and data destruction at the end of a research project. Our nursing research team will be available to continue as beta testers for these and any other new developments, procedures, or protocols created for the enclave.

Finally, there will still be operational issues to resolve between OIT and users. For example, the process of data backup within the enclave needs to be addressed. Currently, there is no mechanism to back up research data; however, there is a method for restoring the VM to a prior date. Ongoing discussions are needed to reconcile the needs of researchers and technical limitations of the enclave. Overall, the future challenges of operating in the secure enclave represent opportunities for nurse scientists to partner with OIT in the development of cutting-edge research infrastructure.

Impact on Nursing Research

Our work in the development of a secure data infrastructure, which took approximately 2 years (Figure 2), has the potential to impact future nursing research. As nurse researchers increasingly need to access sensitive healthcare data for their research, they must engage with their Information Technology offices to ensure that their needs are met. Nursing must have a voice in data storage and security so that our data owners have confidence that our data environment is safe. Without those data infrastructures, nurse researchers will not be granted access to important data needed to ask critical nursing questions.

Acknowledgments

Funding: This publication was made possible by Grant Number R01NR017848 from the National Institute of Nursing Research. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the National Institute of Nursing Research or National Institutes of Health.

References

1. Samadbeik M, Gorzin Z, Khoshkam M, Roudbari. Managing the security of nursing data in the electronic health record. Acta Inform Med. 2015;23(1):39–43. [PubMed: 25870490]
2. Herold R, Beaver K. The Practical Guide to HIPAA Privacy and Security Compliance. 2015; CRC Press: Boca Raton, FL.
3. Moore W, Frye S. Review of HIPAA, Part 1: History, protected health information, and privacy and security rules. J Nuc Med Technol. 2019;47(4):269–72.
4. Cozad MJ, Lindley LC, Eaker C, Carlosh KA, Profant T. Debunking myths about health insurance claims data for public health research and practice. Am J Public Health. 2019;109(11):1584–5. [PubMed: 31577482]
5. Foster I Research infrastructure for the safe analysis of sensitive data. Annals AAPSS. 2018;675:102–20.

6. Ruttner L, Borck R, Nysenbaum J, Williams S. Guide to MAX data Medicaid Policy Brief #21. 8, 2015 Mathematica Policy Research.

Key Points:

- Nurse scientists often use highly sensitive patient data in their research.
- Data owners are increasingly requiring more complex data management plans for storage and analysis of sensitive data.
- Nurses are at the forefront of innovative data security such as secure enclave within their organizations and universities.

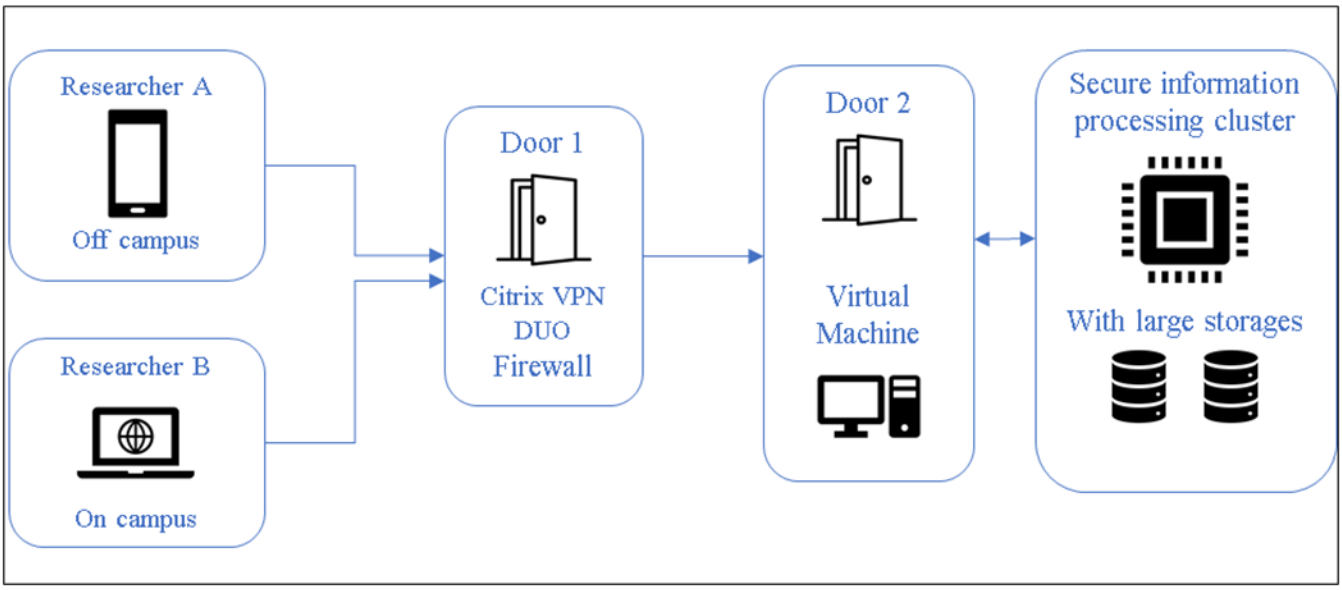

Figure 1.
Secure Enclave Scheme

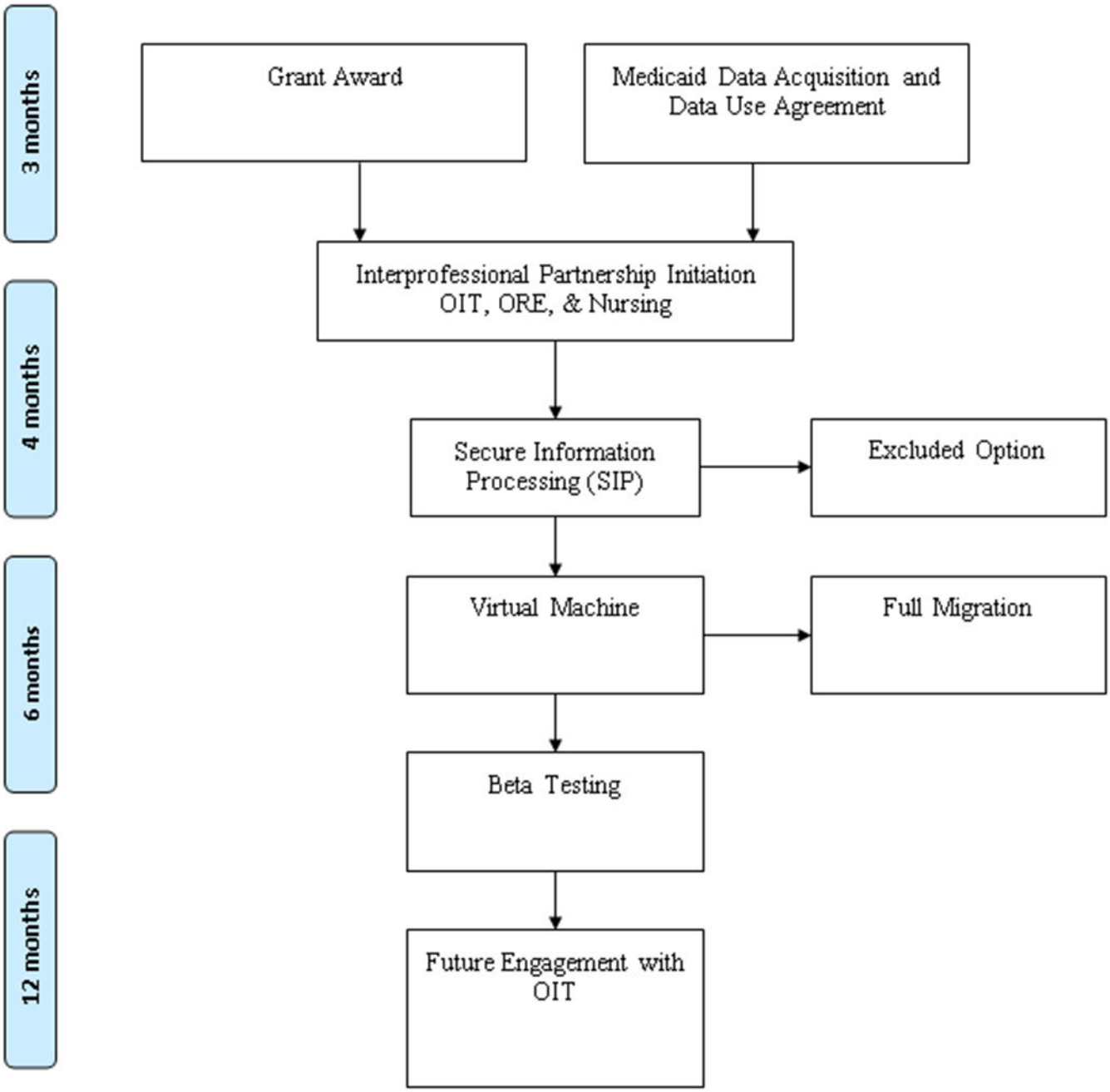

Figure 2.
Secure Enclave Activities and Timeline