Link, M.R., A. Shankar, D.Y. Hancock, R. Henschel, S. Michael, C.A. Stewart. 2021. Security standards compliance and ease of use of high performance computing systems in clinical research. Presented at the Fourth ISC Workshop on HPC Applications in Precision Medicine. A virtual workshop (http://ncihub.org/groups/hapm21) of the International Supercomputing Conference 2021. Available at: http://hdl.handle.net/2022/26637

ISC HPC 2021 Workshop ▸ Overview

application | Big data | exascale | HPC | medicine | precision medicine

Fourth ISC Workshop on HPC Applications in Precision Medicine – Virtual (Rescheduled from 2020)

July 2, 2021 | Virtual

Conference Website: http://isc-hpc.com

Security Standards Compliance and ease of use of High Performance Computing systems in clinical research

Matthew R. Link
Research Technologies Division
Pervasive Technology Institute
Indiana University Bloomington
IN, USA
mrlink@iu.edu
orcid.org/0000-0002-7764-9110

Anurag Shankar
Ctr. App. Cybersecurity Res.
Pervasive Technology Institute
Indiana University
Bloomington IN, USA
ashankar@iu.edu
orcid.org/0000-0001-5878-9217

David Y. Hancock
Research Technologies Division
Pervasive Technology Institute
Indiana University Bloomington
IN, USA
dyhancoc@iu.edu
orcid.org/0000-0001-8082-8980

Robert Henschel
Research Technologies Division
Pervasive Technology Institute
Indiana University
Bloomington IN, USA
henschel@iu.edu orcid.org/0000-0001-5878-9217

Scott Michael
Research Technologies Division
Pervasive Technology Institute
Indiana University
Bloomington IN, USA
scamicha@iu.edu orcid.org/0000-0002-0509-1197

Craig A. Stewart
Dept. of Computer Science Luddy School of Informatics, Computing, and Engineering
Indiana University
Bloomington IN, USA
orcid.org/0000-0003-2423-9019

***Abstract*— Precision health research and personalized health therapies involve analysis of protected health information. In 2007, Indiana University established the ability to analyze protected health information (HIPAA alignment) as the minimal and default security level for its research High Performance Computing (HPC) systems and research storage systems. This resulted in a dramatic increase in the use of IU HPC systems by clinical researchers. Security levels were later upgraded to FISMA Low as a default. We recommend that, within the US, FISMA (Federal Information Security Modernization Act) Low compliance be the default minimal level of security for large-scale HPC systems. This would facilitate precision medicine research and enable higher education HPC resources to be used in response to future civil health emergencies.**

Keywords*— *Precision Medicine, HPC, Protected Health Information, PHI, HIPAA, FISMA, clinical research, COVID-19

I. Introduction

Strategies for securely handling protected health information (PHI) are essential in any organizational setting where High Performance Computing (HPC) systems are used to enable precision health research, clinical treatment, or both. Within the US, handling of such protected health information (PHI) must be secured in a way that is compliant with relevant laws and regulations. The purposes of this report are: 1) to describe Indiana University's experience with respect to enabling analysis and storage of protected health data on university-owned High Performance Computing (HPC) and research storage systems and, 2) on the basis of this experience make generalizations and recommendations that may aid other institutions in supporting precision health research with HPC systems. In particular, we make recommendations about the minimum security standards that should be adopted by any university or college that has significant HPC resources, given their value in precision health research as well as in supporting research related to future civil health emergencies as has been demonstrated in the US response to the COVID-19 pandemic.

II. Background: HIPAA and FISMA

HIPAA – the US Health Insurance Portability and Accountability Act – was initially passed in 1996. It provided the US Health and Human Services Secretary with rule-making powers leading to the promulgation of the HIPAA Security Rule in 2000. The rules have been amended since then, with the last significant revision in 2013 [1]. The HIPAA Security Rule specifies 18 patient identifiers that must be protected, including patient names, addresses, and telephone numbers (genomic data are not yet included, but likely will be in the future). HIPAA applies to most traditional providers including health insurers, clinicians, and hospitals who curate PHI, including universities that contain a medical school. HIPAA is non-prescriptive – it does not detail how each safeguard should be implemented. Rather, it asks for "reasonable and appropriate" implementation consistent with available resources, budget, organization size, and other such factors. HIPAA is self-asserted; this means that the relevant authorities at an institution agree with an HPC or storage system provider's interpretation and implementation of the HIPAA Security Rule safeguards. HIPAA implementations may thus vary from one organization to another. One speaks of a system as being "aligned" with HIPAA, as there is no security threshold specifying what compliance entails.

FISMA – the Federal Information Security Modernization Act, passed in 2014 – is different [2]. FISMA mandates compliance with the National Institute of Standards and

Technology (NIST) Risk Management Framework (RMF) and security control baselines defined in NIST Special Publication 800-53 [3]. Every US government agency and its contractors must implement FISMA. Unlike HIPAA, FISMA sets out specific control baselines for how to secure systems, and FISMA compliance can be at one of three levels: Low, Moderate, or High. FISMA is more comprehensive than HIPAA, with nearly fifty times as many controls. Adopting the FISMA Low standard generally enables an institution to align with HIPAA and most other cybersecurity guidance affecting research. NIST provides a HIPAA to NIST 800-53 mapping in NIST 800-66 [4].

III. History of HIPAA alignment and FISMA Low compliance as default for IU HPC Systems

In 2000, The Lilly Endowment, Inc., a private charitable trust operating within the State of Indiana, provided a $105M grant to Indiana University (IU) to fund the INdiana GENomics Initiative (INGEN). INGEN's purpose was to accelerate adoption at IU of what was then called genomic-based medicine (now called Precision Medicine). Enabling such research was important to the IU School of Medicine (IUSM) for several reasons, including belief in the future importance of genomic medicine in research on diseases that are areas of strength for IUSM, such as cancer, inflammatory diseases, and alcoholism.

The INGEN grant award included $7M for the acquisition and support of HPC resources. These monies were budgeted for the Research Technologies Division of University Information Technology Services [5] (affiliated with the IU Pervasive Technology Institute [6] since 2008). Research Technologies leadership went to work immediately hiring new staff to support biomedical researchers. Research Technologies initially set a very simple policy regarding use of IU's research HPC systems and related storage systems: biomedical researchers were welcome to use these systems as much as they wanted, without any limitation, other than that they were required to de-identify any patient data before such data were stored on IU storage systems. The result was simple: most medical researchers working with patient data did not make use of IU HPC systems and initially regarded the $7M budgeted for advanced computing within the INGEN project as a waste of money.

In 2007, Research Technologies leadership launched an initiative to make HIPAA compliance the default minimum security stat of all of its HPC and storage systems. This was done in response to encouragement from researchers and leadership of the IU School of Medicine, as well as the discovery that at least one storage system had to be aligned with HIPAA in order to meet a commitment made by IU to support fetal alcohol spectrum disorder research. Given the need to enable PHI storage and analysis on one system, it made sense to maximize the resources available to the IU medical and health research community and align all centrally provided computational and storage systems with HIPAA. An external consultant performed a gap analysis and risk assessment. Considerable effort was then spent within Research Technologies to establish and document security practices. In 2009, the IU Compliance Office confirmed Research Technologies' assessment that all Research Technologies HPC and storage systems were suitable for storage and analysis of PHI in alignment with the HIPAA Security Rule. In 2013, Research Technologies decided to graduate to the next security level by adopting FISMA Low as its minimum security standard. The NIST RMF was put in place and the NIST 800-53 Low security baseline was implemented. HIPAA compliance was achieved via control mapping as per NIST 800-66. Currently, Research Technologies personnel consult with privacy experts within IU and with individual research teams to help them develop secure workflows for handling research PHI. To the best of our knowledge, IU was the first US public-sector HPC center to make HIPAA alignment a default condition for research computing cyberinfrastructure systems.

IV. Results: Growth of use of IU HPC Systems

As of the announcement of the INGEN grant award to IU in 2000 with $7M allocated for HPC systems, research storage, and consulting support, usage of IU supercomputers was somewhere between "none at all" and "trivial." Between 2000 and 2007, usage of IU's primary supercomputer, an IBM SP, began to increase. However, initial usage was almost entirely limited to database functions, taking advantage of proprietary IBM software that had, for that time, very sophisticated capabilities for querying and joining across multiple disparate databases. This meant trivial CPU usage; in fact, there was relatively little use of IU HPC systems by clinical researchers until 2009 when IU announced HIPAA alignment. This announcement was made only within IU so as not to present IU systems publicly as an interesting challenge to malicious individuals and communities.

Statistics on use of HPC systems by IU School of Medicine researchers are somewhat limited until 2012. Figure 1 below shows the CPU-hours consumed by users of IU's HPC systems from 2012-2020 in two categories: research teams that affiliated with the IU School of Medicine, and all other researchers.

Use of IU central HPC systems by IU School of Medicine researchers was near 0 as of 2005, and usage was light up to 2009, but there was a big jump from then to 2012. From 2012 to 2020, use of CPU hours by these researchers increased in absolute and relative terms. CPU hour use per calendar year rose from 2,251,237 in 2012 to 32,083,356 in 2020. Since 2012 CPU use by medical researchers has constituted a significant fraction of the total. Medical School CPU averaged of 8% of total usage from 2012 to 2020, with a peak of 12% in 2016. Usage in 2020 was 9% of total. This usage is significant, given the CPU-hungry applications of the communities that traditionally use HPC systems at IU, including physics, astronomy, geology, atmospheric science, and chemistry. Demand for use of HPC systems for medical research can be bursty; Research Technologies has allowed reservations, for medical research groups, of up to 75% of a single system for two weeks, and up to 50% for a month. Some medical researchers would happily use an entire HPC system for a month when analyzing data. While we have not yet done this out of consideration for other researchers, we could in an emergency.

The number of researchers using IU HPC systems within the IU School of Medicine has grown substantially from 211 in 2012 to 728 in 2020. Still, only 6% of IU School of Medicine researchers use IU HPC resources. Those who do use HPC systems tend to be prodigious users: 2 of the top 10 individual users of CPU hours in 2020 were IU School of Medicine researchers. In contrast, 63% of researchers in the IU Bloomington Department of Physics use IU HPC systems, with

3 among the top 10 users of CPU hours for 2020. Our informal observation is that there is currently a generational change going on among IU's medical researchers that came earlier in other sciences. Medical researchers also seem more likely to "use" HPC systems by proxy. That is, faculty are likely to have another researcher in their group run jobs in support of the lab's research efforts, whereas in the physical sciences it is more common for faculty leaders of research groups to run jobs personally.

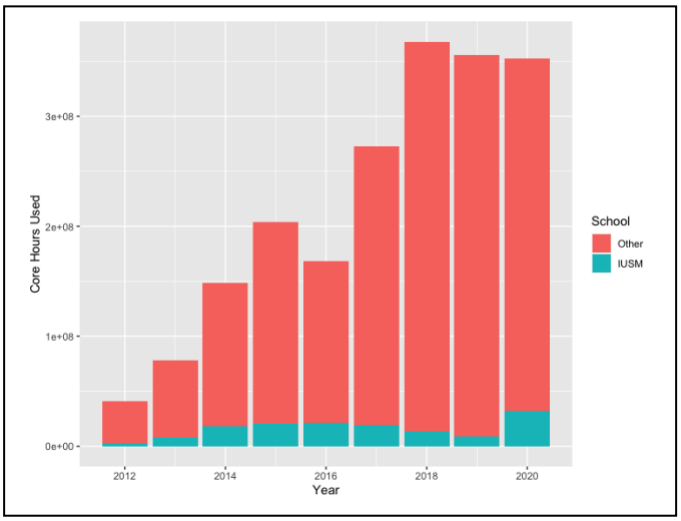

Fig. 1. Total HPC system CPU hour consumption per year, showing CPU hours used by research teams affiliated with the IU School of Medicine (in blue) and all other researchers at IU (in red).

V. Success Story Anecdotes

The general growth in use of IU HPC resources has been associated with a number of significant success stories in supporting medical research. We outline three below, representing a small sampling of now hundreds of interactions.

The Collaborative Initiative on Fetal Alcohol Spectrum Disorder (CIFASD). CIFASD is a collaborative research organization funded via a suite of grants from the US National Institutes of Health. The Research Technologies leadership was funded to operate the Informatics Core for CIFASD in 2003. Only well after receiving the grant award did we realize that operating the Informatics Core entailed storage of 3D images of the faces of child patients – PHI as defined by HIPAA and impossible to de-identify. As a result of HIPAA-aligned storage, IU was able to contribute significantly to the understanding of Fetal Alcohol Spectrum Disorder, supporting research leading to new diagnostic and therapeutic tools [7].

Indiana Alzheimer's Disease Research Center. When Research Technologies personnel began working with the Indiana Alzheimer's Disease Research Center several years ago, they were using a lab-based cluster to analyze patient data, taking two days to analyze a single patient. Migrating this workflow to Research Technologies HPC systems reduced processing to a few hours per patient. IU supercomputers have been now used to analyze genomes and brain images of more than 1,000 patients. Today the Indiana Alzheimer's Disease Research Center is one of just 32 national centers designated by the National Institute of Aging [8], and is one of the lead organizations within the US National Alzheimer's Disease Neuroimaging Initiative (ADNI) [9]. This latter role is due partly to the excellence of IU's researchers and partly to IU's ability to store the brain images, genetic sequences, and medical histories of hundreds of Alzheimer's patients.

A personal anecdote: an ill-behaved cancer. Co-author Stewart and his doctors were very surprised when he was diagnosed with Stage IV colorectal cancer in 2017. Shortly thereafter, Stewart underwent the standard treatment regime for such a situation, including colon and liver resection. Stewart had a recurrence of cancer in his liver in spring of 2018 leading to another resection. In 2019, Stewart's cancer was back again, so a sequence of his tumor genome was ordered. It was interesting for Stewart, as a patient and the person who made the final decision to have all Research Technologies HPC and storage systems aligned with HIPAA, to watch one of his doctors analyze his own genome interactively on an IU HPC system. Stewart was enrolled in immunotherapy and remains functionally healthy today, thanks to HIPAA alignment of IU supercomputers. Stewart is happy to be alive. More importantly, analyzing genomes on IU HPC systems is now a routine part of patient treatment by doctors of the IU School of Medicine and its Simon Cancer Center.

VI. Discussion

Indiana University has revolutionized its medical research and treatments this century thanks to the start provided by the INGEN grant award from the Lilly Endowment, Inc. The IU School of Medicine now includes a vibrant program in Precision Medicine [10]. The Research Technologies Division of University Information Technology Services has aided this and revolutionized the nature of its relationship with medical researchers in 2009 by making HIPAA alignment the minimum security standard for HPC and research storage systems it provides to the IU community. It is also of note that IU offers these services on a "first come, first served" basis in which there are no applications, no usage fees (for default storage quotas), and no obstacles to use of IU's research cyberinfrastructure for any researcher – medical researchers included. Central HPC systems are now budgeted at the university and CIO budget levels as a common good. Medical researchers are responsible for working within the School of Medicine to ensure that their workflows are consistent with local guidance for alignment with HIPAA, but this is a small obstacle given appropriate security of central HPC and research storage systems.

The results of IU's "HIPAA as default" policy in terms of medical researcher use of HPC resources has been dramatic. The two most important factors governing choices of whether or not to adopt technology are the perceived value that technology and its perceived ease of use [11]. We also know that a low perceived ease of use can cause potential adopters not to adopt new technology even if such technology can be of net benefit. By changing policies from "sure, use the system, just deidentify your data first" in 2007 to "default is HIPAA alignment, go ahead and store and analyze your PHI data as you have them in your workflows" in 2009, Research Technologies changed both perceptions and reality of ease of use of its HPC and research storage systems. Groundbreaking new research related to many diseases has been facilitated and accelerated as a result.

We have learned several lessons in supporting biomedical research involving PHI. Achieving FISMA Low compliance is a one-time heavy lift which involves picking applicable security controls from a list of 124 NIST 800-53 controls. At many institutions most of the required controls may already be in place, particularly technical and physical controls. For many HPC centers, most effort in implementing FISMA low will likely be involve instituting administrative controls such as governance and creating required documentation. One concept was and remains a learning experience for us: risk management does not mean risk elimination. HIPAA involves reasonable and appropriate risk response, not risk elimination. Risks can be accepted so long as documented justification is provided. For instance, the risk of a system not being behind the institutional firewall may be accepted given need for high speed data transfers and mitigating controls such as host firewalls, two-factor authentication, and encryption at rest. The experience gained from aligning HPC systems can also be applied to research systems needed to accommodate non-HPC clinical use cases such as databases and survey administration. It is possible to achieve HIPAA alignment or FIMSA Low compliance on HPC and storage systems without unduly burdening users not dealing with regulated data. Raising the security baseline can be done unobtrusively with careful risk management techniques.

We believe that the lessons we have learned are generally useful for other HPC centers, those within the US in particular. There is nothing unique about IU in terms of its capabilities. IU is simply farther along in supporting clinical research on its HPC systems than many other institutions. IU's experience shows that HIPAA alignment and/or FISMA Low compliance accelerates precision health research and enables research that would otherwise not be possible. FISMA compliance may also serve as a competitive advantage when pursuing grant funding. IU could meet current average level of demand by specifying only a portion of our systems for PHI, but not peaks in demand. A "FISMA Low everywhere" policy has several advantages over dedicating a portion of our systems for analysis of PHI: it allows researchers to choose which systems to use, it accommodates the sometimes bursty nature of medical research workflows, it provides a consistent level of security overall, and it keeps us prepared to dedicate all our HPC systems to medical research in an emergency. The COVID-19 pandemic has demonstrated how important it can be for major HPC centers to be prepared to support research related to major civil emergencies [12].

In order to accelerate precision medicine research, and to prepare to be in a position to aid in the case of future civil emergencies, we recommend that major HPC centers at US universities make FISMA Low their minimal basic security stance for all of (or at least most of) their HPC and research storage systems. This would provide a solid, well documented security baseline and enable storage and analysis of PHI in case of a civil emergency. So doing would be good for the research programs of individual universities as well as civil emergency preparedness in the US overall.

These recommendations regarding US laws and regulations apply of course only to HPC centers within the US, but the recommendation about security and ease of use for enabling analysis of protected health information is generalizable and important to HPC centers around the world. For example, the NIST RMF and NIST 800-53 are worldwide security standards and are potentially useful to HPC centers worldwide interested in enabling secure storage and analysis of regulated data.

VII. Conclusions

IU's policy of supporting research by clinical researchers analyzing clinical research data on all of its research HPC and storage systems has created a perception and a reality of good ease of use of these systems. The result at IU has been an acceleration of progress in biomedical research and new breakthroughs medical research that might otherwise not have been possible – or at least would not have been possible at IU. The approach of HIPAA alignment or FISMA Low compliance can be replicated at other US research institutions. We recommend that other institutions with significant HPC resources adopt FISMA Low compliance as a default, enabling them to aid research and enable their resources to be used in future civil health emergencies.

Acknowledgment

This work has been supported by the Lilly Endowment, Inc. and the IU Pervasive Technology Institute. Opinions expressed here represent those of the authors and may not reflect views of these sources of financial support. We thank colleagues involved in HIPAA alignment of IU systems, including Dr. William K. Barnett now of Harvard University. We thank anonymous reviewers and Dr. Terry Nickolas of the IU McKinney School of Law for comments on earlier versions of this manuscript.

References

[1] U.S. Department of Health & Human Services. 2013. Combined Regulation Text of All Rules. https://www.hhs.gov/hipaa/for-professionals/privacy/laws-regulations/combined-regulation-text/index.html.

[2] Cybersecurity & Infrastructure Security Agency. 2014. Federal Information Security Modernization Act. 2014. https://www.cisa.gov/federal-information-security-modernization-act.

[3] National Institute of Standards and Technology. Security and Privacy Controls for Information Systems and Organizations. NIST Special Publication 800-53 Revision 5.

[4] National Institute of Standards and Technology. 2008. An Introductory Resource Guide for Implementing the Health Insurance Portability and Accountability Act (HIPAA) Security Rule. NIST Special Publication 800-66 Revision 1. https://csrc.nist.gov/publications/detail/sp/800-66/rev-1/final

[5] UITS Research Technologies. 2021. Home Page. https://rt.iu.edu.

[6] Pervasive Technology Institute. 2021. Home page. https://pti.iu.edu.

[7] Collaborative Initiative on Fetal Alcohol Spectrum Disorder. 2021. https://cifasd.org.

[8] Indiana Alzheimer's Disease Research Center. 2021. Home page. https://medicine.iu.edu/research-centers/alzheimers.

[9] Alzheimer's Disease Neuroimaging Initiative. 2021. Home page. http://adni.loni.usc.edu.

[10] Precision Health. 2021. Home page. https://precisionhealth.iu.edu.

[11] Viswanath, V., M.G. Morris, G.B. Davis, and F.D. Davis, User Acceptance of Information Technology: Toward a Unified View. MIS Quarterly, 2003. 27(3): p. 425-478.

[12] COVID-19 HPC Consortium. 2021. Home Page. https://covid19-hpc-consortium.org.