
There are significant concerns for those working in medical archives, health sciences collections, and history of medicine collections regarding the provision of online access to medical records. Archival control is required to de-identify all records containing protected health information (PHI) as defined by the Health Information Portability and Accountability Act (HIPAA), but 2013 alterations to the privacy rule included a change in the definition of PHI that excludes information about individuals deceased for more than fifty years. This exploratory study investigates how archivists in medical and health collections approach the use of this newly available unrestricted health information (UHI) in digital environments. Interviewed participants address how they conceptualize UHI, view the practice of digitization, impose standards on their practice, and manage access to collections online. Findings suggest that archivists do not perceive significant differences between online access to PHI and UHI online, exerting archival control by de-identifying all online data.

Headings:

- Archives
- Archival Materials -- Digitization
- Medical Archives
- Privacy
PRIVACY IN THE PAST: A STUDY OF THE USE OF UNRESTRICTED HEALTH INFORMATION IN DIGITIZED MEDICAL ARCHIVAL COLLECTIONS.

by

Sarah E. Almond

A Master’s paper submitted to the faculty of the School of Information and Library Science of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Master of Science in Library Science.

Chapel Hill, North Carolina

April 2019

Approved by

__________________________________________
Denise Anthony
CONTENTS

INTRODUCTION ......................................................................................................................... 2
LITERATURE REVIEW .............................................................................................................. 5
METHODOLOGY ..................................................................................................................... 16
FINDINGS ................................................................................................................................ 22
DISCUSSION .......................................................................................................................... 41
CONCLUSIONS AND FURTHER RESEARCH ...................................................................... 49
BIBLIOGRAPHY ..................................................................................................................... 52
APPENDIX A ............................................................................................................................ 55
APPENDIX B ............................................................................................................................ 58
INTRODUCTION

In the practice of archivy, few questions loom larger than those related to digital objects. Born-digital items present their own challenges, but the preservation of and access to digitized materials are also problematic. The Society of American Archivists has endorsed the Well-intentioned Practice for Putting Digitized Collections of Unpublished Materials Online (W-iP), prepared by OCLC Research. This initiative, however, focuses primarily on access, with a nod towards potential copyright issues. For those working in medical archives, health sciences collections, and history of medicine collections, however, there are significant concerns that combine the question of access with others regarding privacy and the use of protected health information (PHI).

The introduction of the Health Information Portability and Accountability Act (HIPAA) law in 1996, and its subsequent implementation in 2003, immediately brought these issues to the forefront, and archivists in the medical field have responded in kind; much has been published regarding the archivist’s essential role as both an educator and a protector of privacy for countless records. But what of PHI not covered by HIPAA such as historical medical data and identifiers of the sort found in government archives, medical sciences libraries, and medical history libraries? In the case of some state archives, sunshine laws have guaranteed public access to the records of institutions such as mental asylums and special schools, whereas other states insist that these records be restricted in perpetuity. Medical sciences and medical history libraries must consider these issues from a grayer standpoint still, taking donor agreements and the like into
account as they process born-digital collections and decide which material holdings might benefit from digitization. Further, precedent set by the social history movement years earlier largely recommends the anonymization of health records one or two generations removed from HIPAA constraints, suggesting that archivists take not just patient lives into account, but those of their extended families.¹

Pressure is being applied from other angles as well. In an impassioned article published in the fall 2014 issue of RBM: A Journal of Rare Books, Manuscripts, and Cultural Heritage, Jeffrey S. Resnick, Chief of the History of Medicine Division of the U.S. National Library of Medicine, implored historical medical libraries to “be relevant and valuable in the digital age,” as “unique stewards of the past whose collections can speak directly to the temporal human condition.”² He goes on to argue that “[t]he expanding digital world and commensurate cultural expectations are driving processes of learning about the past increasingly through virtual surrogates of physical material.”³ Similarly, the growth of Digital Humanities (DH) practice—with the study of social history in particular—has led to new and expanded approaches to large collections of data that may be found in the archives. As digital objects and digitization become more and more commonplace, precedents set in earlier decades become farther and farther removed from the reality of the present. In the age of DH and online access—all things that archivists are being encouraged to promote—what are effective guidelines or best practices to use when making sensitive, yet unrestricted health information available in

---

³ *Ibid*, 112.
digital formats? If such guidelines do not currently exist, how can we as archivists go about creating them?

Due, perhaps, to the rapidly changing landscape of digitized and born-digital materials, DH practice, and online data preservation and safety, there is very little material published on this subject save Gustainis and Letocha’s “The Practice of Privacy,” which focuses primarily on the management of PHI in analog collections. Using the methodology of an exploratory study, the intent of this work is to investigate archival ethics as they apply to the access and use of unrestricted health information (UHI), defined for the purposes of this study as medical records featuring any of HIPAA’s eighteen identifiers for individuals who have been deceased for fifty years or longer, and are thus non-covered by HIPAA regulations at the federal level. More specifically, this is an investigation of how archivists and librarians in medical libraries, health science collections, and history of medicine collections approach the use of UHI in digital environments. Do archivists perceive an ethical difference between providing access to traditional archival collections containing UHI and those that are digitized or available online? What, if any, criteria do archivists consider when deciding what materials to make available digitally? And finally, when digitizing records containing UHI, do archivists have processes in place to limit or restrict online access and establish control or context? If so, where and how are these processes implemented?

---

4 The eighteen HIPAA identifiers are listed in Appendix A.
5 Born-digital acquisitions, due to their contemporaneous nature, most often contain restricted PHI, rather than UHI, which goes beyond the limited scope of this study.
LITERATURE REVIEW

There are several aspects of this inquiry to take into account. With that in mind, this literature review consists of four separate points of view pertinent to the questions at hand: precedent and current context for decision making in the medical archives, issues and perceptions of access and potential audiences, specific challenges of DH projects, and finally, the question of research and archival ethics within a digital context.

PRECEDENT AND CURRENT CONTEXT

The most widespread use of UHI occurred during the social history movement of the late 1970s, and the practices established during that period are still widely used by historians and archivists alike. John Harley Warner explores many of the questions historians and researchers in particular have brought to their research in his 1999 article “The Uses of Patient Records by Historians: Patterns, Possibilities, and Perplexities,” pointing out that there is an “absence of uniform policies” regarding medical records in the United States, often leaving “the management of confidentiality up to the historian.”

David Wright and Renee Saucier point out that protecting privacy has been paramount for most medical historians since the 1970s, and assert that the decision to anonymize historical patient records has been widespread. The authors also point out that social historians have used keys and other “neutral” identifiers to verify the veracity of their

---

research, whereas some scholars have chosen to protect the identities of their subjects completely, choosing to use pseudonyms exclusively or purposefully avoiding footnoting relevant patient information with identifying characteristics.\(^7\)

For their research, Angela McCarthy and her fellow authors decided to use “lightly” anonymized public records in a database of 19th century mental institutions in Australia and New Zealand. They argued that “the use of historical records involves historians navigating questions of ethics,” and that “although...many names of patients are in the public domain...this article refers only to initials of surnames...and our references provide a clear trail to identity.”\(^8\) Still other scholars are arguing that the anonymization of public health records obscures the history of chronically underserved populations and robs these populations of the chance to reclaim their stories.\(^9\)

In his article “Privacy in a Public Place,” Paul J. Sillitoe examines the question of privacy from the perspective of the archivist, arguing that “without a clear definition of privacy limits, archivists are unable to fully consider when an invasion of personal privacy might be occasioned.”\(^10\) Sillitoe’s work is especially useful in its exploration of the meaning of “sensitive” information, and the continuing difficulty of defining the term.

The most thorough assessment of the current landscape as pertains to archives, however, is Gustainis and Letocha’s “The Practice of Privacy,” which draws on the authors’ experience managing CLIR-funded processing initiatives of significant medical

---

\(^7\) Wright and Saucier, “Madness in the Archives,” 70-72.

\(^8\) Angela McCarthy, Catharine Coleborne, Maree O’Connor, and Elspeth Knewstubb, “Lives in the Asylum Record, 1864 to 1910: Utilising Large Data Collection for Histories of Psychiatry and Mental Health.” *Medical History; London* 61, no. 3 (July 2017), 368.


collections containing PHI. The authors point out the difficulty encountered by researchers and archivists alike in understanding the landscape of restricted PHI, noting that “variations in whether and how repositories are covered by HIPAA and differences in state laws result in much confusion.” They offer up best practices developed in the administration of the “Private Practices, Public Health” grant that are designed to educate researchers about the variety of legal restrictions and promote increased access to archival materials for the purpose of research. While the study is thorough and quite informative, the authors focused on creating applicable standards for helping researchers navigate the complexity of PHI restrictions and statutes; they do not address potential ethical complications of access per se, digital or otherwise.

ACCESS

Perhaps unsurprisingly, the most vocal argument for the widespread digitization of materials is that of access. Reznick sees digital access as an act of “stewardship” on the part of medical records repositories, wherein such institutions may broaden their areas of influence and increase their value for research purposes. Several archivists, however, point out that medical libraries and archives are not necessarily intended for the general public. In general, there appears to be an assumption amongst archivists as well as medical historians that the term access refers to a very specific audience—that of the

---

12 Reznick, “Embracing the Future as Stewards of the Past,” 112.
researcher. Gustainis and Letocha advocate for “more robust descriptive information” in order to attract researchers.\textsuperscript{14} Susan C. Lawrence refers to the reluctance of libraries to collect or acquisition potentially sensitive medical information as “access anxiety” wherein archivists are ultimately more concerned with HIPAA compliance than making records available online; it is inferred that online accessibility should be a primary concern.\textsuperscript{15} Studies show that historians are becoming increasingly comfortable with online repositories, and are beginning to view them as trusted resources on par with physical archives. Further, the ease of access that digital collections provide enables researchers to cite sources in more transparent and easily referenced ways.\textsuperscript{16} Wright and Saucier predict, however, that the increasing influence of DH projects, many of which offer previously unthinkable levels of data analyzation and visualization, will “provide hitherto untested waters in terms of access” that will “renew the debate over anonymization and confidentiality in the years to come,” once again bringing ethical considerations of privacy to the fore.\textsuperscript{17}

One such project, the Central State Hospital Project, acts as a fascinating case study for Wright and Saucier’s prognostication. The project employs tiered access to a digital archive based on familial and researcher relationships to the material. This anticipates both the capability of digital archives to negotiate access as well as a view of

\textsuperscript{16} Donghee Sinn and Nicholas Soares, “Historians’ Use of Digital Archival Collections: The Web, Historical Scholarship, and Archival Research,” \textit{Journal of the Association for Information Science and Technology} 65, no. 9 (September 2014), 1803.
\textsuperscript{17} Wright and Saucier, “Madness in the Archives,” 82.
access based in what Michelle Caswell and Marika Cifor refer to as “an ethics of care.”

In this unique approach to archival practice, Caswell and Cifor encourage archivists to embrace a “radical empathy” that acknowledges and privileges the bond between archivist and record creator. By digitizing and providing access to UHI in the form of the patient records of a segregated insane asylum, participants in the Central State Hospital Project acknowledge that, while there are not yet “best practices” to inform questions of digital access, an empathetic and thoughtful reimagination of collection access would serve the needs of stakeholders and archivists alike.

“Discussions among archivists in the past ten to fifteen years about community archives have made it clear that communities are very much interested in being treated fairly with reference to records kept by others about them and held by archives,” project participant Patricia Galloway asserts, “while historians and medical researchers have different concerns.” Lorrie Dong, in her article about the project, identifies stakeholders as “patients’ relatives, the current and former staff members at the hospital, the state archives, and the state’s Department of Behavioral Health and Developmental Services,” and notes that “records will be contextualized in the digital archives,” promising “a level of privacy protection” for the records “that goes beyond existing state and federal laws.”

---

21 Lorrie Dong, “Exploring the Reaches of Privacy and Technology: Central State Hospital Digital Archives Project,” Archival Outlook, August 2016, 10.
Dong identifies the technology behind this promise as the development of “a set of digital tools that would help archives make collections with challenging privacy issues more readily discoverable and potentially available for use,” describing “custom digital tools…able to search and redact information according to current legal restrictions and the identity of the information seeker.”22 This kind of tiered access in digital archives is new technology indeed, but has already been implemented in the Murkutu content management system (CMS), developed by the Center for Digital Scholarship and Curation at Washington State University for use by Native American communities. Claiming that the “first priority” of the software is “to help build a platform that fosters relationships of respect and trust,” the Murkutu project enables content managers to determine “cultural protocols” that dictate “levels of access to…digital heritage materials based on…community needs and values.”23

Tiered access to digital collections is not only a response to theories of radical empathy in the archives and digitization as remediation. It also enables the archivist to exercise archival control and illustrate context in an environment previously considered a free for all. These technologies are still new; nevertheless, they may have serious impact for concepts of digital access in the future.

**DIGITAL HUMANITIES**

Unsurprisingly, the DH field is still new enough that a clear definition of the term is hard to come by. Östman and Turtiainen succinctly (but rather dryly) describe DH as

---

22 Ibid., 10.
“humanistic research appropriating…technological methods.”

Claire Brennan, quoting Burdick, Drucker, Lurenfeld, Presner, and Schnapp’s 2012 Digital_Humanities, describes DH as “a production-based endeavor in which theoretical issues get tested in the design of implementations.” Methods and implementations aside, the practice of DH has resulted in the digitization of thousands of records, and has given rise to fascinating online projects utilizing collections to enrich experiential learning and archival engagement. Digitized texts are a boon to access, as they afford the opportunity for remote work as well as analytical tools not previously available to analyze physical records, and digital data visualizations, databases, and text mining reveal previously hidden patterns in large amounts of historical data.

DH practice has had an elevating effect on the archivist’s role, as well, wherein archivists may see themselves as more than an educator or facilitator, but rather a full partner in the creation of specific projects. Thus, the ethical responsibilities of the archivist and researcher are entwined in unprecedented ways. Brennan argues that the rise of digital history has led to historians “creating their own virtual archives” that “have severely restricted access conditions” as well as “new, online, public collections;

26 One such (fantastic) example is the NYPL Green Book project: https://publicdomain.nypl.org/greenbook-map/. For an example specifically related to UHI, visit The William Osler Profiles in Science Collection at https://profiles.nlm.nih.gov/ps/retrieve/Collection/CID/GF, which provides access to portions of the William Osler collections of the Osler Library of the History of Medicine and Alan Mason Chesney Medical Archives.
27 Lisa M. Given and Rebekah Willson, “Information Technology and the Humanities Scholar: Documenting Digital Research Practices,” *Journal of the Association for Information Science and Technology* 69, no. 6 (June 1, 2018), 808; McCarthy et al., “Lives in the Asylum Record,” 370-75.
regardless, “new access to archives made digital mean an awareness of archival practice is at the heart of digital history.” Still other scholars see the technology at work in DH as an impetus to re-examine traditional research practices, pointing out that such technology can challenge assumptions and encourage researchers to “dig beneath the surface” of received knowledge and scholarly habits. Part of this challenge involves “relinquishing control...with the possibility that [data] might be misinterpreted, recontextualized inappropriately, or even misrepresented altogether.” This is a marked difference from traditional archival practices of control, most notably the origin of the finding aid. Beyond control, however, issues of context loom large in DH, with several authors pointing out that the lack of archival context available in online collections leads to exactly the results listed above. Dix et al. point out that “a traditional scholar would need to travel to the physical collection in order to consult an archive. Its location...would itself tell something of the origins and reliability of the material...The index might be clearly printed on 3x5” cards...or be non-existent, leaving the researcher to leaf through unsorted papers without guidance...The physical form, location, and disposition of the artefact tell as much as the word written on it.” Notions of how to replicate that context in a digital setting would go far to provide an ethical environment wherein patient information and privacy might be better understood by the intended audience. But how to provide it? Yup.

Dix et al. claim that DH projects “should be more perspicuous,” but fail to consider the ethical framework behind the technology that supports such projects. Digital technology is still relatively new when compared against traditional methods of scholarly publishing and archival arrangement and, as James H. Moor points out, “revolutionary technology generates many ethical problems” before encouraging readers to “anticipate [technological] consequences and establish policies that will minimize [its] deleterious effects.”

If, as Sillitoe points out, “archives and records services are about to be caught in new legislations for which they were not the primary target,” it may do archivists well to strongly consider the deleterious effects of digitizing historical medical data and making it freely available.

The most elemental of ethical concerns includes protecting people from harm. Jeroen van den Hoven describes the negative effects of information scraped, culled, hacked, or otherwise provided by virtue of being online as “information-based harm” and reasons that “protecting identifying information...diminishes epistemic freedom...but also diminishes the likelihood that some will come to harm.” He then goes on to identify specific instances of harm based on personal information, such as reputational harm.

34 Sillitoe, “Privacy in a Public Place,” 13.
leading to a diminished career, and harm inflicted by “classifying people in such a way that their chances of getting some good are diminished.”

In the case of historical medical data, the personal information being shared is not that of a living individual, but rather a relative, once again calling to mind Wright and Saucier’s assertion that current ethical attitudes mandate that medical archivists must take the lives of patients’ extended families into account. How to establish the potential for harm to a deceased patient, or their distant, living relative? What’s more, how to apply conceptions of consent to historical health records made available online? Dean Seeman, in a fascinating exploration of ethics in the attribution of names in metadata describing digital objects online, offers that, in cases where potential for harm or consent are not easily established, context acts as a key consideration. Further to contextual consideration, Seeman recommends that archivists consider questions such as a subject’s reasonable expectation of privacy, the creation of a record in a private or public setting, the relative vulnerability of the subject, and the age of the artifact in question. In conclusion, though, he asserts that “a balance must be struck between the rights of the individual and the obligation to society and the community.”

Östman and Turtiainen conclude their study of DH ethics by suggesting that practitioners shift focus from research ethics to researching ethics [emphasis added]. Where is the emphasis?

---

35 Jeroen van den Hoven, “Information Technology, Privacy, and the Protection of Personal Data” in Information Technology and Moral Philosophy, ed. Jeroen van den Hoven and John Weckert (Cambridge; New York: Cambridge University Press, 2008), 311-12; Carole L. Jurkiewicz asserts that “the majority of information that people generate through the use of technology is accessible and available, or hackable, by anyone with access” in “Big Data, Big Concerns: Ethics in the Digital Age.” Public Integrity 20 (January 2, 2018), S48.
37 Östman and Turtiainen, “From Research Ethics to Researching Ethics in an Online Specific Context,” 73.
In an example specific to medical data, van den Hoven succinctly points out the double-edged sword of PHI:

“Many people do not object to the use of their personal medical data for medical purposes...whether these are directly related to their own personal health affairs, to those of their family, perhaps even to their community...as long as they can be absolutely certain that the only use that is made of it is medical...They do object, however, to their medical data being used to classify them or disadvantage them socioeconomically, to discriminate against them in the workplace..., deny them social benefits, or turn them down for mortgages or political office.”

In the case of public records used to benefit humanities research, the risks are the same. When hereditary health issues are taken into account, and also taken out of context—placed online in an uncontrolled atmosphere where the data can be easily misinterpreted or used to classify or disadvantage living relatives—it can be difficult to find the balance between individual rights and obligation to society or community. At the same time, however, while acknowledging that medical historians “have engaged with the question of patient confidentiality on several levels,” Wright and Saucier point out that “despite the anxieties” of earlier social historians in the 1980s, “there has been no predominant scandal that has altered dramatically the landscape of access or research...over the last quarter of a century.”

Though the authors suggest that the risks of exposing sensitive patient information may not be as significant as many fear, technological developments—especially with regards to online publication—over the past twenty-five years must also be taken into account.

---

METHODOLOGY

Given the paucity of contemporary literature examining the decisions that archivists must face in the digitization of potentially identifying health information records, it was determined that a qualitative exploratory study using semi-structured interviews would be the best approach for this research, so as to establish a better understanding of the challenges, workflows, and ethical infrastructure behind the practice. An interview guide (appendix B) was prepared, with questions focused on individual archivists’ perceptions of ethical difference between physical and online collection access, criteria used to identify a collection for digitization, and levels of collection limitation or restriction that may be developed in the digital environment. Using homogenous purposeful sampling techniques, archivists and digital resource managers at medical archives, health sciences collections, and history of medicine collections across the nation were contacted via email in order to establish times for telephone or video-conference interviews, which were recorded and transcribed. Once all interviews were completed, the transcriptions were analyzed and the resulting observational data was inductively coded.

Definitions of what constitutes a medical record vary from state to state, as do laws allowing for the release of UHI. Medical archives, health sciences collections, and history of medicine collections may be covered or non-covered entities under HIPAA requirements. The variables present in this study, combined with the paucity of research on the subject, do not lend themselves to a quantitative approach. Also important to the
nature of this exploratory study is the ability of the archivists in these repositories to express the reasoning behind their decisions in their own words. Qualitative research is well-suited to providing a strong understanding of context, and, for the purposes of this study, context is essential to any attempt to create replicable workflows or guidelines.

RECRUITMENT

Due to the narrow focus of this study, a homogenous purposeful sampling approach was used to gather data from archivists and digital resource managers employed in medical archives, health sciences collections, and history of medicine collections. The specificity and potential variables of the study, discussed in more detail above, required a purposeful, rather than a probabilistic approach.

To that end, seventy-five potential institutions were identified from twenty-seven states and the District of Columbia. These institutions were selected based on information gathered from the Harvey Cushing/John Hay Whitney Medical Library resources at Yale University, as well as the National Library of Medicine’s online directory of medical history collections. In order to maintain the integrity of the sample, preference was not given to institutions based on their identity as covered or non-covered entities under HIPAA. After visiting the websites of each institution, contact information was gathered in the form of a direct email address, with a preference towards staff with a title of archivist, curator, or digital resource manager. In the event that specific staff email addresses were not available, contact information was established as the institution’s

---

public email account. If there was uncertainty about whom to contact, more than one employee’s email address was collected, resulting in seventy-nine emails being sent to seventy-five institutions.

The email (Appendix C) addressed the intent of the study and requested a brief interview with the addressee, provided they met specific requirements of being an archivist or librarian with over three years of experience that has implemented or planned to implement a digitization project using UHI within the last five years. Snowball sampling was also employed wherein addressees were asked to recommend a colleague that might be able to speak to the topic in the event that the addressee did not fit the criteria described above.

Of the seventy-nine emails sent, eleven responses were received, resulting in seven interviews. Three of the respondents did not fit the criteria; one other was unable to be interviewed due to the time frame of the study. The relatively small size of the study sample discourages generalizations regarding the homogenous group of interest, but still offers significant insight into the specific processing decisions made by archivists when determining what to make available digitally, and what barriers to access they might put in place. Further, additional research into this subject would most likely result in additional interview requests, potentially increasing the response rate.

The participants interviewed for this study share a mean average of ten years’ experience at their current positions, and represent institutions at the federal level, as well as public and private universities. Six of the interviewees hail from entities not covered by HIPAA; one participant was from a HIPPA covered entity. All fit the selected criteria of being an archivist or librarian with over three years of experience that has
implemented or planned to implement a digitization project using UHI within the last five years. What follows is a brief description of each participant, identified and numbered consecutively in the chronological order of the interviews.

The first interviewee (participant 1) is an archivist and manager of digital resources for a health sciences library affiliated with a public university and school of medicine. The second (participant 2), holds the title of archivist/curator in an institution with the same characteristics. Participant 3 is an archivist in a private library and museum with a focus on the history of medicine. Participant 4 is a representative of the only HIPAA covered entity in the study, and manages collections in medical archives affiliated with a private university and medical school. The fifth and sixth interviewees (participants 5 and 6, respectively) are each archivists representing educational institutions with schools of medicine, the former a public university and the latter a private one. Participant 7 is also an archivist and manager of digital resources, but at a federal, non-covered institution focused on the history of medicine. Combined, the participants offer a comprehensive geographical survey of the United States, and an even division between public and private institutions of higher learning.

THE INTERVIEW PROCESS

Prior to conducting the interviews, permission to record was requested and, once granted, the proper recording format was decided upon. Due to geographic and schedule constraints, all of the interviews were conducted remotely, via videoconferencing software or telephone.

41 The following interview options were offered: videoconferencing via Zoom or Google Hangouts, or recorded phone conversations using the Rev application.
The interviews were semi-structured, and loosely followed the course of an IRB-approved interview guide (appendix B). This guide focused on the identification of individual collection users, limitations or restrictions placed on either digital or physical collections available to the public, and inquiry into decision making workflows with regards to digitizing collections. Further questions asked archivists to clarify “sensitive” information beyond the eighteen HIPAA identifiers and addressed the availability of UHI, defined as health records of individuals who have been deceased for fifty years or more, and are thus non-covered by HIPAA regulations. In order to provide context, the researcher offered her personal experience with the digitization of North Carolina’s public mental health records as an example. Upon discovering the differences in state legislation as pertains to the availability of certain medical records, the researcher also asked all participants to outline what medical records were considered publicly available in their archive based on state laws.

DATA ANALYSIS

Each discussion was transcribed and analyzed following completion of the interview. Once all interviews were transcribed, observational data from the transcriptions were analyzed qualitatively. Data were then organized thematically, using inductive coding techniques, into the following classifications: conceptions of unrestricted health information, digitization as practice, imposition of standards, and managing access.
LIMITATIONS

As mentioned above, the sample size of the study, even with the application of homogenous purposeful sampling, is too small to act as a representative for the community of practice as a whole. Similarly, there is the potential for bias, given that one researcher both coded and conducted interviews with participants. At the same time, however, the following documented strategies were employed to more fully support the validity of the data collected.

In an effort to recognize the iterative nature of an exploratory study, semi-structured interviews were employed at the same time that data were concurrently collected and analyzed. Collected data were triangulated between interviews and extant literature and, further, given the small size of the sample, every effort was made to provide a broad swath of experience, from lone arrangers in medical history collections to a large archival staff handling one thousand or more unique requests per year. The goal of this study is to provide a general framework of archival decisions and attitudes pertaining to use of UHI in a digital environment, a framework that may ultimately spark conversation and further the goal of establishing effective guidelines and best practices for the field.
FINDINGS

CONCEPTIONS OF UNRESTRICTED HEALTH INFORMATION

For the purposes of this paper, UHI is defined as medical records featuring any of HIPAA’s eighteen identifiers covering individuals who have been deceased for fifty years or longer and are thus non-covered by HIPAA regulations. One of the first things that became clear in the interview process, however, is that several factors influence the definition of UHI and how it is perceived and utilized by archivists. Six non-federal entities are represented in the study, covering five states, all with their own state-level definitions of a medical record, and different record restrictions. In one state, for example, all medical records are restricted in perpetuity, whereas another has restrictions applicable to the records of state-run psychiatric institutions only: these are, in the words of one archivist, “basically restricted forever” (participant 6).

Such variance by state creates significant complications in archivists’ approach to UHI in general, particularly with respect to digital collections. Though Susan C. Lawrence originated the term “access anxiety” to describe the reluctance of archivists to collect or acquisition potentially sensitive medical records due to HIPAA compliance concerns, the term can be extrapolated to apply to a reluctance to engage with or digitize historical medical records based on similar apprehensions and uncertainties. For those working in repositories located in states where medical information is restricted in perpetuity, UHI, in effect, does not exist at all.
For some, uncertainties can be as simple as determining whether all persons represented in a medical record died fifty years prior. Having spent the past ten years working on the digitization and transcription of handwritten nineteenth and early twentieth century hospital records into a searchable database, one archivist, the only employee working on the project in her department, mused, “the people must not still be living…but there is no way we could embark on the mission to ascertain whether they are dead or not,” citing a dearth of both time and resources. The archivist, who, after consulting legal counsel, has made plans to restrict HIPAA PHI identifiers such as name, address, and dates in the database before publishing it online, expressed concern that “if we hide the address, we are losing so much that you could use for tracking the localizations of illnesses,” noting that the ledgers contain records chronicling an epidemic of la grippe of interest to researchers.

Other archives have adopted policies that ease the interpretation of what is and is not UHI. An interviewee at a non-covered entity pointed out that her institution has adopted their “own restriction policy that is loosely HIPAA-based” that “restrict[s] health information for a period of eighty years from the date of record creation.” Explaining that the restrictions exist for “pragmatic reasons,” the archivist noted that the policy is applied “consistently” and that any material “beyond the eighty year restriction is totally open access” (participant 6). An archivist at another, covered institution admitted that they are “quite content to stay clear of [HIPAA] stuff and not think about it unless [they] have to,” but also conceded that old hospital records from the mid-nineteenth century housed at their institution are “fine” and that they “certainly allow access and use of that material” (participant 2).
Precedent and issues of sensitivity, however, also inform archivists’ decision making process with regards to the use of UHI and how the dissemination of such information might affect the privacy of present generations. When asked if they considered some historical health records to be more sensitive than others, all of those interviewed replied in the affirmative. “Of course there are materials that are more sensitive than others,” one archivist pointed out, describing “inconsequential” yet restricted historical health records pertaining to basic medical complaints. “It’s not the same thing as records related to a family’s history of sexually transmitted diseases or psychiatric conditions where there could be a social stigma attached to certain kinds of health information” (participant 4). Discussing newly digitized UHI records that are in the pipeline for online exhibition, another archivist explained that, though “those folks are long gone,” the archivist has still taken care to confirm that none of the patients could be “comprehensively identified” by digitized data provided online. “Say they had a genetic disorder of some sort that was identified,” the archivist posited, “an insurance company could discriminate against [future generations] if they realized that they had a predetermined genetic issue” (participant 3).

Yet another interviewee, a digital archivist at a covered entity, working closely with collections focused on the history of AIDS and HIV, revealed their department’s approach to sensitive material, access anxiety, and vague interpretations of UHI. “Anytime we as archivists encounter protected health information in any of the collections that we’re processing, either physical or digital, we would separate it into a restricted part of the collection, and note that.” When asked if the restrictions applied across the board, they further explained that “if someone wanted to look at restricted
material, especially patient health material, we would first review that material to try and decide whether or not it was covered under HIPAA because of time periods. And if for some reason we did not believe that it was covered…we would consider giving access to them” [emphasis added] (participant 5).

Though their non-covered institution applies blanket, HIPAA-inspired restrictions on PHI that falls within a period of one hundred years from the date of the record or death of the individual, the last archivist interviewed suggested alternative practices that might better protect the live descendants of patients whose UHI is available digitally. Their suggestions take the potential for sensitive UHI into account, managing different risk factors with corresponding time boundaries. Noting that genetic and psychiatric data “might have larger impacts on [a patient’s] children or grandchildren,” they proposed “a longer [restriction] period for those kinds of data” as “an option to think about.” Noting that his team “haven’t pitched” their ideas as “formal policy,” they concluded that “I think one hundred years is too long. That seems excessive…but maybe I’m wrong. It may all change, because with genetic data maybe it’s different. Maybe that should never, ever be public, you know?” (participant 7).

Though each archivist’s approach to and conception of UHI was different, all were in agreement that, with regards to digitization projects potentially featuring UHI or PHI, it’s best to “err on the side of caution” (participant 3) in terms of compliance and to “take all…things into consideration” (participant 2), including the privacy of living relatives. One archivist’s observations, specifically addressing UHI in a DH project featuring historical asylum records, suggest that such urge to caution is about more than a fear of causing offense or simply violating HIPAA protocol. “[I]t’s important to engage
the community that was impacted by the practice of institutionalization,” they said.

“There are trust issues in terms of custodians of the records and who are the stakeholders…whose families are represented in those records…[E]ngaging the…community is important…so that they’re not feeling like they’re being re-exploited again, that their family members are being treated with respect.” “Not everything needs to be online,” they concluded, “you can digitize collections and there’s more that you could do with that in terms of data analysis and all, but when you’re putting this kind of information online without a thoughtful process and engagement with the stakeholders, I think there’s all sorts of unintended consequences” (participant 4).

**DIGITIZATION AS PRACTICE**

Perhaps “not everything needs to be online,” but what are the guidelines that inform the decisions of archivists in medical archives, health sciences collections, and history of medicine collections when selecting collections for digitization? Researcher interest, available funding, and access are the factors most readily identified as a result of this study, but a few of the archivists interviewed also expressed that the decision to digitize certain collections has been based in preservation rather than access, resulting in digital collections that will not be made available online. The question of how or whether to actually publish digital collections online at all also proved a point of interest for the participants in this study, with six of the seven interviewees agreeing that unrestricted online access differs from physical access to collections in key ways.

For several of those interviewed, online access was perceived as the equivalent of lost archival control, whereas the mediation, decorum, and ritual of the reading room
provided context and a chance to clearly communicate collection use restrictions. The vast, undefined audience for online collections is a source of concern, particularly for archivists at non-covered entities. “Once [an unrestricted collection] goes online, you lose control,” one archivist stated. “It just opens up a whole…barrel of worms instead of a can of worms” (participant 3). “We are more sensitive about putting things online,” another said, before noting that their few published digital collections do not usually contain any sort of identifying information, historical or otherwise, and are aimed at “a general audience, say high school, college audience” rather than researchers (participant 7).

Those at covered entities had similar reactions. “The difference is probably meaningless,” one interviewee offered, “But I guess I just feel that there’s always more control in the reading room…allowing [researchers] to download materials from an online digital collection …you just feel that there’s a loss of intellectual control there” (participant 2). At another HIPAA covered institution, the archivist pointed out that data presented online “is available just by searching. It is highly retrievable and sortable and indexable” before expressing that digitized online data is “less mediated in most cases…if someone comes into the reading room…, they might sign something, we might be having a face to face conversation and that might dictate how they meet these materials” (participant 5).

For another archivist at a non-covered institution, the differences between online and physical access also include processing workflows. “We use a more flexible, efficient processing approach which results in us not reviewing every single record in a collection. We already know there may be things in there that we missed. The fact that someone can only access it in the reading room lowers the overall risk of wide exposure of something
sensitive.” They went on to point out that archives staff requests that visitors to the reading room “sign a sensitive materials agreement that explains this approach,” noting that “we don’t have those same measures in the online environment” (participant 6).

In a dissenting opinion, one archivist at a covered institution argued that adherence to HIPAA standards of restriction resulted in an equivalence between online and physical collections. It bears pointing out that their particular institution differs from others in the study in that the laws of their state restrict all PHI, regardless of the age of the records, rendering redactions and restrictions a necessary and essential part of all processing and digitization workflows. “We still look at materials on a case-by-case basis, and then review them with respect to the definition of protected health information under HIPAA and then what is protected under [state] law. That’s our standard. We’re not basing it on some kind of personal, ethical standards; we go by what the law allows” (participant 4).

When asked how their staff go about determining what collections to digitize, the archivist offered up a variety of criteria, mentioning that physically delicate or damaged collections may be digitized as a result of their condition (participant 4). Heavily requested materials are also good candidates for digitization, and the repository offers digitization on demand so that researchers may request items that interest them. “In terms of mass digitization,” though, this archivist suggests different considerations. “Something that’s going to have a lot of privacy concerns—that needs to be heavily redacted—is going to be more costly and labor intensive to digitize,” they point out, before asking “How many people are going to use this particular collection…and who’s paying for it?
How wide is the research community that would benefit from this particular digitization project?” (participant 4).

Other participants echoed these criteria. One described a collection of roughly two hundred linear feet containing UHI that “has to be accessed digitally” because of red rot and dormant mold (participant 3). Yet another reflected that user interest was the first of several facets taken into account when considering digitization: “If it’s a high use collection we know it’s a high interest topic.” Institutional priorities were also a contributing factor, with the archivist offering examples such as upcoming exhibits or “an event coming up that we want to support with digital materials” (participant 6).

The presence or absence of funding however, was identified by most archivists as a key component in digitization efforts and decision making. In most cases, archivists described projects that follow the money that is allotted to them; the presence of UHI or PHI was not a significant consideration in these decisions. Archivists also pointed out that, in such situations, it is not unusual for digitization workflows to be influenced by donor interest and, in some cases, collections are digitized that would have been too labor intensive to do so otherwise. Describing the digitization of a “large contemporary faculty papers collection” with multiple restrictions, one interviewee expressed doubt that her institution would have embarked on the project without the influence of outside interest and funds. The donor, whom the archivist described as a “partner-funding institution,” encouraged experimental workflows that they admit “wouldn’t be a very strategic way for us to go about digitizing” were there more “limited resources” (participant 6).

The archivists at covered institutions, faced with the heavier workload of complying with HIPAA restrictions, were a bit more frank. When asked about
digitization criteria, one interviewee replied, “Honestly, one of the first things is what is going to get funded because we don’t have any money to digitize those things…it breaks my heart a little bit to admit that” (participant 5). Another archivist explained that donors want digitization projects “to be something sexy” and, reflecting on their most recent large scale digitization project, related that they didn’t “think there was any reason to digitize [that collection]…or to digitize as [much] as we did” (participant 2). Though an archivist (participant 5) did go on to mention other criteria for digitization, such as collections with high research value, “compelling” collections, and “what we are able to provide in terms of access”), their covered counterpart lamented that, though they “have a queue of materials we would like to digitize that are smaller collections…much more interesting, [and] probably have much more research value as well…there’s just no money there to do those yet.” They concluded that “we would rather that our projects are generated from the ground up as opposed to the top down” (participant 2).

The only clear example of a “ground up” digitization project discussed in this study, however, is proceeding very slowly, which the archivist blames on a lack of funding. Describing the hospital records digitization and transcription project that they have been working on for the past decade, this archivist related that “we tried to apply for a grant…but it turned out that, for most of the funds that we could apply to, we weren’t eligible because we didn’t have the gift of deed from the hospital.” They went on to describe how, after securing a letter of stewardship for the ledgers, the project was still ineligible for most digitization grants. Though the bulk of the database is transcribed and the archive is at present in the process of cleaning the data, the archivist expressed a desire for funding that could provide “a person…devoted to the project,” explaining that,
“in my case, there are always so many more urgent things that [the project] is always pushed aside” (participant 1).

**IMPOSITION OF STANDARDS**

As a panacea to the indiscriminate murkiness of determining UHI, as well as a guide for non-covered entities, several of the archivists interviewed recommended and supported the uniform application and extension of HIPAA and HIPAA-influenced standards to all potentially identifying medical records, be they historical or contemporary. Moreover, having established in the previous sections that all were in agreement that certain forms of UHI—such as data pertaining to psychiatric health, sexually transmitted diseases, and genetic material—are more sensitive than others, it is should be noted that, even amongst those with dissenting opinions, none of the archivists questioned advocated for unmitigated online access to UHI. “Somebody saying, ‘Oh, this information needs to be free’…well, would they say the same thing about their grandmother?” one archivist at a covered institution offered. Discussing a recent genetic testing initiative to analyze the saliva on the stamps of unrestricted historical letters found in archives, they commented, “those dead people are dead, but their living relatives are still around and whatever’s found in there could have implications for them.”

Personal feelings aside, however, this archivist is nonetheless a strong advocate for the application of homogenous standards of restriction that incorporate HIPAA policies—including the eighteen identifiers that make up heart of the privacy law—

---

42 This article by Sarah Zhang, published online by *The Atlantic* on March 1, 2019, may be found here: https://www.theatlantic.com/science/archive/2019/03/dna-tests-for-envelopes-have-a-price/583636/
regardless of the age of a medical record. Speaking to unrestricted UHI in the repositories of other states, they asserted that “it’s important that archivists follow what the law is and not impose their own moral understandings of what is sensitive or not…archivists get into trouble [when] they stray from legal definitions and get into their own personal ethical considerations.” Similarly, they identified the review process and privacy board required by HIPAA to access restricted collections as a “real benefit.” “Having a review process is important, where the researcher has to thoughtfully engage with how they will be using this health information, how much do they need the identifiers and identities of the individuals, what is their intent towards that information, what is their plan to protect that information? The HIPAA privacy rule [spells] out this review process, it defines what the eighteen identifiers are, it sets standards for protection” (participant 4).

The comments of another archivist at a covered institution reiterated these statements, describing a process of establishing archival access that privileges privacy in PHI as well as UHI. “We take the same measures with all the health information, no matter what,” they point out, explaining that their institution initially restricts all identifiers, reviewing the records’ potential for unrestricted access only at the researcher’s request and not before (participant 5).

Observing that “there’s not a whole lot of guidance out there…to make informed decisions,” one archivist at a non-covered entity underscored the legal need for consistently applied standards. “We still have policies that say, ‘oh, at the archivist’s discretion,’” they said. “That’s not a real metric…[There] should be documented criteria that you can go back to and say, ‘well, you know, if someone sues you what business practice do you have to defend yourself?’” They went on to observe that, with the rapidly
changing nature of data dissemination and risk of an accidental privacy violation, archivists cannot make access policies “up on the fly” and advised using “the eighteen HIPAA elements…as a roadmap” (participant 7).

HIPAA is not the only policy that these archivists are referencing for potential UHI standards, however. Both of the archivists above expressed a desire for increased communication between archives and health informatics practitioners, pointing out that resources shared between the two could have a positive influence on the practice and maintenance of restrictions in medical archives and health sciences collections in particular, thus enabling a more efficient implementation of standards. “It’s just kind of amazing to me that here [on campus] we have this huge, lavishly funded institution which does all the development of technology around storing and then providing HIPAA compliant access to patient data for researchers,” mused the former, “and the archives don’t really fit into that in any way…It’s kind of a wonky setup” (participant 5). The latter archivist, employed at a non-covered entity, pointed out that “there are private and secure identifiable data sets out there…there’s all this [data] that’s all identifiable, and that’s a great trove of data for studying, say, health disparities, or economics of healthcare…[T]hose data are available for research and use, but they’re in very controlled environments. It’s not for general consumption…We’re kind of outside that research environment.” “[I]f we got into that system,” they continued, “what would the impacts be?…You would think you would want to be part of that a little bit, just to, again, to cover bases and for your own risk management” (participant 7).

Risk management for copyright infringement is also a field being mined by medical archivists for potential standards and increased archival access. When asked if
their repository applies additional restrictions to UHI used by researchers in online environments, or to restricted material made accessible through the IRB, an archivist responded that setting “restrictions on what [a] person can do after they’ve accessed the records [is] not something…that the archivist does” in their institution. When asked if there was department protocol for dealing with UHI that could be considered harmful to a patient’s relatives or descendants, they replied that, though such an issue hadn’t arisen, they “would always be open to considering closing something if somebody was a direct descendent, [and] if it was reasonable to assert that the information could be damaging or embarrassing.” Speculating that such consideration would be applied on a case-by-case basis, they added that “we would also weigh [the decision] against the value of the thing they’re asking us to close. Obviously not financial value but research value, if it’s something that there’s going to be an informational or research loss if we were to close [it],” before concluding “We have a take-down policy for copyright stuff…we digitize stuff [where] there’s some level of copyright risk, but we always have a take-down policy…So in a similar spirit we’d be willing to consider it [for potential UHI complaints]” (participant 6).

Archivists in smaller institutions exhibited a slightly different view concerning standards implementation, however. The sole archivist in a total staff of five, one interviewee mentioned that UHI data for a certain collection “are over one hundred years old,” but nevertheless noted that “with HIPAA concerns…the individuals [in the records] are not…identified…[they] can’t be traced.” When referencing a collection featuring UHI where there were legal concerns, however, the application of restrictions was revealed to be a bit more arbitrary. The archivist pointed out that they had restricted
access to the collection and noted it in the finding aid; researchers “would have to have a good reason” to request access to the data, and access was provided “at the discretion of the executive director” rather than a privacy board or IRB (participant 3). Another digital manager at a non-covered institution discussed an ongoing decision-making process within the department involving the scope of hospital records to be made available in an online database, and in the process related that they had spoken to both legal counsel about potential HIPAA issues as well as historians and researchers, in order to see what they would like to get from the data. A final decision was still pending, but the archivist offered that, in their “private opinion,” they favored full online access for researchers. “If you are going to digitize something, you should make it available, with everything” (participant 1).

MANAGING ACCESS

Once perceptions of UHI have been parsed, digitization workflows explicated and identified, and recommended standards for restriction accepted, what remains is likely the biggest ethical challenge to archivists in the health sciences and history of medicine collections—access. Though some, following HIPAA and HIPAA-informed standards, leave the ultimate access decisions in the hands of a privacy board or IRB, all participants interviewed for this study believed that the acts of identifying users, disseminating collection metadata and description, and applying archival control in the digital environment were pertinent facilitating access and that these actions closely mirror the user-centric priorities of the current archival field as a whole. Descriptive access to
collections, particularly regarding restrictions, still appears to be occurring at the level of the finding aid.

Archivists from smaller, non-covered repositories identified their users as primarily academia-affiliated—students, professors, researchers, and occasional genealogists (participants 1 and 3). Representatives from larger, covered institutions report a similar user profile. “It’s a heavier percentage of academics and students and then a much smaller percentage of people just with personal interests,” one archivist responded (participant 2). Still others identified users more specifically aligned with significant collecting areas. One archivist, also at a covered institution, replied that, though their collections had a variety of users such as “historians of medicine…historians of science, [and] people who are collecting materials…to make documentaries with,” the collecting areas for which the repository is most widely known are “the ones that tend to get the most usage” (participant 5). An archivist at a non-covered, federal repository that “collects the humanities side” of things added that “professional historians of medicine [are] probably our biggest kind of user group” with the addition of “people working in the areas of public health and the social impacts of health and medicine on humanities and science” (participant 7).

Interviewees included two archivists at significant medical repositories, one covered, the other not. Each identified a wide variety of users from “around the world” and mentioned scholars, artists, documentarians, and journalists (participants 4 and 6). The archivist of the covered entity went on to identify additional users, such as “marketing and communications people who are wanting to write publicity pieces about [the institution], some…history aspect” as well as “a particular additional category [of
users] that are medical genealogists, people who are interested in their family’s medical history” who may have had relatives treated at the institution, and wish to explore their records (participant 4).

Even when digitized copies of UHI records are not available online, many of the archivists interviewed explained their commitment to a thorough description of collections through finding aids available on the institutions’ websites. In the case of repositories that restrict both UHI and PHI, reviewing access requests on a case-by-case basis, restrictions are noted in the finding aid as well. The finding aid “would have a folder listing,” explains an archivist from one of these repositories, “and it would say restricted” (participant 5). For restricted collections containing PHI that may be available in digital formats, several archivists pointed out that they were making a concerted effort to relate those restrictions in different areas of the collection finding aid, most often on the folder level, as recommended by Gustainis and Letocha.43 “We’ve started doing our finding aids a little differently,” responded one archivist. “We’ve moved to ArchivesSpace recently, and we’re beginning to make sure that those restrictions are upfront, that people know in the very early pages of the finding aid, that portions of the collection may be restricted” (participant 2). Another explains, “We have tried to include restriction information as much as possible at the folder level. So, one of the things we do is…we will screen the material to identify any protected health information or other confidential personnel information and we have been including notes on the folder level record identifying the documents that have restrictions on them” (participant 4). Yet another archivist describes the attempts of their repository to drill down further into the

file level, while also providing helpful information about the purpose and duration of the restrictions. “We typically do have a file inventory in our finding aids and so, if a particular file is closed, we will say it’s restricted,” they point out. “We’ll note the term of restriction…then we always use a scoping content note to explain what is the content of the file that is restricted. We don’t want to just say, ‘It’s restricted…’ we’ll say, ‘It’s restricted because the file contains sensitive health information. Those are two things we make sure to always communicate” (participant 6).

Determining access to restricted information differs between institutions. Some provide access on an ad hoc basis. Initially, one archive “would basically have an access copy of a collection from which all the restricted data had been removed” that they would provide to researchers, and “then, if they had a specific request, we would just consider on a case by case basis whether to provide access to the restricted data” (participant 5). Another archivist noted that their department would “often” restrict use, but “might not restrict access depending on what people are here to research.” Describing a digital collection that is redacted publicly, they explain that the original, unredacted scans are still available in the archive. “If someone phoned us and said, ‘hey, I’m searching to see if my relative was a part of the [collections],’ we would probably let that person see one of those old records” (participant 2).

Other repositories, both covered and non-covered, have more formalized policies in place that involve applications to privacy boards or IRBs to determine researcher access to restricted digital or physical materials. In that event, archivists are still working to enable access by myriad methods. An archivist at a large, covered repository described the efforts of their department to place access information in a centralized, obvious
location on their website. “We have a section related to HIPAA and privacy policies in 
general, and our whole suite of application forms is on there, all the different HIPAA 
routes for access,” they explain, but also point out that they take this access one step 
farther. “Also, on the finding aids themselves we will link to our policies page” 
(participant 4). When asked if the privacy board would take different forms of usage into 
account, such as a public-facing DH project or a personal genealogy, they replied in the 
negative. “The standards of protection are the standards of protection…We assume 
anything can go on the web.” Another archivist offered a different access solution, stating 
that “you have to actually apply for access [and] the policy defines the access route…or 
alternatively we say we will provide you copies of [redacted] materials.” “It’s so labor 
intensive,” they also admitted, “I’ve never actually done that” (participant 7). An 
archivist at a non-covered institution offered yet another route, providing help to 
researchers navigating the application process. “If somebody comes to us and they want 
to see records with restricted information,” they said, “then we will help them prepare 
their IRB proposal” (participant 6).

When asked to consider other ways that they might digitize and provide access to 
collections with UHI, queried archivists responded with several different options, all of 
which featured some manner of archival control. “We haven’t digitized anything that has 
individually identifying information,” one archivist confessed. When asked if they 
thought that they would do so in the future, the response was “eventually…[W]hat I 
would see us doing is making it what we call “dark.” We would have the finding aid so 
that people know the records exist, but they might only be accessible within the library 
itself, like someone would have to come to our library and use our computer to look at
them, rather than…being able to download various things. They might have access to the information but they couldn’t reproduce it” (participant 3).

The idea of tiered access for digital collections came up more than once, though the technology is not well established within medical archives, health sciences collections, and history of medicine collections. An interviewee working with a database of hospital records that has not yet been published online mused, “if we decide to hide the [identifiers] and keep the originals that would be great, like maybe create some kind of masks” (participant 1). Another, who praised the “generative” process of “thinking about what it might actually look like to try to think about different models of providing access to information,” mentioned Murkutu CMS as an option. “One of the main things that it does is allow for different kinds of protocols for who can access what kinds of collections, which are mostly based on accounts…you see all these [digital] objects, but you can’t see a whole different metadata layer [without access]” (participant 5).
DISCUSSION

The findings from this study illustrate the myriad issues archivists contend with when making decisions about the use or digitization of unrestricted health information found in their collections, ethical and otherwise. Each of the seven archivists interviewed had to have specific knowledge of federal and state privacy restrictions as they pertained to their institution before they could even speak to the presence of UHI in their collections, much less how they chose to provide access to this information. Archivists’ anxieties about privacy restrictions in medical archives, health sciences collections, and history of medicine collections are well documented in Gustainis and Letocha; in the course of this study, it became clear that those working in digital initiatives are no less susceptible.44

Though Reznick defines the “stewardship” of historical medical libraries and archives as the task of promoting and preserving their holdings, each of the archivists interviewed for this study has expressed an additional sense of stewardship, to their users and the communities represented in their holdings.45 This particular sense of stewardship informs much of their respective decision making processes, be that an adherence to and promotion of HIPAA-based standards for UHI collections, or an extension of archival control such as redaction, restriction, or tiered access in digital environments. Such

44 Gustainis and Letocha, *The Practice of Privacy*, 165.
45 Reznick, “Embracing the Future as Stewards of the Past,” 114.
choices echo the affective responsibilities to records creators, records subjects, records users, and larger communities identified by Caswell and Cifor as radical empathy.46

Ethically, there is a consensus among those interviewed that there is a significant difference between providing access to collections containing UHI in the reading room versus doing the same in an unmediated online environment—a confirmation of Moor’s assertion that “revolutionary technology generates many problems.” In an effort to “anticipate the consequences” of the “deleterious effects” of online collections featuring UHI, archivists are paying particular attention to materials that they deem sensitive, “…a family’s history of sexually transmitted diseases or psychiatric conditions where there could be a social stigma attached to certain kinds of health information” (participant 4) or “a genetic disorder of some sort” (participant 3).47

The statements above evince a reasoned understanding of information considered to be sensitive by Sillitoe—“that which would induce substantial distress in a reasonable person if made publicly available”—as well as a clear conception of the potential for information-based harm were UHI from their collections released online in either an unrestricted or uncontextualized manner.48 Yet it is also clear that each archivist understands the users of their repositories intimately, and works to create access in other ways to ensure that research materials are available to them. In short, many of the participants expressed a desire to balance the needs of their users with an ethical

47 Moor, “Why We Need Better Ethics for Emerging Technologies,” 32.
obligation to those featured in the records, in keeping with Seeman’s assessment of the digital landscape.  

The digitization priorities identified by the subjects of this study by nature reflect this balance. Almost all of those interviewed reported using HIPAA or HIPAA-influenced standards and restrictions to facilitate or mediate digital access, but several provide access—not only to UHI but also restricted material containing PHI—on an ad hoc basis, particularly to family members seeking information about their relatives. Even those archivists who apply their standards more consistently still make an effort to assist users in their research.

It is the virtually unanimous embrace of applying HIPAA-influenced standards to UHI revealed by the respondents, though—as well as the arguments in favor of this decision—that is the most informative for the purposes of this study. In an acknowledgment of the anxiety around both the determination of UHI as well as the wide-ranging impact of making medical documents available online, three of the seven archivists interviewed expressed a desire for “documented criteria” for an online policy, regardless of whether they had established one or not (participant 7). Another reiterated that guidelines for online access cannot be based on archivists’ “personal ethical considerations” (participant 4). Perhaps because HIPAA is already there, or because repositories are already following HIPAA guidelines for materials with PHI, more than half of the archivists in states that make UHI publicly accessible are embracing their role as full partner in the creation of digital exhibits and collections, and are redacting HIPAA identifiers from historical documents while they do it.  

There are a few different applications of this practice, any of which could be considered as guidelines to be used with digitized collections made available online, or digital copies of documents added to online catalogs or finding aids. The most opaque, of course, is the option of restricting all medical records containing HIPAA identifiers in perpetuity. The archivist at this institution does so by necessity; their repository resides in a state that restricts all medical records, even beyond the fifty year restriction period required by HIPAA. The archivist at this covered repository therefore applies the HIPAA identifiers to the entirety of their medical collections, and digital records are treated no differently from physical records. Redacted material may be made available digitally, but heavily redacted collections are not a top priority for digitization or online access.  

Researchers must apply to the institution’s privacy board to gain full access to medical records, which creates a significant barrier to access. In their 2015 report, Gustainis and Letocha surveyed sixty-three researchers, historians, and archivists as to whether they could apply to a review board for collection access, and only fifty-six percent reported having access to such a board. Of those who did have access, only fifty-six percent of those with access submitted a research proposal. Once a research proposal is accepted, the privacy board determines the protective measures required of the researcher, such as but not limited to coding data or keeping digital material on a closed network. Though access is more difficult to come by, the archivist at the above institution offers a convincing argument for their procedures that seems designed to “do no harm” and

51 “Something that’s going to have a lot of privacy concerns, that needs to be heavily redacted, is going to be more costly and labor intensive to digitize…so that can be a factor in setting priorities” (participant 4).  
52 Gustainis and Letocha, *The Practice of Privacy*, 166.
encourages researchers to consider the full breadth of the information they are requesting (participant 4). 53

A less restrictive application of HIPAA-inspired policies takes place at a non-covered entity. Records containing HIPAA identifiers are restricted for a period of eighty years from the date of creation; following that period, materials are “totally open access.” When asked, the archivist readily described “sensitive” collections that “made [the staff] have a…strong reaction” such as “photographs of incarcerated individuals who had committed suicide” and photographic material “studying the impact of the long term effects of starvation on folks from concentration camps.” “Ultimately we have not had a case where we decided it was necessary to apply a longer restriction period,” she concluded, though they did not make clear whether the disturbing collections they mentioned were made available online (participant 6). Again, access to physical and digital material was the same.

Yet another application of this practice is specific to online and digital collections containing UHI. After considering the dates of HIPAA and restrictions on medical records in their state, one archivist redacts HIPAA identifiers in all material published online, regardless of whether the records are publicly available or not (participant 3). Their reasoning, too, seemed designed to avoid information-based harm, explaining that the relatives of identified individuals might experience a loss of health insurance coverage based on pre-existing genetic conditions. For other historical collections in the pipeline for digitization that could not be easily de-identified, the archivist proposed

---

53 “I think having a review process is important…the researcher has to thoughtfully engage in how they will be using this health information, how much do they need the identifiers and identities of the individuals, what is their intent toward that information, what is their plan to protect that information” (participant 4).
making the collection “dark,” or restricting researcher access to the reading room of the institution.

It is worth noting, though, that while each of these three scenarios employ HIPAA-influenced restrictions, differences in scale and application are not only applied as a result of outside influences. Nor are they arbitrary applications applied at the whim of the archivist; these restrictions also speak to the varying sizes and resources of the institutions they represent—an important and necessary consideration in the attempt to identify a set of best practices for the profession as a whole. The first two applications of HIPAA identifiers are from large, well-funded institutions with significant endowments. The third is a solution devised by an archivist at a significantly smaller institution—they are one of a total staff of five. Letocha and Gustainis, speaking of physical access to medical collections, point out that “not all archives have the resources to support access, such as privacy boards, institutional review boards, or informed legal counsel.”\(^{54}\) Online publication stresses these slender resources even further.

Resources determine what material is digitized or placed online, as well. Many participants considered The process of providing redacted digital copies of medical records as too labor intensive to seriously consider as an access measure. Extensible processing methods, which require fewer labor hours and, in the words of one archivist “results in us not reviewing every single record in a collection,” enable physical but not digital access to records, due to the calculated risk of missed identifiers being made available online. Funding is also a significant issue in digitization criteria. With the exception of two repositories, all archivists questioned mentioned funding as a significant

\(^{54}\) Gustainis and Letocha, *The Practice of Privacy*, 165.
consideration in the digitization process, some even expressing doubt that their donor-funded, recently digitized collections have as much research value as those on the backburner (participants 2 and 5). Donors see digitization projects as a “sexy” use of new technology, but many of the archivists queried despaired being able to make their own criteria for significant online exhibits (participant 2).

In the spirit of expanding limited resources and reimagining online access to medical records, archivists expressed several different ideas for the immediate future. Two of the archivists interviewed suggested collaboration between the practice of health informatics and the archives (participants 5 and 7). “It’s like from a records management…perspective…we’re doing the same thing, so why is one real elaborate thing set up over here and then chronically underfunded things set up over here?” one archivist asked, before asserting that,

“The perspectives have a lot to offer…and we don’t tend to let that happen…just having the resources that are devoted to health informatics as a perspective to offer to archives…[and] some of the really cutting edge work that archival practitioners have done around negotiating consent among the people who are documented really has a lot to offer to the medical field. And those connections don’t get made a lot” (participant 5).

Proposed collaboration with health informatics and the emerging technologies of that field is not only coming from a financial perspective, however. Several archivists expressed an interest in engaging with different kinds of online access, facilitated through what Moor identifies as “revolutionary technology” in order to anticipate potential consequences of making UHI records digitally available.55 From the perspective of the above interviewee, however, archivists’ engagement in the ethical considerations of this access could be valuable to health informatics practitioners, as well. In the meantime,

---

55 Moor, “Why We Need Better Ethics for Emerging Technologies,” 32.
Archivists expressed their awareness of grant-funded academic explorations into online content management systems with tiered-access solutions such as Murkutu and the digital archive built for the Central State Hospital Project (participants 4 and 5). Other technological possibilities, such as the development of virtual reading rooms or specific database permissions, were mentioned as well, clearly expressing archivists’ interest in the extension of archival control to the online environment. Managed online access of this sort has the ability to solve many of the issues brought up by this study, but the focus, for the time being, will remain on digital collections with single access points and unmediated audiences.

Ultimately, this study has established that there are significant discrepancies between the regulations and material resources available to medical archives, health sciences collections, and history of medicine collections, discrepancies that directly affect the ability of these repositories to digitize and make accessible valuable collections containing UHI. Funding discrepancies cannot be addressed through the development and implementation of best practices for digital access, nor can cross-institutional collaboration, but the reflections of each of the archivists interviewed for this study indicate that basic guidelines could: help educate archivists about UHI in their collections thus alleviating digital access anxiety and ethical concerns; and reduce the labor and effort required for each institution to put their own best practices in place, a boon for repositories with limited resources. Ultimately, the implementation of such practices has the ability to foster increased communication within the field, as well as an ideological alignment between medical libraries, health science collections, and history of medicine collections nationwide.
CONCLUSIONS AND FURTHER RESEARCH

The archivists interviewed for this study constitute a small sample size of the medical archives field; nevertheless, several different types of repositories are represented, such as covered and non-covered entities, as well as institutions subject to different federal and state privacy laws. Therefore, while this study cannot be considered representative of the field as a whole, it contains information useful to further inquiries and attempts to create standardized guidelines for the digitization and online publication of UHI.

Interestingly, what emerged most clearly from this study is that medical archivists, tasked with the dual duty of facilitating access for researchers and protecting patients and their relatives from information-based harm, see very little difference between the online publication of unrestricted health information and that of restricted health information. Though they are committed to providing information and data to researchers, the medical code of perpetrating no harm is paramount, and the removal of HIPAA identifiers in health information provided online is a common practice, regardless of whether that information is accessible by the public. This provides some level of archival control beyond the finding aid, while at the same time removing the question of whether some identifying information is “safer” than another by virtue of its age. Further, this practice serves to remove at least one of the facets medical archivists must consider when deciding what to make available digitally, facets that seem daunting when taken as
The blanket application of HIPAA identifier restrictions has other benefits as well—it encourages archivists at covered and non-covered institutions to educate themselves about privacy law, regardless of the status of their repository, and also establishes some level of homogeneity within a wildly diverse field. Though Gustainis and Letocha do not address the concerns of online access in their work, the three steps that they identify as essential to enabling access—raising awareness of collections, focusing on collection description as a chief access point, and advocating for more comprehensive federal and state privacy protection laws—can all be accomplished through a collaborative effort to establish guidelines and best practices for the digitization and online publication of medical records.

Future research could certainly be extended to include a more comprehensive sample of medical archives, health sciences collections, and history of medicine collections. Based on information gathered thus far that suggests medical archivists perceive little difference between the online publication of unrestricted health information and that of restricted health information, the scope of the study could be extended to include all digitization and online publication projects. It would also be useful for future studies to focus on the differences between covered and non-covered entities, in order to develop best practices for each; the variety of restrictions found in the medical archives field does not suggest that one set of best practices will prove applicable to every institution.

Regardless, archivists interviewed for this study overwhelmingly expressed a desire for a consistent set of criteria to apply when making decisions about online access.
The implementation of such will protect archivists and their institutions, not only legally, but ethically, and set an important precedent that can move the field into what promises to be a rich and technologically-motivated future.
BIBLIOGRAPHY


McCarthy, Angela, Catharine Coleborne, Maree O’Connor, and Elspeth Knewstubb. “Lives in the Asylum Record, 1864 to 1910: Utilising Large Data Collection for Histories of Psychiatry and Mental Health.” Medical History; London 61, no. 3 (July 2017): 358–79.


Reznick, Jeffrey S. “Embracing the Future as Stewards of the Past: Charting a Course Forward for Historical Medical Libraries and Archives” RBM: A Journal of Rare Books, Manuscripts, and Cultural Heritage, 5, no. 2 (Fall 2014): 111-123.


APPENDIX A

HIPAA IDENTIFIERS

1. Names

2. All geographical subdivisions smaller than a State, including street address, city, county, precinct, zip code, and their equivalent geocodes, except for the initial three digits of a zip code, if according to the current publicly available data from the Bureau of the Census:
   a. The geographic unit formed by combining all zip codes with the same three initial digits contains more than 20,000 people
   b. The initial three digits of a zip code for all such geographic units containing 20,000 or fewer people is changed to 000

3. All elements of dates (except year) for dates directly related to an individual, including birth date, admission date, discharge date, date of death; and all ages over 89 and all elements of dates (including year) indicative of such age, except that such ages and elements may be aggregated into a single category of age 90 or older

4. Phone numbers

5. Fax numbers

6. Electronic mail addresses

7. Social Security numbers

8. Medical record numbers

9. Health plan beneficiary numbers

10. Account numbers

11. Certificate/license numbers

12. Vehicle identifiers and serial numbers, including license plate numbers

13. Device identifiers and serial numbers

14. Web Universal Resource Locators (URLs)
15. Internet Protocol (IP) address numbers

16. Biometric identifiers, including finger and voice prints

17. Full face photographic images and any comparable images

18. Any other unique identifying number, characteristic, or code (note this does not mean the unique code assigned by the investigator to code the data)
APPENDIX B

INTERVIEW GUIDE

1. Who are the users of your digital and physical collections?

2. Beyond what is required by HIPPA, do you limit or restrict access to any of the historic health information in your physical collections belonging to those deceased for more than 50 years?
   a. If so, why?
   b. If so, how do you communicate or indicate access restrictions?

3. In both physical and digital collections, is there historical health information that you consider more or less sensitive than other historical health information? Why or why not?

4. Do you perceive a difference between providing access to physical collections containing historical health information and making those collections available digitally or online?

5. When considering what material from your collections to make available digitally, does the presence of historical health information affect your decision-making process?
   a. What other criteria do you consider when deciding what to digitize?

6. Beyond what is required by HIPPA, do you limit or restrict access to any of the historic health information in your digital collections belonging to those deceased for more than 50 years?
   a. If so, why?
   b. If so, how do you communicate, indicate, or implement access restrictions?

7. Does your institution have policies in place requesting that researchers anonymize or code data that they collect from historical health information in your collections?
   a. If so, do these policies differentiate between published research and projects made available digitally?

8. Do you have any advice for archivists working with historical health information and/or planning digitization projects?
APPENDIX C

PERSONALIZED EMAIL CONTACT TEMPLATE (SNOWBALL SAMPLING)

Dear [contact name],
My name is Sarah Almond, and I am a master’s student in UNC Chapel Hill’s School of Library and Information Science. For my master’s paper, I am exploring archival ethics as they apply to the access and use of historical health information, defined as the health records of individuals who have been deceased for longer than 50 years, and are thus non-covered by HIPAA regulations. More specifically, I am examining how archivists and librarians in medical libraries, health science collections, and history of medicine collections approach the use of historical health information in digital environments.

I am conducting an exploratory study and seeking to interview archivists and librarians with over three years of experience who have implemented or planned to implement a digitization project using historical health information within the last five years. I have been studying the websites of several institutions that have active digital collections in the areas of medical history or health science with an eye towards faculty or staff that might fit this description, and came across your name. If you feel that you would be able to speak about this topic, I would greatly appreciate the chance to interview you. If you don’t feel that you can speak to this topic, but know of another colleague that might, I would welcome the recommendation.

My study has been approved by the IRB, and I am happy to answer any questions you might have regarding my research project. If you are willing to participate in my study or have questions, I can be reached at [email address].

Thank you very much for your time.

PERSONALIZED EMAIL CONTACT TEMPLATE (RECOMMENDED CONTACT)

Dear [contact name],
My name is Sarah Almond, and I am a graduate student in UNC Chapel Hill’s School of Library and Information Science. For my master’s paper, I am exploring archival ethics
as they apply to the access and use of historical health information, defined as the health records of individuals who have been deceased for longer than 50 years, and are thus non-covered by HIPAA regulations. More specifically, I am examining how archivists in medical libraries, health science collections, and history of medicine collections approach the use of historical health information in digital environments.

I spoke with [name] regarding my topic, and they mentioned that you would be a good person to talk to regarding my research. I am conducting an exploratory study and seeking to interview archivists with over three years of experience who have implemented or planned to implement a digitization project using historical health information within the last five years.

If you feel that you would be able to speak about this topic, I would greatly appreciate the chance to interview you. If you don’t feel that you can speak to this topic, but know of another colleague that might, I would welcome the recommendation.

My study has been approved by the IRB, and I am happy to answer any questions you might have regarding my research project. If you are willing to participate in my study or have questions, I can be reached [email address].

Thank you very much for your time.