# The Opportunities of Engagement: Working with Scholars to Improve Description and Access at the Center for the History of Medicine

Emily R. Novak Gustainis, Head, Collections Services, Center for the History of Medicine, Francis A. Countway Library of Medicine

### **Abstract**

The Center for the History of Medicine has used its two CLIR-funded grant initiatives to engage researchers at all stages of their professional careers in an effort to understand how archival description can be improved or modified to the benefit of scholars while remaining attentive to workflows that speed processing. This paper considers findings from the Center's "Foundations of Public Health Policy" (2008) and "Private Practices, Public Health: Privacy-Aware Processing to Maximize Access to Health Collections" (2012) initiatives.

s the Center for the History of Medicine's (the Center) head of collections services, my day-to-day work focuses on the mechanisms of discovery: systems, standards, and practice. While I occasionally staff the reference desk, it requires planning on my part to meaningfully interact with our researchers and find out if what we are doing on the back end is meeting their needs. Researchers are our prime movers, the reason we kick the gears into action and then examine the cogs, so it is no wonder user assessment is instrumental to trying out new ways to do our work. Over the past seven years and through two CLIR grants, the Center has posed a number of questions to its constituents: How do you prefer to learn about collections? What makes a finding aid useful? What is the minimum amount of information you need to determine if a collection is worth your time? And, most recently, how can we help improve access to records containing health information about individuals? These opportunities to engage fuel our data collection, both on our processes and our products.

I will say up front that there are only two ideas to take away from this paper, and I expect they

are not revolutionary. One, repositories should increase the depth of description for collections that have limited access; and two, approaches to description should better recognize the needs of different disciplines. In my experience, our inclination as archivists is to provide much more description for open collections because they can be used right now, rather than to expend energy helping our users make tough choices about whether or not to go through the process—when it is even possible—of appealing for access to unprocessed collections. Similarly, our urge to suppress catalog records for unprocessed holdings can drastically reduce opportunities for our researchers to discover collections, resulting in missed opportunities to gauge what is of user interest.

With the Center's 2008 grant, "Foundations of Public Health Policy" and 2012 grant, "Private Practices, Public Health: Privacy-Aware Processing to Maximize Access to Health Collections," which was proposed under the auspices of the Medical Heritage Library and conducted in partnership with the Alan Mason Chesney Medical Archives of the Johns Hopkins Medical Institutions, the Center had the opportunity to explore, respectively, the

descriptive needs of a wide variety of researchers and those of medical and social historians needing access to records containing health information about individuals held by both HIPAA (Health Insurance Portability and Accountability Act) and non-HIPAA covered repositories. The results were polar opposites of one another.

### Foundations of Public Health Policy Grant

For "Foundations of Public Health Policy," the Center committed to building a community of interested public health practitioners, scholars, and students to support the acquisition and use of public health collections and to advise us on collections description and outreach. For descriptive assessment, this included recruiting 52 public health and information professionals, students, and historical researchers to evaluate the use of box and folder lists for the collections we were processing, those of public health and public health administration leaders Leona Baumgartner (1902-1991), Allan Macy Butler (1894-1986), Howard Hiatt (1925-), and David Rutstein (1909-1986). The box and folder lists were grouped by series and delivered online as spreadsheets through a take-home exercise and a post-exercise interview. Our objectives were to:

- Assess how useful researchers found box and folder lists containing select metadata independent of any top-level descriptive information associated with a finding aid (such as biographical notes, subject access point, or series descriptions)
- Ascertain whether or not researchers could perform routine information-seeking tasks associated with identifying materials of interest for research use with just a spreadsheet

 Determine whether a spreadsheet provided enough descriptive information to engender "trust" in the resource researchers were using; that is, to find out whether researchers were comfortable with this resource alone as a determining factor to schedule a research visit

Additionally, the scholarly engagement exercise and interviews were helpful to observe the potential challenges MPLP (or at the Center, what we would consider "appropriate level" processing) and other innovations may present to users. Participants, who were divided across the four collections processed for the project, were asked to complete an exercise (Appendix A) consisting of both general and quantitative multiple choice questions and collection-specific, qualitative questions about how they used two versions of an Excel spreadsheet to answer a number of questions. This included an "A" version spreadsheet with nine elements (box number, folder number, series, subseries, sub-subseries, folder title, begin date, end date, and in the case of Hiatt alone, access restrictions) and a "B" version with additional descriptive information provided in a notes field, such as to indicate the presence of photographs. The average completion time of the exercise was one hour, with the post-assignment interview averaging 30 minutes. In retrospect, it was impressive we had 52 respondents, though we did need to eliminate some of the non-U.S. participant data because of differing interpretations of survey terminology.

From our interviews, we gleaned the following:

By overwhelming consensus, participants agreed that a spreadsheet was a good enough tool to make preliminary selections about what to look at during a research visit, particularly in conjunction with a collection-level record available through HOLLIS, the online catalog of the Harvard University Library.

- Folder lists without contextual groupings (such as series or subseries) would have been less helpful to discovery. Comments from participants confirmed that folder titles can be "deceptive" on their own. Knowing that a folder named "Meetings with Jane, 1990" is part of a series of teaching records or a series of patient files substantially changes a researcher's interpretation of a folder title.
- Inexperienced researchers tended to want more descriptive information and preferred spending less time at a repository. Experienced researchers cared only about access. Moreover, experienced researchers stated that regardless of how an archivist handled a collection, they wanted to conduct their own records review. In other words, they were not going to take anyone's word on what was in a collection or if it was relevant to their work.
- No one was concerned about subject access and only a handful of people mentioned an interest in having more contextual information (beyond groupings) or contentrelated information.
- Participants cared about record formats. Most liked the "B" version of the spreadsheet with the notes field because the occasional notes provided by processors indicated the presence of photographs or other non-textual items.
- Participants liked the idea of applying a minimal controlled vocabulary to a folder title ("qualifiers") to help disambiguate folder transcriptions, such as "Writings" or "Correspondence." However, multiple qualifiers would have to be employed; if you qualify a folder with "writings," the implication is that the whole thing will be writings, not writings intermixed with correspondence and reference material to support said writings. One individual thought indicating whether

- or not a preceding or succeeding folder that had a more descriptive title (such as activities) contained the same type of records as a vaguely titled folder.
- Unless archivists are going to provide a substantial number of folder-level scope and content notes, it does not seem worth providing them at all. Most people did not think folders with notes were "more important," and very few people assumed Center archivists were making judgment calls on content.
- While most researchers assumed the spreadsheet could be manipulated, such as to sort by date or add columns, they did not do so. They scrolled and keyword-searched the spreadsheets. For two collections, we provided one massive spreadsheet, and for two collections we provided folder lists tabbed out in Excel by series. The tabbed versions of spreadsheets ended up confusing a few participants, particularly those uncomfortable with Excel. Spending time on novel ways to use the spreadsheet was not of much interest.
- Thoughtful accessioning practices can facilitate access in advance of a full finding aid. Public health processors roughly sorted the records into series and subseries in advance of box and folder listing. The more astute we are about doing this at the point of acquisition or accessioning, the faster we can enable meaningful access in lieu of full finding aids.

As a result of the grant, our then acquisitions archivist—and what is now our acquisitions team—spends much more time grouping and packing records up front and on-site with the donor when feasible. Because all acquisitions are listed before being sent to off-site storage, the result is a higher-quality collection inventory that is easier for public services staff to navigate and a resource that speeds processing planning. We are also far

more comfortable allowing researcher access to unprocessed collections.

### **Private Practices, Public Health Grant**

In 2012, the Center partnered with the Alan Mason Chesney Medical Archives of the Johns Hopkins Medical Institutions to process the papers of seven leaders in public health research and advocacy. The grant, "Private Practices, Public Health: Privacy-Aware Processing to Maximize Access to Health Collections," addressed issues of access in both HIPAA and non-HIPAA covered entities. seeking best practices for archivists facing challenges and confusion over making health-related records in their collections available. Specifically, the project sought to: a) understand the informational needs of scholars seeking to use restricted records and what they consider most valuable to research; b) evaluate how language used by the Center and Hopkins to communicate restrictions in finding aids correlates with the perceived utility of the finding aid and the potential of the collection to satisfy a research need; and c) get feedback on the process for applying for access to collections with restricted records.

To do so, the Medical Heritage Library, represented by the Center's Kathryn Hammond Baker, Scott Podolsky, and Emily Novak Gustainis, and Hopkins's Phoebe Evans Letocha, led discussion sessions, launched an online survey, presented at professional conferences, and ultimately distributed Novak Gustainis and Evans Letocha's jointly authored "Recommended Practices for Enabling Access to Manuscript and Archival Collections Containing Health Information about Individuals" to the research and professional communities for feedback. While this engagement work was under way, Center and Hopkins staff were simultaneously processing collections, including the Center's collections of the papers

of Oliver Cope (1902-1994), Stephen W. Lagakos (1946-2009), Erich Lindemann (1900-1974), and Arnold S. Relman (1923-2014), as well as the records of the Harvard School of Public Health's Department of Biostatistics, 1981-2009.

### Efforts included:

- A SurveyMonkey survey (Appendix B) on access to health records, "Research Access to Protected Records Containing Health Information about Individuals," was distributed to the Medical Heritage Library governance committee and circulated to professional and discipline-directed listservs. In total, 63 people responded. Data obtained as a result of the survey were analyzed for engendering discussions between archivists and historians and informing the creation of the Recommended Practices document.
- A workshop for Harvard University's History of Medicine Working Group, comprising graduate students and faculty from the History of Science Department, which included an interactive review of finding aids for collections containing restricted records and a discussion of the type of descriptive content researchers might need to evaluate the usefulness of the records for research.
- A lunch workshop at the 2014 annual meeting of the American Association for the History of Medicine (AAHM) to elicit the information historians need to determine whether or not it is worth applying to an Internal Review Board (IRB), what information is most useful to them, and what they think is missing from finding aids. As part of the session, historians Janet Golden, Rutgers University, and Cynthia Connolly, University of Pennsylvania, shared with the audience their research experiences and difficulties using patient records to inform

their research and launched a discussion of how to improve the user experience.

 A presentation at the 2014 Annual Meeting of the Society of American Archivists in Washington, DC. The session, "Partners in Practice: Archivists and Researchers Collaboratively **Improving** Access Health Collections," offered perspectives of both historians/researchers and archivists on the importance of making a wide variety of records that contain health information about individuals discoverable. The session was moderated by Susan Lawrence of The Ohio State University and included a presentation by John Harley Warner of Yale University, "Why Patient Records Matter to the Historian."

As a result of these efforts, it became clear that because of the complexities related to applying for access to records containing health information about individuals-whether closed to comply with HIPAA, state law, or institutional policydifferent kinds of descriptors were necessary. When survey participants were asked whether or not they could apply to a review board to obtain access, only 56.14 percent (32) had access to a review board, and of the 32 individuals who did have access, only 56.25 percent (18) actually went through the process. Respondents said it took too much time-especially when they learned about the restrictions only when they were already on-site, were not convinced they would actually get access if they went through the process, or believed that they lacked support or guidance. Respondents stated that the most significant barrier to using records containing confidential/ protected health information held by special collections, archives, and museums was lack of information about the records themselves and access procedures, stating, "I see records that look interesting in catalogs or collection guides, but I can't tell if they will be useful" and "The process takes too long."

Researchers simply needed more information to help them make decisions. As professionals, we cannot control what the Internal Review Board or Access Board applications require or how frequently they meet, but we can provide more description. Offering samples of variables found in the records, such as patient diagnosis or condition; the age, gender, and race of patients; and the types of medications or procedures, can make a difference. Better indicating the kinds of records in a collection (patient histories vs. case files vs. medical records, rather than just "patient records") and providing more explicit statements at the folder level indicating why the folder is restricted, and for how long, can also help. For example, all of the Center's finding aids authored for the grant include descriptive information about the type of restrictions found in the collections, why they were imposed, and how to obtain access. At the folder level, the Center provided a transcription of the full folder title (redacting patient names), the year the records will open to the public, and a qualifying description in the form of a folder-level scope note to convey the intellectual contents of the folder without revealing protected information. Because the Center is a non-HIPAA covered entity, access restrictions were determined by sampling the content of records in each folder. Hopkins, a HIPAA-covered entity, has a similar descriptive approach, except Hopkins staff must screen documents at the item level for protected health information. Readers are encouraged to review the paper "The Practice of Privacy" by Emily Novak Gustainis and Phoebe Evans Letocha, in this volume, for more detailed information about the descriptive process.

So how do we balance the needs of users who are seemingly satisfied with a structured box and folder list on one side of the continuum and those that want the kind of information that can only be gleaned by labor-intensive sampling? Perhaps for public health and other to-be-determined collections with little or no patient records, we should scale back on descriptive processes, letting the collections do the talking while we ramp up efforts to describe records that do not easily present. Archivists can also become more comfortable describing boxes of access-protected records only at the container or series level, focusing on the records that comprise a series more holistically. Flexible and appropriately applied approaches to processing can be built only through concerted efforts to understand the needs of our multiple constituencies. The Center is now testing the best practices developed for the description of collections containing health information about individuals on the Dwight E. Harken papers, 1930s-1990s, and should be completed concurrent to the symposium. We look forward to sharing the finding aid with members of the community for feedback.

### References

Gustainis, Emily. 2015. Now Available!
Recommended Practices for Enabling Access to
Manuscript and Archival Collections Containing
Health Information About Individuals (Feb.
12). Available at http://www.medicalheritage.
org/2015/02/now-available-recommended-practices-for-enabling-access-to-manuscript-and-archival-collections-containing-health-information-about-individuals/.

## APPENDIX A: Example User Study: "Foundations of Public Health Policy" (2008)

Center for the History of Medicine, User Study, 2009–2010

s part of its Foundations in Public Health Policy grant work, the Center for the History of Medicine, Countway Library, is conducting a study to document and analyze how researchers use and respond to different access tools developed for delivering information about its collections to the public. This grant, as funded by the Andrew W. Mellon Foundation and administered by the Council on Library and Information Resources (CLIR), has enabled the Center to experiment with ways to make collections available to researchers over shorter periods of time. Your participation in this study will help us determine the efficacy and utility of these tools.

The study consists of two parts:

- 1. An exercise consisting of both multiple choice survey questions and detailed questions designed to assess how participants use two different versions of a spreadsheet to answer questions about a collection. We estimate that it will take about one hour to complete the study exercise.
- 2. An interview (either at the Center or via conference call) that will focus on content and usability issues related to how you used the spreadsheets to complete the exercise. We estimate the interview to take between thirty minutes and one hour.

We sincerely appreciate your taking the time to participate and provide feedback. We are happy to answer any questions you might have about the project. Please contact Michael Dello Iacono, Project Archivist, MPD13@hms.harvard.edu, and Emily R. Novak Gustainis, Collections Services Archivist, ERN6@hms.harvard.edu.

### 7. Which is more important to you?

- a) Being able to spend less time at a repository because I have very detailed information about a collection and know exactly what to have pulled for me, even if it means I might not be able to access a collection in the immediate future while this type of information is being collected.
- b) Having access to a collection, even if it means I may need to spend a lot of time at a repository searching for the information I want.

## 8A. If you could only have ONE of the following discovery tools, which would you rather have available to you online and in ADVANCE of your visiting the Center for the History of Medicine?

- a) A detailed summary of a collection's content, biographical/institutional information about the creator(s) of the collection, and information about what kinds of documents are in the collection, including date spans for all materials.
- b) A spreadsheet containing a list of every folder "title" in every box of the collection as it was originally labeled by the person who created or assembled the collection.

### 8B: Why did you pick A or B?

### 9. Which is more important to you?

- a) Being able to simultaneously search for subjects or people across many finding aids in order to discover which collections at a particular repository may help me with my research.
- b) Being able to print an inventory for, or guide to, a collection for personal reference use from a list of collections posted on a repository's website.

### **B. Spreadsheet-Specific Questions**

#### Part B-I

To complete Part B-I, please copy and paste the following location into your web browser: http://repository.countway.harvard.edu/xmlui/handle/10473/3600

Under "Sample Submissions," select the file: CLIR\_baumgartner\_boxlist\_A.xls

Under "Associated Files," click on "View/Open" for the file: CLIR\_baumgartner\_boxlist\_A.xls

Please take a few minutes to look at the spreadsheet and then answer the following questions. Please feel free to cut and paste answers from the spreadsheet into this Word document.

1. What are the five major groups of records found in the collection?

- 2. Pick two of the groups you identified in question 1. What kinds of information would you expect to be in these groups of records?
- 3. What types of materials would you expect to find in the collection?
- 4. Why?
- 5. In what year did Baumgartner travel to Russia?
- 6. Please name two of Baumgartner's published articles from the 1950s.
- 7. What steps did you take to answer to questions 5 and 6?
- 8. How much material is there in the collection related to speeches given by Baumgartner?
- 9. What steps did you take to answer question 8?
- 10. If you were looking for letters between Baumgartner and her family members, where would you expect to find them?
- 11. Please list three individuals Baumgartner corresponded with.
- 12. From 1954-1962, Leona Baumgartner served as Commissioner of Public Health for the city of New York. Where would you look for records in the collection that relate to Baumgartner's public appearances as Commissioner?
- 13. You are researching the activities of the American Public Health Association (APHA) in the 1950s and 1960s. Which boxes would you ask to see?
- 14. How would you find out if there were items from the 1970s?
- 15. How would you determine how much material from the 1970s is in Series 4?
- 16. Did you need to print the spreadsheet in order to answer the above questions?
- 17. Did you save the spreadsheet to your desktop or local drive before working on the questions?

#### Part B-II

To complete Part B-II, please copy and paste the following location into your web browser: http://repository.countway.harvard.edu/xmlui/handle/10473/3600

Under "Sample Submissions," select the file: CLIR\_baumgartner\_boxlist\_B.xls

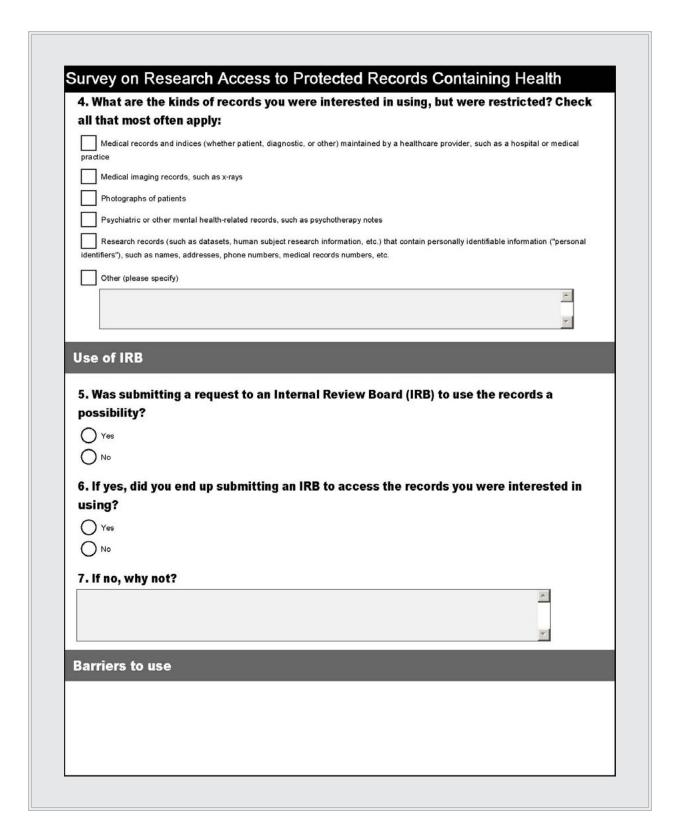
Under "Associated Files," click on "View/Open" for the file CLIR\_baumgartner\_boxlist\_B.xls

Please take a few minutes to look at the spreadsheet and then answer the following questions. Please feel free to cut and paste answers from the spreadsheet into this Word document

- 1. Summarize the kind of information provided in the "Notes" column.
- 2. What types of materials would you expect to find in the collection?
- 3. Baumgartner delivered many speeches and lectures throughout her career. Please name two people who delivered a talk along with Dr. Baumgartner.
- 4. Please explain how you found the above answer.
- 5. Do folders with notes in the "Notes" column contain more important documents?
- 6. Please list the question numbers in section B-I that you would now answer differently having seen the "Notes" column.

### **APPENDIX B: Private Practices, Public Health Survey (2012)**

About th	e survey
Medical Archive collections cont	eing conducted by the Center for the History of Medicine, Francis A. Countway Library of Medicine, and the Alan Mason Chesney s of the Johns Hopkins Medical Institutions as part of a joint effort to develop best practices for enabling access to special aining protected health information (PHI) and other types of access-protected ("restricted") records. For the purposes of this survey, on (protected or otherwise) is defined according to the HIPAA Privacy Rule (1996) as:
<ol><li>the provision that identifies th</li></ol>	cluding demographic information, which relates to: 1) an individual's past, present, or future physical or mental health or condition of health care to the individual, OR 3) the past, present, or future payment for the provision of health care to the individual, AND are individual or for which there is a reasonable basis to believe can be used to identify the individual. Protected health information common identifiers (e.g., name, address, birth date, Social Security Number) when they can be associated with the health addove."
	er HIPAA, a medical record, laboratory report, or hospital bill would be PHI because each document would contain a patient's nam intifying information associated with the health data content.
	o this survey, you are helping libraries and archives improve how they describe records and make hidden collections available to lore useful ways.
Thank you!	
$\circ$	ou used manuscript collections or archival records as part of your research?
O Yes	
○ No 2. If yes, I	have you used a collection guide ("finding aid") that included information about or not patient or other health-related records in the collection had access
2. If yes, I whether corestriction	or not patient or other health-related records in the collection had access
No 2. If yes, I whether o	or not patient or other health-related records in the collection had access
2. If yes, I whether crestriction Yes No  3. How ha	or not patient or other health-related records in the collection had access
2. If yes, I whether corestriction  Yes  No  3. How has collection	or not patient or other health-related records in the collection had access ins?  The second
No  2. If yes, I whether corestrictio  Yes  No  3. How has collection  A libraria	or not patient or other health-related records in the collection had access ins?  Eve you learned about the presence of restricted records for the majority of the its you have used (or were interested) in using?
No  2. If yes, I whether orestrictio  Yes  No  3. How hat collection  A libraria  Online u	or not patient or other health-related records in the collection had access ins?  Eve you learned about the presence of restricted records for the majority of the ins you have used (or were interested) in using?
No  2. If yes, I whether orestrictio  Yes  No  3. How hat collection  A libraria  Online u	or not patient or other health-related records in the collection had access ins?  Eve you learned about the presence of restricted records for the majority of the ins you have used (or were interested) in using?  In or archivist is sing a finding aid is sing a library catalog record



$\sim$	
The IRB proc	ess takes too long
Too much pa	perwork is required to get access to restricted records
$\tilde{\sim}$	that look interesting in catalogs or collection guides, but I can't tell if they will be useful
Nothing is di	
$\sim$	aren't open when I have time to do my research
$\tilde{\sim}$	Il be able to quote, reference, or use the records in publications  Il me if I can use the records or not
$\sim$	apply to me. There are no barriers to my use of records
Other (please spec	
Other (prease spee	
collection g most useful	criptive information do you think is missing from library catalog records or uides (such as those for a manuscript collection)? What information would to you in deciding whether or not a collection has information relevant to y
collection g most useful	uides (such as those for a manuscript collection)? What information would
collection g most useful research?	uides (such as those for a manuscript collection)? What information would
collection g most useful research?	uides (such as those for a manuscript collection)? What information would to you in deciding whether or not a collection has information relevant to y
collection g most useful research?	uides (such as those for a manuscript collection)? What information would to you in deciding whether or not a collection has information relevant to y
collection g most useful research?	uides (such as those for a manuscript collection)? What information would to you in deciding whether or not a collection has information relevant to y
collection g most useful research?	uides (such as those for a manuscript collection)? What information would to you in deciding whether or not a collection has information relevant to y
collection g most useful research?	uides (such as those for a manuscript collection)? What information would to you in deciding whether or not a collection has information relevant to y

nealth information?	Not very useful	Somewhat useful	Very useful	Does not apply to my
Average age of patients at time of treatment	0	0	0	research
Date span of records	$\circ$	0	0	0
Diagnosis/condition	$\sim$	ŏ	ŏ	ŏ
Duration of treatments	$\sim$	ŏ	~	
Genetic information	Ŏ	ŏ	0	0000
Geographic region covered	ŏ	ŏ	ŏ	ŏ
Names of medical devices used	Ŏ	Ŏ	Ŏ	Ŏ
Prescribed medications	0	0	0	0
Race of patients	0	O	0	Ŏ
Sex of patients	0	0	0	Ŏ
Surgical procedures	0	0	0	0
Treating physician/surgeon	0	0	$\circ$	0

	Not very useful	Somewhat useful	Very useful	Does not apply to my research
Admission/registration records	0	0	0	0
Autopsy records	0	0	0	0
Billing information	0	0	0	0
Case files	0	0	0	0
Consultation files	0	0	0	0
Correspondence	0	0	0	0
Diagnostic indices	0	0	0	0
Family medical histories	0	0	0	0
Genetic testing records	0	0	0	O
Graphs and charts	0	0	0	0
Hospital policies	0	0	0	0
Immunization records	0	0	0	0
Insurance claims	0	0	0	0
Informed consent records	0	0	0	0
Lab notebooks	0	0	0	0
Microscope slides/specimens	0	0	0	O
Photographs/medical imaging	0	0	0	0
Patient histories	0	0	0	0
Patient questionnaires	0	0	0	0
Patient summaries	0	0	0	0
Prescription books/logs	0	O	0	0
Research protocols	Ō	Ō	Ō	0
Surgical logbooks	0	0	0	0
Other (please specify)				
omments and D	emographics			
0-2				
12. Do vou have co	mments regardin	g access to special	collections cont	aining health

↑13. How do	you identify yourself?
A student (any	field)
A professor/ins	tructor of history, the history of medicine, or the history of science
A professor/ins	tructor of another Humanities sub-discipline or a different discipline
A researcher (r	no academic affiliation)
A physician or	healthcare provider with an interest in the history of medicine or science
A librarian or a	rchivist
14. How long	have you identified as the above?
1-5 years	
6-10 year	
11-15 years	
Over 15 years	
○ № 16. If so, plea	se provide your name, affiliation, and email adresss:
0	se provide your name, affiliation, and email adresss:
16. If so, plea	