

RBMS REPODATA TALK (June 22, 2018)

- Context for our Work -- Eira (5 min)

[SLIDE of RBMS journal]

In the first issue of the Rare Books and Manuscripts Librarianship journal published in 1986, Columbia University School of Library Service professor Terry Belanger excerpted several reports his students wrote about the state of Special Collections Libraries across the country.¹ Belanger's students, who completed their reports between 1976 and 1985, frequently remarked on the poor physical state of many of the collections they visited - noting visible environmental damage due to bad HVAC systems, fragmented conservation practices, and strained labor conditions for staff.

When Belanger's students started their reports in 1976, scientists were establishing that CFCs, methane and ozone were major contributors to greenhouse gases, and that deforestation had major consequences for future climate scenarios.² *Science* magazine reported that scientists were beginning to "[express] growing concern to each other and to government officials", but that "In trying to convey their message to federal policymakers, however, those concerned with climate have had little apparent success."³ 1985, the year Belanger's students wrote the last reports quoted in his 1986 article, marked the last time that the global average temperature for the world was cooler than normal.⁴ [SLIDE of NOAA Global Temperature Report for April] Society has now notched over 400 consecutive months where "average global temperature ha[d] been warmer than normal".⁵

While many major research libraries have undoubtedly improved their situations since Belanger's student reports, we know that far more special collections libraries continue to struggle with ensuring appropriate staffing and resources for collection preservation. The fact is that while the voices of well-resourced institutions play a major role in our professional discussions, there are many small and community institutions that hold a wide array of archives and special collections materials that we rarely hear from or realize even exist.

[SLIDE of NOAA Global Time Series bar chart]

For as long as I have been alive, the average temperature of the planet has moved in one direction - upwards. So if we're headed to a future where the climate outside our building walls

¹ <https://rbml.acrl.org/index.php/rbml/issue/view/1>

² <https://history.aip.org/climate/timeline.htm>

³ <http://science.sciencemag.org/content/192/4236/246>

⁴ https://www.washingtonpost.com/news/capital-weather-gang/wp/2018/05/18/april-was-earths-400th-warmer-than-normal-month-in-a-row/?noredirect=on&utm_term=.251883cebcf4

⁵ https://www.washingtonpost.com/news/capital-weather-gang/wp/2018/05/18/april-was-earths-400th-warmer-than-normal-month-in-a-row/?noredirect=on&utm_term=.251883cebcf4

will continue challenging the holdings inside our buildings, and if many archives and special collections still don't have the level of resources and staffing they need, we need to radically rethink how we deal with concepts of sustainability, stewardship, and preparation within our institutions. The question is not if climate change will affect archives. The question is how, and to what extent.

And until fairly recently, no one could really answer that question.

[SLIDE of New York Times front page]

Tomorrow marks the 30-year anniversary of NASA scientist James Hansen testimony to Congress urging sharp reductions in fossil fuel use. We know that many of climate change's effects are, at this point, inevitable because a certain amount of harm is now inevitable due to atmospheric feedback loops that have resulted from political failures to reduce greenhouse gases over the last 3 decades. According to the National Climate Assessment, the United States will experience an increasing number of climate-change related trends that will influence residential patterns, agriculture, natural resources, and future investments in infrastructure. Many of these changes will be due to increasingly severe weather and rising sea levels that will pose significant dangers to most of the population in the country (USGCCRP, 2014).

Climate change is one of the greatest contemporary threats to archival repositories and the records in their custody. Increasingly severe disasters like hurricanes, floods, and wildfires pose immediate dangers. Archives preserve historical records in multiple formats that are critical for legal matters, administrative accountability, and documentary cultural heritage. When these records and documents are lost following extreme weather events, for example, their absence severely handicaps socio-cultural and economic reconstruction efforts. At best, archives affected by such events may be able to evacuate certain holdings, to move collections to safer parts of buildings, or to salvage materials using disaster-response teams. At worst, a disaster may result in total loss, with collections of records or even a repository's entire holdings damaged or lost beyond recovery. Longer-term trends such as human migration and rising sea level may necessitate decisions concerning the geographic relocation of archival records.

When I began working on the intersection of archives and climate change, one of the major issues I kept running up against was being able to contextualize risk to archives. It is one thing to be able to gesticulate wildly at coastal areas and say "your archives here are probably in danger!" But.... so is every other building in the neighborhood. If entire cities are threatened by rising seas and stormier weather, then what are the appropriate ways to measure risk to archives? I was frustrated by our collective professional inability to quantify this in a way to get up and make my fellow archivists pay attention.

So then I discovered another issue: if you ask "Where is every archive in the US?" no one has a good answer. The archives profession does not maintain a comprehensive data set of US repositories. For a profession whose mission is to organize and preserve the data for all aspects of society, we do a frighteningly terrible job about maintaining data about ourselves and the locations in which we do our work.

[SLIDE: Problem Statement]

For the past two years, Eira and I have been working to contextualize and quantify climate change risks to archival repositories. We began with a simple research question: “**which American archival repositories might suffer exposure to the future impacts of climate change?**”

Disaster is often the lens through which our profession views significant weather events. An example of this we kept coming back to in our research was the experience of Jeffrey Williams, associate director of the Health Sciences Library at NYU, whose library suffered considerable loss as a result of the 14 foot storm surge caused by Hurricane Sandy in 2012. He has repeatedly stated that what happened to them was not a failure of planning -- they had a disaster plan and even experience with hurricanes before -- so much as a failure of imagination. Think of our project as an effort to imagine more concretely the types of extreme weather events that could affect specific American archives locations.

But it's also important to remember that *disaster* is not the *only* threat to libraries and archives as climate changes. It's also a sustainability issue for our operations. As outside temperatures rise, for example, the costs of maintaining constant temperature and humidity will also increase. And as Eira has already reminded us, this is especially concerning for repositories with poor environmental controls or limited resources to begin with.

[SLIDE: Methods]

Our methods were to:

1. Map 1,232 archival repositories in the continental U.S. using locations data helpfully supplied by OCLC. We should note the mapping is in two dimensions only -- we don't know anything about the archive facilities or where within the facility resources are stored
2. Map current and future environmental disturbances using the best available data -- for our project this included disturbances related to catastrophic flooding, like surface runoff, storm surge, and sea level rise, as well as climate disturbances such as temperature and precipitation increase
3. Geographically intersect the two
4. Try to make sense of the results

[SLIDE: Data Quality and Future Projections]

This project would not have been feasible without the expertise of Nathan Piekielek, geospatial services librarian at Penn State, and Tara Mazurczyk, a Ph.D. candidate in the Department of Geography, who performed the actual data analysis. Throughout this process, we learned a lot about the quality and confidence behind climate models, and how different models exist to track

different future scenarios. I won't get into all the details here but the main thing we want to convey is that no data are perfect, but there's a lot of intentionality and care behind how we've used and interpreted the data.

[SLIDE: Exposure, Not Vulnerability]

Finally, we want to emphasize that this study was only designed to quantify the likelihood of *exposure*, which here refers to a change in environmental conditions. We cannot, from the available data, adequately assess or quantify an individual repository's ability to absorb or respond to those changes, which here we define as *vulnerability* -- that depends entirely on the unique characteristics of each repository -- its personnel, resources, facilities, and plans.

[SLIDE: Findings: Flooding Disturbances]

- 30 locations within a 500-year floodplain. Houston, you might know, has experienced three 500 year floods in the past three years
- 18 coastal locations that would likely be inundated by 1.8-meter (6 foot) sea level rise by 2100
- 84 coastal locations that would likely be inundated by a current category 4 storm surge event if it were a "direct hit" to their coastline
- 219 coastal locations were found to be at some risk of inundation from the combined effects of future sea-level rise and storm surge

[SLIDE: Findings: Climate Disturbances]

- 92 locations may experience a 10-degree or greater annual temperature change by 2100
- 93 locations that may receive an additional 10" of rain annually over current levels

[SLIDE: Overall Findings]

In aggregate, we found that nearly every repository mapped is exposed to some amount of climate disturbance, but the combined risks were quite low. We feel this is good news, overall, but even one exposure can significantly impact a library or archive, and as we saw with Hurricane Sandy, an institution's exposure cannot always be predicted, especially when past weather is no longer a reliable predictor for future weather.

[SLIDE: Conclusions]

We have many takeaways from this study, and our article offers some specific ideas for future professional effort, some of which Eira and I plan to explore further. But the main takeaway we'd want you all to receive from this is that we as a profession and we as professionals working in specific institutional contexts, can do more to understand and prepare for the likely risks. As Eira mentioned in her introduction, an essential first step is simply understanding where our

institutions even exist. The 1,232 repositories mapped in our study is a ***fractional*** representation of the documentary and print heritage in the U.S., and we believe it only represents those institutions with the resources to even participate in an aggregated discovery service like WorldCat, which means we haven't assessed the impacts to local historical societies and community archives and the many other types of the institutions most likely to be vulnerable to extreme weather.

Eira and I decided to try and address this gap, and last summer we received funding from the Society of American Archivists Foundation to find, gather and standardize archival repository locations data that exists in many pockets throughout our profession. Our research assistant on this project was Whitney Ray, who will now share with you some details on the work she did and what we've accomplished thus far.

[HAND OFF TO WHITNEY]

[Slide: Repository Data]

Overview

- Goals
- Methods
 - Data Gathering
 - Data Processing
- Findings
 - Limitations
 - The current set

I'll be discussing the goals, methods--including data gathering and processing--and the results of the project--including the limitations and current data set.

[Slide: Where's the Data?]

Archives don't have anything comparable to IMLS's Museum Universe Data File. As Ben discussed, the original dataset of 1,232 repositories provided by ArchiveGrid under-represented small and community archives

[Slide: Goals]

Our goals were to identify, gather, and disseminate data on the geographic locations of archival repositories in the United States. We also intended to categorize the data by location and type, and share the data in an open and reusable format.

[Slide: Methods: Data Gathering]

Initially, we focused on two lists: the directory of state archives on the Council of State Archivists website and regional archival organizations on Society of American Archivists' website. Coincidentally, each list had 51 items, so initially there were 102 organizations to contact. Often a contact at an organization would provide information about other groups in the area who they thought may maintain a directory or list of some sort. Using this snowball discovery technique, we reached out to many more organizations. We also, in turn, had several individuals hear about our work and reach out to provide us with data or suggest other groups.

[Slide: Methods: Data Gathering]

We contacted a total of 151 archival organizations, the majority of which were State Historical Records Advisory Boards (SHRABs), state archives, and regional groups. I'd like to acknowledge the tremendous work of archivists, volunteers, students, librarians, and memory workers to make archives visible and to protect cultural heritage. This project would not have been possible if not for that work that preceded us. Members of archival organizations were incredibly generous with their time and data. We learned much about the landscape of archival organizations. Some, such as Charleston Archives, Libraries, and Museums Council, were formed in the wake of natural disasters.

The bulk of this was collected between August and December of 2017. We experimented with other methods, including a Google Survey. There was eventual non-response from 10 organizations, out of the 151 data sources. We were also as consistent as possible in contacting organizations three times. This led to a rough estimate of 34,533 data points. This included data on all 50 states and DC.

[Slide: Methods: Data Processing]

This data was wildly variant in content, format, currency, and even whether the buildings identified were archives. We were reluctant to eliminate any data which might point to the existence of an archives.

The SAA defines an archives "any type of organization that holds documents, including business, institutional, and government archives, manuscript collections, libraries, museums, and historical societies, and in any form, including manuscripts, photographs, moving image and sound materials, and their electronic equivalents." In practice, we found that many organizations gathered information on cultural heritage institutions. For example, in North Carolina a now defunct project called NC ECHO, housed at the State Archives, included landmarks and schools which could feasibly have archives. Since the sheer volume of data presented challenges on contacting all the organizations listed, we went for being overly broad rather than possibly eliminating organizations that do have archives but do not describe themselves as such.

[Slide: Methods: Data Processing (Historical Society)]

Organizations maintain data for their own researchers' and internal use. Here's an example of a pretty common workflow challenge. This directory from the Delaware Valley Archivists Group came in as PDF. It contained repositories from multiple states: Pennsylvania, New Jersey, Delaware, and Maryland. For this repository, it contained a PO Box only. Initially, I made best effort to track down location of archives, often digging through individual repositories' websites. This historical society in particular had multiple locations, historical houses they managed, and a museum. I also ranked our sources on a scale of 1-5 for how "actionable" a source is; initially the process was state by state, beginning with the most easily processable data.

[Slide: Methods: Data Processing]

During processing, we sorted into spreadsheets by state, using Google sheets and coded by repository type and location type.

[Slide: Summary info]

In sum, we contacted or gathered information from 151 archival organizations, had initial data on over 30,000 repositories in the United States, and currently have 15,758 "clean" data points.

[Slide: Findings: Limitations]

Our limitations stem from the fact that we used extant data sources. For example, despite our best work, some personal addresses may have slipped through. Often, there are inexact addresses--PO Boxes, addresses that are intersections as opposed to a street point, and university and mail service center addresses. We know there are gaps, but we were working off existing sources to show to range of information out there on archival repositories. There's a wide range of specificity: some addresses included information about the location of the archives in the building.

Another issue to point out is that some organizations did not want their address made public. These included small and community archives without resources for staffing hours and corporate archives that are only for internal use.

Finally, I'd like to stress that this was an iterative process. We haven't gotten to all of the data. We have not processed multi-state organizations data sources; for example, the directory of corporate archives that is hosted by SAA, National Park Service's locations. We have 20 states to go, and some states have other less easy data than others to process.

[Slide: Findings: Current Set]

Have we gotten to the answer of where is every archives in the United States, as Eira posed? No, but we are a lot closer to it; it's never been attempted to this extent. In total, we now have 15,758 data points, compared to the initial 1,232. We have identified 2,608 in the Gulf States, compared to 104 in the initial ArchiveGrid data set.

[Slide: Findings: Gulf States]

Eira has made this story map for the cities of New Orleans, Houston, and Miami. If we zoom out, we can see that most of the Texas data is historical society repositories, most of Mississippi is public library repositories, and most of Alabama is government repositories. This of course informs us about the data we were able to collect, rather than the actual composition of repositories in these states. Eira will talk more about where you can find the results and why they matter.

[HAND OFF TO EIRA]

- Why our Work Matters/What's Next -- Eira (5 min)

[Slide on Advocacy and Planning]

So why does the work of gathering 30,000 data points for US archives matter? Because it is hard to say anything about the geography of archives without the data on where archives are and where archivists work. While the methodology we used to gather this data had little to do with environmental effects, I want to return to our original motivation for undertaking the repository data gathering that was inspired by the shortcomings of the first data set - the impacts of climate change. Climate change has enormously disruptive potential in threatening the vital records that ensure citizens' rights, as well as the documentary memory of local cultures and communities.

By spring time, we had enough data from our efforts for each coastal state and I began creating prototype maps to see how many geographically vulnerable frontline archives exist. I am by nature exceptionally skilled at catastrophizing, and I wanted to see what available "worst case scenario" data could show me about sea level rise and archives.

There are many open data tools out there depicting the effects of climate change. Unless you really know what you're doing, it's important to use these as high-level reference tools, and not as specific site-planning tools. Here is an example of what you can do with the data we've released: if you map the data we have for archives along the coast, you can begin comparing it to NOAA sea-level rise map layers.

The next several slides depict 6 feet of sea-level rise above the current average high water tidal levels. The orange dots represent archives from our Repository Data gathering project.

[Slides - southern Louisiana]

You can see that even if a specific archive isn't inundated, the region around it will look radically different. The dark blue areas are where the current water is - like the Gulf and Lake Pontchartrain. The light blue areas show where current land masses would be inundated.

[Slide of Prototypes list]

If you would like to explore some of these maps, we have made a few prototypes available.

While our work started with needing better data for our own climate change work, we know that what we've gathered is useful for the profession writ large. Disaster response networks may wish to have better information on archives in their region, researchers may wish to know about all archives in a particular state, and archivists themselves benefit from understanding the breadth of where archives are located across the country.

[Slide of "What's Next"]

We have always intended to make our data as publicly available as possible, and we strongly encourage our colleagues to reuse this data for any purpose they can think of. As of today, we have completed data for 30 states plus Washington DC. This data is available in a github repository - the URL is on this slide. As we finish more states in the coming months, they will be made available as well.

For now, all changes and updates to this data will remain the responsibility of Ben and I. We are hopeful that a professional association like Society of American Archivists will take responsibility for managing it and keeping it up to date, but so far this does not yet seem to be possible.

[Slide of solidarity sunflowers]

The reality is that this threat, although unevenly distributed, is something that potentially effects all of our holdings. To return to Bellanger's student, one wrote the following astute observation about a university collection that appeared to be in otherwise good shape at the time of the report:

"[A] collection such as this one cannot be judged safe simply because it is in relatively good shape today. It is meant to last as long as the University lasts, as a resource for a posterity which may well find new ways to extract knowledge from it we cannot now even begin to envisage. Its condition must, consequently, be judged from the perspective of a very long span of time indeed. And in this world, time passes faster than we realize. Threats of physical decay, of theft, of ignorant mutilation-these are unlikely to be visible at any one given moment, but they persist for the long haul. Given the basic condition, a decision to do nothing becomes a decision to leave the Collection vulnerable to such threats."