

Columbia
Journalism
School 

TOW CENTER
FOR DIGITAL
JOURNALISM

Public Record Under Threat: News and the Archive in the Age of Digital Distribution

April 2018

Nate Hill

With an introduction and conclusion by Sharon Ringel and Angela Woodall

Platforms and Publishers: Policy Exchange Forum IV

February 23, 2018 | Stanford University

Hosted by the Tow Center for Digital Journalism and the Brown Institute for

Media Innovation

Introduction: The Dark Age of Information?	2
Executive Summary	3
Discussion I: Emily Bell in Conversation with Jake Orlowitz and Mark Graham	3
Source Shift	3
Versioning and Totality	4
Discussion II: Digitization and Preservation of the News Archive	5
How to do it right	5
Magnetic Media Crisis	5
Software and the Web: also a Crisis	6
Discussion III: Who Keeps the Public Record? Access and Transparency	7
Public / Private	8
Open(ish) Government Data	8
Persistence of FOIA Data	8
Whose job is this, anyways?	9
Key Challenges and Opportunities	10
Conclusion	10

The Policy Exchange Forums are a critical component of the Tow Center's Platforms and Publishers research project. In these sessions, participants representing both the platforms and publishing sides of the news industry can engage on issues related to the ethical and civic values of journalism. The forum focuses on the relationships between technology, business, journalism, and ethics, and brings together diverse stakeholders to discuss current issues and surface potential new ones.

The project is underwritten by the John D. and Catherine T. MacArthur Foundation, with additional support by the John S. and James L. Knight Foundation, the Foundation to Promote Open Society, and The Abrams Foundation.

Introduction: The Dark Age of Information?

By Sharon Ringel and Angela Woodall, Research Fellows at the Tow Center for Digital Journalism

The evolution of the internet has created a vast storehouse of information, both current and historical, and all at the fingertips of the general public as well as journalists. But as we find ourselves apparently saturated by information and overwhelmed by its sources, we face a potential crisis of preservation as we seek—and often fail—to archive all manner of digital content.

News organizations have been slow to recognize and respond to the preservation challenges presented by digital technology. As a result, newsroom discussions about preservation and archiving are few and far between. However, professionals such as librarians, archivists and technologists outside of news industry are having these conversations, about retaining both conventional news content and online news, as well as about the problems inherent in trying to preserve the multitude of digital and data projects.

This shift from paper (or film) to digital record prompted the title of the conference that is the focus of this report: “Public Record Under Threat: News and the Archive in the Age of Digital Distribution.” The conference, the fourth and last in this phase of the Tow Center for Digital Journalism’s Platforms & Publishers series, arose from a recognition of the problems this shift has created in the archive..

While archivists and librarians generally agree that preservation of news content is an acute concern that must be addressed quickly, platforms and publishers often express doubt about whether they have anything to contribute to the discussion. Getting this crowd together in the same room—archivists, journalists and technologists—is a first step toward defining the problem, and subsequently mapping out solutions.

A note on formatting: This report, written by Nate Hill (Metropolitan New York Library Council), was produced for the Tow Digital Center for Journalism Policy Exchange Forum (PEF) series. The subject of the April 13, 2018 PEF forum will be expanded upon by the Tow Center’s forthcoming study of news archiving practices, led by Sharon Ringel and Angela Woodall.

The four-hour conference was open to the public and divided into four sessions: Tow Center Director Emily Bell led off the first section with Mark Graham (The Internet Archive) and Jake Orlowitz (Wikimedia Foundation). Their talk was followed by a conversation, “Digitization and Preservation: The News Archive,” between the moderator Katherine Boss (NYU Journalism Library) and Stephen Abrams (California Digital Library), Michael Corey (Reveal), Anu Paul (Digital Archivist - Consultant) and Evan Sandhaus (The Atlantic). Next was a presentation by Brown Institute Magic Grant fellow Francesco Fiondella, “Data Interrupted.” The third session, “Who

Keeps the Public Record? Access and Transparency,” was moderated by Nate Hill (Metropolitan New York Library Council) and included Victoria Baranetsky (Reveal), Karen Cariani (WGBH), Alex Howard (Sunlight Foundation), Regina Roberts (Stanford University Library) and Geoffrey Samek (YouTube).

Executive Summary

This event featured three panels of journalists, technologists, librarians, and engineers who discussed how to preserve the first draft of history in an era of newsroom cutbacks, ephemeral social media feeds, and disappearing data. The conversation addressed the importance of the public record, the ways reporters currently use archives, and what practices, policies, and strategies might affect access to archived news in the future. Among the key findings:

- Traditional news organizations have been displaced from their historical role as primary sources of reliable firsthand information when news breaks; the open web and social media platforms have made everyone with a device into a potentially credible source.
- As a medium, the internet is theoretically capable of supporting versioning and near-total coverage of any news event. This creates new opportunities and challenges for preserving, accessing, or displaying many stories. New organizations that are “of the web” like the Internet Archive and Wikipedia are working to address this. However, even here these organizations are aware that they do not capture everything.
- Traditional newsrooms rarely see preservation and allowing users access to their archive as a business priority. Organizations that have successfully created a public-facing archive often chose to focus on user experience and integration of archival material into the rest of the business ecosystem.
- Some older forms of information storage, particularly magnetic media, become unstable with time. There is not enough funding or labor available to transfer all of this material into more stable and consistently maintained storage environments.
- Data visualizations, interactive graphics, and stories that depend on APIs (Application Programming Interfaces) or proprietary platforms or applications also come with significant preservation challenges. Some solutions and strategies are emerging, but the challenges currently outpace the solutions.
- Government data, which is often critical to reporting, is not always as freely available as it should be. Privatization of public data collection and storage, FOIA exemptions, local use of the “Glomar” denial (in which an agency subject to a FOIA request says it will “neither confirm nor deny” the existence of records sought), and the slow pace of technological

change at government agencies all stand in the way of timely, lawful response to public inquiry into public data. It can be unclear what organization is most clearly responsible or capable of preserving public data.

- War, conflict, or unrest affect the collection of data. Gaps in data collection are irreversible; they affect our ability to inspect, analyze, or calculate a response to an event.
- Machine learning algorithms are one tool used to prevent vandalism or manipulation. The algorithms may flag some news content that is inappropriate, inaccurate, or violates the terms of service of a web platform, but regulating and moderating contributions is difficult and ultimately can require human intervention.

Discussion I: Emily Bell in Conversation with Jake Orlowitz and Mark Graham

Mark Graham is the lead for the Internet Archive's [Wayback Machine](#), and Jake Orlowitz manages the [Wikipedia Library](#). Their conversation with Emily Bell, founding director of the [Tow Center for Digital Journalism](#), set the stage for the forum.

The Wayback Machine is a massive archive that backs up as much of the public web as possible, currently at 327 billion pages and counting. The Internet Archive offers a variety of digitization, archiving, and library services. The Wikipedia Library is one project in the greater portfolio of Wikimedia Projects and makes subscription databases available to Wikipedia editors, which allows them to cite trusted, reliable sources in their articles.

Source Shift

In the past, newspapers depended on “clipping files,” an archival solution that feels practically impossible now. Articles were literally cut out of newspaper pages and filed away in folders, arranged in such a way that with some luck a researcher or writer might be able to retrieve them. Now, web archives like the Wayback Machine have largely replaced these practices, in part because publishers themselves are doing far less archiving. Graham largely attributed this shift to practical concerns: shrinking budgets, shifting priorities, and technological challenges like content management system (CMS) migrations, which often cause problems like dead links. It simply isn't a financial priority to preserve old material and make it useful when news is constantly breaking. So the Wayback machine currently archives 1.5 billion URLs per week, 150 million of which are news-related urls. The goal is to archive *all* news, but as Graham said, “we are doing an OK job.”

When archivists seek to preserve the news in a holistic manner on the web, it becomes clear that traditional news sources are only a small piece of the puzzle. For example, Bell asked Orlowitz if

Wikipedia would consider Twitter a news source. Preservation becomes problematic when politicians' Facebook pages are where they contextualize breaking news stories. Orlowitz said that Wikipedia is an archive in and of itself, since every edit and change is recorded in each page's history; Wikipedia is versioned.

Orlowitz went on to speak about the importance of verifiability in his platform's work, meaning that if an editor makes a claim, someone else should be able to check it and see that it is true. This is a guiding principle for Wikipedia, so with Twitter as an example, they have to question whether the source of the tweet is the right source to make the claim. If our president tweets about history, he might not be a reliable source, but our president's tweets are a great source for what he said. This makes social media a primary source, rather than secondary or tertiary source.

Versioning and Totality

Versioning as a means of achieving transparency and reliability was a theme that emerged in Graham and Orlowitz's conversation, and it was revisited throughout the day. Both Wayback and Wikipedia "version" the web in a manner analogous to the way that [Github](#) or [Subversion](#) manage version control for software developers working with code repositories. This, they say, makes their product especially reliable.

Wayback and Wikipedia are unique because they attempt to address the totality of platforms and press on the web, not just individual sources. This approach provides multiple ways to understand an event; it allows us to see it and analyze it from multiple perspectives.

The later panel presentations and resulting discussions largely focused on the challenges endemic to archiving dynamic news content, a subject that can be counted on to exasperate journalists and archivists alike. Orlowitz suggested that a more transparent versioning process that makes the flow of time visible would be useful. In his example, a slider bar underneath a news story that has been revised and published in multiple versions would allow a reader to view and understand the changes made along the way. Similarly, as was illustrated in a later presentation by Francesco Fiondella, an interactive map can show the spread of an idea, perspective, or disruption across both geography and time.

Discussion II: Digitization and Preservation of the News Archive

The first panel featured a group of practitioners who outlined the challenges and the opportunities associated with archiving news, each of the speakers showcasing their own practices with a selection of archival content. Together, they addressed much of the history of mass media distribution formats, and those formats' associated playback and display technologies, as well as the associated legacy broadcast and transmission strategies. Among participants, there was unanimous agreement about the threats posed to the collective news archive by a lack of funding, an inadequate workforce, and the unwillingness of institutions to engage in long term preservation or discovery strategies for their news archives.

How to do it right

Evan Sandhaus presented a success story: [TimesMachine](#), a project at the New York Times that makes nearly their entire run of publications from 1851-2002 available to their digital subscribers. At its inception, the project had to unify the presentation of three different periods of the Times archives: Papers from 1851 to 1960 had been scanned but suffered from mediocre OCR (Optical Character Recognition, the process scanning software uses to identify the image of a letter or numeral and “read” it as text); papers from 1960 to 1980 had been more accurately transcribed and after 1980, transcription was rarely a problem, but formatting was far more varied and tricky than it had been in prior eras.

Sandhaus told the gathering that he believed three key ingredients were necessary in any recipe for a successful large scale digital archive. The first and great overarching priority should be a laserlike focus on user experience. The second, and linked, priority ought to be the creation of useful metadata in order to make material discoverable; echoing Jennifer Schaffner's 2009 OCLC Research paper, "[The metadata is the interface](#)". Third and finally, Sandhaus spoke about curation. His team wove archival content into the entire information and business ecosystem of the New York Times; rather than allowing Times Machine to sit like a remote vacation resort island.

Magnetic Media Crisis

Anu Paul, a Digital Archivist, spoke about the challenges of the high costs associated with digitization and preservation. Paul shared details about her project digitizing the popular, long-running public radio broadcast Fresh Air with Terry Gross. WHYY, the Philadelphia-based radio station where Fresh Air began in 1976, received support from The Council on Library and Information Resources (CLIR) to digitize the show in its entirety and to make the [archive](#) accessible via WorldCat. The archival challenge Paul and her colleagues faced is perhaps most powerfully summed up in a quote from Gross herself in a [brief article](#) promoting the project: “I

remember the days when I'd record, on cassette, the interviews that I wanted to keep for myself and store them in shoe boxes..”

Paul began the discussion with an overview of what he characterized as a magnetic media crisis, a theme that Karen Cariani of WGBH-Boston revisited in the second panel. A [report](#) from CLIR states that “According to manufacturer's data sheets and other technical literature, thirty years appears to be the upper [age] limit for magnetic tape products, including video and audio tapes.... Recently, articles have been appearing which suggest that the life expectancy of magnetic media is much shorter than originally thought.”

There are projects and organizations working to address this challenge. The IMLS-funded [Memory Lab Network](#), a project led by the [Memory lab at the DC Public Library](#), is an effort to bring community-based media migration facilities, technology, and training to seven public libraries nationwide. In addition, the [XFR Collective](#) is “a non-profit organization that partners with artists, activists, individuals, and groups to lower the barriers to preserving at-risk audiovisual media—especially unseen, unheard, or marginalized works—by providing low-cost digitization services and fostering a community of support for archiving and access through education, research, and cultural engagement.”

Software and the Web: also a Crisis

Stephen Abrams, Associate Director at the [California Digital Library](#), added to the list of technical hurdles: He steered the conversation toward [data degradation](#). Abrams also reiterated that we have a resourcing challenge, both human and financial, and while technical challenges exist, they are not insurmountable so long as someone is willing to pay to address them. Abrams, along with Michael Corey from Reveal, laid out another spread of archival challenges related to software and the web.

The open internet facilitates massive, incessant, instantaneous, networked publication of all media worldwide, it can make anyone connected into an author-publisher, and it gives every author-publisher a global audience of other author-publishers. This means that the scheduled, punctuated production of daily print newspapers or the 6:00 evening news on television has been washed away in a continuous, flowing river of news that is produced and distributed as easily by a clever eight-year-old as it is by the New York Times. How can we document, segment, interpret, preserve, and contextualize the river in perpetuity? Right now, Abrams said, the best we do is take snapshots of the river.

While the abundance of material may make archiving the entirety of the open web daunting, the good news is that at least we can access that material. The next problem Abrams mentioned is that we have no way to archive and preserve material featured on proprietary apps. As Mark Graham mentioned earlier, there is also no way to archive content hidden behind a paywall. And as Emily Bell said earlier in the day, when a politician contextualizes current events on their own

Facebook wall, this becomes important news material. What can we do when the politician deletes it 10 minutes after they published it?

Further complicating the problem of contextualization, many popular platforms utilize complex algorithms to mass-customize results for their users, meaning that no two individuals see the same news even when they use the same platform. Algorithmic bias and individually targeted distribution of content complicate the question of what news should be archived. Is there some canonical version of the news to preserve? Can we preserve all perspectives?

Michael Corey is a Senior Data Editor at [Reveal](#), who addressed problems with software dependencies and legacy APIs. The web works like an operating system that can power many rich applications, with data and services shared across the network, all of which are supported by a variety of businesses, governments, and nonprofit entities. Many of the interactive, data-driven web stories readers find in online news today depend on third-party services like mapping APIs, javascript libraries, and a variety of ways of supporting animations. Even the CMS migrations that are so common at organizations of all types result in degradation and alteration of stories. While Corey carefully documents and pursues solutions to the necessary software dependencies to keep up his stories, there is a great deal of labor and foresight involved. He gave an example about the history of the butter cow sculptures displayed at the Iowa State Fair for which he had created a micro-site for the Des Moines Register in the late 1990s. After he left that job and the paper went through a series of CMS migrations, the story that he had worked so hard to research and display had degraded to the point that much of the information was lost.

Despite the many challenges, the panel still had some optimism. Just as there are organizations and individuals fighting the magnetic media crisis, there are also people, projects, and technologies focused on software preservation and web archiving, among other issues. [The Software Preservation Network](#), with support from IMLS, the Andrew Mellon Foundation, and the Alfred P Sloan Foundation has a project underway supporting software preservation in libraries and archives. New strategies like [Emulation as a Service](#), and container technologies like [Docker](#) suggest new ways for digital archivists to approach this work. Additionally, projects like Rhizome's [Webrecorder](#) and the Internet Archive's [Archive-It](#) empower more individuals than ever to archive the web. Both projects make it possible for more individuals to participate in web archiving; Archive-It is a subscription service that helps organizations collect, catalog, and manage their collections of archived content, while Webrecorder is a free service that anyone can use to capture and save web pages.

Presentation: Data, Interrupted

Francesco Fiondella, along with Catherine Vaughan and Amir Imani, were recipients of a Brown Institute [Magic Grant](#) for a project called [Data Interrupted](#). Fiondella comes from the [International Research Center for Climate and Society](#), an organization founded by NOAA and

Columbia University in 1996 to apply climate science in the service of society. *Data Interrupted* is a unique view of the tragic Rwandan genocide through the lens of data collection.

Fiondella presented a map of all of the weather stations collecting data in Rwanda. He showed that as the civil war and genocide progressed, active weather stations disappeared from the map one by one until finally there were none. While this showed the loss of one kind of data, Fiondella told us that this weather data could act as a proxy for nearly any other kind of public data collection. *Data, Interrupted* is not a project about weather data alone, it shows the impact of conflict on a country's ability to collect any kind of data.

Discussion III: Who Keeps the Public Record? Access and Transparency

The last panel added the discussion of government records and private platforms to the already-complex set of preservation and access problems surfaced in the first panel. Local, state, and federal government authorities collect and produce a tremendous amount of data. This data is about the public, so it is for the public and ultimately belongs to them. Despite this, there are quite a few barriers to access in place, with new ones emerging all the time. At the center of this discussion was a tension between the government as both the source and keeper of the public record versus the new private platforms which host a great deal of discourse. One panelist summed this up by asking “should we demand the same kind of transparency and accountability of private platforms that we expect from our government?”

Public / Private

Alex Howard, Deputy Director at the Sunlight Foundation, spoke of the trouble with public discourse happening on private platforms. Sunlight Labs used to build tools that made private platforms more transparent by making the content more durable. For example, [Politwoops](#) is an interface they handed off to ProPublica that displays and archives deleted tweets from public officials. Other tools, such as those made available through [Documenting the Now](#), are meant to have similar impact. Increasingly we find ourselves relying on interventionist tools built by activists, journalists, academics, and librarians to extract important information from these platforms, sometimes via APIs, but often with scrapers and other, kludgier means.

Open(ish) Government Data

On the government side, a related problem that both Howard and Victoria Baranetsky addressed is the notion of “opacity through obscurity.” Because public records are poorly described, undiscoverable, and housed on dated platforms (see: filing cabinets, PDFs), there is a significant barrier to access. Organizations like [Code for America](#) and projects like [Endangered Data Week](#) advocate for reform and build interventionist tools, but an essential problem with applying the

same philosophy used to extract data from private corporations to the government is simply that the government is *not* Silicon Valley. Government is designed to persist; slow change is meant to be a system feature, not a bug. Yet this “feature” has enabled opacity through obscurity. Baranetsky reminded the audience that for a [FOIA](#) request, an agency is only bound to do a “reasonable search” for the requested information. The FBI uses the same search tools now that they used in the 90s. If government technology is embarrassingly, dangerously outdated, what does a “reasonable search” even mean?

Building on this, Baranetsky went on to talk about the use of the [Glomar response](#) in state and local settings, specifically as it was used by the NYPD to “neither confirm nor deny” a counterterrorism investigation that was challenged by journalists as a racial profiling incident. This response was upheld in the New York Court of Appeals on March 29, 2018. The Glomar response itself originated in the 1970s, when the CIA said they could “neither confirm nor deny” an attempt at retrieving information from a sunken Soviet submarine via a ship called the USS Glomar. The CIA argued that simply acknowledging such a project poses a threat to privacy and national security. Baranetsky argued that this was not meant to be used in state and local settings, that [FOIA](#) is a federal law and states have their own freedom of information laws, for example [FOIL](#) in New York State. By extending this as a legitimate local response, government transparency and accountability faces further challenges.

Persistence of FOIA Data

Regina Roberts is a Collection Development Librarian at Stanford University working on the [Open Policing Project](#). At Stanford, the University researchers have been collecting data about traffic stops across the country. The researchers do the FOIA requests, and then they clean the data and use algorithms to harmonize data across the state. The repository at Stanford becomes a refuge for the data. This project forces archivists to address important questions: Who should have access to this traffic data, and what exactly are the privacy concerns? What is ethical use and reuse of this data?

These questions go beyond the Open Policing Project. Journalists making FOIA requests are typically doing so in order to gather important data that supports a story they are working on. Once the story breaks, what becomes of the data they requested? In many cases, there isn’t an answer to that question. Once the story has broken, it is possible that the data may end up on Github, but there is no industry-wide process for preserving either the raw data or the analyzed, cleaned, and modernized data. This loss of both data and labor is problematic—a concern central to forum’s second discussion, as well—in addition to data preservation, the losses affect the promise of the versioned web, and the opportunity to analyze across platforms.,

Whose job is this, anyway?

Karen Cariani from the [WGBH Archives](#) and Jeff Samek of [YouTube News](#) each spoke a bit about their role in news preservation and dissemination. They both clearly stated that they feel responsible for preserving the news archive, and tried to address what their responsibility might be given their unique projects. Cariani oversees OpenVault, a collection of video, audio, images, and searchable transcripts produced by the public television and radio station [WGBH](#), all of which are available for individual and classroom learning. Samek is the Product Manager for the Youtube News Channel, a “news destination featuring comprehensive up-to-date coverage on the latest top stories, sports, business, entertainment, politics, and more.”

On April 1, Ethan Zuckerman had written a piece on Medium called [Four Problems for Media and Democracy](#). In the article, he wrote “News may be too important to leave to the whims of the market.” For Cariani, this rang true, given her work in public media. Samek told the audience that optimizing for clicks to keep people glued to their platform is an outdated practice; that YouTube’s recommendation and prioritization algorithms are tuned to encourage longer-term, healthier relationships between the platform and the users for the good of both the users and society. While neither believed they were in a position to present a solution to “financing the news,” as Zuckerman put it, both individuals said their organizations remain committed to presenting news as well and as accurately as possible, choosing the most pertinent stories to tell, and doing all of this for the health, betterment, and ongoing engagement with their user base.

There was general consensus throughout the day that the responsibility to preserve the public record is and always has been distributed; earlier Howard spoke of the way that monks served as a government record retention unit centuries ago—a practice that both preserved much of human knowledge across a period of political turmoil and also had its own set of complex political ramifications in work that was altered or left out of the canon. Repeatedly throughout the day there were calls for more transparency from private platforms. The panel received a tough question that Samek answered about using automated processes to remove violent videos, and the occasional errors or oversights that are made in the process. Samek admitted that the problem was a difficult one and said that YouTube takes it seriously. In reference to distinguishing between video about Syrian refugees of ISIS recruitment videos, he said “the disambiguation there is very hard for humans, essentially impossible for machine learning.” Samek said AI was only a piece of a solution also involving human intervention. The exchange recalled Orlowitz’s explanation of how Wikipedia distinguishes quality article edits from vandalism by first using layers of natural language processing algorithms and then finally human editors.

Key Challenges and Opportunities

- The responsibility to preserve the first draft of history is distributed between government, private companies, community organizations, individuals: everyone. This work is not “someone else’s job.”
- AI, machine learning, and automated processes of any sort are imperfect, prone to bias, and require human intervention. In other words: robots will not be solving this problem for us.
- News on the web gives us new opportunities to examine it in totality, across platforms, and from multiple perspectives. While this may feel “[Too Big to Know](#),” it creates new opportunities to analyze human knowledge at scale.
- Government is not well equipped to manage the amount of information it collects, and privatizing information management reduces or eliminates transparency and creates greater potential for abuse.
- Accelerated technological change and progress is creating preservation problems faster than we are solving them, and preservation is rarely if ever a business priority or a robust, taxpayer-supported priority.

Conclusion

By Sharon Ringel and Angela Woodall, Research Fellows at the Tow Center for Digital Journalism

Asking questions about disappearing records can open up other, larger conceptual questions about the preservation of material, whether that is what we traditionally think of as journalism, or multimedia projects and social media: What should be preserved, who should preserve it, and how to preserve it.

Such deliberations can be summed up in these three questions surfaced during the conference:

1. What information is worthy of preservation? In the analog era, the choice was obvious. Newsrooms kept the final version of the printed paper. Today, the choice is not so clear. In fact, it’s a major challenge that faces newsrooms and other institutions every day. When digital information is ephemeral, and at the same time we are surrounded by an unprecedented amount of information, what should we keep and what version?
2. Who is responsible for the preservation of information? When the president can delete his tweets, who is responsible for the preservation of social media, or even who should hold him accountable? Journalism has claimed the role of accountability but has not yet accepted the responsibility for preservation. And platforms are motivated to collect data because of the potential for profit, rather than as a public service.
3. The last question is how can we preserve this information? Not everything should be saved, and archives have always been incomplete and selective. However, the issue cannot be so easily dismissed, as the assembly of scholars and specialists made clear during the conference. Each in their own way grappled with challenges to preservation including algorithms that produce personalized online experiences, sites that are

migrated to new content systems, and rotten links. What happens to Flash-based apps, data visualizations, and virtual-reality projects? Even if they are archived, they will not function.

If there is one primary takeaway from the conference, is it is that these challenges may seem like technical puzzles, but they are also a matter of priority. Technology may prompt specific technological responses, but the way we answer is cultural. According to Stephen Abrams of the California Digital Library—which houses, among other collections, the Online Archive of California—some problems are easy, but either too low on the to-do list, or are considered expensive. Budgets to pay for preservation of records rise with volume. “The cost of storage has not gone to zero,” he told the audience during the conference. To muster the necessary funding, organizations have to first see preservation as a priority.

The work of changing their perspectives requires advocacy, strategic coalition building, and patient diplomacy.