AMELIA ACKER

How "Archive" Became a Verb

Debates over a proposed national data center in the 1960s reveal how anxieties around privacy and digitization led to the shift toward corporate control of personal data.

Each morning, like so many other people, I start writing by firing up my word processor. Recently, the executable program that I use on my laptop became an online application. It is now just one app in a suite of software tools that my institution licenses. When I shut down my computer for the day, I instinctively go to save my file—but there is no longer a "save" option. The app instead has a message in its menu that reads "Where's the Save button? There's no Save button because we're automatically saving your document."

The banality of this dialog response always intrigues me because it elicits questions that the drop-down menu has left conspicuously unanswered. Why is the document saved automatically? Where is it saved? How is it stored? Who made the decision that saving and accessing information automatically on platforms that are managed in the cloud rather than as a file on my laptop's hard drive is a good idea?

The answers to these questions lie in the evolution of *archiving* from storing files to managing digital information. Today's archives do not belong to the person who creates them. Instead, they are "born networked," or created within connected information infrastructures, because corporate data collection is now seamlessly integrated into the networked technologies that we use most. Where our data was formerly managed and controlled by public institutions, such as government agencies, libraries, or universities, today it is the domain of private platforms. This transition from public institutions to corporate firms explains how the noun *archive*—which once meant a place or set of records—became a verb denoting data handling, collection, and accumulation, infused with the power to control what information we can save and the means to access it.

Cell phones, apps, and social networking platforms all depend on information management systems that archive data and manage access to data collections. Despite their ubiquity, these data archives are increasingly hard for people to find, know, or even envision. Born networked data is distanced from users, and access points are becoming more distributed and invisible. Much like the automated processes that have replaced the "save" button for my word processing software, the means of accessing data that we create are now largely unknown to end users.

The beginnings of this shift in power can be found in the transition from manual data transcription using punch cards to machine-readable data and automated recordkeeping systems. One story in particular sheds light on some of the crucial issues involved in this transition. In the 1960s, experts proposed a national center for government information, generating contentious public and congressional debate before ultimately failing. This pivotal moment in twentieth-century data management was when *archiving* became a verb—and when access to collected data became a concern for scientists, bureaucrats, and the American public. At stake was whether personal information gathered from citizens by the state could be processed and transformed into aggregated data resources for access without violating Americans' privacy rights.

"The privacy-invading technology that would surpass all others"

During the postwar period, more personal information was being collected about Americans by institutions than ever before, including school records, military service records, and personnel files. As computers became faster and more efficient at processing data at scale, recordkeeping migrated to massive sets of punch cards processed by machines. Punch card computing opened new possibilities for any sector that relied on collecting and aggregating information for operations, such as scientific inquiry, government administration, and business management.

As more people started to learn about computers in industrial and organizational applications, the power of personal information, once aggregated into machine-readable data, began to loom large in the consciousness of the American public. By the early 1960s, the Internal Revenue Service, the FBI, the military, and colleges throughout the country used mechanized automated recordkeeping systems, giving data aggregation the patina of government secrecy and administrative control. So as recordkeeping systems became computerized, the privacy of citizens' information became political. For many Americans, the new capabilities of computers and public descriptions of them conjured secret dossiers and fears of Big Brother.

Social scientists and other researchers saw that computational processing had the potential to transform methods of inquiry in ways that could help society understand itself. The advantages of a centralized national repository for data were clear to these scientists: Access to more data would be more efficient and valuable. It would be available for more people and from more sources, instead of being spread across a patchwork quilt of separate data banks. But before a centralized repository of government data could be developed, the case for a centralized data archive in support of governance and administration needed to be made by bureaucrats to administrative officials and elected representatives in Congress.

A series of exploratory case studies were conducted throughout the 1960s to research centralizing access to government data. In response, social and behavioral researchers—both within and outside the federal government—proposed what came to be known as the National Data Center. The proposal prompted several congressional hearings in the House and Senate throughout 1966. Led by Congressman Cornelius Gallagher and Senator Edward V. Long, respectively, the hearings addressed the possible invasion of privacy that would result from a data center using computer technology and automated recordkeeping to manage data gathered from the public. According to privacy scholar Priscilla Regan, "Congress's first discussions concerning computerized record systems cast the issue in terms of the idea that individual privacy was threatened by technological change." But, as the hearings continued and critiques in the press began to circulate, concerns shifted from focusing on the potential impacts of new computing technology on data processing to the sheer volume of information being collected about individuals—some three billion records, according to a Senate subcommittee report.

By the end of the year, the congressional inquiries exploded into a full-blown controversy, and as one observer wrote in 1967, the plan for the National Data Center "acquired the image of a design to establish a gargantuan centralized national data center calculated to bring Orwell's *1984* at least as close as 1970." These fears about files with personal information being aggregated into dossiers and made accessible through computers would shape data protections in the United States for decades to come.

> Cell phones, apps, and social networking platforms all depend on information management systems that archive data and manage access to data collections.

In government we trust?

On one side of the National Data Center debate were boosters of a centralized national archive for data that prioritized access and standard processes. For years, government economists and statisticians, academic researchers, librarians, data professionals, and federal bureaucrats pushed for centralized access to reliable, authentic, and prepared statistical data for scientific study and analysis. Access to federal data at scale could alleviate society's greatest problems by measuring needs and assessing ongoing government services. In proposing a national center, the boosters believed that adequate privacy safeguards already existed; government agencies that aggregated data ensured confidentiality when collecting information from individuals.

On the other side of the debate, media pundits, legal scholars, and congressional representatives raised the alarm about coordinated access efforts for fear of government overreach. They predicted the abuse of networked dossiers of every American citizen, using data collected from cradle to grave, that would bankrupt American citizens' privacy rights. As claims about the capabilities of automated recordkeeping and computers circulated, criticisms of the national data bank plan began to appear in newspapers and magazines warning of invasions of privacy.

Proposals for the National Data Center did not consider the abuse of individual privacy in their recommendations. Despite the significance of planning for data access, boosters rarely raised limitations of who would be given access to the National Data Center and under what terms. Once data was acquired and aggregated into a data bank, proponents had assumed that ensuring confidentiality would not be a concern. But critics of the center assumed the opposite—that privacy would be placed at serious risk by aggregating personal information into population-level datasets, easily read by machines and accessible to all-knowing computers.

What the proposals did not say about privacy and the potential for abuse reveals the National Data Center proponents' trust in government recordkeeping processes. When scientists and research data managers were imagining what a social science data archive would be, they spent more time describing the processes of preparing and distributing data for analysis than how the data was collected in the first place, and the impact data banks would have when used in research. Conversely, the public controversy over the proposals and criticism from Congress spread awareness of the power in population-level data and the potential for abuse.

Looking back, two versions of the problem of how to protect citizens' privacy using computing technologies are now clear. Scientists trusted that government-assembled data would preserve confidentiality at its point of collection, whereas the public distrusted how data, once centralized and stored in machine-readable formats, could be recombined, programmed, and manipulated.

In reality, the privacy and confidentiality that the lawmakers demanded in the hearings had been effectively ensured by mechanized punch card tabulation, which was incredibly time intensive. The transition to data stored on magnetic tape reels potentially made abuse more likely by cutting down on processing time, which made aggregated records easier to search, merge, and match. The only way to guard against such abuse would be through oversight. Ironically, such oversight couldn't be ensured without centralizing data sources across the government into a data bank.

Ultimately the alarmists won, ensuring privacy legislation and laws that limited the federal government's ability to aggregate computer records. However, public fears about data integration and the collection of personal information would eventually come true—not through the federal government, but through universities and corporations that were collecting digital data in enclosed private contexts, and which would eventually create a market for personal data collection by third parties to thrive.

Concerns shifted from focusing on the potential impacts of new computing technology on data processing to the sheer volume of information being collected about individuals.

The realities of automated recordkeeping

Public concerns and expert testimony in the congressional hearings about networked computing and the potential for data abuse were perhaps overwrought. Orwellian visions from critics that predicted "*1984* by 1970" were highly implausible. Numerous studies of government agencies' data operations commissioned around this time found that most did not have comprehensive information about data sources, storage locations, or formats. Most of these congressional hearings and subcommittee reports—including reports from federal agencies and the National Academy of Sciences—argued that the invasion of privacy and abuse of data banks were not inherent in the transition to computerized systems and automated recordkeeping.

Still, many legal critics' predictions, as well as the public's concerns, turned out to be sounder than social scientists and data managers of the time believed. For decades, stakeholders inside and outside government lobbied for more information about the government's recordkeeping practices and data archives. Throughout the reports and rebuttals, experts and data specialists struggled to locate where the risks of accessing data in aggregate actually resided, and where safeguards should be placed in the data lifecycle to preserve privacy.

Many bureaucrats and public servants pushed back on the characterization of government overreach because existing laws already prohibited the misuse of personnel records and individuals' information. They pointed out that collecting data from citizens was necessary for the essential work of government, as well as for statistical analysis, economic forecasting, and enforcing laws that relied on aggregated personal information from states and federal agencies. But legislation mandating population statistics and data collection from citizens, like many other laws governing information and communication technologies, could not anticipate what digital data and networking technologies would soon bring.

In his 1971 treatise *The Assault on Privacy*, which gave an account of the National Data Center's failure, legal scholar Arthur Miller wrote that "the apparent victory against the National Data Center by the defenders of privacy is largely a Pyrrhic one." Miller had been a staunch critic of the center, publishing editorials during the hearings calling for a more robust legal framework to protect citizens from data mishandling. He was concerned about the ways that privacy was threatened by the transition to computer-driven recordkeeping—not only in government, but in industry as well.

Miller's point was as significant then as it is now. During the 1966 congressional hearings, there was very little scrutiny of organizations outside of government that were collecting population-level data, such as insurance companies, credit bureaus, polling firms, and advertising agencies. By failing to create a National Data Center, the government opened space for corporations to aggregate data for their own use and profit. Preventing a nationalized data center enabled a new market for privatized data banks and data integration solutions to take hold and thrive.

The computing advances of the 1970s—including automated processing and time-sharing with minicomputers and the rise of the database and data interchange—catalyzed the networking technologies and personal computing devices that became the foundation of today's wireless internet. At each of these junctures, there were people concerned about who would own data once it was accumulated, how it could be leveraged over time, and how such evidence could be both made accessible and controlled with data management administered by third-party firms. Gradually, as archiving became a verb, these data management techniques transferred active personal information management from individuals to machine-driven, automated processes.

Many legal critics' predictions, as well as the public's concerns, turned out to be sounder than social scientists and data managers of the time believed.

The data privacy fears of alarmists turned out to be correct. But it is now primarily private corporations instead of public institutions that collect and assemble data archives and act as gatekeepers to access. Contemporary data archiving in corporate contexts is a continuation of the historical processes of data dispossession, distancing creators from their data and implementing automated processes of managing it without their awareness. As Americans have ceded the ability to manage their data in exchange for using information and communication technologies, the implications of this shift become increasingly obscured, leaving corporations with unprecedented power over personal data and the information that society creates today.

Amelia Acker *is an associate professor in the School of Communication and Information at Rutgers University. She is the author of* Archiving Machines: From Punch Cards to Platforms *(The MIT Press, 2025), from which this essay is adapted.*