



# *The Collection and the Cloud*

By AMELIA ABREU MARCH 9, 2015



## *In the future, will platforms own our pasts?*

In February 2015, computer scientist Vint Cerf, known widely for developing the TCP/IP internet protocol standard, gave a lecture at Carnegie Mellon University's Silicon Valley branch campus in which he spoke of a coming "digital dark age." This sound bite was subsequently picked up by the BBC and made the rounds on wire services, and was forwarded and tweeted in digital-preservation circles. Cerf, who holds the title of "chief Internet evangelist" at Google, warned about the incipient reality of lost and incomplete digital archives. "Figuring out how to hang on to things is important," Cerf argued earnestly. "We don't have a standard way of hanging on to the software as well as the disks."

As an archivist, what shook me so deeply in Cerf's comments was his disconnect with the political economies of internet content production, or content

production in history. The concept of a “dark age” implies that there is a singular body of canonical knowledge worth preserving, evoking an idea of erasure from history that’s familiar to anyone else besides a distinguished white man accustomed to historical centrality. For everyone else, erasure from history is political struggle. Even today, women, people of color, transgender and disabled people, sex workers, care workers are all struggling to have their stories told and represented fairly. Any discussion of internet archiving has to, at some point, confront this problem: How do we talk about the politics of cultural records? If we cannot preserve everything, who defines what is worth saving?

The internet and social media platforms, we are told, have changed all of this. We can share our way to an archival legacy in the form of a massive data profile compiled passively and automatically, and we can just search among everything! Sadly, this seems less and less like a solution. There is a huge difference between searching and finding.

Contemporary technologies have given us platforms that look a lot like personal archives, yet the archival functionality of platforms feels like an empty promise. In the culture of surveillance, evidence collection is never-ending, a shadow to every activity. These daily interactions with (and production of) data, however unwitting they may be, are mediated by some interface somewhere. Their codes of transmission permeate our offline lives: the hashtags, likes, swipes, physical and metaphorical embodiments of intimate data exchanges.

Many platforms cease to be relevant and tend to go away wholesale. In engineering terms, the data becomes no longer relevant. As a result, our pasts can exist piecemeal in distributed systems, in more or less moribund conditions, with no consistent means of access.

Unless you feel a desire to engage with your past self, it’s easy to leave these platforms alone and forget those versions of yourself ever existed, along with the troves of data associated with them. In the early 2000s, on breaks between processing archival collections, I monitored the Friendster profiles of my friends, boyfriends, and others I was connected to socially, but now Friendster has been largely lost. I might have deleted my Myspace profile in a fit, maybe after a breakup, but it still might be there on a server somewhere. I’ve forgotten

at this point the jokey fake name I put on these profiles to protect myself from prying or judgment. I quit Facebook in 2012. But in “Hotel California” fashion, my data will greet me if I check back in. (It’s part of what keeps me away.)

It is becoming harder to maintain one’s digital archive without relying on cloud services; personal computers are now typically sold without the memory capacity to store the richer data we can generate on a daily basis. And these clouds are situated in a complex and obscure climate of for-profit data centers and server farms. As Paul Jaeger points out, “geographical considerations” such as physical location, environmental resources, and legal jurisdiction are key questions for evaluating the integrity of cloud services for archival storage. Cloud computing, he argues, represents “centralization of information and computing resources in data centers, raising the specter of the potential for corporate or government control over information.”

The personal data we supplied remains archived without our conscious intent, according to protocols we have no input in shaping. Loss of control of the personal archive means a loss of societal control of the cultural record. As archival theorist Sue McKemmish writes, “a personal archive considered to be of value to society at large is incorporated into the collective archives of the society, and thus constitutes an accessible part of that society’s memory, its experiential knowledge and cultural identity — evidence of us.”

Jasbir Puar uses the term “trace body” to describe a relationship between the physical and digital realms: When physical bodies cross through checkpoints, so do trace bodies. Both are subject to scrutiny and examination. Archived data guides the scrutiny, feeds the algorithms that rank and identify the outliers and system errors. Like physical bodies, trace bodies can be profiled, targeted or surveilled using categorical metadata: purchasing records, physical whereabouts, or the ethnic origins of one’s last name. Data collectors are building their own archives, and building schemas for surveillance, tracking, and segmentation that shape our everyday existence. We feed the algorithms by response.

Our relationship with interfaces is social: we’ve become so familiar with interfaces that we now expect them to behave themselves, in part to set an example for us. Yet, our data traces are often pushy, rude and clueless: the

banner ads reminding us of the shoes we looked at yesterday, the “customers who bought this” recommendations, the suggestions to friend or follow our exes. We politely ignore this evidence of ourselves. If anything, our relatively smooth relationships with devices, platforms and interfaces have been the result of lowered expectations. As our relationships with interfaces progress, we may find that no amount of design improvement can allow the platforms to address our needs automatically and anticipate our changes of heart.

Platforms collect and structure our personal data in such tidy compartmentalized ways, yet the utter lack of context in data mining continually catches us off guard. We communicate for the algorithms, not with them. Thus, sharing and saving is a continual schematic shift: what may be informal and under the radar one day might sharpen the mechanisms of data collection the next. Moreover, this line between authentic sharing and surveillance-aware communication is wavy at best. danah boyd uses the term “social stenography” to describe the aware-of-the-surveillance way of communicating we’ve come to adopt.

The job of an archivist is work outside interfaces, working with the documents outside their native context, often to make sense of a previous generation’s cast-off data for the interest of the current and future ones. “Processing” a collection is an exercise in not only “saving” the important things and anticipating the interests of future interested parties, but in weeding out the stuff that no one wants to see. My first job as a processing archivist was in a mathematics archive. One obscure Romanian mathematician would send a box of junk quarterly, using the archives as a junk drawer. My boss gave me curt instructions on throwing away his bills and divorce papers and told me to be very selective with his vacation photos.

But who would be qualified to make such decisions about the archives of activists who in part have protested their erasure from the historical record? I think of all the brilliant, give-no-fucks activists I follow on Twitter. I would never want to speak for bad\_dominicana, or those on the ground in Ferguson — so how could I begin to speak for their archives?

Often, the only fully realized archival subjects are the figureheads: those whose papers are collected, the institutional collectors and those for who they profess to collect for, an unknown researcher from the not-so-distant future who sees the same value in the collection as the collector.

Later in his talk, Cerf brought up the Internet Archive, founded and run by another of the Internet's distinguished men, Brewster Kahle, who developed WAIS, another early internet protocol in the early 1990s. The Internet Archive is housed in a former Christian Science temple in San Francisco's Presidio district and has maintained many of the features of the old church, including its sanctuary. Along the back wall of the sanctuary, where one might find a choir loft in a church building, are tall glowing server towers. The physical towers are powerful reminders of the infrastructure needed in order to collect and preserve digital things — one more accurate than “the cloud.”

From a collections standpoint, it's clear that the Internet Archive isn't the Internet Archive, but an Internet Archive, very much built and collected from a certain standpoint and position of power. Those who are actively collecting in the digital realm represent a specific set of values, a perspective, and as in traditional archives, this perspective reflects a certain hegemonic order of knowledge. The Internet Archive's Grateful Dead collection is vibrant and exhaustive, developed in the image and enthusiasm of Kahle, an avid Deadhead. Archival institutions tend to have a point of view. University archives collect records of their institution; governmental archives collect government records. The Internet Archive, and other collections of its ilk, collect from the standpoint of old-guard Internet culture.

No one I know of is collecting and preserving from a position that stands to counter this. For the generation of artists, citizens and activists who has come of age in the era of social media platforms, the power of archives is deployed in the banality of surveillance. Distance from one's data is a design feature, and ownership of one's data profile seems impossible. What from our digital environments can become historical and archived?

Contemporary archival practices advocate a hybrid of two approaches first some interpretation of keeping the original order of things: respect des fons, and an a

posteriori organizing stuff into sensible categories. Of course, many of the collections that have been in archives for decades had been organized in ways that simply did not work. I'd been asked several times to "reprocess" a collection and organize it in a way that made more sense to me or my bosses. How often do archives get shifted now when algorithms adjust?

I wonder if the data collected by platforms will at some point become more transparent, and at what cost or contextual shift. Will my daughter be able to sift through my dark data profiles and learn about the egregious number of times I looked at someone else's profile? Will there be a new round of data mausoleums, offering to sell us peeks at the past? Is data like defaulted debt, ready to be bought and sold at a fraction of the price and subject to a secondary market?

Where are the future archives? Moreover, where are the future points of canonical extinction?

---

**SUBSCRIBE** \$2 PER MONTH



**ENTER YOUR EMAIL**

**CONTINUE**

---

## FURTHER READING



FEATURES



FEATURES



FEATURES



ESSAYS &  
REVIEWS

## *A Declaration of the Dignity Image*

By AMERICAN  
ARTIST

In art that defines  
itself by what it  
withholds from  
public view lies  
the challenge for  
us to do the  
same

## *Editors' Note, Vol. 38: Futures*

By THE NEW  
INQUIRY

Who has no  
future, and  
whose future is  
guaranteed by  
the present?

## *The New Inquiry Vol. 51: Blogroll*

By TNI

An editor's note  
about your blogs,  
and ours.

## *Seeing Things*

By JB BRAGER

A review of At the  
Edge of Sight:  
Photography and  
the Unseen by  
Shawn Michelle  
Smith

---

**The New Inquiry is a 501(c)3 organization.**

**CONTACT**  
**SUBMIT**  
**DONATE**  
**ABOUT**

**SUBSCRIBE**  
**BROWSE THE ARCHIVE**  
**TERMS OF USE**  
**NEWS**



SUBSCRIBE TO NEWSLETTER

**SIGN UP**