PROJECT MUSE®

Hacking the Academy

Cohen, Dan , Scheinfeldt, Joseph T
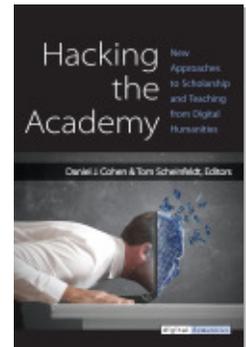
Published by University of Michigan Press

➡ For additional information about this book

https://muse.jhu.edu/book/22907

Reimagining Academic Archives

Christopher J. Prom

> 'Does the past exist concretely, in space? Is there somewhere or other a place, a
>   world of solid objects, where the past is still happening?'
> 'No.'
> 'Then where does the past exist, if at all?'
> 'In records. It is written down.'
> 'In records. And—?'
> 'In the mind. In human memories.'
> 'In memory. Very well, then. We, the Party, control all records, and we control
>   all memories.'
>
>   —GEORGE ORWELL, *Nineteen Eighty-Four*[1]

Archives are rarely created for the express purpose of being preserved, but develop organically as people live their—typically chaotic—lives. Archivists—many of whom serve in university archives and manuscript libraries—are dedicated to identifying, preserving, and providing access to a selective, authentic, and usable record of that messy human experience. People from all walks of life use archives to generate new ideas—or test existing ones—to confirm rights, to hold others accountable for their actions, to gain personal depth of understanding, to establish a connection with society or to the past, and to perform functions that help preserve democratic institutions, sustain civil society, or ensure social justice.

The archivist's charge was difficult enough to fulfill before the advent of networked computing technologies. Many people make overblown claims that a "digital dark age" is now upon us—that all of the electronic files we are creating will someday vanish. At first blush, we instinctively wonder how this could be possible: if there is one thing our lives do not lack, it is access to information. People demand, and are constantly developing better ways to control, index, and sort massive stores of information, but few believe that it will all someday vanish, or perhaps slowly rot away.

It is trite to say that e-mail, websites, blog entries, digital photographs, textual records, database files, and other electronic records are very susceptible to accidental loss, deletion, or decontextualization, even if we do not accept the premises of dystopian predictions that civilization will col-

lapse after the oil runs out, or a catastrophe besets humanity. Nevertheless, records become more fragile and vulnerable as individuals, business, and even governments outsource data storage and management to the warm embrace of commercial vendors, ostensibly under the rubric of cost cutting and efficiency. Also, most people now create records using a wide range of tools, services, and hardware, leaving interrelated records strewn across hard drives, shared servers, social-networking sites, and cloud applications. These documents reside under the care, custody, and control of many different people and organizations—not simply the person or organization that created, and has a vested interest in, their content.

Leaving aside the factors mentioned above, every set of electronic records is itself a constructed and contested entity. The person who creates or assembles the documents molds them into an archive through their activities, interests, and sometimes, their malfeasance, subterfuge, or inertia. Those who control its means of access also have a chilling ability to shape how that record is presented to the public, as certain citizens of the People's Republic of China know all too well.

However one wishes to slice or dice technical issues related to the creation and management of records, we know for certain that it is impossible to construct accurate histories without accurate and faithful evidence of people's actions. Those who use archives can reconstruct or understand those actions only when records are maintained in an intellectually coherent fashion. The contextual relationships between the individual documents that comprise an individual or corporate entity's intellectual output must be preserved. Similarly, future users of archives need to know how the records they are using are related to records produced by other records creators. Given these facts, what types of organizations are best placed to serve as the long-term, trusted custodian of authentic, verifiable, and accurate electronic records?

It is tempting to think that the preservation of digital heritage can be left to those who provide the service of storing and disseminating the thoughts that we distill using keyboards, video cameras, or other digital devices. But to do this would leave the records at extreme risk of loss. At the eighth European Conference on Digital Archiving, Steve Bailey described this problem using an apt metaphor: Imagine if we had trusted the preservation of the records left by Samuel Pepys—the eighteenth-century London diarist—to those who produced his communication media: the stationer who sold him his notebooks, the tanner who sold him his vellum, and the cartographer who sold him the maps he carefully annotated.[2]

Of course, each of the businesses Pepys patronized has long since passed gently into the night. We believe that the same fate will not await Google, Facebook, or Twitter, but even if they manage to survive, what will happen to the content stored in minor services, on contracted webhosts? Tellingly, the terms of service for nearly every free platform or low-cost web host make absolutely no promises regarding digital preservation, or even the return of content to users in case of business failure. Catastrophic business failure is hardly beyond the realm of possibility, as a shareholder in Arthur Andersen will point out. Over a fifty-year period, Google is as vulnerable to social or economic change as the newspaper industry, or perhaps a revolt over its privacy policies may mortally wound it. Even now, its revenue stream is highly reliant on a single source of income: advertising sales.

The recent archiving deal announced between Twitter and the Library of Congress may or may not portend a partial solution to the problem of relying on commercial entities to preserve information needed for historical research. But let's not kid ourselves: the Library of Congress is extremely unlikely to strike deals with every commercial entity providing social-media services, much less every web host, in the country. Other factors will undermine the effectiveness of mass archives. Users, quite understandably and predictably, have already begun to assert a—self-declared—right to remove content from the Library of Congress. The Twitter terms of service put into effect on September 10, 2009, provides Twitter express permission to make tweets available to anyone they choose, and the disposition of public tweets made prior to this date, as well as all of the private tweets, should be an interesting issue for the California judicial system to resolve.

Even if the mass archiving of materials from millions of records creators did not face significant legal hurdles, the methods that libraries use to catalog and make information available are not well suited to preserving the full context necessary to make individual records understandable. To oversimplify at the risk of stereotyping: libraries deal well with items, such as books, or consistent runs of uniform media, such as serials; archives deal well with aggregations of mixed media, and with preserving the contextual information that make them understandable. While large repositories such as the Library of Congress can use cutting-edge tools to mine and repurpose large volumes of data, most tweets cannot be understood without extensive recourse to other online materials, such as blog posts or videos.

Using their professional principles of provenance, sanctity of original order, collective appraisal, and active custodianship, archivists possess the conceptual tools to preserve and make accessible the raw materials of future history: e-mail, digital photographs, and other electronic records. Unfortunately, most archives have made little systematic progress in identifying, preserving, and providing access to electronic records.

Why have most archives failed to effectively address electronic records issues? The reasons are many, but in the end the typical answers are that "digital preservation is hard," and "we don't have enough money to do it properly."

Nevertheless, working closely with university faculty, staff, and students, archivists must reorient archival programs toward electronic records, and to appropriate a set of low-cost tools and services to preserve digital information in a trustworthy fashion. The exact way in which local archives may choose to rethink, reconceptualize, reconstruct, or re-create itself will vary and must be shaped by local context, but almost any institution can cobble this together with existing open-source software. Ultimately, traditional archives must be reimagined in an act of constructive transformation.

Notes

1. George Orwell, *Nineteen Eighty-Four* (New York: Plume, 1949).

2. Steve Bailey, "In Whose Hands Does the Future of Digital Archiving Lie?," presented at the eighth European Conference on Digital Archiving, http://www.vsa-aas.org/de/aktuell/eca-2010/2010-4-29/.