The Future of Saving the Past

ARCHIVAL TECHNOLOGIES IN THE DIGITAL ERA

Radcliffe Magazine
By Corydon Ireland

Saving the past is what archivists do. In quiet places, piled high with records—at universities and museums, mostly—archivists bring order to the soul of a culture: diaries, pictures, books, ephemera, and (increasingly) masses of digital data. They select, organize, and describe collections with a summarizing tool called the finding aid. After that, they ensure access and provide reference services to historians and other users.

Meanwhile, these seldom-seen guardians of conserved culture are grappling worldwide with new technologies to navigate an emerging, quickening, and expanding digital age. That struggle was the focus of a two-day Radcliffe Workshop on Technology and Archival Processing this spring, where 65 experts gave glimpses of the future and deliberated on the implications.

They heard, for instance, about computer programs that crawl through terabytes of data looking for word patterns; facial recognition software that mines video captions for metadata; and new ways to scan ancient manuscripts, defeat the puzzle of handwriting, and provide key words for searching scholars. All the while, most futurists agreed, archivists themselves will remain irreplaceable: the judges of content, the organizers of artifacts, and the providers of historical context that scholars need. Even the finding aid will survive, though perhaps with new ways for users to add comments.
A professor of artificial intelligence, Lambert Schomaker tries to crack the code of handwriting. Relying on samples of handwritten characters, the machine-learning program called MONK employs “shape analysis” to identify letters. Schomaker said that someday this technology will be used to identify individual writers.

The workshop, at the Knafel Center, was the third such gathering at Radcliffe in three years, all funded by Radcliffe’s Academic Ventures. The first was in May 2011, just a few months after a related revolutionary event: the 2010 Radcliffe Workshop that resulted in the Digital Public Library of America (DPLA), a free online archive of American culture, which was established in April 2013.

The technology workshops for archivists, sponsored by Radcliffe’s Schlesinger Library and this year cosponsored by the Association of Research Libraries, represent a somewhat slower revolution. Over three years, they have spurred collaboration, showcased new technologies, identified research challenges, and brought together “diverse holders of archival connections,” said Higgins Professor of Natural Sciences Barbara J. Grosz. “I’m delighted by the evolution of the first workshops into an ongoing series.” A computer
scientist, Grosz was dean of the Radcliffe Institute at the time of the 2011 workshop. She and a team of archivists from the Schlesinger’s experimental archives project realized that libraries everywhere faced a growing logjam of unprocessed archival materials. Perhaps, they thought, a series of network-building workshops—periodic think tanks on emerging technologies of automation, visualization, and information processing—might help.

On the first day of the recent workshop, 120 listeners crowded into Knafel to hear the historian Dan Cohen, executive director of DPLA, deliver the keynote. He argued that DPLA will replace the model of static, unconnected repositories with a collective, multi-source, interactive digital platform that has interoperable standards for information retrieval. Archived culture will flow from “ponds to rivers to oceans” of information like the DPLA, a national archive with analogs already in Europe and elsewhere. The future
promises a worldwide “unified catalog,” said Cohen, which will link archives across the world into a “global digital library.”

Other workshop speakers touched on seemingly impossible projects as large in scope as this global archive, including handwriting recognition software. Lambert Schomaker, a professor of artificial intelligence at Groningen University, described a decade of work on a pattern-recognition and machine-learning program he calls MONK. “All day long, the machine is looking at handwriting to make sense of it,” Schomaker said. Primed with samples of handwritten characters, it uses “shape analysis” to identify letters (someday it will recognize individual writers). Centuries of documents, script variations, and obscure ancient languages make perfection—and transcription—impossible, said Schomaker. But in the end, he said of cracking the elusive handwriting code, “this will work.”

Archivists of the future will untangle audio and video files that right now are not easily searchable, said Larry Goldberg, director of community engagement for WGBH in Boston. Present-day voice recognition software still needs work, and digitally embedded videos are often comically inaccurate, he said. Playing word-mangling examples from YouTube, he quipped, “This is the entertainment portion of the program.” Meanwhile, other technologies—such as movie descriptions for the visually impaired—are untapped resources for archivists in search of ways to write video finding aids.

A facial recognition program for archivists is the quest of Conrad Rudolph, a professor of medieval art history at the University of California, Riverside. His prototype computerized evaluation system matches “similarity rates” in fine art paintings and sculpture from eras when realism ruled, he said. When computers learn both the style and the identity of an artist, and when collected art is scanned and summarized in quantifiable data, future scholars will be able to tap a global reservoir of figural information. “You can see the archival value of this,” said Rudolph.
Trevor Owens likened digital tools for the archivist to a means of enhancing traditional analog skills. When thinking about digital tools, he said, imagine “a mechanized shirt of armor that extends your capabilities.” Owens is a digital archivist in the Office of Strategic Initiatives at the Library of Congress.

In a wrap-up, the Auburn University library technologist Aaron Trehub wondered out loud, “Is the rapture upon us?” He was talking about “the dream of the universal library” proposed by Dan Cohen. But unsettling questions abide regarding the seeming rapture of emerging technologies, he said—questions about authority (who will describe collections of the future?), crowdsourcing (who will check everything for accuracy?), and access, affordability, the fate of privacy, and control over metadata afloat in the Internet. In the end, humans will continue to rule the archival roost, said Trehub. “There is still virtue in slogging through the material.”

The changes ahead in the digital era are “evolutionary and transformative,” added Richard Pearce-Moses, a former president of the Society of American Archivists. Embrace the idea that technology has something to offer, he told the assembled archivists, or “become extinct.” Meanwhile, take comfort in the fact that the fundamentals of collecting, organizing, describing, and offering access remain the same, and still require time and effort despite digital tools. “Sorry, Harry Potter, there is no magic,” said Pearce-Moses. “Search is never going to be a piece of cake.”
Corydon Ireland is a staff writer for the Harvard Gazette.

Photos by Webb Chappell

Illustrations by James Yang

Search Year:
2014

RELATEDEVENTS

Conferences & Symposia
Take Note
Thursday 11/1/12 to Friday 11/2/12

Conferences & Symposia
Why Books?
Friday 10/29/10

NEWS

Books Meet Bytes
The world of libraries is being shaken by the digital age, changing patterns of readership, information retrieval, perhaps even brain circuitry. The dance toward the digital drew archivists from around the world to Harvard's Radcliffe Institute for a workshop on technology and archival processing.

Radcliffe Workshop on Technology & Archival Processing
Richard Pearce-Moses reflects on the insights and ideas discussed in a recent workshop on technology and archival processing.

No New Backlogs
The Schlesinger Library has assembled a wide range of scholars for interdisciplinary collaboration on how to deploy technology to streamline archival processing.

Claire Bond Potter
When asked what her favorite thing is about being a historian, Claire Bond Potter, who recently joined the Schlesinger Library Council, says with some mirth, “Reading other people’s mail.”