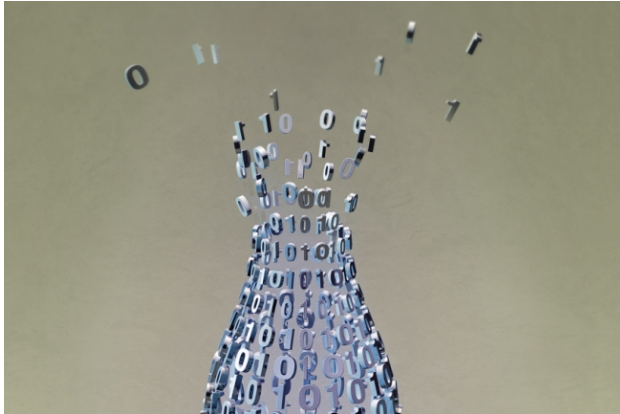# THE CHRONICLE OF HIGHER EDUCATION

THE CHRONICLE REVIEW

The Risk of Digital Oblivion

What knowledge should we save?

Randy Lyhus for The Chronicle Review

By Abby Smith Rumsey | MAY 04, 2016

Technology has been good to the pursuit of knowledge. Each advance, from cuneiform to computer chip, spurs us to push the limits of knowledge further. The benefits of the newest innovation — the digital — are obvious: more evidence. A lot more. The world's largest radio telescope, the Square Kilometre Array (SKA), is expected to produce "up to one exabyte (10 to the 18th bytes) of data per day, roughly the amount handled by the entire Internet in 2000." The radical reduction of barriers to reading and publishing online has resulted in an abundance of cultural expression in audio, video, textual, and numeric formats. The horizons of knowledge are receding not simply because we have more evidence. We also have powerful tools to analyze data at scale, see beyond the limits of human perception, and discern patterns invisible to the naked eye.

Thanks to an ever-expanding collective memory, we have become the dominant species on Earth — dangerously dominant. Yet paradoxically, in this age of digital abundance it is harder, not easier, to secure knowledge for future generations. Where will the astronomers of the SKA telescope store their exabytes of data? How much will future generations know about today's online culture when the average webpage lasts just 100 days? We have relied on durable physical objects to carry knowledge across space and time. Among the oldest of memory technologies, cuneiforms date back 5,000 years and are still legible. You need expertise in Semitic languages, but you do not need a machine to read them. By contrast, digital data are ephemeral, easily overwritten, dependent

upon hardware and software, decipherable only by machines.

Higher education is based on the achievements previous generations have deeded to us. What will we pass forward? If we learn to manage the abundance of digital information as quickly as we learned to manage print, solutions will come within several generations, approximately when the first few cohorts of digital natives mature, age, and begin to reckon with their own legacies. By that time, though, much will be lost, not by choice but by default.

Preserving data, crafting polices for their use, and paying for the benefits of future access are all formidable problems, but the challenges to long-term access are not only technical, political, and economic.

The primary challenge is moral and goes to the very core of our humanity: What do we owe other people, including people not yet born?

We live in a technology-dependent world of our own making. The moral responsibility for how we preserve knowledge is greater now than ever before, with more at stake. Given our expanding powers to modify and manipulate natural processes, if we fail as stewards of that knowledge, it will not only lead to our own undoing as a species; we risk unraveling much of the web of life of which we are a part.

Information technologies have always posed moral challenges. Innovations in memory storage divide people into the Techno-optimists, who praise them as disruptive and creative, and the Nostalgics, who condemn them as disruptive and destructive. Both groups know that something profound is being taken out of our direct control and entrusted to others. Socrates, the ur-techno-curmudgeon, warned that the invention of writing would lead to ignorance and ultimately the death of memory itself:

> *For this invention will produce forgetfulness in the minds of those who learn to use it, because they will not practice their memory. Their trust in writing, produced by external characters which are no part of themselves, will discourage the use of their own memory within them. You have invented an elixir not of memory but of reminding; and you offer your pupils the appearance of wisdom, not true wisdom, for they will read many things without instruction and will therefore seem to know many things, when they are for the most part ignorant and hard to get along with, since they are not wise, but only appear wise.*

Socrates got it wrong. It was the societies that entrusted their memories to papyrus, paper, and now computer chips that have immensely expanded the scope of human knowledge. Writing and audiovisual technologies have allowed individuals to know things that lie outside the scope of their own experience. Sharing in our collective memory greatly expands our powers of empathy and imagination. Besides, Socrates, ever the ironist, knew that Plato's writing about him was the only way we would know Socrates ever existed. But he was correct to point out the moral hazard of taking memory — an intrinsic part of ourselves, our identity — and moving it outside of our direct control. Today, digital natives often encounter this hazard when they share personal information on the web.

The creation of knowledge can be very expensive, but the expense is deemed a good investment in the future. Preservation of that knowledge is cheap, by comparison. Yet most research universities, the ones whose scientists and scholars generate and use data, are not making the basic investments in archiving that are necessary.

In the near term, we have two urgent tasks. The first is ensuring that the wealth of knowledge in our analog collections can be found on the Internet, either by putting them online or by creating digital records to indicate where those artifacts are. In 20 years, if a collection cannot be discovered through a web search, people will effectively not know it exists. We need to accelerate digitization efforts.

Conversion of sources into databases actually creates new knowledge. There is a wealth of detailed information about the ocean, the atmosphere, the flora and fauna of recent centuries found in maritime logs. When people had to read each log page by page, the sum of that information was impossible to grasp. Now that they are being digitized, they form an invaluable database of oceanographic, atmospheric, and biological trends.

As part of the digitization effort, we will need to stabilize analog sources that are rapidly deteriorating — notably 20th-century audiovisual formats — and create digital reference copies. Other resources long considered artifacts from a bygone era, such as the drawers full of bugs and birds in natural-history museums, are now being rediscovered as gold, these museums the Fort Knox of genomic history.

The second task is to rescue the present from oblivion. Oblivion can begin as soon as the next software update. The challenge here is posed by scale. Given the upfront expenses of publishing books or making movies, our model of stewardship has been to ask what we can afford to save. Now, given digital abundance, we must ask what we can afford to lose. We can feel certain about the long-term value of government records that hold officials accountable to their citizens, records of where hazardous and nuclear waste is stored, longitudinal atmospheric and oceanographic data, and the genome data bank, among others.

Put simply: Information has value to the extent that it can be reused. That said, we can't predict the uses that may be possible in the future.

But for most data, it is hard to grasp which content has long-term value and which can be let go. Put simply: Information has value to the extent that it can be reused. That said, we can't predict the uses that may be possible in the future, any more than in the 19th century people could predict that all those maritime logbooks would one day enable scientists to study a problem that 19th-century people did not know would exist. For that reason, we should err on the side of including more rather than less, even if we have to keep most data at very low levels of curation until such time as people may be able to assess its value, 50 or 100 years hence.

We have too many lessons from history to choose any other course. Early exemplars of information technologies are highly vulnerable to infant mortality. In the case of silent films — of which only 20 percent survive — collectors faced the technical issue of preserving the highly combustible nitrate film. Perceiving little long-term value in movies created as popular entertainment, it was very tempting, if not downright sensible, to scrape the silver nitrate off the substrate for reuse. Only with the passage of time can we perceive how much information those films carried, unintentionally.

The faster the rate of cultural and ecological change, the more unpredictable will be the value of any given piece of information. Borrowing an analogy from biology, we can say that the more culturally diverse our memory bank, full of seemingly outdated, obsolescent, or backward corpora of knowledge, the greater the chances that we will survive abrupt change through cultural adaptability.

Given the importance of the historical record for the entire enterprise of higher education, it is startling to realize how many structural obstacles there are to meeting our obligations as stewards of inherited knowledge. For all that it grants prestige to the creation of "new knowledge," the academy is a conservative, hierarchical entity, not adapting easily to major structural changes.

The physical and life sciences are the domains most responsible for inflating the data universe. So-called Big Science builds instruments like the Large Hadron Collider, whose primary output is petabytes of data. Scientists gather and catalog data for analysis, from deep-sky telescopes to deep-sea dives, from buried archaeological sites to ice cores drilled deep into the Antarctic shelf.

When it comes to data management, scientists face perverse incentives to create a lot coupled with disincentives to manage it for the long term. As a rule, funding agencies and universities support discovery and analysis. The National Science Foundation and the National Institutes of Health now require grantees to provide data-management plans that include the possibility of reuse where appropriate. But those requirements have no effective enforcement mechanism. On the contrary, reliance on relatively short-term grants and the hypercompetitive nature of forging a career mean that many scientists are stuck on a treadmill, seeking new funding before they've even closed out

their old grants. Besides, the prestige economy tells them that data management is mere "housekeeping," by definition a second- or third-order profession, someone else's job (typically a postdoc or grad student, in fact.)

Humanists have similar reward structures that keep them at a remove from libraries, museums, and archives. In the 19th century, building and editing collections was the core work of history and literature disciplines. Over the course of the 20th century, a division of labor emerged between the scholars who were users of collections and the librarians who were builders and custodians of collections. Today libraries and archives are positioned within higher education as service organizations, not laboratories of discovery.

> **Today libraries and archives are positioned within higher education as service organizations, not laboratories of discovery.**

But things are changing. Scientists increasingly study how nature organizes itself and how dynamic systems change over time, from the history of the cosmos to the evolution of species. Retrospective data — evidence from the past — is now at a premium. In the humanities, scholars are engaging anew with primary sources as humanities corpora are digitized. The digitization of the human record demands — and elevates the status of — critical editing of texts verbal, visual, and aural. A new generation of scholars is confronting the urgent demands of analog and digital stewardship. One of the unintended consequences of the shrinking humanities faculty is that many highly accomplished humanities Ph.D.s are now drawn into libraries and creating rewarding careers.

Universities are building an ecosystem of complementary archiving efforts. Best of all, leading university presidents have committed to building out the Digital Preservation Network, a large-scale service working to ensure future access to scholarly resources.

Across all disciplines, scholars are teaching digital natives. Most students are more fluent than their professors with current consumer technologies. But they have yet to gain true literacy, digital or textual, visual or audio: the ability to assess the authenticity and truth

value of a source. It is from the faculty that they learn about the values that need to be built into the use of data — our own and everyone else's. They may listen to what their professors say, but they pay closest attention to the behaviors they see modeled. We must articulate and model the values we want to embed in the new memory systems.

Technologies will come and go. Fifty years from now, Instagram, Facebook, and Snapchat will be as antiquated as the telegram. What is important is that each generation grows up understanding the value of rights to privacy and control over one's own data, being schooled in the trade-off between security and privacy, between open systems and closed, and having the information at hand to make intelligent choices for themselves.

Rebuilding memory systems for the digital age raises the same questions that Socrates faced: the morality of memory versus the efficiency of writing (or coding), and the need to take responsibility for what you know. The worst thing would be to let legacies of knowledge perish. We in the West think that we do not engage in censorship, but we allow our commercial companies to control access to vast domains of online resources. We pride ourselves on being open and diverse, but we are in danger of replicating the fate of the Library of Alexandria. Classical learning was allowed to decay and disappear because ideological monocultures — first Christian, then Islamic regimes — deemed pagan learning useless or worse.

We live in a culture that gives priority to instrumental, "useful" knowledge, and we apply scientific models of truth to social and cultural problems in no way amenable to scientific verification, such as poverty, mental illness, racism, and inequality. A species forced into monoculture is very vulnerable in times of great change, and we are at risk of creating a monoculture of knowledge in the Western mode. We cannot know now what kinds of knowledge we will need tomorrow. We cannot afford to lose the record of the multiple ways of being human in this world.

Thomas Jefferson, like most of the founders, believed that the growth of knowledge and of liberty go hand in hand. As if foreseeing the Internet, he wrote, "That ideas should freely spread from one to another over the globe, for the moral and mutual instruction of man, and improvement of his condition, seems to have been peculiarly and

benevolently designed by nature, when she made them, like fire, expansible over all space, without lessening their density in any point."

What do we owe the future? The freedom to choose to know or not to know. Understanding the past gives us a sense of how the world works. It allows us to imagine our own future in that world. Students across the country are now claiming the right to make their own history visible on campus. They do so because when they do not see themselves and their history on campus, they cannot understand if they really belong there, the way other people belong. Historical memory is not about the past — it is about the future. Its price is stewardship, and it is a bargain.

Abby Smith Rumsey is a historian and the author of When We Are No More: How Digital Memory Is Shaping Our Future *(Bloomsbury).*

THE CHRONICLE OF HIGHER EDUCATION