14th Blackbird: Digital Preservation as an Environmentally Sustainable Activity

Ben Goldman, Penn State University

PASIG 2016 meeting, October 28, 2016

Part of Session: ENVIRONMENTAL RESPONSIBILITY, SUSTAINABILITY, COSTS, BENEFITS AND RISKS

The title of my presentation is inspired by a 2004 D-Lib article called "Thirteen Ways of Looking at Digital Preservation," which urged us to consider digital preservation "not just as a mechanism for ensuring bit sequences created today are renderable tomorrow, but as a process operating in concert with the full range of services supporting digital information environments, as well as the overarching economic, legal, and social contexts[1]." In the spirit of this article's premise, I'd like to expand the scope of our consideration to include the environmental context and explore how the natural resources of our planet are supporting digital preservation activities, and ways archivists and librarians might consider adapting current digital preservation practices to meet these challenges.

All preservation requires energy. Maintaining constant levels of temperature and humidity in climate controlled stacks is an obvious example of the energy needed to support the preservation of our analog collections. The energy required to preserve digital content, while not as visible to us, is considerable. Information that can be accessed online exists as data stored on a physical machine *somewhere*. In most cases, such servers are clustered together in facilities (called data centers) that require a significant amount of electricity to keep running 24/7. Data centers must also be kept at cool and constant temperatures, so additional energy is

[1] Lavoie, Brian, and Lorcan Dempsey. "Thirteen ways of looking at... digital preservation." *D-Lib magazine* 10.7/8 (2004): 20. Available online at: http://www.dlib.org/dlib/july04/lavoie/07lavoie.html

required to keep these machines from overheating. In some cases, the cooling of these facilities makes them water-intensive as well as energy intensive. Facebook, for instance, asserts that in 2015 they used over 150 million gallons to cool their data centers[2].

The energy appetite of data storage is something that's been examined for nearly two decades. In 1999, one analyst detailed an alarming (yet ultimately overstated) energy consumption rate in an article with the provocative title, "The Internet Begins with Coal[3]." A report eight years later estimated that data centers were using 1.2 percent of all electricity in the U.S., an amount correlated to the total electricity consumption of color televisions[4]. Last year, the environmental organization Greenpeace examined the impact of data centers and concluded that their demand for electricity would rival that of the six largest energy-using countries in the world[5].

We as a society are still largely powering this technical infrastructure using fossil fuels. Petroleum, coal and natural gas combine to supply around 80% of global energy consumption[6], and ninety percent of the power consumed in the United States is generated from non-renewable sources[7]. The United States is the world's largest consumer of energy to power data centers, and as Greenpeace has pointed out, "the internet's growing energy footprint has

---

[2] Facebook. "Our Footprint." 2015. Available online at: https://sustainability.fb.com/our-footprint/

[3] Mills, Mark P. *The Internet Begins with Coal: A Preliminary Exploration of the Impact of the Internet on Electricity Consumption: A Green Policy Paper for the Greening Earth Society*. Greening Earth Society, 1999.

[4] Koomey, Jonathan G. "Estimating total power consumption by servers in the US and the world." (2007).

[5] Cook, Gary, et al. "Clicking clean: how companies are creating the green internet." *Greenpeace Inc., Washington, DC* (2014). Available online at: http://www.greenpeace.org/usa/wp-content/uploads/legacy/Global/usa/planet3/PDFs/clickingclean.pdf. Specifically, the report found that, based on the estimates contained in the SMARTer 2020 report, "the aggregate electricity demand of the cloud (including data centers and networks, but not devices) in 2011 was 684 billion kWh."

[6] International Energy Agency, "Key World Energy Statistics 2016." Available online at: http://www.iea.org/publications/freepublications/publication/KeyWorld_Statistics_2015.pdf

[7] United States Energy Information Administration, "Energy in Brief." Available online at: http://www.eia.gov/energy_in_brief/article/major_energy_sources_and_users.cfm

thus far been mostly concentrated in places where energy is the dirtiest[8]." Virginia, for example, is home to data centers operated by the largest cloud providers, and claims to funnel seventy percent of global internet traffic through its facilities every day, yet powers this energy demand primarily through coal[9]. The Internet, it seems, began and **persists** with coal.

Like other fossil fuels, coal, once combusted, concentrates in the Earth's atmosphere as carbon dioxide. And an increase of atmospheric greenhouse gases, including carbon dioxide, has raised the planet's temperature nearly two degrees in the past one hundred years, leading to a string of serious effects such as melting ice caps, ocean acidification, rising sea levels, increased precipitation, and more frequent extreme weather events.

The infrastructure we use to preserve digital content, then, is largely powered by the same fossil fuels responsible for our environmental crisis. We all agree that preservation is an activity undertaken for the benefit of future generations. But I can't help but wonder whether future researchers will even have access to a digital heritage that is currently dependent on the fossil fuels that *must* be phased out if society is to have a future at all. Our reliance on fossil fuels should be considered yet another digital preservation risk that we plan for. And ethically speaking, how harshly might future generations judge a profession that is preserving cultural heritage using the same resources that created our planet's crisis in the first place?

I know it's challenging to consider these issues when our digital collections are likely growing at a faster rate than our staffing and funding. There's a good reason why most discussions of

---

[8] Cook, Gary, et al. "Clicking clean: how companies are creating the green internet." *Greenpeace Inc., Washington, DC* (2014). Available online at: http://www.greenpeace.org/usa/wp-content/uploads/legacy/Global/usa/planet3/PDFs/clickingclean.pdf.
[9] Ibid.

sustainable digital preservation have focused mainly on the economics of an activity that requires an unceasing commitment of resources. And there have been some excellent professional conversations around the environmental impacts of digital preservation, but these have tended to focus around the “greening” of servers and data centers. These are essential sustainability considerations for us to continue discussing. I’m just not sure it’s enough. By viewing sustainability solely through an economic lens we risk obscuring the long-term human, social, and ecological impacts of our professional decisions. And as Greenpeace has pointed out, energy efficiency gains in the technology sector have continually been canceled out by unabated demand for storage. I think if we’re going to address sustainability relative to the environmental context, it will require us to critically examine our own professional assumptions and activities.

> **“Technological decisions often have extremely long-term consequences, especially regarding the extraction of non-renewable resources and the depositing of toxic substances. Ethical adequacy would require that in fairness the risks and the costs of all such decisions be considered over the full-time frame of the impact. One more conclusion may be drawn about cost- and risk-benefit analyses: by their anthropocentrism they tend to ignore considerations of ecological justice.”**
>
> **– Frederick Ferre, *Philosophy of Technology***

So what might that look like? I can think of a few areas that might warrant consideration, but with limited time I’ll focus briefly on practices that support authenticity and integrity. Forensic imaging, for example, is a practice, now widely adopted in the archival profession, that results in a full capture of a hard drive's or computer disk's bitstream, which retains significant characteristics and metadata from the computer environment used to create and manage those files. But this practice also accumulates ancillary data present in the bitstream, some of which may be unwanted or unnecessary, which can make considerable demands on storage. And as we heard this morning, forensics is a building block for emulation, which requires additional computing demands. Once ingested, we endeavor to protect all digital objects from change. Frequent verification of fixity increases the likelihood that we can detect and address unwelcome alterations. However minimal, though, every integrity check performed over time against every digital object, racks up a sizable energy receipt for our storage environments. Strategies for digital preservation and access routinely lead to the duplication of the same content across multiple objects. We migrate files to different formats but maintain the original formats to ensure fidelity of future migrations, or to preserve an unaltered original, or employ tools that automate the generation of multiple derivatives over time. These practices are embodied in our core principles, such as the OAIS model, and though they be conceptually sound, in sum they likely have a significant carbon impact.

It's possible to view practices that support authenticity and integrity as an over-correction to the increased ephemerality and mutability of cultural heritage in the digital age. Whereas paper can remain stable for decades, even centuries, under the right conditions, our experience with technology is one of constantly managing, accommodating, and adjusting to change, when what we desire as professionals -- what we assume will enable future use -- is **stability**. But perhaps there is no natural stability for digital objects, as much as we might want there to be one. If instead the nature of digital resources is to be changed, transmitted, copied, and reconstituted in new contexts, then perhaps we might consider evolving our theories and practices to embrace acceptable levels of mutability... or to support, in the words of Bethany Nowviskie, ***a graceful degradation***. *Lossiness* is a word with negative connotations in the digital preservation context, but perhaps it can be a sustainability strategy if we account for the environmental context of our work.

> "There is an implied ideal of 'natural stability' on the part of both climate and heritage, and any alteration of this 'natural stability' is perceived through a paradigm of 'threat'. Embedded within these ideals is an implied notion that our possession of heritage is a positive thing, which impels us to undertake actions to 'preserve' the integrity of (physical) heritage assets for the sake of future generations. The implication is that change is threat, and we must prevent and/or mitigate this threat."
>
> – Harvey and Perry, *The Future of Heritage as Climates Change: Loss, Adaptation, and Creativity*

My fundamental premise is that short of a complete societal shift to renewable energy or something politically infeasible like a tax on carbon, digital preservation practitioners have a responsibility to consider how our professional activities, not just our resources, and not just our data centers, can be made more sustainable. This would require us to consider ways our current practices and theoretical frameworks can be adapted to lessen the carbon impact of our work, which may introduce new preservation risks to be managed, but could also increase the long-term durability of digital information. My hope is that we'll at least begin to engage in a debate about the viability of our current theories and practices given these ecological externalities.