

Digital Preservation Plenary Session, Wednesday, July 24, 10am

Green Bytes: Sustainable Approaches to Digital Stewardship

David Rosenthal
Stanford University

Kris Carpenter
Internet Archive

Krishna Kant
George Mason University

Joshua Sternfeld, Moderator
National Endowment for the Humanities

As our digital cultural and scientific heritage grows at an exponential rate, it is often easy to overlook the underpinning material costs. Data, of course, are not “virtual” or “ephemeral”; rather, every byte requires resources to ensure its reliable storage and accessibility. Recent reports suggest that data management currently taxes upwards of 2% of total global energy consumption.¹ The data center is quickly emerging as a counterpart to the “analog” storage facility as one of the central infrastructural components of our preservation ecosystem.

Current research into the sustainability of data centers, especially in the commercial sector, would suggest that there is plenty of room to improve energy efficiency. The greening of data centers has led to innovations in every facet of their operations, from retrofitting buildings in order to recalibrate air flow and cooling, to adopting computational strategies that reduce the load on spinning hard drives. This panel will explore how approaches to achieving green sustainability already underway in other industries could be adopted by the digital preservation community.

Green digital preservation raises complex questions about stewardship. Is the noble goal of preserving all data for the long term actually sustainable? If not, what data should we preserve and for how long? How do we manage basic preservation functions such as selection, curation, and redundancy when those processes have expanded beyond reasonable human intervention? Is digital preservation in its most currently ideal form a case of the perfect being the enemy of the good? While institutional repositories large and small address these issues on a daily basis, placing them within a larger context of *environmentally sustainable* preservation practices may help reorient priorities, expose deficiencies that otherwise may go unnoticed, and supply new incentives for taking responsible action.

A comprehensive examination of digital environmental sustainability requires an interdisciplinary perspective that merges material and access needs, and brings together groups

¹ Glanz, James. “Data Centers Waste Vast Amounts of Energy, Belying Industry Image.” *The New York Times*, September 22, 2012, sec. Technology. <http://www.nytimes.com/2012/09/23/technology/data-centers-waste-vast-amounts-of-energy-belying-industry-image.html>.

Green Bytes: Sustainable Approaches to Digital Stewardship

that too frequently have been content to remain in isolation. Digital preservationists, IT specialists, administrators, computer engineers, environmentalists, and many others all have a stake in reducing the environmental impact of digital repositories. Metrics by which to evaluate the operational costs of data centers, such as those produced by the non-profit organization Green Grid (<http://www.thegreengrid.org/>), or the JISC-funded Greening Information Management Assessment Framework, offer ways in which digital preservationists may conduct preliminary, quantifiable assessments. These are only first steps, however, and a much more comprehensive, interdisciplinary approach is needed.

While there is still plenty of work to be done, the possibilities already at our disposal for drastically reducing the carbon footprint of digital preservation -- many of which will be covered in the course of the panel -- prove tantalizing and may warrant further investigation. Some institutions have already discovered cost and energy efficiencies in moving towards shared or distributed network systems, while others are experimenting with cloud storage. File compression, tiered storage, streamlining fixity checks, and eliminating duplicate or redundant data represent methods for reducing the size of a digital collection, which in turn could produce potential energy savings. Finally, many data centers are reconceptualizing environmental storage by loosening traditionally rigid standards for heating and cooling systems while also renovating the building envelope. Far from advocating one measure over another, this panel emphasizes the need for every digital repository to pursue its own path for striking the appropriate balance of access, reliability, and sustainability.

About the Panel

Representing a diverse cross-section of preservation and research interests, the three presenters for this plenary panel will outline the basic challenges and current efforts to find practical solutions. David Rosenthal will frame the problem of green digital preservation as a tension between the expectation of constant online access and the need to reduce energy consumption. Drawing upon his own work in developing a low-power storage network called DAWN (Durable Array of Wimpy Nodes) and evidence from the commercial storage industry, he will discuss some of the economic barriers repositories face when trying to implement sustainable measures. Kris Carpenter will highlight the work of the Internet Archive as well as other U.S. and international institutions in applying creative workflows and practices. And finally, Krishna Kant will share his knowledge of several scientific and engineering breakthroughs that have the potential to minimize the carbon footprint of data centers.

Given the limitations of a single plenary session, this panel is intended to raise more questions than provide answers and to serve as a conversation starter among NDSA participants and others. The Digital Preservation conference affords an opportunity to explore how

Green Bytes: Sustainable Approaches to Digital Stewardship

environmental sustainability may generate new areas of research and development, build partnerships, establish preservation networks, and lead to advocacy. Conference participants are encouraged to read in advance some of the latest literature on the subject from the provided bibliography (which is by no means comprehensive or exhaustive) and come prepared with questions. Slides and other output produced as a result of this panel will be made available following the conference. We plan to continue the conversation in social media (#greendigpres) and other outlets such as the National Endowment for the Humanities Division of Preservation and Access homepage and the Library of Congress' Signals blog.

Environmentally responsible stewardship of digital collections necessarily draws into question basic assumptions about preservation and access. The unrestrained growth of born digital materials, coupled with the material limitations for adequately maintaining such a system, exerts pressure on our notions of selection, scale, automation, arrangement and description, migration, and data quality. Consideration of this convergence of digital infrastructure with stewardship practices can lead to new breakthroughs in our management of the data lifecycle, from how we collect and ingest raw data, to how we provide access to curated information far into the future.

Joshua Sternfeld

Senior Program Office in the NEH Division of Preservation and Access

jssternfeld@neh.gov



Presenters' Websites

David Rosenthal: <http://blog.dshr.org/>

Krishna Kant: <http://www.kkant.net/>

Kris Carpenter: <http://archive.org/index.php>

Select Bibliography

Anderson, Paul, Gaynor Backhouse, Daniel Curtis, Simon Redding, and David Wallom. *Low Carbon Computing: A View to 2050 and Beyond*. JISC, November 2009.

http://www.jisc.ac.uk/media/documents/techwatch/jisctsw_09_02d.pdf.

Blanchette, Jean-François "Viewpoint: Computing as if Infrastructure Mattered," *Communications of the ACM*, Vol. 55 No. 10, 32-34. (Preprint available:

<http://polaris.gseis.ucla.edu/blanchette/papers/caiim.pdf>)

Green Bytes: Sustainable Approaches to Digital Stewardship

Blue Ribbon Task Force on Sustainable Digital Preservation and Access. *Sustainable Economics for a Digital Planet: Ensuring Long-Term Access to Digital Information*, February 2010.
http://brtf.sdsc.edu/biblio/BRTF_Final_Report.pdf.

Climate Group, The. *SMART 2020: Enabling the Low Carbon Economy in the Information Age*. Global eSustainability Initiative, 2008.
http://www.smart2020.org/_assets/files/02_Smart2020Report.pdf.

Cook, Gary. *How Clean Is Your Cloud?* Greenpeace, April 2012.
<http://www.greenpeace.org/international/Global/international/publications/climate/2012/iCoal/HowCleanisYourCloud.pdf>.

Google's Green Data Centers: Network POP Case Study.
http://static.googleusercontent.com/external_content/untrusted_dlcp/www.google.com/en/us/corporate/datacenter/dc-best-practices-google.pdf.

Masanet, Eric, Arman Shehabi, and Jonathan Koomey. "Characteristics of Low-carbon Data Centres." *Nature Climate Change* 3, no. 7 (July 2013): 627–630.
doi:10.1038/nclimate1786.

Rosenthal, David S. H., and Daniel L. Vargas. "Distributed Digital Preservation in the Cloud." *International Journal of Digital Curation* 8, no. 1 (June 14, 2013): 107–119.
doi:10.2218/ijdc.v8i1.248.