Lots (though Fewer) of (and More Hosted) Copies Keep Stuff Safe

Nicholas Taylor (@nullhandle)
Program Manager Emeritus, LOCKSS and Web Archiving
Stanford Libraries

Internet Archive Web and Data Services Partner Meeting
20 September 2019
What threatens the long-term persistence of digital information?
Amazon cloud crash wipes out customer data; Will users be compensated?

By Molly McHugh
Posted on April 28, 2011 10:57 am

Yahoo Quietly Pulls The Plug On Geocities

Posted Apr 23, 2009 by Leena Rao (@leenaraa)

Sorry, new GeoCities accounts are no longer available.

Current GeoCities customers:

- After several complaints, we have decided to shut down GeoCities later this week.
- If you don't have a backup of your GeoCities site, please sign in before the site goes down for more information.
- Sign in to GeoCities

Save 50% on a web site with Yahoo! Web Hosting.

- Get a personalized web address (domain name) for only $9.99/yr.
- Design a great-looking site with Yahoo! site builder.
- Never worry about security on your site.
- Get unlimited disk space to store all your pictures and audio files.
- Receive as many emails as you like. Each email is automatically sent back to your inbox.
- Get answers even at 3 a.m., with 24-hour phone support.

Future Perfect

“Climate change” and “global warming” are disappearing from government websites

The deletions follow a pattern of policy changes on climate change under the Trump administration.

By Umair Irfan | Updated Jan 11, 2018, 12:30pm EST

Moon landing tapes got erased, NASA admits

Maggie Fox, Health, Science Editor

WASHINGTON (Reuters) - The original recordings of the first humans landing on the moon 40 years ago were erased and re-used,
understand preservation threats

- long-term **data integrity** is hard
- need architecture informed by **actual leading threats** to data
- don’t underestimate:
  - people making **mistakes**
  - **attacks** on information
  - organizational **failure**
mitigate preservation threats

- lots of copies is **necessary but not sufficient**
- central points of failure can **undermine all copies at once**
- **multi-organizational** preservation storage provides:
  - **resilience** against organizational failure
  - **compartmentalization** against operator error
  - **decentralization** against attacks
  - **diversity** in technical infrastructure

"Domino's" by david pacey under CC BY 2.0
LOCKSS Networks

- a dozen networks
- hundreds of institutions
- all types of content
- preservation via diversity of:
  - technologies
  - institutions
  - networks
network model evolution

• in original, Global LOCKSS Network, **all nodes stored copies**

• communities running private LOCKSS networks are **consolidating nodes**
  • **only a subset** run nodes
  • **whole community** is served, participates in governance, provides funding

• **under (co-)hosted service model**, Stanford + trusted partners may serve as anchor storage hosts
IPLC Preservation Network pilot

- **5x nodes**: Harvard, IA (x2), Princeton, Stanford
- **all IPLC Archive-It collections** ingested + under active preservation
- ingest + audit/repair **take place automatically** via LOCKSS software
Questions