

Designing Storage Architectures for Digital Preservation

Library of Congress

Meeting Location: Sofitel Hotel Washington, D.C.

Meeting Dates: September 27–28, 2010

(Questions collected from participants during the meeting)

How many back-up copies do we need?

Can ice sculpture be preserved for a century? What does “can” mean?

How often should checksums be verified to optimize their use in preserving data?

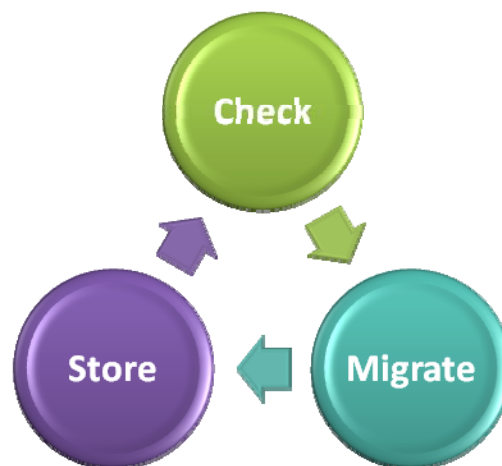
How about vinyl for long-term storage with new access tools?

What are tiered solutions to be considered for the current time?

Preserving is not the problem; retrieval is the problem. How to access 100 yrs. from now?

Answer: migrate

Cycle of Preservation



How can I link my digital preservation files to my dynamic metadata?

If you discover a problem in your storage system, how do you bound the problem so the whole system isn't suspect?

How does one recover from data loss or corruption? Too much emphasis on avoidance as one questioner noted.

How can we optimize media format migrations over many iterations?

Why do we build systems ourselves but find ones run by others to be more trustworthy?

How can we make sure that the bit that flips is not the most critical one?

What can we learn from forensics of migrations?

Are we learning from on-line, off-line, near-line needs?

David says you can't verify high reliability—prove him wrong.

Why permanence? Isn't selection more important than permanence?

What is our appetite for risk?

What is an acceptable risk level?

If all the hardware solutions were solved (permanent media, accurate checksum sets), what would be the biggest challenge?

What is good enough preservation?

What drives data migration?

- Cost
- Technology
- Preservation
- Regulation

Isn't the core issue economics, not technology or guaranteed preservation?

Small institutions don't have \$ or your expertise. How do we get our data preserved? How are you breaking down silos?

How do we deal with long-term encryption?

How can we standardize checksums if it's not financially viable?

Designing Storage Architectures for Digital Preservation
Questions Worth Asking
September 27, 2010

Whose responsibility is it to initiate the process of checksum/checksum validity standards?

How compare strategies to ensure data integrity in preservation archives?

- More copies?
- Verify all files every X years
- Invest in more reliable technology

Would you rather have more stuff preserved or more reliability for less stuff?

How do brain-damaged accounting practices (o.k., GAAP) affect technology acquisition, and thus development, deployment, and use?

I am not convinced that a system to my standard of elegance yet exists.

Given that practical limits of storage will always exist, is it socially possible to create a consensus strategy to determine which content should be preserved and which discarded?

Does everything need to be saved? Shouldn't we get rid of what isn't necessary, by having authors and creators curate their stuff?

Agreed upon standard file formats for interoperability across platforms & disciplines-- is it possible?

What do we mean by "born archival"?

How do you determine when something is not worth keeping?

If you noticed some of your data was damaged/missing, could you figure out how widespread the problem was? Could you figure out the cause? How much detail on your process is sufficient to allow you to answer "yes"?

How much data is being a) thrown away, b) not preserved? And is it right to focus so intensely on checksum solutions and "5 9's" of data integrity?

How can we foster more cooperation among archives/libraries/corps & alliances of vendors to maximize budget & other resources?