

Tales from the crypt: What are the emerging practices of large-scale storage?

Sponsored by the NDSA Infrastructure Working Group

NDIIPP/NDSA Partners Meeting Workshop

July 20, 2011

11:15 a.m.

Facilitators: Trevor Owens (Library of Congress); Karen Cariani (WGBH)

Attendees: 42

Over the last six months, the National Digital Stewardship Alliance Infrastructure Working Group has been talking *storage*. After listening to several presentations on shiny new decentralized, distributed and or cloud storage platforms we quickly realized that the diversity of our members experience meant that we were thinking around each other a fair bit. At that point we decided it would make a lot of sense for the group to start by documenting NDSA member's approaches to large scale storage for digital preservation and stewardship.

Ultimately, we intend to share the results of this project with other organizations looking for guidance about preservation and stewardship storage. We are working to pin down what our collective experience suggests are key elements to consider when planning, implementing and maintaining these systems. In the process, we hope to articulate some of the principles that guide the members approaches to storage systems and architectures.

At the Make it Work conference, the infrastructure group was happy to host Tales from the Crypt: How are we meeting the challenges of large scale storage?, a workshop focused on refining the preliminary results of our work on storage. During the workshop, we shared the responses of an open ended questionnaire from eight members, who represent the diversity of the NDSA membership. During the workshop, we broke into groups to discuss how the responses from these members resonated with the workshop participants experiences with large scale storage. Discussions and commentary from the different groups was quite lively. At one point, there were 35 different individuals collaboratively editing the document!

Some areas of discussion included:

- Member approaches to disk and tape
- Simple file systems and the value of control
- The “when” and “why” of format and system migration
- Member approaches to vendor systems
- Current and upcoming critical characteristics of storage systems for digital preservation

As next steps, the infrastructure group has been refining this document by adding the comments from workshop participants. Ultimately, we intend to broadly share the finalized document, particularly with institutions looking to establish their own large scale storage infrastructure for digital stewardship. Also, we imagine that the resulting document will be of interest to vendors that want to know more about what drives institutional approaches to large scale storage.

Before finalizing this document on the NDSA member's approaches, the working group plans on a few substantive additions to the current draft. Specifically:

- **Creation of a matrix laying out trade offs in different member's approaches:** There was general interest in creating a matrix from some of the documents key points, laying out the pros and cons with different choices and approaches to preservation and stewardship storage. The group plans to draft and integrate a matrix like this into the document.
- **Short targeted survey of the membership:** There was general consensus that a targeted survey of the NDSA membership focused on generating some descriptive statistics to back up or complicate some of the claims we are making in the document would be useful. We set a target to launch the survey in November.
- **Addition of a brief section focused on definitions and terms:** The diversity of the group's membership reinforces the need for some tight working definitions to facilitate common understanding. For example, something as fundamental as the overlapping definitions for "storage system", "preservation system", "storage architecture", "storage management system", and "IT infrastructure". Working group members are also creating a short addition for the document that tightens up our use of terminology.

The group will share, and invite further discussion and comment on, this work in progress at the next "Designing Storage Architectures for Preservation Collections" meeting this September. After that, and some more work from the working group, we will more broadly distribute the resulting report. You can expect to hear more about this project, and any report that comes out of it, here on this blog. Any NDSA member interested in being involved in this initiative can contact the infrastructure working group co-chairs. If your organization is not currently an NDSA member, please consider [joining the Alliance](#).