A Digital Repository Year:
One Museum's Quest for the Basics

Paula Jabloner
Computer History Museum
1401 N. Shoreline Blvd.
Mountain View, CA 95128 USA
1 (650) 810-1016
pjabloner@computerhitory.org

Katherine Kott
Katherine Kott Consulting

katherinekott@katherinekott.com

Keywords

Digital Repository, Digital Preservation

1. INTRODUCTION

The Computer History Museum (CHM) had its own mini deluge of digital data. Our in-house produced high definition oral histories, lectures and exhibition videos were usurping our available server space at over 60 terabytes, with another 10 terabytes of historic digital artifacts including images and software. With the aid of grant funds from Google.org, CHM took on the work of creating a prototype digital repository in one year. The digital repository working group is excited about the possibilities the new repository represents for expanding our digital collection while putting the Museum in the forefront of small cultural institutions creating digital repositories and we hope to share what we have learned with other similar organizations.

We needed to find solutions that could be managed by an organization of our size (less than 50 employees), yet offered the flexibility to handle the wide range of content we collect. The assumptions we used were based on the museum's immediate needs and time constraints.
They include:

- The digital repository will use existing tools and systems
- CHM will not add staff to build a custom solution
- Open source software will play a significant part in the digital repository management solution
- The preservation layer will be built on top of common commodity storage components that are modular and extensible
- The creation of a digital repository is an on-going commitment by CHM

So far we have created policies, have selected and are implementing software and storage hardware. We are now in the fourth quarter of our year odyssey and are ingesting a small sample set of digital objects to test the prototype. We achieved this in a year carefully defined by quarterly phases.

.

2. CONFRONTING THE PROBLEM: PREPARATION

Here we:

- Defined the problem and the catalyst
  - 60 terabytes of Museum produced high definition (HD) video
  - with no sustainable back-up or preservation methods
- Cultivated permanent stakeholders from senior management and the Board of Trustees
- Engaged cross-departmental working group of four digital preservationists

3. CREATING THE PROBABLE SOLUTION: PLANNING

Here we:

- Hired a digital repository consultant
  - The 'authority' she gave the project in the eyes of the stakeholders was invaluable
- Created a concise Project Charter defining scope, objectives, roles, roadmap, and assumptions
- Surveyed the Museum's 'archival' digital objects
- Performed a current literature survey and wrote best practices guide

4. CURATION: POLICY & FRAMEWORK

Here we:

- Wrote digital repository management software functional requirements
- Surveyed and test drove open source digital repository management software
  - Selected *Archivematica*
- Recruited and hired a storage infrastructure consultant
- Explored storage options, configurations, and pricing
- Completed a policy document

5. COMPLETING THE PROTOTYPE (ONGOING)

We are:

- Testing the DIY storage infrastructure (hardware & software stack)

o installing storage infrastructure and *Archivematica*
- Ingesting test digital objects while creating procedures document
- Writing a 5-year sustainability plan
- Exploring avenues for year two funding for ingest, full deployment, and prototyping an on-line interface

6. STORAGE INFRASTRUCTURE

We firmly believe the straightforwardness of the storage infrastructure will guarantee the sustainability of the digital objects entrusted in its care. This DIY infrastructure is comprised of:

- Working space for backups of non-ingested digital objects and archive space on the same infrastructure totaling 256 terabytes of raw storage.
- Supermicro storage using 3 TB SATA drives running either *NexentaStor* or *FreeNAS*.
- Two backup servers and Supermicro storage with fewer terabytes. One on-site and the other at our off-site storage facility.
- LTO 5 tape backups with tapes stored in both locations.
- Main server running Archivematica, rsync and other software.

7. CONCLUSION

The working group is excited about the possibilities the new digital repository represents for expanding our digital collection while putting the Computer History Museum in the forefront of small cultural institutions creating digital repositories.

Our lessons learned are that the three most important ingredients were setting the correct expectations from the beginning, adequate planning with an emphasis on quarterly results, and having the right team in place that was both dedicated to the project and with the right mix of experience, talents and abilities. This truly took a team effort.

8. ACKNOWLEDGEMENTS