

A Digital Repository Year

“For the purposes of this paper, and to meet CHM needs, we will define the digital repository as the systems and workflows that support *digital asset management* and *digital preservation*.”

Digital Repository Best Practices for Cultural Heritage Institutions

Katherine Kott 2012



Paula Jabloner

Over the last 10 months, The Computer History Museum, is finally seriously dealing with preservation of digital data - though we've been collecting for 25 years – we are truly a work in progress.

Our goal is to build a prototype digital repository in a year. So is the Computer History Museum attempting the impossible, only a year? We're a hair's breath away from achieving our goal? So how did we get there?

our journey unfolds



out of chaos and into orderliness & learning??



The museum was at a **precipice**. We were creating lots of high definition video productions [that is at the rate of 20+ TBs year or 1 GB per minute of filming, with 40+ already in the can] but had no sustained method to back up these extremely large files - let alone preservation methods.

Over the last year we've "**misplaced**" some digital files maybe never to be located again.

We all realized this needs to be a **priority**, our video production was not going to lessen and in fact seems to ever increase.

Computer History Museum by the numbers



1st permanent exhibit opened in 2011

60+ terabytes of digital objects including our truly unique historic software collection

15 terabytes anticipated yearly rate of growth

500+ oral histories

3,000 moving images

5,000 linear feet of archival materials

10,000 photos

35,000 artifacts



In the entrepreneurial and innovative spirit of silicon valley we dove in hoping for at least a bronze medal with minimal splash.

Where do you start such a project?

Upper management did some networking. I helped write a grant. We got funding for 1 year – but not the 2 we requested at $\frac{1}{4}$ of the anticipated project cost. Time to rethink our strategy! Let's build a prototype only.

First concentrated on backing up our current assets while engaging in the planning & documentation process. The year was all about creating policies then building & testing the infrastructure. Ingest into the digital repository would only be a small representative set of digital objects

As our software curator - a 20 year survivor of engineering at Apple computer - once said: "This is the most complex project I've worked on [not for the technical issues] but because there are so many moving parts."

Aug. 1st represented 10 months into the project. So what have we learned, and where are we heading.

The 2 Ps



PLANNING

- Quarterly goals & reports
- Documentation
- Sustainability

PEOPLE

- External & Internal Talent
- Support



We started by surveying the museum's ecosystem, there were 2 major ingredients in our Digital Repository.

First up

PLANNING

An quarterly emphasis of **review**, reporting & goals to keep all focused
Documentation – making sure that we have clear and concise written explanations of why we went forward in a certain way.

I put a special emphasis on these two because--

Sustainability -- in a museum that likes to think of itself as a “start-up” --hasn't come easily and the whole project revolves around long-term thinking. Taking the long view can be very hard in our short attention span cycle. We need to be more than the flavor of the week --which essentially was how the project was started.

PEOPLE – A digital repository is not a solitary endeavor!! All those ‘moving parts’ precludes this.

Even with all the computing power available this is really about right people who have keen analytic abilities & the desire to get preservation done.

Quarterly Roadmap

Q1= planning

Q2 = Framework & requirements

Q3 = set-up the hardware & software infrastructure

Q4 = Implementation

Q4 2011 - completed	Q1 2012 - completed	Q2 2012 - completed	Q3 2012
Complete digital preservation literature survey, HY	Software system (DAM) survey, HY & team	Select & install software system (DAM) solution, HY & team	Final policy & procedures manual, HY & team
Inventory of current digital holdings, HY & TL	Draft storage infrastructure requirements & recommendation, TL & TBD consultant	Select, purchase & install storage infrastructure, TL, TBD consultant, & team	Implement prototype system (infrastructure & software), team
Back-up HD video collection, TL	Stakeholders functional needs survey, HY & team	Stakeholders report, PJ & HY	Training & education program, team
	Best practices document, KK	Final consultant's report, KK	5 year sustainability plan, PJ
			HD video & software collection on stable storage environment, TL



Ist up Planning

Our roadmap has tasks defined by each ¼. The 1/4ly reporting provides the feedback mechanism to make sure we're all traveling the same road and forces us to agree on the next set of deliverables. We also meet weekly, always at the same time to discuss all current issues & objectives.

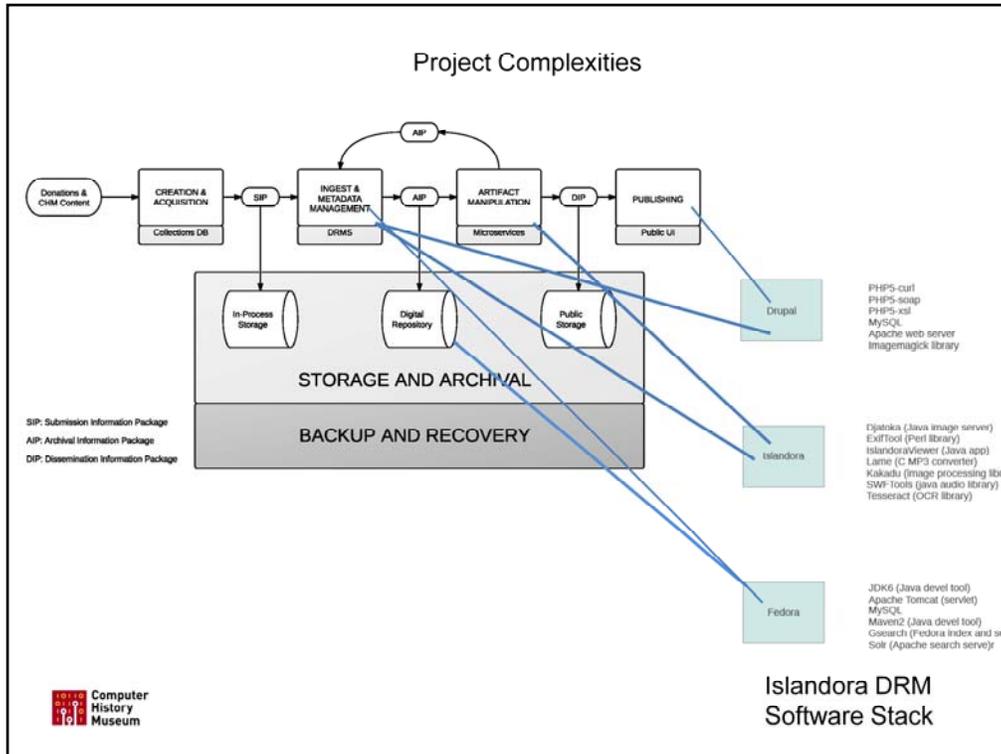
Roughly the 1/4ly objectives are:

Q1= planning

Q2 = Framework, requirements

Q3 = set-up the hardware & software infrastructure

Q4 = Implementation



Documentation is important, as the next layer in our carefully crafted cake. Without documenting our decisions and mechanisms to move forward we would just be lost considering all the complexity of the project. For example, something I had never realized prior to this project, as you can see in the diagram, was the project stack of how all parts need to fit together precisely. We can't just buy any disk drives (cheap expensive, how many terabytes etc.) and assume they will take care of what we need. Likewise it is the same for software. This is only 2 of many project diagrams we've created.

Assumptions- based on immediate needs & time constraints:

- use existing tools and systems
- Open source software
- No additional resources
- preservation layer will be built on common commodity storage hardware that are modular and extensible
- creation of a digital repository is an on-going commitment by CHM



No punch cards!



Computer
History
Museum

Our most important document to date, is our **project charter** which sets expectations over the project year. The charter includes, the road map, assumptions going into the project, staffing, objectives & deliverables, purpose, and scope. As an organization with a short attention span the details and assumptions are very important. The 2 page charter was purposely shared with as wide an audience as possible with the expectation that everyone has the time to read & review 2 pages.

The most important element of the charter is to be as clear as possible about the **assumptions** going forward. For us that meant basing it on CHMs immediate needs & time constraints, which dictated the use of:

existing tools and systems

Open source software

Internal resources available will be constant but not expanding

The preservation layer will be built on stuff relatively readily available that is common commodity storage components that are modular and extensible

This is an on-going commitment by CHM

This is a prototype system only

Project Charter Year One

Objective		Deliverable
Understand the scope of CHM digital collections	✓	Inventory of current digital holdings
Stabilize interim storage for digital collections	1/2	New interim storage system installed
Identify and document best practices	✓	Consultant's report
Establish policies and procedures for managing and preserving digital collections	✓	Policy and procedure manual
Ensure staff are familiar with digital collection management policies and procedures		Training & education program for collections staff based on policy & procedures manual
Evaluate and select software (digital asset management) system	✓	Core Team software recommendation
Evaluate and select digital preservation (storage infrastructure) management system	✓	Core Team preservation system recommendation
Model and test implementation of digital object management and preservation workflow	✓	Prototype system in place to support all phases of digital object lifecycle
Share project findings with other organizations	✓	Present papers, share findings through blog posts, etc.



The concise objectives and clearly follow on from the assumptions about what is actually possible. We've completed in some fashion 8 of 9 in the last 6 months.

Following on from the roadmap –our most important accomplishments to date:

best practices & literature survey

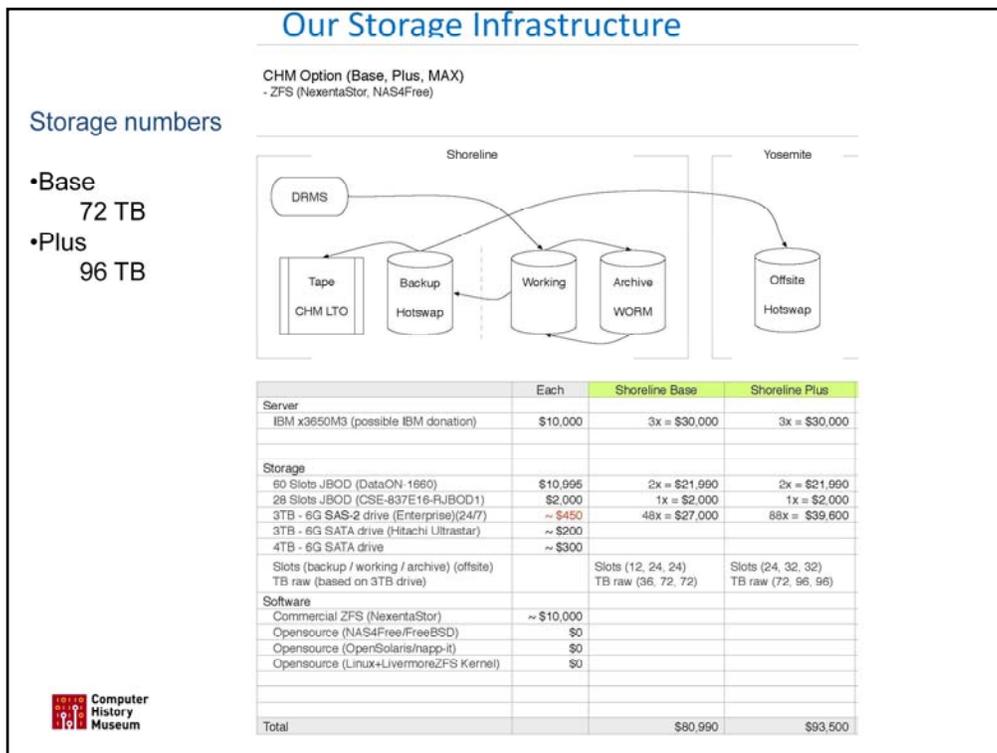
created a **project plan**,

Conducted **informal meetings** or interviews with all the stakeholders, we have buy-in

Chosen **archivematica** as our DRMS after completing a functional requirement report,

Created a **policy document** that defines and guides the DR1 far into the future. With an emphasis on appraisal and ingest into the DR. Establishing clear collecting guidelines at CHM as been an issue in the past. Therefore nothing can be ingested into the repository unless we have clear legal rights with a signed deed of gift or a Memorandum of understanding.

Currently **testing & building** our storage infrastructure



And what is the storage stack? A storage consultant provided us both a framework for thinking about our system architecture and a sounding board for our proposed ideas throughout system development. But the architecture was created by in-house by CHM's IT professionals.

Pros of CHM-DIY Solution:

Cost effective;

Offers **scalability** and relies on affordable commodity hardware;

CHM will have "**ownership**" of the solution and the knowledge that comes with this ownership;

Open architecture ensures flexibility over time;

Open source software ensures longevity, ease of migration and upgrades, and community support;

And on top of it all we're using the open source software

Archivematica because:

Closest match to CHM needs, meets >80% of our requirements, in particular

Automatic **metadata extraction** & encoding tools;

Digital preservation tools;

Batch import & export features.

Easy to install and administer, in-house talent

Active community of developers and users

Comprehensive **Documentation**

Can **contract** with primary developers to add needed features

The 2 Ps



PLANNING

Quarterly goals & reports
Documentation
Sustainability

→ PEOPLE

External & Internal Talent
Support



Computer
History
Museum

INFRASTRUCTURE ONLY TAKES YOU SO FAR....

2nd up People (core team mention?)

Having the **right team**, is the most important ingredient in creating the DR, more so than planning. Considering all the moving parts and our novice status, we've drawn talent both **internally & externally**.

Our inspiration - The Speak & Spell Design Team, 1978



Gene Frantz, Richard Wiggins, Paul Breedlove and Larry Brantingham.



Computer
History
Museum

The In-house core team pulls from across the institution centered around collections & IT staff, all with an interest in preservation.

The roles we perform were based on our level of interest and abilities & defined clearly in the project charter. I oversee all things related to the DR making sure everything goes smoothly and within our time table. My goals are:

Respect the views of those involved as experts, but you still question.

Engage stakeholders to ensure continued institutional support and participation in digital repository development;

Lastly, build visibility for DR

It boils down to **coerce & communicate with everyone as much as possible.**

Other Core team members are the:

Digital media archivist Heather

DRMS software;

IT manager -

storage **architecture** implementation

Software Curator -

Designed & tested the prototype system

Hiring outside consultants was critical based on our time constraints & staffing situation.

It might be, our **best decision** was hiring, Katherine Kott, as our DR consultant. She has great people skills to dialogue with & understand the concerns of all the participants and her background {SDR} provides the knowledge base to claim authority over the entire project especially for high level stakeholders. In a word she has been invaluable.

For me the most important outcome is making sure all the talent you have available is being **listened too, respected and incorporated** as required.

Of Course You Can



Communications

Communications is as important as technical know how in this endeavor. Its all about building support across the board. CHM has a very informal organizational structure meaning we inform the CEO, chair of the board and an executive committee member regularly.

Sustainability



A homeless dignitary visits the Computer History Museum



Sustainability

My current communications challenge is to get all the stakeholders to agree what is needed to sustain the new repository.

For those of you who are old enough to remember the Pets.com from the internet boom years we had a visit from the sock puppet - who is currently **homeless** – earlier this month. We don't want our data to fall into the sock puppet's fate.

I hope to:

articulate a **funding** stream to allow for years 2 & 3 of development. I always seem to be writing grants lately.

create a prototype that works within our small organization making it **easy to maintain & manage**. At all major decision points, this is the fundamental question.

make people happy by providing **access** - so funding is easier to procure

Promoting the work of the DR both internally and externally.

If she can only cook as well as Honeywell can compute



Computer
History
Museum

Conclusion

Therefore we need to last longer than the Kitchen computer. Neiman-Marcus's 1969 Christmas catalog featured the Kitchen Computer for recipe storage & retrieval. For a mere **\$10,600** you got the computer, a cookbook, an apron, and a two-week programming course.

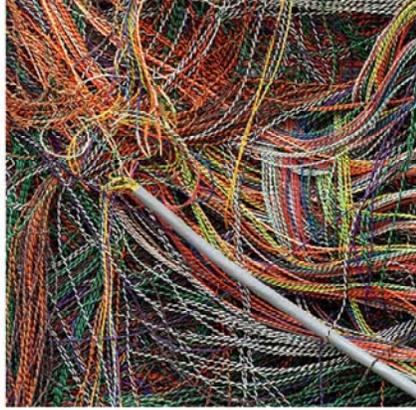
Surprisingly, not one was sold –40+ years later many Ipads are in kitchens. But the Honeywell 316s only exist in museums.

In conclusion –

the museum is unique in our **short timeframe** for implementation
& **small size**.

Thank you

Paula Jabloner ---- pjabloner@computerhistory.org



Backplane wiring of the CDC 7600 supercomputer, c. 1969

The wiring was done by hand and had to be precise to within a fraction of an inch to optimize the speed of the machine



UNIVAC metal tape reel, 1950s
Each metal tape held about 3MB