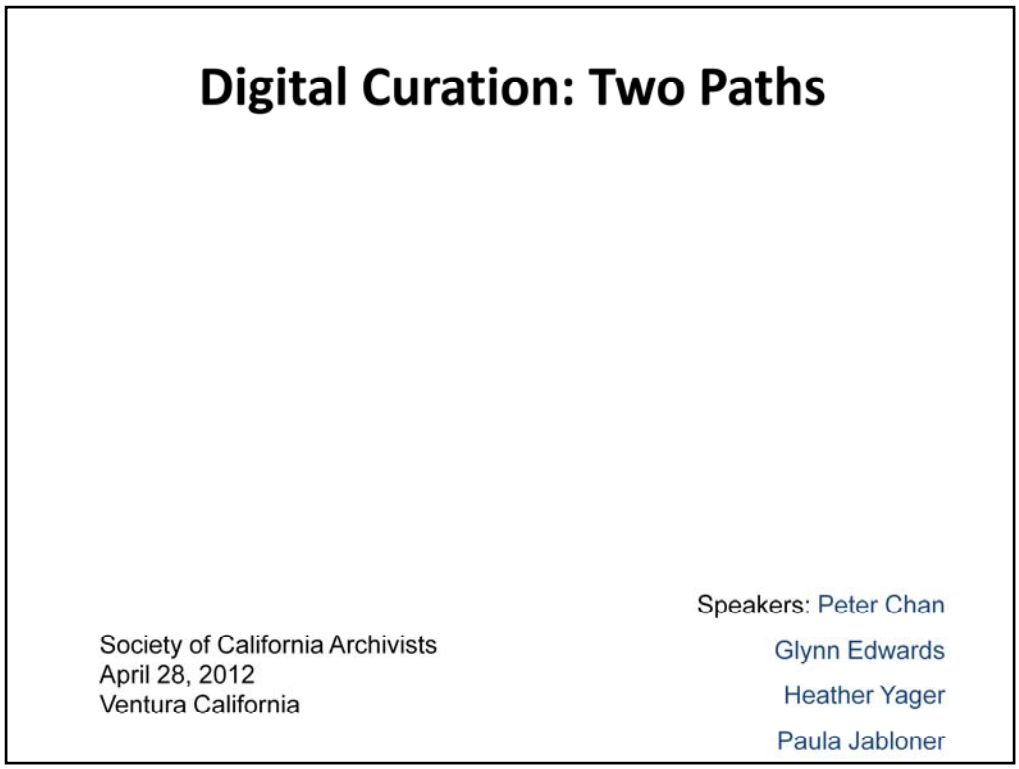


Heather is Digital Media Archivist at the Computer History Museum engaged in building stable storage and a software platform to house the Museum's digital assets. Previously she worked at the Academy of Natural Sciences in Philadelphia as a database programmer and system administrator and the Xerox historical archive.

Glynn is in charge of collection services and metadata for both the manuscripts division and the born-digital program in Stanford Universities Department of Special Collections & University Archives. She came to Stanford in 2006 from Harvard's Schlesinger Library for the History of Women, Radcliffe Institute.

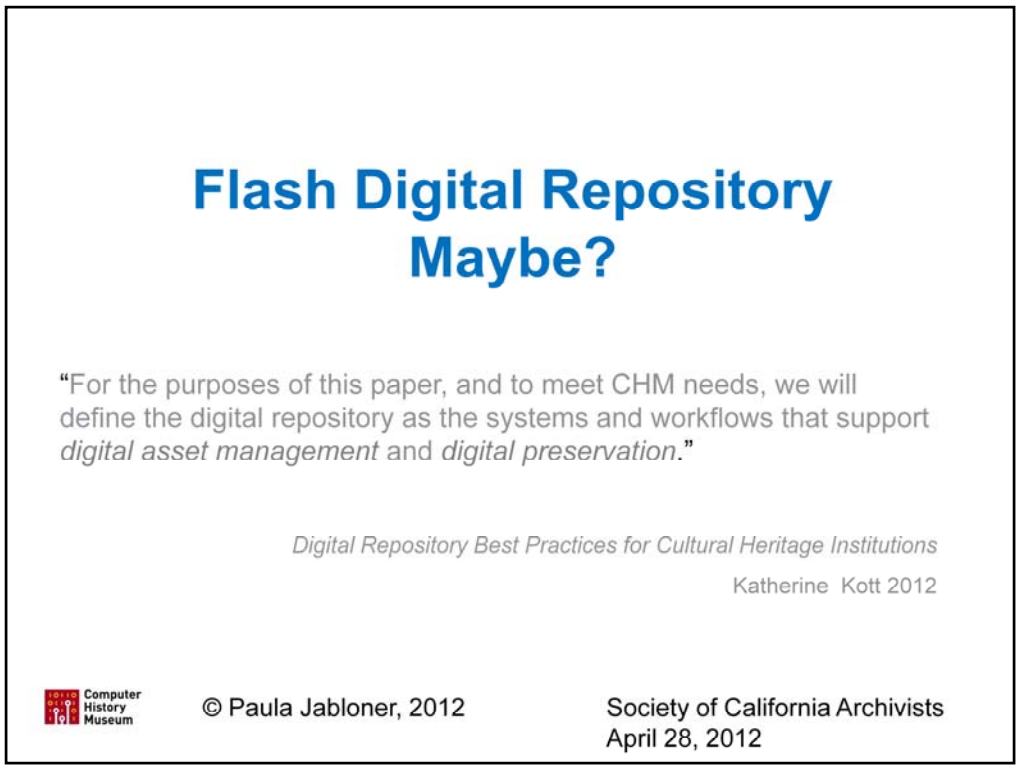
Both Glynn & our next speaker Peter Chan were part of the AIMS Born-Digital Collections grant awarded by the Andrew W. Mellon Foundation – a four-institution collaboration with the University of Virginia, Yale University, and the University of Hull.

Peter is responsible for developing workflow to process born-digital archives in the Department of Special Collections & University Archives at Stanford University. He also manages the digital forensics lab.

As **Director of Collections**, I am in charge of the Computer History Museum's collection in all formats- artifacts, paper, AV & digital objects. Previously, I started the archival program at CHM. Prior to CHM, I was project director for the collaborative Silicon Valley History Online digitization project. I've worked as an archivist at History San Jose, GLBT historical society, and Princeton University.

You might be wondering what the definition is of digital curation? The book *Digital Curation* lists over 4 different definitions. That said the simplest definition is that it looks across the entire lifecycle of digital data from creation through preservation, reuse or disposal. We'll present 2 different stages & approaches in the digital curation lifecycle. We'll start with CHM as we begin seriously dealing with preservation of digital data even though we've been collecting for 25 years – we are truly a work in progress. The Stanford folks have been in the preservation business longer and will outline their high level workflow for the born-digital program followed by a more hands on approach to working with the various physical media in their forensic lab.

At the end or our presentation we'd really like to hear your stories, concerns with, attempts at or anticipated progress in curating & preserving the digital objects in your collections.



Sometimes I feel the Computer History Museum is attempting the impossible. Build a prototype DR in a year while digitally curating our in-house produced video. Is it possible? We'll see.

our journey unfolds

http://blog.emeals.com/2010/11/cyber-monday-gift-giving/

out of chaos and into orderliness & learning??

Computer History Museum

The museum was at a precipice. We were creating lots of high definition video productions but had no sustained method to back up these extremely large files - at 1 gigabyte a minute of filming - let alone preservation methods. In fact over the last year we've "misplaced" some digital files maybe never to be located again. But that is another session topic. Hopefully I can provide somewhere between a teaspoonful to a pot of helpful advice from our Digital Repository recipe.

Computer History Museum by the numbers

© Mark Richards, **Soroban, with Sharp EL-429 Electronic Calculator, Computer History Museum** 102626585p

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

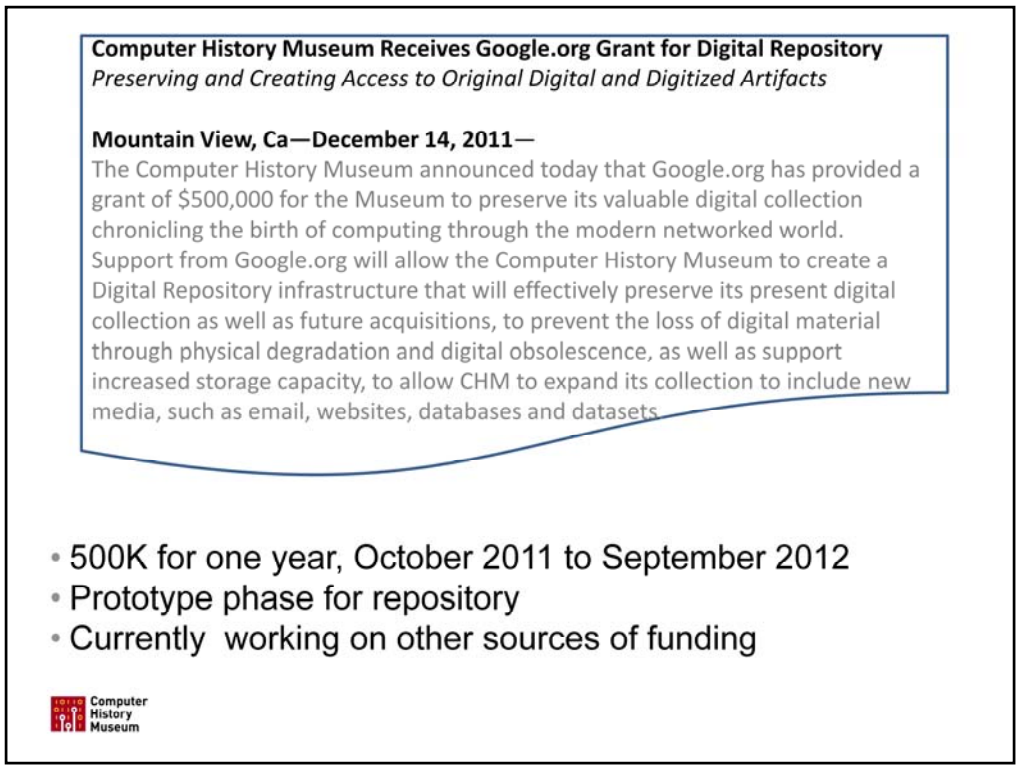
5,000 linear feet of archival materials

10,000 photos

35,000 artifacts

Computer History Museum

The museum itself is relatively small with 50 employees and operates much more as a "start-up" at only 15 years in Silicon Valley - unlike our venerable and long standing neighbor to the north, Stanford. In the entrepreneurial and innovative spirit of silicon valley we jumped in hoping for the best.

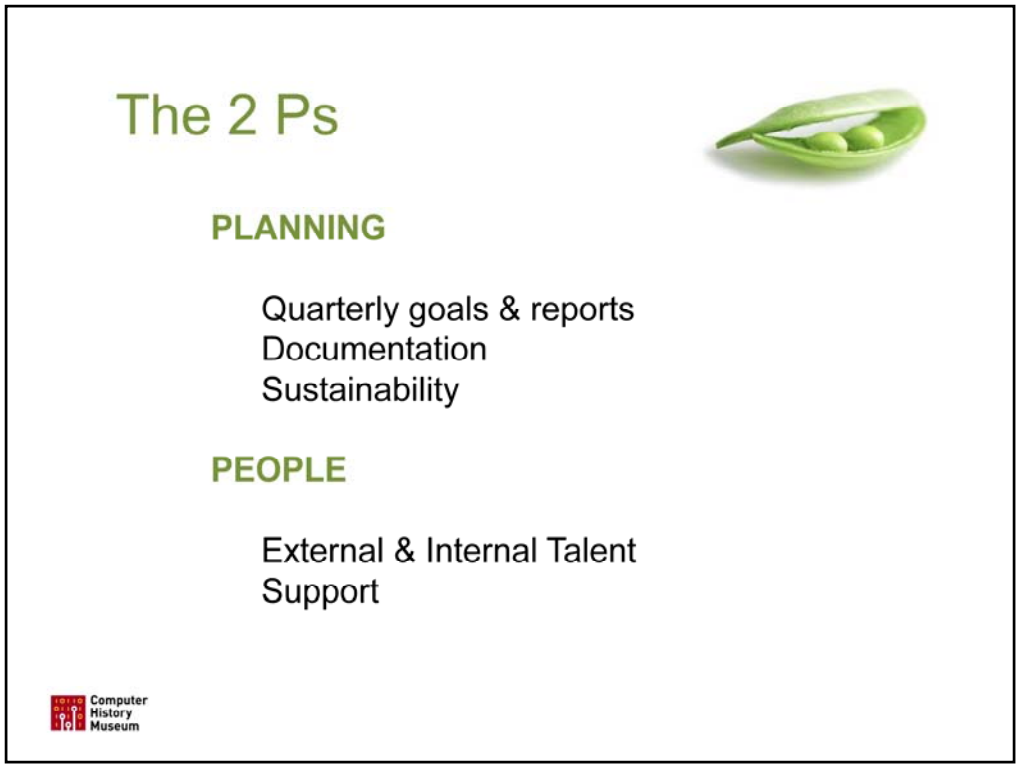


Upper management did some networking. I helped write a grant. We got funding for 1 year – but not the 2 we requested at ¼ of the anticipated project cost. Time to rethink our strategy! Let's build a prototype only.

We concentrated our scope on backing up our current assets while engaging in the planning & documentation process. Ingest into the digital repository would only be a representative sample set of digital objects. In essence we're assessing what our potential is as we entered the DR marathon. So far we're keeping pace. But I must say, this is the most ingredient heavy recipe, 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."

April 1st represented 6 months into the project.



In surveying CHM's ecosystem, there were 2 major ingredients in cooking up our Digital Repository recipe.

First up

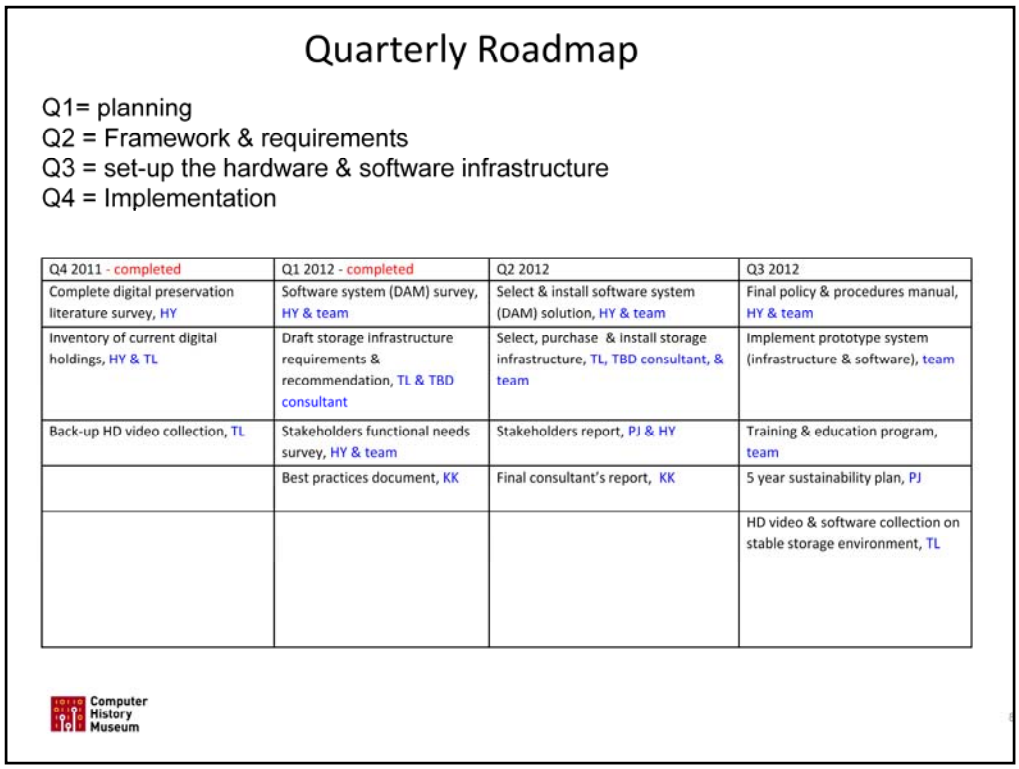
PLANNING

An quarterly emphasis of review, reporting & goals

Documentation – making sure that we have clear and concise written explanations of why we went forward in a certain way.

Sustainability - in a museum that likes to think of itself as a "start-up." 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.

PEOPLE – 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	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

Computer History Museum

Ist up Planning

New concept & project, short time frame, competing demands – It was about the planning process - but not necessarily the actual plans. We all had to be involved & agree as much as possible. I was hoping for no surprises but expecting many twists & turns, which has turned out to be true.

The 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.

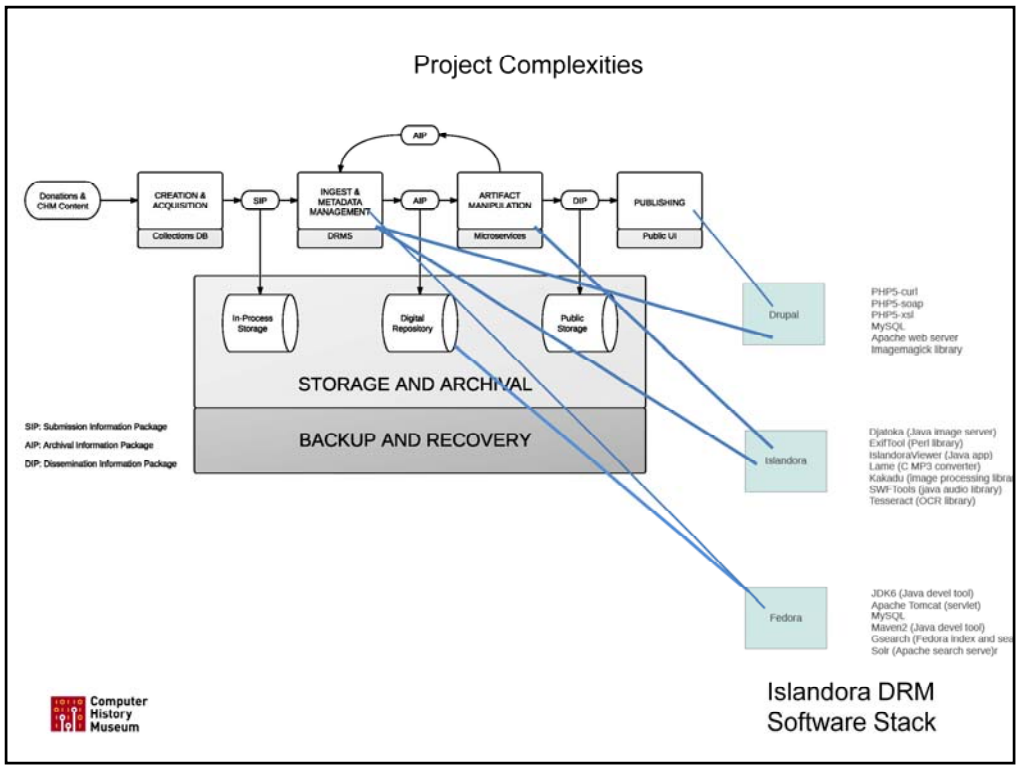
Roughly the 1/4ly objectives are:

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 by any disk drives 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.

I think 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

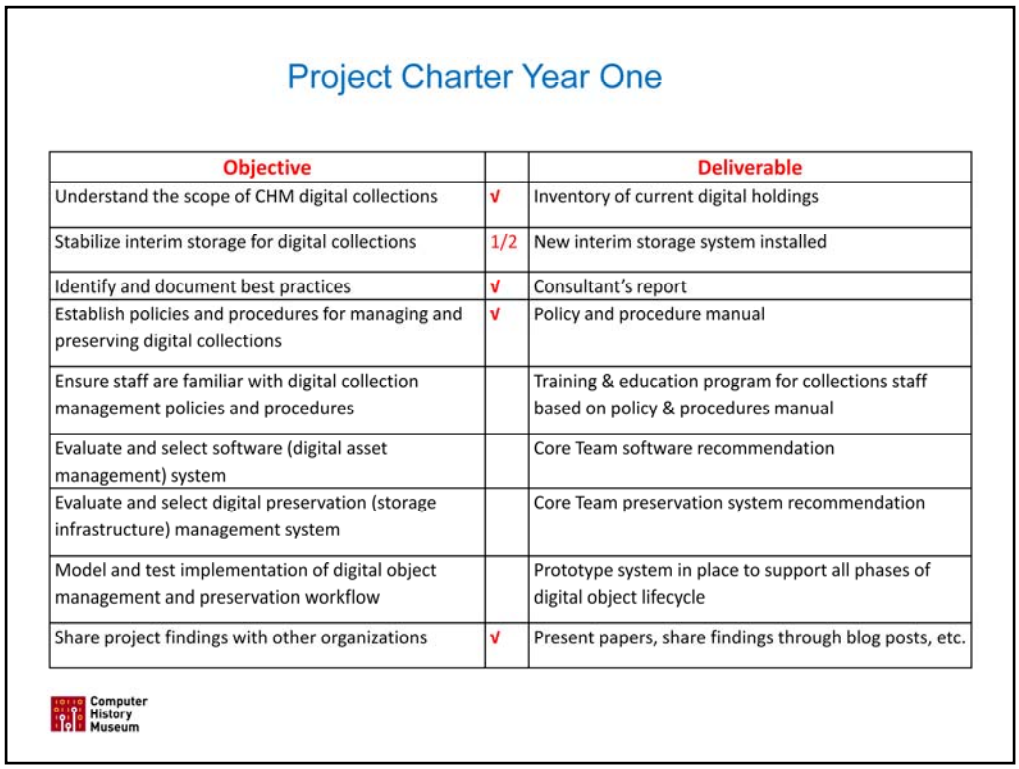
Internal resources available will be constant but not expanding

The preservation layer will be built on stuff relatively readily avialable

that is common commodity storage components that are modular and extensible

This is an on-going commitment by CHM

Prototype system only or a planning & proof of concept phase



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.

Computer History Museum

The objectives as outlined here are meant to be concise and clearly follow on from the assumptions about what is actually possible. We've completed in some fashion 4 of 9 in the last 6 months.

Following on from the roadmap – in Q1 & 2 our most important accomplishments are :

best practices

created a **project plan**,

Conducted **informal meetings**/interviews with all the stakeholders

functional requirement report, outlining the needs for hardware & software

Created a policy document that defines and guides the DR1 far into the future. The document creates a nimble structure that will allow for small changes over time - 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.

Planning - Sustainability

For those of you who are old enough to remember Pets.com in 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. This is probably my most important current objective but the least fleshed out.

I hope to:

articulate a funding stream to allow for years 2 & 3 of development. We're currently waiting on 2 grant opportunities

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.

Finalize a Policy document that looks toward the future.

complete a report to stakeholders - outlining the needed resources over the next 5 years

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

Promoting the work of the DR both internally and externally.

The 2 Ps

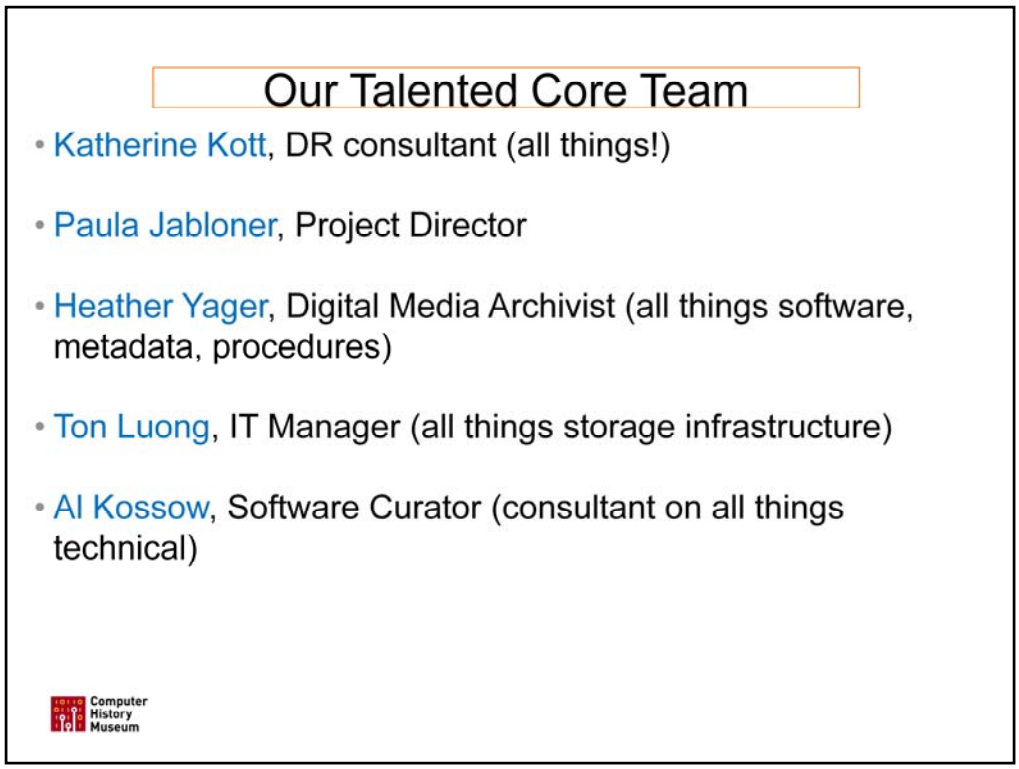
PLANNING

Quarterly goals & reports
Documentation
Sustainability

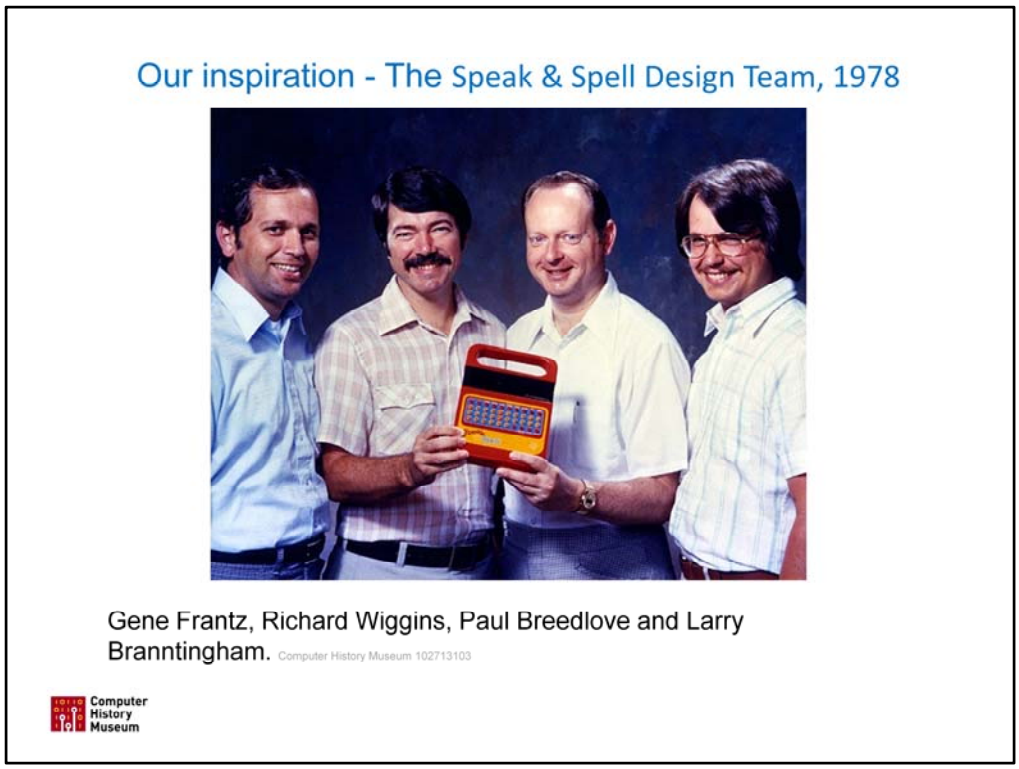
→ PEOPLE

External & Internal Talent
Support

Computer History Museum



Having the right team, is the most important ingredient in creating the DR, more so then planning. Considering all the moving parts and our novice status, we've drawn talent both internally & externally from many sources depending on availability and our needs. The In-house core team is not just collections staff but pulls from across the institution. No one was forced into this project and all are interested in preservation. In reality, the project is comprised entirely of consultants - that we brought together creating a common road forward. Luckily the core team members are all interested in learning and respect the talents of the others, something that has come with working together over the past few years before the project started.



The roles we perform were based squarely on our level of interest and abilities & defined clearly in the project charter. As the executive chef, 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 can 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 include:

Heather (our software huntress) is the very busy sous chef,

Take lead role in reviewing, surveying and testing possible DR management software;

report on literature in the DR field to help guide work of team

will Draft procedures for processing digital artifacts;

IT manager -

Oversee selection of the storage hardware with input from the storage consultant

Develop standards and policies for backups, and data recover

Software Curator -

25+ plus years of technical expertise, in effect our in-house technical consultant

He is our compass & sounding board – continually proding us to think creatively & in-depth

Hiring outside consultants was critical based on our time constraints & staffing situation and they needed to have long-term experience & expertise with digital repositories or storage infrastructure.

It might be, our best decision was hiring ex-Stanford employee, Katherine Kott, as our DR consultant. She has great people skills to dialogue with & understand the concerns of all the participants and her background provides the knowledge base to claim authority over the entire project especially for high level stakeholders. In a word she has been invaluable. I understand not everyone can afford consultants or may have more in-house expertise but much of our project's success comes with hiring & collecting the right people.

We've also brought on a paid storage infrastructure consultant, & on a board member's introduction we've met with a number of pro-bono consultants. We've visited with staff at other repositories as well. Many times these discussions confirmed we were on the right course giving us confidence as we move ahead.

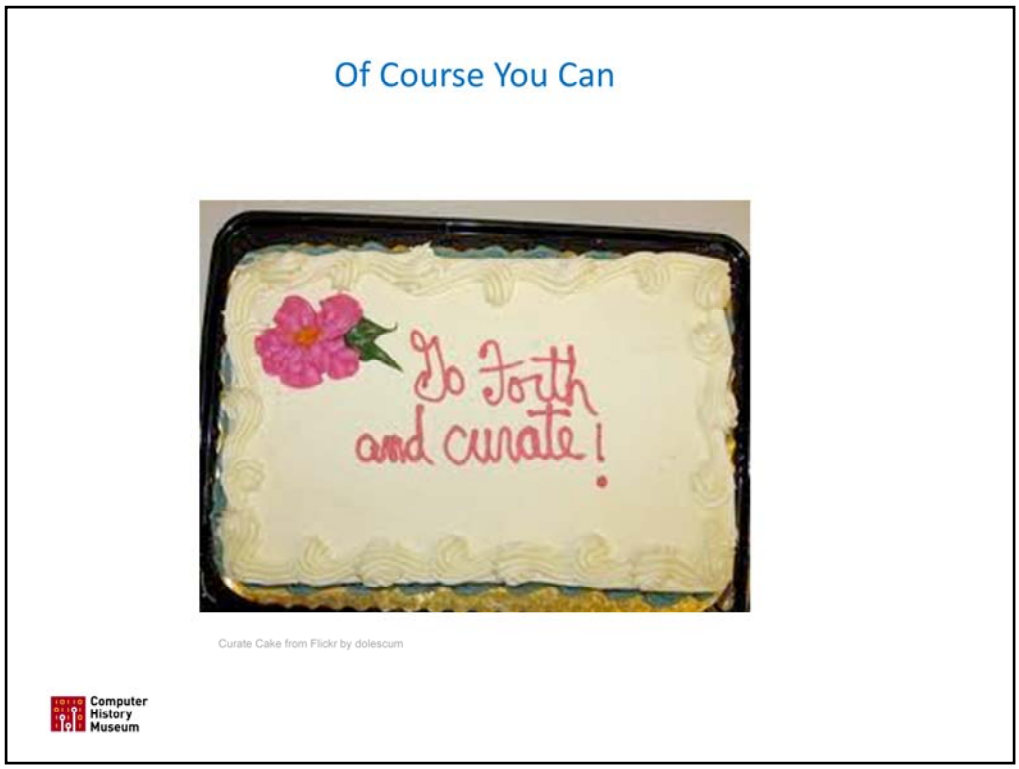
For me the most important outcome is making sure all the talent you have available is being listened too, respected and incorporated as required.

Support

It seems building support might be is the most hazardous part of our roadmap. People's on the fly expectations can be very different from the reality you deliver. Likewise you "assume" everyone is in agreement only finding out too late they're not.

CHM has a very informal organizational structure. So we inform the CEO, chair of the board and an executive committee member regularly. At first impression this seems like a distraction, but I'd argue it helps guarantee the long-term success of the project. You want to have the right people on your side at every level. With such fluid and open communications setting expectations is required, & as mentioned earlier, our project charter was great for this.

This is still a work in progress, so many of the assumptions can be derailed when we're not the shiny new thing. Support is an on-going cultivation.



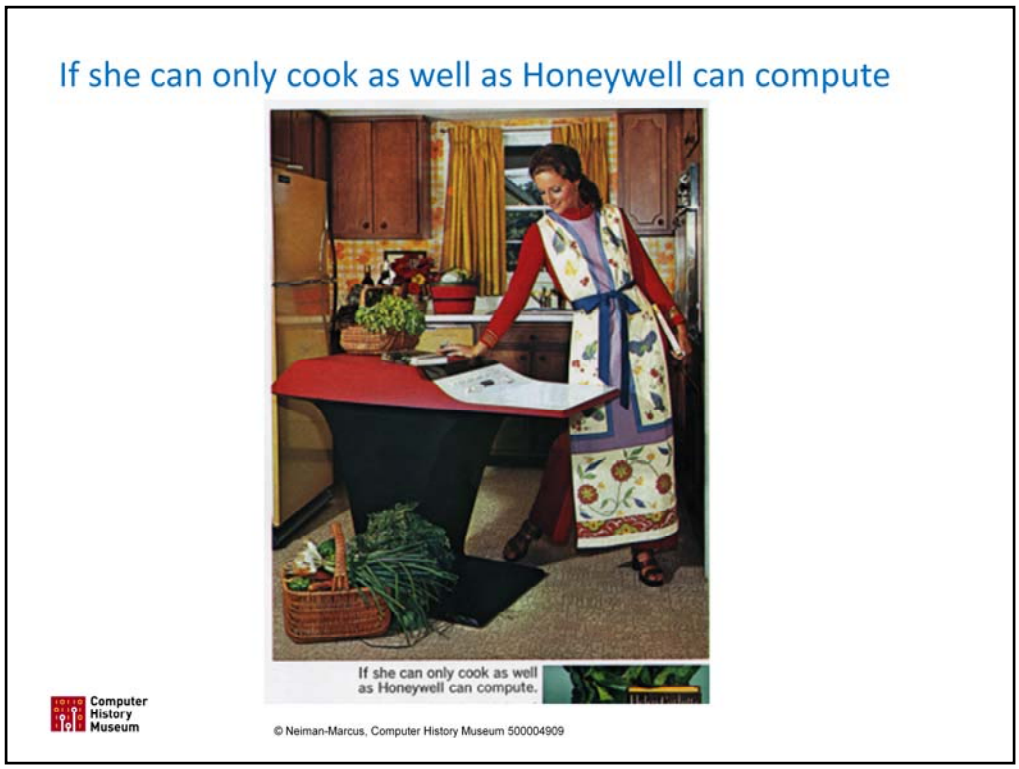
Current steps

If our plans and people continue to be productive the recipe for the next 6 months calls for:

Implementing & choosing DRM software & storage hardware

Ingesting sample digital objects from our collection

Lastly, I'll say it again, communicate openly & continually with everyone involved.



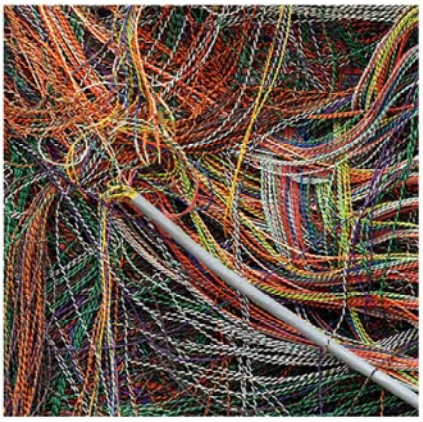
Conclusion

The cover of the Neiman-Marcus's 1969 Christmas catalog featured the Kitchen Computer for recipe storage & retrieval. For $10,600 you got the computer, a cookbook, an apron, and a two-week programming course.

Not one was sold –40+ years later many Ipads are in kitchens but no Honeywell 316s. Paper's prevalence for storing recipes and archives has passed into history. I'm hoping you at least got the flavor of how you can go about creating your own recipe for protecting the digital objects in your collection.

Thank you

Paula Jabloner ---- pjabloner@computerhistory.org



UNIVAC metal tape reel, 1950s
Each metal tape held about 3MB
Computer History Museum 102673949

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
Photo © Mark Richards, Computer History Museum 102695455

Please feel free to contact me with any questions you might have on creating a digital preservation framework in your institution. I'm happy to share documents and many I mentioned in this session will be available soon on CHM's new blog.

Now Heather will give you a window into her magic road trip selecting digital repository open source software.