



**"The Attic & the Parlor:
A Workshop on Software Collection, Preservation & Access"
Computer History Museum
Mountain View, CA
May 5, 2006**

List of Participants

Moderator: Henry Lowood, Stanford University

Participants:

Lee Courtney, Member, CHM Software Collections Committee
Margaret Hedstrom, Associate Professor, University of Michigan
Sellam Ismail, Private Collector & CEO, Vintage Tech
Paula Jabloner, Archivist, CHM

Linda Johnson, Volunteer, CHM
Al Kossow, Robert N. Miner Software Curator, CHM
Paul Lasewicz, Archivist, IBM Corporation
Henry Lowood, Curator for History of Science & Technology Collections, Stanford University

Paul McJones, CHM Software Collections Committee & Adobe Systems
Bernard Peuto, Chairman, CHM Software Collections Committee
Paul Pierce, Private Collector
Michael Powell, Member, CHM Software Collections Committee

Tim Shoppa, Private Collector, Trailing-edge.com
Len Shustek, Chairman of the Board, CHM
Olin Sibert, Multics project, MIT
Dag Spicer, Senior Curator, CHM

Amy Stevenson, Archivist, Microsoft Corporation
Kirsten Tashev, VP of Collections & Exhibitions, CHM
John Toole, Executive Director & CEO, CHM

Introduction

The suddenness of the leap from hardware to software cannot but produce a period of anarchy and collapse, especially in the developed countries.

--Marshall McLuhan

Not everything that can be counted counts, and not everything that counts can be counted.

--Albert Einstein

This outline describes several themes explored during a [Workshop on Software Collection, Preservation & Access](#) that was held on May 5, 2006 at the Computer History Museum. The workshop's purpose was to gather a small group of leading practitioners in the area of historical software preservation methodology to help shape the strategic direction and practical next steps for long-term software collection and preservation efforts, particularly at CHM but also elsewhere.

What is meant by "The Attic and Parlor?"

We suggest that the issue of software preservation may have a dualistic ethos: first, that of a broad, community-based effort to collect for preservation with little or no curatorial/interpretive layer; and, second, that of a more narrowly-focused effort to collect, preserve *and present* the more seminal instances of software, that is within a framework for understanding beyond that of raw primary source.

The broad-based effort we here term 'the attic;' the narrowly-focused one, 'the parlor.' The analogy, firstly, is to that of a home's attic, where many undifferentiated, unsorted items often are placed, awaiting further examination at some later date, perhaps by later generations; secondly, the parlor as that area of the home in which the most important items of a person's life are kept lovingly-preserved, and shown with pride as embodiments of the owner's collecting prowess, interpretive expertise, and desire to enlighten others.

In our view, both philosophies are necessary and mutually-reinforcing but we offer the contrasting approaches as a useful way to open up a conversation about collecting and strategy.

The Attic

The attic represents a community-based collecting function. It is a mission to collect items of relevant interest in as broadly-based manner as possible. This effort makes use of traditional methods of artifact collecting (both passive and active) by which donors approach collectors and collectors approach known holders of software, but which is also, excitingly, able to use the Internet to attract donors from around the world, greatly increasing the reach and quantity of materials that can be collected.

The goal is the creation of a massive repository of software of all kinds and in all forms. There are obvious scalability and quality control issues in such a strategy, which this workshop can explore in detail.

The attic is a messy place; among the gems will be duplicates, trash, counterfeits, incomplete works, and objects of questionable provenance. One allied view of the attic is that of the archaeological midden¹ from which objects can be extracted and promoted to the Parlor, now or later.

The Parlor

¹ Midden: "A mound or deposit containing shells, animal bones and other refuse that indicates the site of a human settlement. A midden can be a rich source of information about the prior human activity at the site."

The distinguishing features of the parlor are that of thoughtful selection and interpretation. Items move into the parlor only by careful, deliberate volition of their owner. The analog to the parlor in the Museum is its curated and cataloged collection, which provides the source material needed for exhibitions and historical research. The Museum ought to pursue in depth the origin and development of specific, highly-important software in a scholarly fashion, as items that befit its own parlor.

The parlor is not a messy or necessarily voluminous place. It needs only a small number of significant collections or objects — ones that reflect the deeper archeology of software — in order to be a resource for quality research and exhibition. High priority is placed on collecting not only the software itself, but a wide variety of other related objects and documents that contribute to the historical record.

Common features

Museums are often unfairly accused of being a 'black hole,' a place where many physical items that are donated seem to not re-appear for use. To be fair to museums, this usually reflects a poor understanding on the part of the plaintiffs of what a Museum really is (viz. the fraction of items that are collected vs. those that are displayed) as well as the staffing challenges that museums typically face.

Because software is different, both the Attic and the Parlor can serve as a means of countering this problem by making all contributed material accessible to the extent permitted by law and technology. This directly serves the public mission of the museum, and helps in attracting still more donations of software (particularly among scholarly donors) by establishing itself as an institution which will make meaningful use of such donations.

Both collections encourage and enable the exploration of the history of software and its 'impact on the human experience.'

The actors

Creating both attics and parlors takes considerable effort. Who does the work? It is tempting to be simplistic and declare that parlors are created by professionals within institutions, and attics are assembled by enthusiastic and independent amateurs. In fact this is neither true nor desirable.

Many exquisite and comprehensive parlors (shrines, actually) for software systems have been created with passion and enthusiasm by knowledgeable users or creators of those systems. Conversely, many institutions – the Computer History Museum among them – have, at least as part of their archive/attic, poorly documented and incomplete assemblages of randomly-acquired but potentially valuable software.

Perhaps the institutions' best course is to embrace and encourage both activities while not expecting to hold a monopoly on either. The institution can uniquely provide stability, organization, resources, and the promise of long-term preservation. Non-institutional participants can provide energy, expertise, passion and access to otherwise hidden materials.

As a distributed group effort, both parlor and attic collecting can learn techniques and pitfalls from other collaborative initiatives such as open source software development and Wikipedia-like resource creation.

Organization of the Workshop

In preparation for the workshop, participants were asked to consider the stimulating questions below. Participants answered one or more (or new questions inspired by them) during 5-15 minute presentations. A free-ranging discussion period followed each presentation. At the end of the workshop, main conclusions of the workshop were developed, including action-oriented policies for use by the Computer History Museum in software preservation.

The Questions:

- What is the structure of, and what are the arguments for, the Attic?
- What is the structure of, and what are the arguments for, the Parlor?
- Which notion is a better metaphor for how to put together a collection now?
 - Should we just try to create a repository for collectors to deposit bits, or do we focus on carefully crafted projects?
 - What is the balance of attic and parlor?
 - Is it necessary/reasonable/possible to do both?
- Who does the work?
 - What roles should/can institutions like CHM (ACM, LOC, IEEE, etc.) play?
 - What roles should/can individuals play?
 - What roles should/can software companies or corporations play?
 - If collecting software is a distributed (“open source”) activity,
 - How are rules established for selection, standards, cataloging, etc.?
 - How is quality maintained?
- In the year 2050, what should the software archive covering the era from 1950 to 2000 look like?
 - What can we do now to create a foundation for such an archive?
- What existing collections are endangered and could be rescued?
 - How do we conduct these rescue operations?
- How do we manage the absorption of existing attics and parlors into ours?
 - (There are organizational as well as technical versions of this question.)
- How do we encourage and organize more informal/amateur collecting activity?
- Is software collecting an activity where collaboration among multiple institutions is helpful, or would the coordination overhead impede progress?
- Do we need to create a new institution or independent initiative to succeed, or would that be counter-productive?
- Who are likely collaborators? If CHM were to take the lead in creating a network of institutional collaborators right now, who would they be?
- How do we ensure international participation?
- What is your advice to CHM?
- What role would you like to play?
- What are the next steps?

Program

<u>Time</u>	<u>Speaker, Affiliation, Presentation topic</u>
<i>CHM Introduction</i>	
08:30 – 08:45:	Len Shustek, "Welcome & Introduction"
08:45 – 09:00:	Kirsten Tashev, "Collections & Exhibitions Overview"
<i>Session 1: The Museum Context</i>	
09:00 – 09:20:	Al Kossow, "CHM's Software Collection Philosophy"
09:20 – 09:40:	Dag Spicer, "The Parlor"
09:40 – 10:00:	Len Shustek, "The Attic"
10:00 – 10:20:	Lee Courtney, "Organizing the Attic, Furnishing the Parlor - Considerations for Moving Forward"
<i>Session 2: Applications/What's Going On Now?</i>	
10:35 – 10:55:	Olin Sibert, "What is the Best Metaphor for how to put together a Collection now?"
10:55 – 11:15:	Paul Pierce, "The Structure of my Online Attic"
11:15 – 11:35:	Tim Shoppa, "The PDP-6 and PDP-10 Software Archives"
11:35 – 11:55:	Margaret Hedstrom, "Emulation as a Digital Preservation Strategy"
12:30 – 12:50:	Amy Stevenson, "Building a Model for Sharing Proprietary Software"
12:50 – 13:10:	Paul McJones, "A Case Study in Software Collection, Preservation, and Interpretation"
<i>Session 3: Philosophy</i>	
13:10 – 13:30:	Sellam Ismail, "Contemplating a Standardized Software Media Preservation Methodology"
13:30 – 13:50:	Paul Lasewicz, "What Would Bill Do? The Business of Preserving Software"
13:50 – 14:10:	Bernard Peuto, "The Open Software Collecting Movement and its Impact on Collecting Institutions"
<i>Session 4: Wrap-up/Conclusions</i>	
14:30 – 16:30:	Wrap-up/Conclusions

Proceedings

Len Shustek: Welcome and Introductions

Shustek provided background on the museum's founding & history as well as an update on the current status of various Museum activities such as exhibits, fundraising, growing number of employees, etc.

He stated that CHM currently has three of the four things needed for a museum: 1) collection, 2) physical home, and 3) lecture series. The fourth missing piece is museum-quality exhibits. He stated that CHM received a \$15M gift from the Bill and Melinda Gates Foundation which will help us achieve the missing ingredient. He further stated that CHM is financially self-sufficient and has been so for the past 11 years. The fundraising goal is to raise \$100M and CHM has \$25M to go. \$50M will be for a permanent endowment.

Kirsten Tashev: Collections & Exhibits Overview

Tashev provided an overview of the last 6 years of collections and exhibitions activity and a view to the next 3 years;

- There are some 50,000 artifacts and 4,000 linear feet of documentation in the collection;
- Collection is growing by ~1,000 square feet/year;
- Collect diverse types of materials to assemble a complete view of computing history;
- CHM database uses Dublin Core metadata standards; one set of metadata across all of our collection types;
- File naming structure has been created for ease of searching, e.g. Google-type searching of file names;
- Visible Storage exhibit created in 2003, primarily an object-based display or "a warehouse with labels;"
- Computer chess exhibit opened in 2005. It is a narrative-based display & also a prototype for software topics, engaging "all types of audiences". Visitor evaluation underway;
- A 14,000 square foot exhibit ("A Timeline of Computing History") is scheduled for the fall of 2009;
- In addition to timeline, the exhibit may include theme rooms (storage, software, networking, input/output and processors) as well as topical exhibits;
- Museum's web presence includes a searchable collections catalog (~25,000 records);
- 50% of the collection has been cataloged; the goal is 80%, of course there will always be a backlog;
- In addition to searching the catalog, the museum's role is also to interpret computer history for the general public, so there are online exhibitions with curatorial interpretation;
- New computer chess exhibit is available online;
- Online Reading room includes "Selling the Computer Revolution" exhibit, access to 260+ digitized marketing brochures;
- All online exhibits link to the catalog and source material; exhibit meets archive;
- People ask, how do you exhibit software? No differently than other concept-based (not artifact) exhibits, e.g. ecology, human biology, civil rights, etc. You do this with stories about people with multimedia and interactives that explain concepts;
- Larger question is who is the audience? Posterity is a vague sort of audience. Concerned with the here and now. A museum's primary function or "value-add" is interpretation. We can't be all things to all people, however, how best do we serve our communities given our resources are not unlimited? What does this mean for software preservation and access?

AI Kossow: CHM's Software Collection Philosophy

Kossow provided an overview of the current CHM software collection, status, priorities and goals. He stated that there is a current bias towards consumer software (1980s to present).

Current holdings (approx.):

- 1,000 ½" tapes
- 500,000 punch cards
- 2,000 8" floppies
- 9,000 5" floppies
- 2,000 3" floppies

Some interesting things in the collection:

- DEC Large Computer Group (LCG) Archives (DEC System 10/20)
- HP1000 Software Archive
- B5700 Source Tapes
- Sigma 5 Software (cards and tapes)
- DECTape Collections (DECUS 12-bit and PDP10)

Collecting Priorities:

- Systems in CHM collection (restoration and display)
- Systems that are simulation targets
- "The Software 100" listing (programming languages, operating systems, diagnostics, applications)
- Obtaining copies of other existing software collections (i.e. bitsavers)
- Evangelism and proactive collecting for materials pre-1975, e.g. in danger of extinction

Media recovery lab goals for FY07:

- Paper Tape Reader
- Punched Card Reader
- DECTape Reader
- 7 and 9 track ½" Tape Reader
- Media and Formats Working Library

Software Archive Plans:

- Linux/RAID system for short-term data archive;
- Local copy of external web content (e.g. bitsavers.org);
- Visibility to the world!

Timeframe:

- One year to get the infrastructure of the lab in place;
- What is more at risk? Don't have to worry about media degrading but need to make sure you're collecting and the format of (documentation).

Discussion

Shoppa: We have been interpreting data on media and so far it is easy. The risk factors are finding people with expertise to interpret the "content" including thousands of paper tapes with only a number on them and no idea of what is on them.

Courtney: Sigma 5 & [??] no longer exist; who has the knowledge to parse that tape & the understanding?

Shoppa: Sigma 5 is easy; worst case example is disk pack you want to read.

Kossow: The CHM collection's worst case scenario is the Whirlwind paper tapes.

Lasewicz: Copyright is a question? Lawyers concerned about access and use and there is also the possibility of reverse engineering & what they don't know.

Peuto: CHM has been negotiating with several parties for NLS/Augment source code copyright and the experience is not scalable.

Hedstrom: Corporate copyright lasts for 125 years but we should not let legal issues become focus of this day.

Lowood: Use donor agreements to try to get a license agreement for usage.

Tashev: Many donors don't own copyright.

Hedstrom: Technically under the law you can make 3 copies for preservation purposes, this is a revision to DCMA; the question is: can display it or make it accessible?

Sibert: A lot of material comes from people who cannot make a claim to own it.

Stevenson: Has CHM considered a business plan for a media conversion lab to allow the general public to read old media?

Pierce: I have been doing tape for individuals; don't charge but ask for a copy in return; worried about wear on drive.

Stevenson: Want a variety of institutions doing this.

Sibert: What is the collection of people in the world doing this & are we going to reach out to these people?

Hedstrom: A significant number of people working on this in their garages, etc. We should know many 7 track readers out there...

Sibert: Should CHM do outreach to people doing this?

Hedstrom: You will find a lot of federal agencies reading old media out of necessity.

Peuto: Would like to create a directory of software readers.

Lowood: Let's move discussion on to software collecting.

Dag Spicer: The Parlor

Spicer presented a "Parlor" point of view for collecting software, e.g. focusing on key items that can be interpreted/presented.

Concealment of Context

- The attic = concealment of context;
- By collecting too many things, contexts are lost. If we didn't interpret or show why something was important in our time, will future people really care?
- Future generations may not have enough information to make sense of what they have, so "collect now, interpret later" doesn't work;
- There may be no "later;" we will probably be on to collecting more things;

Panic as Method

- Panic is not an acceptable way to collect, e.g. fear of losing things;
- While it is helpful to appreciate that 'things are disappearing every day,' and thus collect broadly, can this not also be a substitute for deeper thinking about the issue of why some things should **not** be collected?
- Collecting too broadly has the feeling of the obsessed individual collector who either doesn't understand his field or care about his audience. How useful (relevant) is such collecting? It is useless to the non-expert.

The 501(c)3 Nexus

- 501(c)3 status is conveyed on an entity so that its activities can extend beyond the life of its founders;
- It is not to establish a clubhouse for a small percentage of the population; rather, it is to be relevant and interesting to the general public;
- The Attic, therefore, could have features that place it in conflict with the purpose of a 501(c)3 institution, especially if it overemphasizes collection over access.

Posterity Thinking

- Is our mission statement dysfunctional? ["To preserve and present for posterity the artifacts and stories of the information age"];
- Why are we collecting for posterity when they're doing nothing for us? Seriously, where is the audience in our mission? "For posterity" suggests that Attic-thinking may be in the DNA of the Museum;
- "Posterity-thinking" or collecting via the Attic metaphor therefore distorts the focus of the Museum;
- By believing most of what is collected is to be enjoyed/understood/processed by future generations, it tends to emphasize collecting over interpretation;
- The Attic metaphor may not factor sufficiently into its worldview that large numbers of people, distributed globally, may unbalance Museum resources since staff can never know how successful such Attic-driven "Global Software Hunters" will be;
- Improperly-managed, such distributed collection actually makes things worse: the "hole" just gets blacker, e.g. no time to process collection and provide access, all focused on collecting;
- CHM already battles the perception (incorrect) that it does nothing with its collections. Nonetheless, the Attic approach may merely accelerate the rate at which the black hole absorbs goodwill, by diverting limited resources to increased collecting rather than interpretation or access, viz. "stuff goes in and never comes out."

Growth as Goal – How can we be different?

- Why use a quantitative lens in collecting?
- Growth is a horrible goal unless you are creating value. Otherwise growth destroys value, because you get in an area where you are not unique;
- As more software is produced, the Attic approach dilutes the discernment that is the hallmark of a Museum;
- This can make us less unique with every passing year, in which we are just chasing our tails or actually Google;

- Do we want to be Wikipedia? Google? Why? They already exist, i.e. an undifferentiated resource with no effective mechanism for enforcing quality or relevance?
- Play to our strengths. Our strength is the depth and promise of our interpretive mechanism that will translate such a complex subject to citizens. Who else does that? Or will? THAT is our reason for collecting software.

Proposal

- What would I do?
- I suggest projects that encompass the gestalt of the problem: the collection, preservation and access—taken from end-to-end—of 5 titles in 2 years;
- Collecting is easy: probably less than 20% of the effort space;
- Hunter v. homemaker model, everyone enjoys the hunt and wants to make the big kill but no one wants to clean and mount the carcass, e.g. to do the hard work of cataloguing, providing access and interpreting the collection.

Len Shustek: The Attic

Shustek presented an "Attic" point of view of collecting software, e.g. collecting in large quantities for completeness.

Why do we collect?

- Our civilization runs on software;
- There has been no comprehensive and intentional activity to preserve software artifacts;
- The authors are still alive, the objects are still available. Time is our enemy!

Open wide the software gates

- Collect with minimum censorship;
- We don't know what will be important;
- We have the space;
- A large collection enables evolutionary and statistical studies;
- A large collection ensures diversity and comprehensiveness of the collections;
- It can be accomplished with a collaborative community;
- We don't know what will be needed: "The historical utility of original artifacts is that they are available for interrogation in the light of unforeseen enquiry"—Doron Swade.
- For example, there is the case of Napoleon's waistcoat button now being used to research whether he was a cocaine addict. When they saved his waistcoat we didn't know this question would be important;
- Expansive collecting is critical since we don't know what people will ask for in the future;
- Space is not an issue: at most 300 GigaLines of source code have been written by human beings (2M coders x 5K lines/yr x 30 years); 10% (30M) of all the programs ever written could be stored in a terabyte (\$400 Fry's);
- Much of it is self-identifying; much of it is junk;

Collect the source!

- Need to collect the source code;
- It is a cultural artifact: a form of literature (Dick Gabriel), beautiful programs are works of art (Don Knuth).
- It provides a view into the mind of a designer: intentions, assumptions, abstractions, mistakes, humor. Little of this gets captured in any written form;
- This is the embryonic first 50 years of millennia of software development; the transition from cave painting to impressionism;
- A voluminous source repository can be analyzed to teach us about the evolution of software engineering.
- The parlor is an unrepresentative sample;
- We need to collect binaries: for use on restored, reconstructed or simulated old computers;
- We need to collect documentation: manuals, notes, papers, email;
- We need to collect stories: interviews, reminiscences, websites.

Other uses for the attic

- An archaeological midden for future additions to the Parlor (now or later).
- A legal resource for discovering and documenting prior art for software patents.
- An historical resource for establishing credit and understanding influences.

Tough issues

- Copyright and ownership;
- Protecting trade secrets;
- Provenance;
- Reading and interpreting bits (transcoding);
- Insuring completeness (libraries? Program development environments?);
- Simulation difficulties ;
- Permanence.
-

Proposal for "Global Software Repository":

- Fueled by community contributions (web-based submission);
- Filter to keep signal-to-noise ratio high (panel of judges);
- Accessible to the limits of the law (mechanism for escrow);
- Limited metadata, provided by submitters (Depend on Google-like searches);
- Goal: 10,000 programs in two years.

Discussion Attic v. Parlor

Toole: This is really not a binary problem [attic or parlor], the question is where should we come down on the issue? The answer is based on available resources. We need to pick our fundamental issues to discuss today. We need to move forward on collecting but to what extent?

Lowood: In doing history of computer games research, Google search wasn't getting the documentation, but going to Usenet discussion lists (still on servers) revealed excellent information not picked up by Google. Need to save context of documents if you just remove them from environment you lose so much, i.e. would want to save Usenet discussions as well.

Courtney: If we take just one approach this would distort the reality of the time. Similar to CHM collecting of hardware context but dynamic is different for software, i.e. space not an issue and community model of ownership (very different from "owning hardware").

Ismail: Physical space is a problem for software not just bits but also packaging therefore a middle ground needs to be reached.

Hedstrom: Just collect and later you can sort and interpret; collecting is cheap while selection is very expensive.

Peuto: need set of parameters, important issue of quality of what is being collected, this should include the context, just collecting bits isn't necessarily helpful. Can always create an attic of high quality or parlor of very low quality. What do we want? Issue of institutionalized look – in reality collecting isn't really happening outside the institution. Focus on non-institutionalized collecting?

Pierce: Nobody builds a house with only a 2nd floor, so we should really have all of it [attic and parlor]. When building the attic curatorial efforts will increase ratio of gems to junk.

Shustek: Many people have wonderful attics. CHM has to think of what is the right thing to do?

Hedstrom: This is reminiscent around launching of the Internet archives...no access for 5 or 6 years, this appalled the archival community. But now we can take this "attic" and can curate an actual collection. We collected early years of "umich.edu" (from internet archives) and then had students "interpret & analyze," just go out and collect so you can create parlors later.

Peuto: How do we present items & provide access? Don't like attic v. parlor the issue is more institutional v. community collecting.

Hedstrom: Collecting is cheap but selection very expensive!! National library is collecting selectively; selection process adds 80% on top of the cost and besides can be selective later.

Lee Courtney: Organizing the Attic, Furnishing the Parlor – Considerations for Moving Forward

Courtney discusses how we collect software from corporate sources.

- Challenges to collecting software artifacts:
 - How-to
 - Resources (community, management, repository technology)
 - Artifact Availability (existence, donor willingness/ability)
 - IP Encumbrances (competitive considerations, copyright, license restrictions, patents, non-disclosures, ownership). How do we collect corporate software? Issues include available resources, artifact availability, and IP encumbrances
- Others working on How-to and Resources we need to focus on Artifact Availability and IP Encumbrances.
- IP issues are really a problem.
- To refine Grady Booch's original list of "software 100" we need to filter for weight (importance), e.g. where to concentrate our efforts and Collection Challenges (IP Restrictions and Availability).
- First we rank by weight/importance, then assign likely IP owner and then consider source state.
- Source state could be categorized in following buckets:
 - Closed proprietary: source code not released because proprietary, competitive, or marketplace concerns (e.g. Windows XP);
 - Available strictly encumbered: source code released through agreement strictly restricting use or redistribution (e.g. HP MPE-V source);
 - Available loosely encumbered: source code released after signed agreement loosely restricting use or redistribution (e.g. Educational institution);
 - Available unencumbered: source code released into the public domain with no copyright or other licensing burden (e.g. IBM OS/360?);
 - Open Source: source code for the system under any of the open source licenses (e.g. GPL, LGPL, BSD, Artistic, etc.);
 - Closed Classified: system owned by government organization for which source code is not available due to security concerns (e.g. DOD AWACS);
 - Unknown: IP encumbrance unknown.
- Using these buckets entire list of top software falls in the following categories: 48% closed proprietary; 31% Available loosely encumbered; 15% Available unencumbered; 4% Open Source; 2% Available strictly encumbered; and 0% Closed Classified;
- Using these buckets Top 20 ranked software falls into the following categories: 40% Closed proprietary; 30% Available loosely encumbered; 25% Available unencumbered; 5% Available strictly encumbered; 0% Closed Classified; and 0% Open Source;
- What does all this mean? Most artifacts/software from Corporate sector, so there are non-trivial IP challenges;
- Majority is closed 40-48%.
- Software is really scary for a corporation; too risky so why bother?
- Risks: insecure, liability exposure, expense and legal hassle;
- CHM should help mitigate these risks for corporations by providing policies & procedures, ownership transfer, i.e. make the donation process easy. Also provide incentive thru recognition;
- Mitigate each of the donor risks:
 - Insecure > documented policies and procedures;
 - Liability exposure > ownership transfer;
 - Expense > make donation EASY;
 - Legal Hassle > Make donation EASY;
 - Why Bother? > Provide recognition and benefit.
- Attic & Parlor must address corporate needs:
 - Demonstrate IP understanding;
 - Flexible approaches to IP issues;
 - Acknowledge corporate requirements (even if unencumbered);
- Therefore an unstructured archive (attic) should be under the legal umbrella of the parlor.

- Create attractive incentives to donate: significant acknowledgement, peace of mind and benefit(s);
- Market exemplars and success stories, e.g. Augment/NLS and Boeing;
- Survey 2-3 vendors regarding corporate donations (Bull, Microsoft, Unisys);
- Develop 2-3 corporate donations "tests" (Apple Mac Paint, Hewlett-Packard MPE, Microsoft PowerPoint, IBM APL).

Discussion

Shustek: Grady Booch "Software 100" list is only one view; for example, it contains only programming languages, no applications; doesn't distinguish between preservation & access, we need to separate these two issues.

Peuto: Once you have it, people expect it to be available on the web, so you need to consider access from the start.

Courtney: Can offer different gradations of what we offer to the donor.

Hedstrom: IP distinction between firms in business and those that went bankrupt. Orphan works discussion going on at US copyright office. Proposed guidelines are for archives to do due diligence and then ability publish. This should be the outcome of the US copyright office discussion. Historians want to see the flops and successes, e.g. David Kirsch's research on dot.com flops.

Peuto: NLS negotiation took 12 to 18 months for legal agreement. 3 companies involved, so no company wanted to accept ownership or sign a letter, so we had to negotiate with high level people (not lawyers). Want to create a blanket type agreement.

Courtney: NLS negotiation is not scalable; so need to repeat and create scalable process.

Toole: Museum would really be interested in this. Corporate software is available after 28 years so things are available after 1978, for example companies are providing access to material like MS BASIC code. Can we create a bandwagon effect?

Courtney: How do you market this product [software preservation]?

Hedstrom: Library of Congress advisory committee on preservation infrastructure is advising printing out first 100 lines and last 100 lines of source code to obtain copyright so that some small parts are being preserved. Would like to demand that they provide ALL the code in electronic form.

Lasewicz: Should consult the Greene/Miessner article on archival processing.

Olin Sibert: What is the Best Metaphor for how to put together a Collection now?

Sibert provides a detailed overview of the MIT Multics History Project and viewpoint on Attic v. Parlor metaphor.

Multics Project Overview

- Collecting Goals: preserve to guard against disappearance, decay, disposal; understand to describe the collection and its significance; complete to guide the collection toward comprehensiveness; and present, to make available and accessible to non-specialists;
- Multics History Project (1964-1990+) underway to collect, collate, preserve historical documents and materials related to Multics project;
- Included archives from several sources: MIT Information Processing Services, Jerry Saltzer office files and documents, Several smaller personal collections and large volumes of material not yet collected (MIT Archives, other personal stores);
- Approach was to digitize everything, preserve paper (some at MIT, others to CHM), created detailed catalog and develop overall summary of collection;
- Project is almost done with material collected so far; includes 45 files boxes to CHM and 45 GB of raw scans;
- Wanted to preserve paper as much as possible; people/donors wanted materials to go to a good home;
- People/donors wanted materials to go to a good home;
- Lessons learned:
 - Essential to have an understanding of material;
 - Keep a good catalog: "really important" choose organization structure tied to material as it is provided and then can be mapped to a more general schema later like Dublin Core;
 - Entering data into catalog is weakest point, was done in Excel and Emacs; cataloguing was biggest time sink;
 - Scanning is not the hard part but optimizing cataloging is what is important;
 - Couldn't scan everything such as oversized;
 - Emphasize cataloging; scanning is not the problem;

Attic & Parlor Metaphor

- "Parlor" is formal, organized, structured, professionally managed, selective, accessible to users and a significant investment;
- "Attic" is informal, unorganized, unstructured, no management, promiscuous, not readily accessible and minimal investment;
- Attic & Parlor is really a continuum, so how about a midpoint?: "Den."
- Den is comfortable, loosely organized, described but unstructured, deliberately collected and identified, minimally but sensibly selective, readily accessible to practitioners, and investment of cataloguing time;
- What if the material is unfamiliar to the collector? This is the essence of "Den" approach; must interview and analyze to obtain a useful high-level understanding;
- Software was created and structured to be worked on and not to be displayed! The context of the development environment is essential to understanding and rarely written down. Context is essential to understanding how to approach the material;
- Techniques
 - Locate and interview the experts while we still can. Used Tom Van Vleck website (Multics pioneer) to help create social networks to understand the software and provide context to this project;
 - Target social networks of company, project alumni;
 - Find and cultivate enthusiasts;
 - Multics History Projects created www.multicians.org;
 - Create a map of software: capture the implicit knowledge of the development organization/process;
 - Describe various ways to approach the software: by source code, by interfaces, by functions, by development history, and by supporting documents (if available);

- Preserve the binaries: software by its very nature is mean to function; read and convert to preserve for the future; document the formats, preserve the bits (capture the knowledge while it's still available), but it takes real work to implement!
- The "Den" metaphor is middle ground, more then the "attic" but less then "parlor." Small additional investment yields big benefits;
- Preserve and understand the material: capture the implicit knowledge as a guide for more intensive future work;
- Rely on the people who know and care: a comfortable chat in grandpa's den.

Discussion

Kossow: There is an understanding that context is needed.

Sibert: Designing software IS a social activity after all.

Lowood: In collecting game history, tools created by players need to be collected as well. Need the context!

Peuto: When people consider oral history they tend to focus on business and not the code creation, and they should.

Sibert: Collect oral histories in relation to the context of the creation but need artifacts to reference this stuff.

Shustek: CHM did this type of oral history with Mac paint source code creators.

Sibert: Would be helpful for CHM to provide more guidance to collect oral histories.

Courtney: Question to IBM & MS archivists, do either of you know about additional context infrastructure kept with source code? For example, at HP we did a videotape of code review meeting; are these preserved?

Sibert: This was written down for Multics.

Stevenson: It just depends on which product; how much context we can provide depends on what gets into the archives. We collect readily used items such as design documents instead of videotapes of meetings. Document procedures but not help desk questions.

Lasewicz: We have a 100,000 software integration collection but no context saved except for product packaging etc. for "product library." I'm uncomfortable with the "attic" at a corporate archives since the company IS the researcher. If you postpone the interpreting it never gets done. Philosophically very uncomfortable with this idea of Museum as attic. Like the parlor for what a corporate archives does unlike National Archives which collects everything. It comes down to your mission and how many people you intend to serve. Is it 1,000s or just 12 people? Cheaper to do interpretation now and results in better quality. Parlor is a better fit for a Museum.

Paul Pierce: The Structure of My Online Attic

Pierce discusses his collecting and online presence, see paper created for this workshop "The Structure of my Online Attic," by Paul Pierce, May 5, 2006.

- Built a website which includes descriptions of machines in his collection;
- Older part of website is more like a "Parlor" and was handcrafted; but hasn't changed in a while;
- Other part of the site is the library or "Attic" which is served from the database. It has quite a bit of structure and is easy to browse. Perhaps it is more like a "garage;"
- Database structure consists of a set of XML documents. Each XML document represents an artifact or artifact classification. Each XML document contains a node in the tree that is the classification hierarchy which makes it possible to browse like a library;
- Created this structure partly out of a love of programming, writing code;
- Would like to show xml code using number of item with Dublin Core field names but haven't gotten around to it; Created just a general purpose "relate to" field, showing how artifacts relate to other artifacts;
- Online "attic" is intended to be a way to present the ongoing inventory of the collection and to expose embedded information in artifacts;
- Hope to make information more accessible to wide range of visitors.

Discussion

Sibert: how do you [Pierce] enter the data into the site?

Pierce: I like to write software so try to automate it all. When reading a tape, works off of a xml file that relates information; ideally would make everything heavily automated.

Shustek: how does this extend to software with comments related to your hardware?

Pierce: I don't represent the software hierarchies yet.

Peuto: what do you think of working with others vs. working alone?

Pierce: good question because representative of most collectors like to work alone; would just like people to use his website

Shustek: what is your collecting policy, collect everything on one machine?

Pierce: selection related to your limited resources, you should collect anything you can properly catalog.

Shustek: leave interpretation to next generation?

Stevenson: important that you have minimally cataloged everything.

Peuto: independent websites good on context because of their deep interest.

Pierce: most websites are made by enthusiasts so that can put the context to it or are motivated to find out about stuff, such as my IBM collecting.

Tim Shoppa: The PDP-6 and PDP-10 Software Archives

Shoppa provides an overview of his website (trailingedge.com) and collecting efforts.

- Involved with DECUS (Digital Equipment Corporation User Society) and working with old machines;
- Created and built website from late 1990s-2000 but now has other interests. Collecting efforts have focused on operating system, language tapes, tools and applications, tools for converting; this is an iterative process;
- Goals of the online archive include: provide tape images usable for emulators, provide tape images that can be reconstituted into physical tapes for actual PDP-10 hardware, provide source files for research, have internal indices for internal cross-referencing and be hyperlinked into the world wide web;
- Also wanted the archive to be relevant in terms of tying into computer history research as a whole;
- Several PDP-10 emulators available today; he uses Usenet for discussions and links to his site; he includes links to DECUS write-ups;
- Currently collecting anything PDP-10 related;
- Would like to see "parlors" link to his site;
- Wants to make website more relevant to the world for history of software and platforms. Wants tape images so they could boot system from tape image; wants source files to be accessible. Wants internal indices to understand the context of all versions. All work to be accessible on web;
- 179 tapes available. Conversion tools used DECUS and VMS now migrated to more modern platform & reverse engineered on some by reading the tapes took cycles. 70,000 files extracted from the 179 tapes;
- Wants lots of indices for a "highly organized attic;"
- Translates line printed formats to html formats to emulate line printer;

Discussion

Kossow: dynamic material causes problems for mirroring a website.

Shoppa: creating a dynamic package for mirroring, would like to see it mirrored in more parlors. How can I do this?

Lowood: from history of technology perspective would like to see included why the program was written? i.e. context, bring in images of catalogs.

Shoppa: have cross links with DECUS write-ups but this isn't working right now.

Lowood: context will drive people to the site. People using the site already have interest in it but this would provide a more user friendly site

Sibert: community outreach is essential for this, get practitioners to talk. What is your point of entry? Need to record these people's contributions?

Shoppa: can't do that type of work but would like people who do, to link to his site.

Margaret Hedstrom: Emulation as a Digital Preservation Strategy

Hedstrom provides an overview on her work at University of Michigan.

- Has been digital archiving for about 30 years. Has made case over the last 15-20 years to save software to use in service of digital archiving;
- Saving content out of original and migrating or turning into some standard format has high limitations as it doesn't preserve the original look & feel;
- National Science Foundation grant for Camillion project at the University of Leeds, IT worked to salvage the Doomsday disk as part of project (software was BBC micro). Has a ten year lifespan with lots of copies available;
- Original hardware and software very challenging, had to develop an emulator. Finally were able to get versions of BBC micro but a lot of work to track this down. Not on internet because of IP issues;
- Also performed research on how subjects (people) would respond to older software on original platforms when migrated and emulated. In emulation color matching, speed was important for the test subjects;
- How important is it to preserve in original state?
- Emulation is not an exact match to original but may be good enough;
- Whenever possible, keep the original bit stream (and physical media, if available);
- Lessons learned from saving software:
 - Peripherals are a real challenge (power supplies, etc.) to get;
 - Not really in business of saving software but saving content;
 - Fairly comprehensive hardware and software platforms but no content that depends on these or vice versa. Software without content or content without software is not useful;
 - Emulation study used hobbyist; couldn't be done without these folks, with BBC micro computer game original users very helpful;
 - A lot of risk and serendipity involved. Countless bugs with Doomsday and had almost given up. Finally found an original programmer. Reinforces that there is NO TIME TO WASTE in saving software;
 - Increasingly content and executables are completely linked – need to run models and data together to understand it all;
 - Understand the connections between content and software, links between museums and archives etc.

Discussion

Lowood: A lot of patches for software; software changes constantly. How to manage various versions of software? Did this come up at all?

Hedstrom: Is it really important to manage it? How many versions of the top 20? Which versions? Critical questions. Version control was really bad. Part of the research explored whether people want to play old games on original platform or emulation. People were not enthusiastic about original platforms, liked emulators because easier and more familiar. Issues in digital archives – how close does what you preserve have to be to the original? Exact color matching, right speed? Subjects noticed interesting things such as sounds are sloppier in emulated versions.

Peuto: Subjects of data [?]

Hedstrom: Problems is the data or model is wrong, so you need both together.

Peuto: In case of games what did people find?

Hedstrom: Emulated version preferred but the people noticed things they had not anticipated. We need to understand use & users much better. Emulation doesn't exactly produce the original "feel" and does this matter? So much more to be done here.

Pierce: common users of software versions not able to tell different versions. Tie bits back to original media to understand the versions and go back to the image of the original media.

Hedstrom: In digital archives world take old stuff and translate into a new media. Keep original bit stream & translate into newer version is good but don't get rid of original since storage is cheap now. Save bit stream but not media its on.

Pierce: For a lot of tapes he reads he keeps an image of original.

Shustek: Is there any value of saving original media?

Lowood: Yes, for authentication.

Hedstrom: but is the label on container right?

Lowood: In the library world, they were microfilming for preservation and throwing out books and they lost all the lost foldout plates that were never microfilmed.

Hedstrom: Google is scanning the University of Michigan library and still saving the books.

McJones: 3 versions of reading technology so have had to run the original tapes 3 times.

Spicer: Just print code out onto acid free paper?

Hedstrom: Stone works pretty good!

Amy Stevenson: Building a Model for Sharing Proprietary Software

Stevenson provides an overview on Microsoft archives and what they are doing with software.

- MS archives are not open to the public; so internal usage drives preservation, therefore can't save everything, don't have the space;
- Would not preserve source code unless there was an internal expected use;
- Goal is to serve the company's needs first, not the greater community but can do on a limited basis;
- CHM, however, can be organized attic for propriety software;
- Benefits of creating a "CHM" Parlor for source code:
 - Creates a safe haven for sharing: Microsoft wants to share software, if many companies offered source code to CHM and there is lots of source code stored here then doesn't make MS such a target;
 - Prior art research;
 - Allows for consistent preservation;
 - Better research;
 - Positive PR.
- Targeting more used software – research impact 1000s of users versus 10;
- Obvious candidates are obsolete software or something that is older and has changed a lot;
- Microsoft candidates for the archive might be that they don't commercially distribute any longer, stuff that is so old not useful commercially;
- Design documents are a whole different animal and not currently being considered for donation;
- Some of the risks:
 - Legal;
 - Financial;
 - Negative PR.
- The Answer? Write a License! Or donor agreement form; exemplary agreement that could be used across the industry;
- There is a volunteer/multidisciplinary effort within Microsoft to create archival agreement whose terms protect the archiving organization and the donating organization;
- Some samples of source code licensing agreements include academic-use licensing program and shared source licensing;
- This wouldn't have been impossible 5 years ago. People are tasked with sharing software because of the court case.
- Goal – license can be a template for CHM;
- Suggestions of issues to be discussed/considered for the CHM software archive:
 - How should the archive be used?
 - Who should be able to access the software within it? Requests in writing?
 - Who will manage this within the organization?
 - What limitations will exist on the use of the material? Can you distribute? Academic use only? How do you screen researchers, only those with legitimate means? How do you define illegal use if provided a copy of code?
 - In what form will the software be stored?
 - How broadly accessible should it be?

Discussion

Hedstrom: in legal cases is source code subject to discovery?

Stevenson: usually we won't give originals, it is really hard to make derivatives and provide.

Sibert: is MS collecting source code? actual code or context?

Stevenson: just the finished code; not the development process. In collecting we ask, what is going to be requested of us? Olin really wants the change history. This isn't captured in a snapshot.

Peuto: would you provide the design documents?

Stevenson: probably would not because of type and IP issues, focused on preserving.

Peuto: internal design documents contained more interesting stuff than revised versions.

Stevenson: documents are a more difficult problem because they haven't approached this yet whereas they license software all the time.

Shustek: do you preserve the development environment?

Stevenson: don't collect the internal tools used.

Shustek: are you concerned about having stuff in another place that lawyers would not have policy to shred.

Lowood: MS flight simulator game changed after 9/11, if someone wanted an earlier version and had a legitimate research reason, how would Microsoft respond?

Stevenson: If you had the right person to sponsor within MS you could. Research always has to get permission through legal or PR. Academic requests PR just says no, only want journalists.

Shoppa: his would contribute to studies on MS products.

Stevenson: periodically gets these requests to improve process. Most interesting document is post-mortem of software.

Paul McJones: A Case Study in Software Collection, Preservation and Interpretation

McJones provides an overview of volunteer work he has been doing for the CHM Software Collections Committee, focusing on FORTRAN preservation.

- Started with the question: What can I find out about the original FORTRAN?
- Next wants to reanimate the software and create an emulator for it. Reacquiring the expertise, reverse engineering, often takes a distributed effort of many people;
- We really want to restore (reanimate) the software;
- Why look at the IBM 704/709/7090/7094?
 - Spanned from Von Neumann to System/360;
 - Hosted original programming languages: FORTRAN, LISP, COBOL, MAD, SNOBOL...;
 - Hosted earliest operating systems: GM/NAA I/O, SOS, IBSYS, CTSS...;
 - Workhorse of scientific and engineering computing.
-
- Created an emulator for FORTRAN:
 - Physical media: finding this can be pretty difficult and ability to read along with documentation;
 - Paul Pierce had the best stuff to do this ;
 - Found a lot at CHM;
 - Remove bits used Paul Pierce – transcription;
 - Some people typing it in manually;
 - OCR of line printing;
 - Read the manuals!!! Kossow & Pierce have them online;
 - Run it on something – emulator;
 - Processors either real or simulated (variety of people doing on different platforms);
 - Reacquiring expertise: get back to IBM in 1962 – the right setting, boot tapes etc., found people working on this;
- Lessons learned in via Rich Cornwell
 - Reference manuals don't always tell the truth;
 - Diagnostics are invaluable in writing simulators.
- How are these projects organized?
 - A multidisciplinary team is required: email, Usenet, blogs, and individuals websites
 - It is often a labor of love;
 - It is often a distributed effort;
 - A variety of web sites contribute.
- How can institutions help?
 - Encourage/coordinate restoration projects: media conversion, host online services as forums;
 - Provide long-term stability: archive software, manuals, simulators;
 - Provide visibility to the public: physical exhibits, website.
- CHM should host online forum to help with these peoples efforts.

Discussion

Shoppa: Did you use Software Collection Committee wiki at CHM to do multidisciplinary approach?

Courtney: What was biggest hurdle?

McJones: Seeing my vision of what info. sources are and who is doing what, unscalable process of finding these people.

Kossow: You can go for years and nobody will contact you and then it just happens.

Courtney: IBM 1620 restoration project didn't know about a Perdue professor w/ 20,000 punch cards and then they arrived after the word got out.

Pierce: Get stuff from Google by making pages crawlable by Google.

Hedstrom: Operating systems & core software, what about applications that ran on these machines?

McJones: Would really like to find these, mostly think they will be at aerospace labs and confidential/proprietary.

Hedstrom: Dissertation on office automation in the insurance industry around this time, PLATO project.

Peuto: There are two large PLATO websites.

Hedstrom: Shift from science & military to business & education around this time (creation of FORTRAN).

Shoppa: DECUS catalogs different places where the software is coming from.

Pierce: SHARE came with abstracts so can see where coming from.

Hedstrom: This also includes the first generation of online library catalogs.

Paul Lasewicz: What Would Bill Do? The Business of Preserving Software

Lasewicz provides perspective on collecting software and provides IBM archives overview.

- Two different groups of software:
 - orphans (no commercial value any longer);
 - materials after 1980 people still want to take care of.
- Therefore 2 solutions:
 1. Orphans: work with attic & parlor methodology;
 2. Non-orphans post 1980.
- IBM collects for business reasons (legal etc.) not history;
- Optimistic because there is business value behind preserving these documents;
- If government gets involved – standards & perhaps align with open source (business models emerged “Red Hat”);
- % of software to be concerned about over time will decrease, will also keep learning over time;
- Different solutions with different strategies;
- Partner with companies to transfer rights years down the road; let these companies do the attic work for us and then CHM can concentrate efforts on the parlor;
- IBM's collection has very little software;
- “The time has to be right” to get IBM to donate software; legal is concerned; hard to tell when the time is right (prior to the model 360 might be okay);
- Question of open source software: what should we be doing today to make future archival efforts easier?
- 100 year record solution – driven by the government (National Archives?).

Discussion

Courtney: What software is in the IBM collection?

Lasewicz: Not much just serendipity, only 40% of collection with any type of administrative control.

Courtney: Are there other reservoirs of software in other parts of company not archives?

Lasewicz: Yes and somewhere in records retention and will find in other areas such as international operations.

Peuto: 1971 software library will be destroyed?; why won't IBM donate?

Lasewicz: The time has to be right, need to find the benefit, maybe for PR: to counterbalance DEC software at CHM or promote IBM as innovator/software creator (i.e. FORTRAN)?

Peuto: When is the time right?

Lasewicz: Legal is nervous and won't invest time but they are supporting open source movement, maybe decades away.

Sibert: Use IBM retirees to push IBM?

Lasewicz: Isn't in a position but they are part of problem with advocating open source. With anniversary coming up maybe more willing.

Lowood: Is open source easier?

Hedstrom: On digital archives side, we were optimistic but what about maintenance of this? Talking about restoring stuff as opposed to being proactive for today's software preservation.

Lasewicz: Optimistic because there is government role for maintaining and preserving government records. Advocating non-propriety software needs pervasive solution to encompass 100 years.

Hedstrom: National Archives is opposed to saving software. Contractor to build a system which they won't be able to populate and there are too many propriety systems.

Lasewicz: This isn't a solution but they need to find one.

Sellam Ismail: Contemplating a Standardized Software Media Preservation Methodology

Ismail discusses his thoughts on creating a universal data standard.

- Magnetic media such as tape and floppy diskette have passed their theoretical lifespan but still readable, we are lucky;
- The bits in EPROM chips (and similar technology) are rotting;
- Punched cards and paper tape are acid-based mediums and are rotting too;
- Something needs to be done now if we are going to save software;
- What do we do?
- What level of preservation?
- Saving the bits is only part of the process. The "Medium is the Message" preserve meta information of the medium itself is necessary;

- Undertaken "FutureKeep" project (work in progress):
 - A standardized, structured file specification for imaging computer data media (image of the holes as they are punched to be extensible and universal like XML.);
 - Preserve bits by describing the physical medium;
 - Features: well documented, universal for all media, extensible (adopted to other media), simple (human readable), free (open source);
 - Format: stored in standard text file, tagged format (XML), human readable (no binary data).

Discussion

Shustek: Basically a block-level description of media?

Ismail: Can be.

Shustek: Is this the appropriate format?

Ismail: Different parts of the file in different levels some tracks & sectors but may need to go to binary.

Pierce: Multiple levels at once of same data?

Ismail: Yes if you want, define as you want, want to continue and develop over time.

Sibert: Where would you store metadata?

Ismail: Would be XML tags.

Shustek: Scan of sticky label?

Ismail: Standard defined.

Hedstrom: Global digital format registry from Mellon Foundation there is 1 million for every format & documentation, e.g. every file format in the world. Is it feasible in a universal format?

Ismail: It is conceivable and we will do it.

Hedstrom: Project in mid-1998 to create universal preservation format. My experience is that this is really hard.

Jablonek: How do you get people to use it?

Ismail: Just trying to present a solution for all these file formats.

Bernard Peuto: The Open Software Collecting Movement and Its Impact on Collecting Institutions

Peuto provides perspective on collaborative process of collecting software and overview of CHM Software Collections Committee activities.

- Multiple collecting models co-existing: some with historical roots and some with non-institutional roots;
- Is there an emerging or dominant model? Numerically it's the non-institutional model;
- Open software collecting revolution taking place, like open source coding, which is redefining the role of collecting institutions;
- CHM aims to foster and join this movement;
- Software collecting is off the scale compared to collecting hardware; the numbers are massive, e.g. 10 to 1,000 to 1,10000 to 1 for each hardware environment;
- Software is a content experience; viewing the content is a large percentage of the exhibit experience; physical artifacts are less important;
- Software collecting is contextual;
- Web is eminently suited for software collecting: little resources need to store and gather, it's all in the interpretation/curation;
- Two models of collecting: reactive (tends to be staff-oriented) and proactive (tends to be volunteers passionate about something);
- What model do we want to choose?
- Reactive is more institutional museum oriented;
- Non-institutional model is driven by passionate websites, 10s to 100s of quality sites are available, no permission needed and better context coverage;
- Wants to maximize collection so non-institutional model (volunteer driven);
- Feedback from CHM Software Collections Committee (SCC) on Attic v. Parlor metaphor:
 - Quality is critical
 - Don't favor either approach, e.g. attic v. parlor
 - Is this the right question?, e.g. attic v. parlor
 - Refocus to more critical questions: who does the work?, how are the rules established and maintained?, what should the archive look like?, what collections can be rescued?, how do we manage this? How do we encourage amateur collecting activity?
- Overview of SCC & their solutions
 - Very pro-active
 - Stable collecting projects: Fortran, Lisp, NLS.
 - In-progress projects: Multics, PDP-1, 1401.
 - New projects: APL, ACM Computer Science Book Project, Resource Directory.
- Case study of proactive collecting: FORTRAN
 - Goal to collect FORTRAN/FORTRAN II compiler source code, documents;
 - Methods include relentless emails, phone calls to people, institutions and blog to record progress and attract comments;
 - 12 month project relentlessly turning over stones, blog available (see Paul McJones comments); did find source code because of persistence;
 - Status: located documents, films, machine-readable source code, created web site at CHM, working with others to get FORTRAN II running on simulator.
- SCC Lessons Learned:
 - Passionate volunteers;
 - Collecting context is critical;
 - Institutional support: museum imprimatur, community training, IP rights;
 - Time is running out;
 - The web has fostered another killer app: open content drive collecting;

- For software it is: distributed, non-institutional, passionate volunteer driven, meritocracy driven, delivers quality based on field knowledge, integrates collection and exhibition, widely accessible.
- Role of institution (museum):
- Geographic role (physical exhibits, events, conferences);
 - Access role (cyber museum, research collections);
 - Community service role (IP umbrella, Imprimatur, community training and support, quality and leadership);
 - "Burial Services;"
 - Provide leadership, guidance, structure, templates;
 - Harness talents of individuals, volunteer energy.

Discussion

Tashev: Who is the audience for all of this? Museum needs to reach the broadest audiences, unlike personal websites and provides longevity.

Peuto: Thinking more about research aspects of the museum.

Tashev: Impressive museums combine primary research WITH education, such as Monterey Bay Aquarium.

Peuto: I think I would choose collecting over exhibits.

Tashev: We have to do both, that's why we are a Museum not a warehouse.

Toole: Missing is the quality, leadership & "people??" we know the community and need to keep them connected such as Pierce who is collecting. During transitions need to keep people going forward.

Peuto: Agree made more simplistic for talk.

Sibert: Talented and motivated people; the role of museum could be to offer an umbrella and provide uniformity across collections.

Shustek: Right, the museum needs to provide across the board services.

Peuto: CHM should host and create these sites.

Sibert: Museum's Imprimatur.

Hedstrom: Institutional v. individual distinction, individuals can harness through the web. "Next generation finding aids" we can't identify everything and people are helping online.

Peuto: Museum fostering these sites.

Wrap-up and Conclusions :

Participants #1 follow-up question, what do we need going forward?

Shoppa: More meetings like this!

Kossow: Electronic forum.

Hedstrom: Outstanding items – oral histories, to document strategy, create a document of all software archivist activities (directory).

Lasewicz: Problem that effort is fragmented, make the museum a leader of this (best practices).

Tashev: Define audience, who we are doing this for? What is the size of this audience? What is the point other than for preservation sake?

Spicer: Consider whether you de-accession stuff? Why are we storing junk? What are the criteria for collecting?

Peuto: Need directory of media conservation.

Peuto: Define what we mean by context in terms of software.

Peuto: What can the museum offer in terms of support?

Sibert: Comprehensive support for non institutional collecting.

Powell: IP issues & desires of museum, provides some models to collectors.

Stevenson: Don't forget the retail packages.

Stevenson: Use technology to make things available more broadly –metadata from outside, wikis, oral histories maybe thru email.

Courtney: Open SCC Plone site to people like Sibert.

Courtney: Identify topics & affinity groups for people to network.

Shoppa: Software engineering and program management groups gain from a software archive.

Shustek: prefer providing “housing” instead of “burial services,” i.e. parlors.

Shustek: creating the repository so it will be preserved

Pierce: As a collector there is only one of me but lots of people want to use my stuff, allowing for wiki between collectors who don't have time to do parloring...we can point people to parlors since he can't create them.

Lowood: Building the digital repository.

Johnson: Find easy way for experts to contribute.

McJones: Think about how corporate donors can benefit from a software repository.

Kossow: What do you think are action items for AK for next 12 months?

Hedstrom: Collaboration with digital content collectors i.e. digital archives.

Hedstrom: Advocacy, need to let people know why stuff should it be saved.

Hedstrom: Build repository no matter what.

Hedstrom: We need research—how do you preserve executables?

Ismail: Attract outside funding to bring hobbyist into the fold.

Jablonek: how is the context going to be captured/created and preserved? Essentially metadata creation is generally 75% or more of a digital project's resources.

Possible Next Steps:

I. Create wiki or communication tool: (Kossow)

Kossow: Creating lists/wikis SCC is oriented toward internal meetings but that shouldn't be the focus. Wiki's are there to build a documents mailing list and to send announcements. Must build the community first before anything. Is this the community to start with? Create a forum of the "software collecting group" and create a wiki and we move forward from here.

Sibert: This is a part-time activity for many of us & mailing lists aren't good if you drop out for awhile (because you don't have time today)...something more toward keeping a permanent record (wiki and threaded discussion). The wiki can be a final product. Create a person who is responsible for list &/or wiki. This is essentially what McJones dusty decks (website) is and that has been successful.

Kossow: Create McJones-type site for the forum. Kossow will be moderator. Have to be a member to post (but easy to become); RSS feed to monitor with a bunch of buckets to sort.

II. Software preservation directory project (SCC & Peuto)

III. What kind of support can the museum offer?

Spicer: Linking important websites after filtering for quality...use of webrings. CHM becomes essential authority for software collection & preservation?

Tashev: Move forward to working with corporate software archives collections? Outreach/relations. How do we persuade people that they need to think about this and take action on it? How do we find the right people? Deliverable would be? White paper? Issues to pursue.

Courtney: What are barriers to donating to the museum?

Lowood: Identify 20 or 30 people who are involved.

McJones: Asked Adobe for stuff but fell into a black hole, lawyers will never let go of the source code. Get 20 ex-CEOs to create this if not commercially viable.

Hedstrom: Need collecting policy and framework.

Tashev: We already have this.

Sibert: Don't make it hard to donate stuff here; need to negotiate with museum.

Courtney: What are the barriers?

Spicer: if you get 20 CEOs, the software can be a Trojan house.

Lowood: collections are from individuals not CEOs, e.g. the managers of the stuff, need to create a network that is interested in this. What group do we want to convene for? Work with HP (2000, 3000) to get permission for certain software tapes and then take agreement to the group.

Sibert: I don't believe the museum would be a help with IP issues at Honeywell. Need to find the right person to overcome apathy & indifference. Who would we target? John Seeley Brown would be an advocate.

Lasewicz: We need to build the business case for why this should be done. Need to create the case first and then find the right people to talk to.

Lasewicz: Business model for Coca Cola & film to Library of Congress. They got preservation paid for in donating the film. Start small and build up to larger companies like HP. Back to helping individual collector, what type of tools are needed?

Sibert: Think like an independent might look? Hosting structure, compare versions of software.

Stevenson: Identifying materials by getting people with content knowledge; using Sharepoint to post on the web?

Sibert: Software is large so need to be able to structure.

Shoppa: Not authors but we need framework.

Peuto: Used Plone to host but problem to write requirements for future users.

McJones: Multics list sending out a call from contributors...is there another community of people who want to think about information design & structure to help with this?

Peuto: Small finite tasks can find volunteers for if do outreach, this group is for brainstorming and then we will find volunteers.

Toole: Resources issue for hosting on CHM yet to be determined what we can do...is there a need do we want to try to do it, maybe we should host here to build community quicker.

Lowood: Create a service to brand.

Spicer: Use bitsavers as a beginning for hosting/mirroring, prototype hosting can be bitsavers.

Kossow: We don't have enough bandwidth, it is too much and should try something with lower bandwidth.

Spicer: We need to investigate what resources we would need to build this.

Stevenson: Maybe Microsoft can sponsor this through the facility in Mountain View. Creating help for outside collectors?

Sibert: Advice on collecting oral histories, what does CHM want when collecting?

Pierce: What best practices are on the website?

Tashev: A lot of this information is on the CHM website.

Shoppa: Plenty of sites that would tell about oral histories.

Pierce: Looking for decent training information put policy documents on the web.

McJones: Move more policy documents from intranet to internet.

Lowood: Most libraries have friends groups and they provide help to these collectors and eventually will give us money and their collections...Can sell groups to people by saying "I have access to these learned people (rare books librarian)".

Tashev: Provide reading lists and links.

Hedstrom: Is there an authoritative source on collecting & preserving software?

Lasewicz: Source on collecting "Records of the High Technology Company" available from Society of American Archivists by Bruce Brummer.

Stevenson: Can provide certification program/branding – junior curator, ad-hoc curator.

Tashev: How much guidance can we provide on software collecting and preservation requirements?

Lasewicz: Need to train them to do stuff at professional standards.

Tashev: Can't even get donors to describe what they want to donate let alone what you are describing.

Stevenson: Rent space to individual collectors?

Hedstrom: Assign interns to a private collector.

Where do we go from here?

Lasewicz: Try giving electronic lists a try first.

Courtney: Once every 2 years we should do a workshop, & next one on a larger scale.

Shustek: Academic symposium & proceedings at some point but doesn't believe the community is ready for this.

Spicer: Create communication mechanism & website. Expectation for us to take initiative on these means.

Hedstrom: Quite a number of existing conferences to put together a panel, there is a bigger community interested in this problem too.

Shustek: There is the history of programming languages (HOPL) conference next year (2007). Maybe we could interest them in this issue?

End of Workshop