
Metaphors for Collecting Software

Olin Sibert
osibert@oxford.com
5 May 2006



Collecting Goals

- Preserve
 - Guard against disappearance, decay, disposal, disaster
- Understand
 - Describe the collection and its significance
- Complete
 - Guide the collection toward comprehensiveness
- Present
 - Make it available and accessible for non-specialists



Multics History Project

- Goal
 - Collect, collate, preserve historical documents and material related to Multics project (1964-1990+)
- Archives from several sources
 - MIT Information Processing Services
 - Jerry Saltzer office files and documents
 - Several smaller personal collections
 - *Large volumes of material not yet collected (MIT Archives, other personal stores)*



Multics History Project

- Approach
 - Digitize *everything*
 - Preserve paper as available (some stays with MIT)
 - Create detailed catalog
 - Develop overall summary of collection
- Status
 - “Almost done” (with material we have so far)
 - About 45 file boxes (to CHM), about 15 others
 - About 45 GB of raw scans
 - 2% of scans in color/grayscale is $> 50\%$ of space



Lessons Learned

- Essential to have an understanding of material
 - Roger Roach and I were lucky – we already knew
- Keep a good catalog
 - Technology is weak here
 - We're using Excel and Emacs for ease-of-use in data entry/editing
 - We'll convert to a database to produce web pages
- Optimize cataloging, not scanning
 - Especially if you're not scanning everything!



Collecting Philosophies

Parlor

- Formal
- Organized
- Structured
- Professionally Managed
- Selective
- Accessible to Users
- Significant Investment

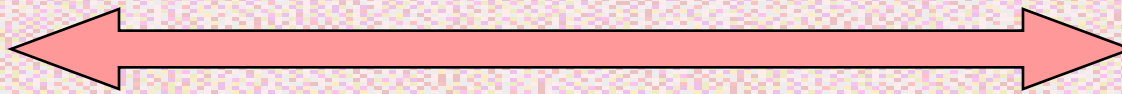
Attic

- Informal
- Unorganized
- Unstructured
- No Management
- Promiscuous
- Not Readily Accessible
- Minimal Investment



Actually, it's a Continuum

Parlor



Attic

- Formal
- Organized
- Structured
- Professionally Managed
- Selective
- Accessible to Users
- Significant Investment

- Informal
- Unorganized
- Unstructured
- No Management
- Promiscuous
- Not Readily Accessible
- Minimal Investment

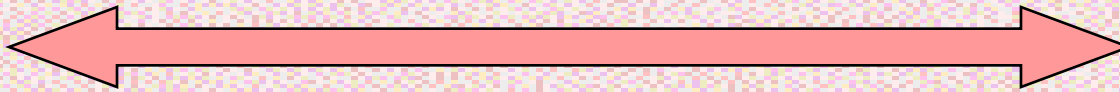


How About the Midpoint?

Parlor

“Den”

Attic



- Comfortable
- Loosely organized
- Described, but unstructured
- Deliberately collected and identified
- Minimally (but sensibly) selective
- Readily Accessible to Practitioners
- Investment of Cataloging Time



Gaining Understanding

- What if the material is unfamiliar to the collector?
 - This is the essence of the “Den” approach
 - Must *interview* and *analyze* to obtain a useful high-level understanding
- Software was structured to be worked on
 - Not to be displayed!
 - The context/environment of the development environment is essential to understanding, and rarely written down
 - Context is essential to understanding how to approach the material



Techniques (1)

- Locate and interview the experts
 - While we still can!
 - Target social networks of company, project alumni
 - Find and cultivate the enthusiasts
 - MHP has **`www.multicians.org`**
 - We should all be so lucky!



Techniques (2)

- Create a “map” of the 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
 - By supporting documents (if available)



Techniques (3)

- Preserve the binaries
 - Software—by its very nature—is meant to *function*
 - Read and convert to preserve for the future
 - Document the formats, preserve the bits
 - Capture that knowledge while it's still available
 - Emulation is a powerful tool
 - Shows the software in its original form
 - But, it takes real work to implement



Summary

- The “Den Metaphor”
 - Software collecting’s middle ground
 - More than an Attic, much less than a 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*

