# Proposed Curation Roadmap
# Computer History Museum Software Collection

Lee Courtney
Lee_Courtney@acm.org
(650) 704-3934

This proposal is a work breakdown structure for curating the existing Software Collection at the Computer History Museum.

The following tasks aim to:

- Define an artifact life-cycle for the Software Collection within the context of the unified artifact database being built by the Museum.
- Catalog and register all media artifacts in the Collection.
- Allow for interpretation of software artifacts in the Collection.
- Achieve a baseline for the existing Collection that will lower the threshold for adding ***and interpreting*** new artifacts.
- Make content of the Software Collection available to the Museum and other communities.
- Make significant progress towards the goals of the Software Collection Committee.

With these in mind the several tasks were identified and after discussion goals for each task identified, approximate durations and schedule created, and brief descriptions of the task written.

**Software Collection Curation Tasks**

**Task:** Initial Survey **Week:** 0 **Duration:** <= 3 weeks

**Goal:** Quantify the set of artifacts in the existing software collection and physical location to determine resource needs for later tasks.

**Who:** Software Curator with assistance from volunteer(s).

**Notes:** Visual survey of the contents of the Software Collection at both Shoreline and Moffett locations capturing location, condition, and some descriptive information. Answer the following questions:

"What type of software is the artifact? (OS, compiler, data, linkeditor...)"
"Media type?"
"Location?"
"What is the physical condition of the media?"

Person doing this should have sufficient domain knowledge to assess if something is 'important' ("Hey, it's Mona Lisa!"), where it fits in the Software Collection ("Oh, this is the security code for Multics"), and where a software artifact fits in the larger collection ("This card deck is BPM for the Sigma-5").

We should approach this task with the knowledge that this data can be thrown away. This is not cataloging, but assessment.

**Task:** Remediation **Week:** 3 **Duration:** 2 weeks

**Goal:** Stop any immediate media deterioration.

**Who:** Software Curator with help from Registrar.

**Notes:** All issues should have been addressed in the move from Moffett, but if a problem is found we should take measures to stop further physical deterioration. I suspect there will be some minor, but necessary, work to be done here. (E.g. I know the Sigma tapes need to be unpacked and hung on tape racks.)

**Task:** Physical Organization **Week:** 5 **Duration:** 12 weeks

**Goal:** Gather ***all*** media together for cataloging, conservation, and transcription.

**Who:** Software Curator with help from Collections Coordinator and possibly volunteer(s).

**Notes:** Once an initial survey identifies all types of media (punch cards, magnetic tape, hard, disk, floppy, ROM, reel-reel magnetic tape, magnetic tape cartridges, source listings, ...) in the Software Collection then: a) coalesce software artifacts in approximately same physical location, b) insure it is stored properly, and c) design and implement cataloging/labeling protocol.

This means finding new space or confirming existing space is adequate; finding, moving, or designing storage (e.g. shelving, tape racks etc.); designing and creating labeling for the storage area; moving physical artifacts as needed. Will need to address the issue of leaving software artifacts at Moffett (not a good idea).

The Initial Survey should identify the physical location all software so that in this step space is allocated and the artifact moved.

I suspect this will impact some other Collections as we will need to make room and assemble storage for media.

**Task:** Database Prep **Week:** 17 **Duration:** 4 weeks

**Goal:** Define a data dictionary for cataloging of the Software Collection within the context of the global Collection database that will before useful to curators, researchers, and historians.

**Who:** Software Curator and SCC subcommittee.

**Notes:** Before any cataloging can take place the essential descriptive data elements must be defined. Working with SCC define a draft meta-data elements needed to fully describe a software artifact. Validate congruence of this definition with the larger overall collection database. Select test cases and validate.

**Task:** Cataloging and Registration **Week:** 21 **Duration:** 16 weeks

**Goal:** 1. Catalog all existing artifacts in the Software Collection.

2. Provide a mechanism for cataloging software artifacts as the Museum's Registrar processes them.

**Who:** Software Curator with help from volunteer(s).

**Notes:** Starting with Initial Survey data, or starting over again, register the Software Collection.

Note - this only catalogs the physical media portion of the artifact in the Software Collection. No actual software content is captured at this step. Cataloging will enter the media component of the artifact into the Museum artifact database. Registration will tag the media portion of the artifact with an accession number.

**Task:** Transcoding Research **Week:** 37 **Duration:** 16-36 weeks

**Goal:** Define and implement a repository for software artifacts that:

1. captures all elements necessary to reproduce a software artifact as encoded on the original media (e.g. reproduce bit stream off a 7-track BCD encoded magnetic tape),
2. allows perusal of artifact content in an organized fashion (e.g. scan through Multics source code),
3. allows search of the repository for specific content (e.g. 'Show me all ALGOL source code in the Repository.'),
4. provides for future content and context extensions defining and allowing interpretation of a software artifact stored in the repository (e.g. adding a 1401 emulator to interpret an artifact).

**Who:** Software Curator with SCC subcommittee.

**Notes:** This is a big problem - maybe we should leave the duration value the same, but change 'weeks' to 'years'? ☺

I believe it will actually take several weeks to crisply define a tractable problem. Once we understand the important issues then can iterate on a solution. It's very important to get this answer right the first time, as the next step in this process, transcoding from original media to the repository, may destroy original media. We don't want to blow our one and only chance to transcode some artifacts.

This is also the stage to thinking about and define special tools and techniques. See http://www.piercefuller.com/collect/a7tv.html for an example. Dave Babcock et al have given some thought to what to wrap around the data to make sure we don't end up with a meaningless bag of bits from an artifact.

Approach here is to segment the problem, and solve some easy low-risk ones first. Prioritize list from the CATALOG/REGISTRATION step and then attack.

***AT THIS POINT WE CAN START DOING SOMETHING WITH ACTUAL ARTIFACT CONTENT (bits).***

**Task:** Transcoding **Week:** 53-73 **Duration:** >= 24 weeks

**Goal:** Move current artifacts from original media into the Software Collection Repository.

**Who:** Software Curator managing ongoing stand-alone volunteer project.

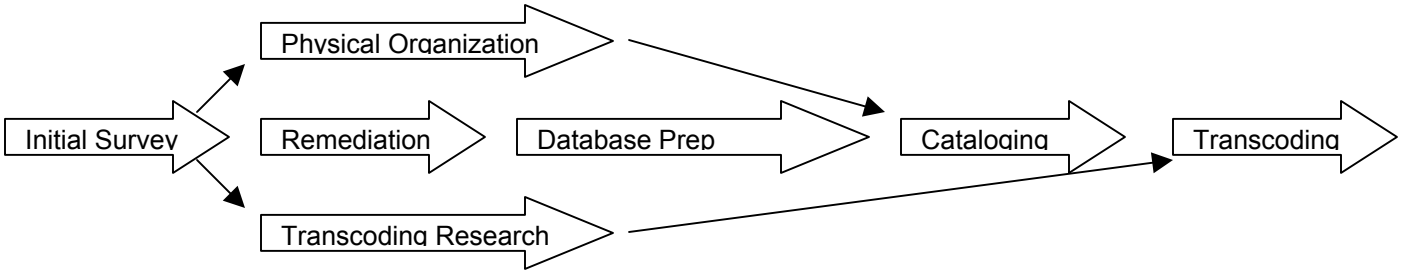
**Notes:** At this point we start populating the Repository with content.

Locate, assemble, test, and maintain required hardware to transcode different types of media.

This will entail taking physical artifacts from the Software Collection, doing data entry for this step, and transferring media content to the Repository.

As part of Software Artifact Lifecycle implementation determine disposition of media.

As with many projects, there is opportunity to proceed in parallel.



**BEYOND THE CURRENT ARTIFACT SET**

All of the above tasks and durations are targeted for artifacts in the current Software Collection. Of course as processes are developed for the existing artifact set, these new processes can be applied against incoming artifacts. Therefore as soon as we have a 'Software Artifact Dataset' within the Collection Database, Alison can start using as part of her Registrar processing.

The next obvious step, and one to keep in mind as we develop processes and procedures for the existing Collection, is how these processes and procedures can be shared with a larger community. That’s a problem left for another proposal.