Structure and Organization of the Software Collection & Preservation Community

Richard P. Gabriel PhD MFA
The Hillside Group, Inc.

Approximate Definition

"A voluntary, collaborative community conserving, collecting, and cataloging software and related artifacts in order to preserve the historical record of computing."

Compare this with:

The CHM Mission: "To preserve and present for posterity the artifacts and stories of the information age."

The SCC Charter: "To develop the policies and procedures necessary to collect, preserve and archive software.

To test those standards and processes experimentally on a number of test cases.

To create an initial list of software that should be preserved in the archive.

To get interested groups and individuals collaborating with us on this preservation initiative.

To augment the resources needed to fulfill the Museum's mission of preserving software."

Mission

- To create a virtually central, as exhaustive as possible, collection of software and related artifacts.
- Help to create tools, processes, techniques, database models, and taxonomies for the collection, preservation, and presentation of software and software-related artifacts.

Components

- Web site
  - collaboration infrastructure such as mailing lists, wikis, weblogs
  - articles, critiques, commentary
  - refereed journal (?, eventually?)
- Collecting and preservation principles
- Repository for collections
  - ask our friends the manufacturers to contribute
- License for contributions
  - contributors retain copyright, if they hold it (we need to worry about legal issues)
  - licensees may catalog, annotate, display, copy, and recontribute the donation to a predefined set of recipients (or perhaps to any organization satisfying particular criteria, or by special negotiation) including the CHM
- Training to be a "Curator"
  - a reward: this title should be one of the rewards for working on the project.
  - we should come up with a nice name for the community ("The Donald Knuth Software History Project," and we should give people who qualify cards that say "Curator, The Donald Knuth Software History Project")

Components

- Governance[1]
  - Executive/Appellate
    - oversee the community, in a sense—mostly concerned with making sure the principles are more or less respected
    - limited size, perhaps 3 people, one from CHM, other two selected by the community
    - dispute settlement
  - Legislative
    - contributors of code or other help
    - leaders of federated sites
    - designees of organizational contributors of code
- Community leader
  - visible leader, perhaps from the governance board
  - leads discussions, performs administrivia, makes routine decisions
  - someone with passion for the enterprise and who really gets what we're trying to do
  - maybe a well-known person, but doesn't have to be

1. Wait for the "Why Governance?" slide

Components

- Federated existing sites
  - general acceptance of principles—perhaps we'll need to refine them at a "constitutional convention"
  - willingness to have their contributions move to the SCPC repository, the CHM, or, for example, the Smithsonian
- Under the CHM umbrella—or "operated" by the CHM

Components

- "History of Ideas" wiki

  This is a pet project of mine. I believe we will misattribute ideas in the future if we go one as we are now. For example, some researchers traced exception handling back to roughly 1972 until some "old timers" pointed out mechanisms from other programming languages in the early 1960s. I believe a wikipedia-type approach would work well for this.

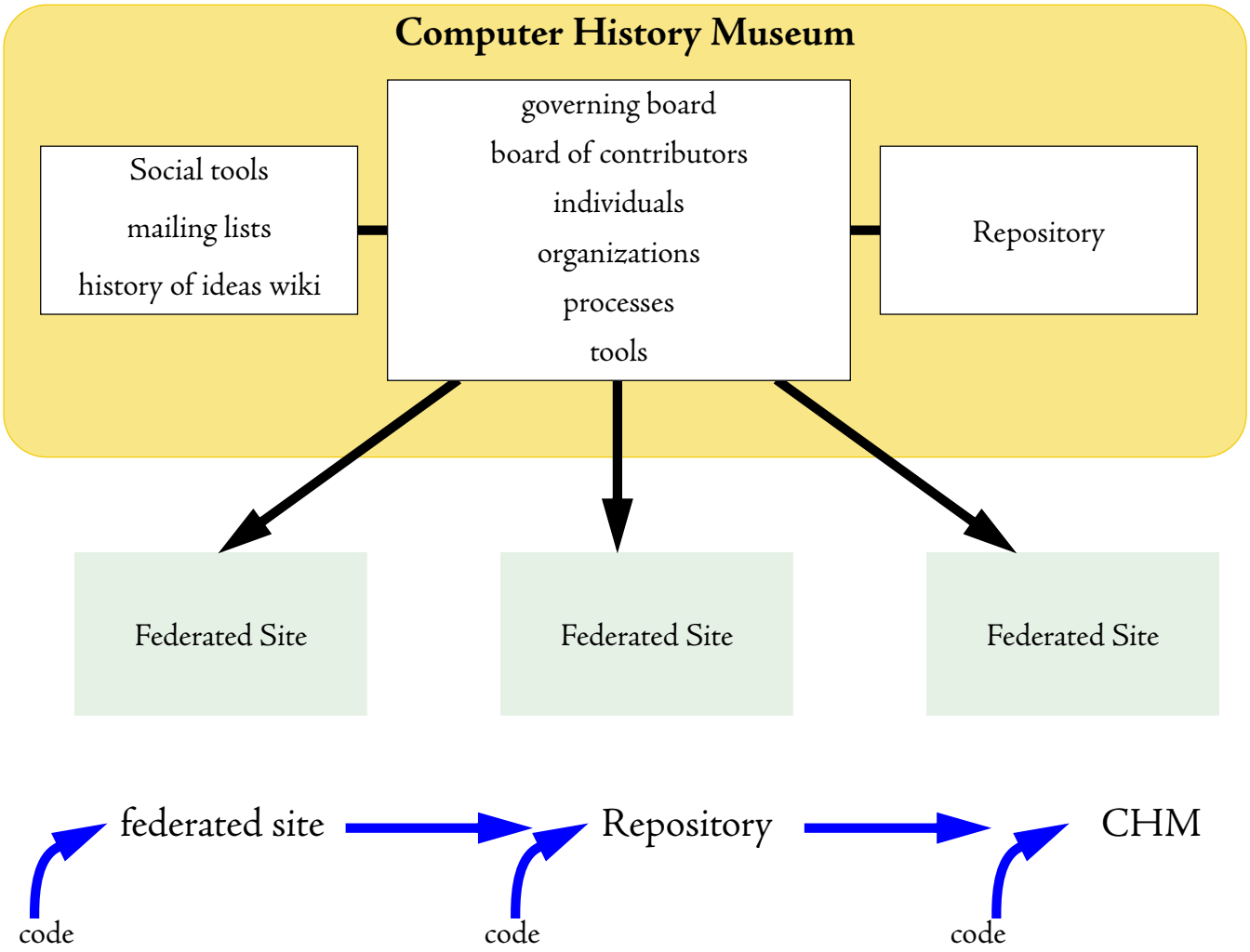
Why Governance?

- Volunteers are doing a lot of work and need to have some control—they don't want to be part of someone else's machine.
- We want to learn from this community, so they need to be able to do things their own way to an extent—the license gives the CHM the right to do things its way
  - we want to encourage them to use CHM tools and processes, but resistance to this could be informative
- For a community to thrive, we need ways for members to grow, and governance is the obvious way.
- The CHM doesn't want to expend a lot of resources on community building, which can be consuming.

Members of the Community

- Individuals
  - collecting software
  - cataloging
  - contributing historical context
- Organizations
  - owners of software that has been contributed
  - owners of software that might be contributed under certain circumstances
  - other sites that are collecting or have collected software that wish to retain their collections—these are federated into the community

Picture of the Community



Contribution Status

- **Conserved**

  Code or other software-related material that is in the repository of a federation member. (To join the federation requires being willing to contribute material to the CHM.)

- **Contributed**

  Material that is in the community repository.

- **Collected**

  Material that has been accepted into the CHM. This material will be minimally cataloged at some point.

- **Archived**

  Material that has been accepted into the CHM and has been cataloged.

- **Displayed**

  Material that is on display at the CHM (either physically or virtually).

Things to Collect

- Reading copies of source code
- Compilable copies of source code
- Ready-to-run object code
- Packaged or Distributed Version of the object
- Documentation
  - User manuals
  - Installation manuals
  - Logic manuals and flowcharts
  - Development design notes
  - Books and research papers
  - Email Exchanges
  - Contract Materials
- Marketing material
- The user's experience
  - Videotapes and photos of the software in operation
  - Interviews with early or important users
  - Profiles of typical (or unimportant) users

Things to Collect

- The developer's experience
  - Videotaped interviews and photos of developers and other people involved
  - Written reminiscences
  - Biographies of key people; Who are all the people
  - Timeline history of milestones and releases
- The business experience
  - Published reviews
  - Information about the companies involved in development, marketing, sales and support
  - Sales history
  - Competitive environment
- The academic experience
  - Publication record and response
  - Information about the researchers, schools, and labs involved in development and distribution
  - Usage history
  - Research environment

Things to Collect

- Questions to Ask:
  - What was the lifecycle of the software & related products
  - How was it designed?
  - For each object, what was the lineage of the objects?
- External Views:
- Commentary, critiques, and annotations
- Simulators
- History of computing ideas (on the wiki)

What Members Can Do

- Contribute material
- Catalog material
- Make lists of possible contributions
- Expand the History of Ideas repository
- Enhance the infrastructure (by working or by contribution)
- Help govern

Services Provided by the Community

- Repository
- Taxonomies
- Cataloging tools
- Principles and processes for conservation and collection
- Training for curators

Why This; Why Now?

- If there is not enough preservation happening right now, then this will help invigorate it
- If there is plenty of preservation happening, then this will bring the collectors together so they can compare and improve their collection, cataloging, and preservation techniques, and discover connections between their collections
- Bringing a community together generally improves creativity through diverse groups and ideas banging against each other

> *Art is a reckless encounter with whatever comes along.*
>
> –William Stafford

- Frankly, if the CHM needs to create tools, processes, techniques, database models, taxonomies, etc, it needs the help of a larger community
- Building a community is not just "establishing a network"—it requires a culture and a way to mature. Often this requires more structure than you would think or like initially. If you don't believe me, believe the expert: "Community Building on the Web," by Amy Jo Kim