

Minutes for SCC meeting of January 19, 2005

by Lee Courtney & Dave Redell

Important Administtrivia

Room for future meetings will be Hopper Room (upstairs north of Admin. area) rather than Noyce Room. See attached CHM map.

Attending

Rudy Batties
Mary Cicalese
Lee Courtney
Bob Fraley
Henry Gladney
Kathe Gust

Phil Gust
Dave Gustavson
Gardner Hendrie
Sellam Ismail
Paul McJones
Chacko Neroth

Bernard Peuto
Michael Powell
Dave Redell
Bill Selmeier
John Toole

New Participant introduced

Rudy Batties physics/math at Stanford, worked at HP & Adobe, now a consultant, Interested in electronic documentation of the history of software.

Action Items and Deliverables (Who; When; What)

All	1/26	Feedback to Sellam on Software Collection Taxonomy
Bernard	1/28	Add Henry, Bill Selmeier, and Bob Fraley to SCC List distribution.
Sellam	2/16	Working set of web collaboration tools available for SCC use by the February meeting.
Sellam	2/16	New iteration of the Software Collection Taxonomy.
Bernard, Lee, Len	2/16	List of Deliverables for SCC for review by committee members

Status updates on specific projects

Fortran (Paul M): More historic documents scanned -- handout attached

CDC (Dave R); No further progress. New item: Jim Thornton (Co-architect of the 6600 with Seymour Cray) dies last week. Carpe Diem.

NLS (Phil G): Difficulties in arranging appropriate meetings have delayed progress for the time being.

Magic (Mike P): No progress over the holidays.

Software Collection Taxonomy - Sellam Ismail

Proposed Taxonomy attached.

He started by saying this is "90% there", although the subsequent discussion may have caused him to revise this estimate. He started from the previous SCC reviews of the ACM and UN taxonomies, and agreed that UN was more suitable. He saw searching as a key driver for needing a hierarchical taxonomy, and devised a 3-level hierarchy, which is presented in detail in his memo.

The top-level categories are Application, Automation, Utility, Programming and Library. Each has two sub-levels giving a uniformly 3-level tree. His plan is to keep the levels (especially the lowest levels) open-ended, and to avoid using catch-all categories ("miscellaneous" or "other") as much as possible.

Many members had comments, including:

- Fine adjustments in categories can have major effects
- A single strict hierarchy is too rigid; alternatives mentioned included
 - a hierarchy with cross-links
 - multiple hierarchies over one repository
 - no hierarchy: just a set of attribute keywords (counter argument: this is too chaotic; perhaps use hierarchy to organize keyword vocabulary?)
 - aspect hierarchies as used by eBay (is this same] as hierarchy of keywords?)
- Need to consider relationship to other taxonomies, especially hardware.
- Should compare with other taxonomies (e.g. UN) as a sanity check.
- Need to track evolving search tool technology and make sure the CHM taxonomy works well with best new tools
- Would we keep adding categories for things that pop up? Current list is not exhaustive. Want to have this as extensible. As time goes on we can refine and add to this structure.
- Doesn't see top tiers being changed over time.
- Similar exercise has been done for hardware. Is there a significant difference between the software and hardware approach? No, shouldn't be. Don't want to have 5 different taxonomy approaches for each different collection.
- Where would embedded SW go? Need structure that supports emerging areas such as embedded (cell phones). Would be a useful exercise to map the list of software targeted for proactive collection (Grady's List) against this taxonomy.
- (Paul) Would it make sense to go to UN list (a flat list) which had a higher level of granularity? Sellam started looking at that, but looks like that list can reduce to this taxonomy. Initial impression is that this taxonomy is single user/PC oriented. Concern about laying this down on a wider range of examples - e.g. Grady's list.
- (Henry) Would like to see citations and in addition a paragraph on each level explaining the how and why. Comment on taxonomy - introduced as something for search, but taxonomy useful in context of hardcopy books. Hierarchy does not lend itself to being useful for modern search (e.g. Google Scholar). This could be a detriment in having a

wider dissemination.

- Guide to searching versus guiding to organizing. This taxonomy is first cut at controlled vocabulary. Would like vocabulary to facilitate volunteers working with the Software Collection. Whenever you have a tree to describe something you invariably have a need for cross-links - i.e. the tree degenerates into a network.

- Sellam is focal point for feedback.

- Is it even feasible or desirable to have a taxonomy? Problem is that often mix purpose of the software, structure of the software, and application of the software. Can we map the universe of software onto this taxonomy, or any taxonomy.

- (Henry) Search tools are not part of taxonomy. All sorts of different types of tools are coming online. Do we have anyone tracking this? YES the Museum does. - Have had discussions in SCC in the past that while we will have single repository, there is the possibility and likelihood that there will be different taxonomies.

- The parallel structure between the different top levels needs some work.

Web tools for SCC - Sellam

Mike Walton has installed Plone. It has limitations, apparently including a limit of one instance per server machine. This is a problem, but Drupal, the only other current candidate, seems to have even worse troubles, including "inexplicably slow" performance. Lacking some breakthrough explanation, Drupal looks like a non-starter. Sellam and Mike will look into this, perhaps in consultation with Eric Smith (who did the 1401 website). This is considered a high priority and the goal is to get a working set of web collaboration tools for SCC by the February meeting.

CHM Software Collection - Sellam

Sorting the software collection Sellam has done some preliminary sorting out of the existing software collection, in advance of any attempt to start actual cataloging activities. This is just physical house-cleaning and weeding out of extraneous stuff.

Testbed - Mary

Mary, Lee and Sellam are working on a Discussion Guide -- i.e. a set of questions to assess usability: how will prospective users want to make use of the CHM software collection. Mary has been studying references provided by Henry G and has found these to be generally quite helpful. In connection with this, Henry passed around a pair of documents from the UK National Archive:

1. Generic Requirements for sustaining electronic information over time
2. Standard for record repositories

He asked that SCC members review these documents, and said that given sufficient positive feedback, he would be prepared to work on adapting the ideas in these

documents to the needs of SCC/CHM.

Repository Info - Henry

Henry circulated and discussed document on British Library. A set of requirements by the (U.K.) National Archives. Deals with long term preservation and authentication. Where should we be in five years. Henry will distribute a URL to the SCC list. Henry can have end-to-end version of this document focused on the Museum as he knows today.

Links to documents mentioned:

<http://www.nationalarchives.gov.uk/electronicrecords/regs2002/>

<http://www.nationalarchives.gov.uk/electronicrecords/generic.htm>

DSpace Comments - Henry

Henry also advised caution regarding adoption of DSpace as a repository for the CHM software collection. While he does not claim to have made a deep study of DSpace, he has heard a couple' of presentations from the DSpace people and has reservations about it. They appear to be spending more time on PR than building product. In particular, he points out that it was built by a collection of universities (raising questions about stability) and seems targeted for large institutions (larger than CHM). He cited two other open-source repository projects: Greenstone and Fedora, which he feels deserve consideration as DSpace alternatives. He also pointed out that commercial offerings such as IBM's Digital Library have demonstrated stability and might well be available to CHM for little or no money.

Upcoming Meetings. NOTE ROOM CHANGE FOR FEBRUARY AND SUBSEQUENT MEETINGS

Day	Date	Time	Conf Room	
Wednesday	February 16	1:00 pm – 3:00 pm	Hopper	
Wednesday	March 16	1:00 pm – 3:00 pm	Hopper	
Wednesday	April 27	1:00 pm – 3:00 pm	Hopper	
Wednesday	May 25	1:00 pm – 3:00 pm	Hopper	
Wednesday	June 22	1:00 pm – 3:00 pm	Hopper	
No July Meeting				
Wednesday	August 24	1:00 pm – 3:00 pm	Hopper	
Wednesday	September 21	1:00 pm – 3:00 pm	Hopper	
Wednesday	October 19	1:00 pm – 3:00 pm	Hopper	
Wednesday	November 16	1:00 pm – 3:00 pm	Hopper	
No December Meeting				