

Minutes for SCC meeting of November 17, 2004

by Dave Redell

Attending

Mary Cicalese (by phone)
Kathe Gust
Sellam Ismail
Doug Jerger

Paul McJones
Bernard Peuto
Michael Powell
Dave Redell

Len Shustek
Ed Taft
Kirsten Tashev
John Toole

Status updates on specific projects

We heard several brief updates regarding progress on the various test-case projects being pursued by committee members.

Fortran Paul McJones gave an update on his collection efforts so far. He has a variety of materials, some in hand and some just located. They are in a variety of formats, including both machine readable media and materials that need scanning (paper, microfiche). He is starting to ask some basic questions about the material he has collected, including:

- What exactly do we mean by a “collection”, beyond just a *de facto* set of artifacts?
- What are the best ways to make collected materials available?
- What is necessary in terms of permissions to use copyrighted material and other forms of protected IP? How should these be obtained?

The last point started a brief discussion of the various types of material and how different the process can be – e.g. for published papers vs. for proprietary source code. Older materials can be both easier to deal with (since they have less presumed commercial value) and harder (since its hard to find anyone with knowledge and authority). Finding the right contacts inside the owning companies (if they still exist!) can be a key. Also, it is important to characterize our efforts as being a Museum activity (rather than an individual activity). It is also important not to promise that the museum will accept, retain or display any particular artifact until a considered decision has been made, so that we don’t accidentally flood the collection with junk.

Control Data Dave Redell gave a brief report on activities following up on the CHM visit by several members of the “controlfreaks.org” group, which is an *ad hoc* online group dedicated to preserving old CDC software and running it on emulators. One activity is to catalog the sizable collection of CDC documents held by the group, in a way that maximizes its potential utility to CHM. Dave originally intended to post the CHM extended Dublin Core document to the group, but was discouraged from doing this by the size and level of detail of the document. Sharon’s “Organizing Principles” document was suggested alternative. It was also suggested that, at the very least, using a CHM-friendly file naming convention would be a start.

IBM 1401 Mary Cicalese reported that in addition to the intensive hardware-related activity on the 1401 restoration project, there is a parallel software effort being spearheaded by Ron Mass. It was suggested that Ron be encouraged to attend some or all of our meetings to keep us up to date on the progress of this work.

NLS Kathe Gust reported that an NLS group has now formed and will soon generate a plan for what to do about conservation of the NLS software artifacts.

Magic Mike Powell reported on his initial work on Magic. He said that a key property of this effort is the wide variety of artifacts, due to the lack of any coherent commercial process behind the original creation of the Magic software. He has found a large amount of fragmentary material, and it is a real challenge just to figure out what is definitive. Mike is thinking hard about what the proper goal for this effort should be. One initial question is: what archival format(s) to use for source files and documents that he finds. He cited burning CD-Rs and even printing hardcopies as possible alternatives. This led to a brief discussion of the general problem, which was taken up again in more detail during Sellam's talk on his Software Collection Plan (see below).

Talk on SCC activities at VCF

Bernard Peuto gave a talk at the recent Vintage Computer Festival about the SCC and its activities regarding historic software. He felt that the talk was very well received, and he proposes to make the slide set available as the nucleus of a generic SCC slide presentation that could be used by any of the members to describe the committee to interested audiences with minimal preparation overhead. Bernard also suggested that one or more white papers be written describing the SCC and its activities.

Outreach Workshop

The current proposed target for this is May 2005. We need to sanity check this as a realistic target and get sign-up by other members of the committee as soon as possible. The immediate goal is to form a subcommittee and draft a top-level plan for circulation and review.

Chess Exhibit

Kirsten Tashev described the Chess Exhibit Project, which is in the early stages of planning an exhibit for initial display targeted at June 2005. The goal is to produce a first instance of a "real" museum exhibit (i.e. as opposed to the current informal visible-storage styles of exhibits). Software aspects that would be highlighted would include the alpha-beta pruning and minimax algorithms.

Software Collection Plan

Sellam Ismail gave a presentation summarizing his current plan for organizing the existing CHM software collection over approximately the next six months. His plan has similarities to many of the discussions that the committee has had before, but it's very encouraging to see a coherent timeline emerging for actually carrying out the various activities.

Sellam proposes the following steps as part of Stage I:

1. Determine the scope of the existing collection
2. Figure out the de facto organization of the material, including:
 - a. Media types
 - b. Taxonomy
3. Database design: define tables mapping to the CHM Dublin Core definition.
4. Data entry facilities, with a focus on web-based tools
5. Physical storage for actual historic media – shelving and physical organization
6. Initial cataloging
 - a. Prune out extraneous material
 - b. Start data entry
 - c. Base entry on external ID – not reading the media (yet)
 - d. Align storage with catalog – allow location of physical media.

The idea is that steps 1-5 would take about 3 months and step 6 would take the other 3 months.

Stage II would begin the actual recovery of the content of the archived media. We will need to establish best-practices procedures for handling and reading of the various media. Once the bits can be recovered, the key question is: what to do with them (cf earlier discussion under Magic). Sellam has invested considerable thought in this question over the years, and proposes a text-oriented archival format based on embedded XML. The idea is to retain not only “the bits” but also all of the “significant” information about lower-level media formatting (e.g physical records, volume labels, etc.) There is obviously a slippery slope here – e.g. one doesn't want to end up saving the analog waveform of the flux transitions on a magnetic medium as part of archiving a Fortran source file. One criterion that was proposed is: retain all characteristics of the physical format that could be sensed by any level of the software. This should support any level of emulation without causing an undue explosion in file size.

One goal for the archival format is “survival for 500 years”. We also need to coordinate with other outside groups that are working on this problem.

Upcoming Meetings

Day	Date	Time	Conf Room
Wednesday	December 8	1:00 pm – 3:00 pm	Noyce
Wednesday	January 19	1:00 pm – 3:00 pm	Noyce
Wednesday	February 16	1:00 pm – 3:00 pm	Noyce
Wednesday	March 16	1:00 pm – 3:00 pm	Noyce
Wednesday	April 27	1:00 pm – 3:00 pm	Noyce
Wednesday	May 25	1:00 pm – 3:00 pm	Noyce
Wednesday	June 22	1:00 pm – 3:00 pm	Noyce
No July Meeting			
Wednesday	August 24	1:00 pm – 3:00 pm	Noyce
Wednesday	September 21	1:00 pm – 3:00 pm	Noyce
Wednesday	October 19	1:00 pm – 3:00 pm	Noyce
Wednesday	November 16	1:00 pm – 3:00 pm	Noyce
No December Meeting			