

**Minutes for SCC meeting of February 18, 2004
by Dave Redell**

Attending

Allison Akbay	Gardner Hendrie	Fritz Schneider
Mike Blasgen	Paul McJones	Len Shustek
Sharon Brunzel	Chris McKay	Dag Spicer
Mary Cicalese	Bernard Peuto	Ed Taft
Lee Courtney	Mike Powell	Kirsten Tashev
Ed Feigenbaum	Dave Redell	John Toole
Dick Gabriel	Anel Rodriguez(?)	Mike Walton

Logistics

Main presentations were Taxonomy and Metadata, followed by a discussion of the relationship between the two. The general plan going forward is to iterate such presentations on foundational issues on a regular basis (every month or two) until they converge.

Task List and Schedule

There is a revision of the task list available. It will continue to evolve, based on priorities and other considerations. It is clear that the activities of the committee will be paced in large part by the availability of both staff and volunteer time.

Software Taxonomy Subcommittee

Goal: Decide on a set of qualifiers defining a hierarchy

Approach: start by reviewing existing taxonomies

Questions: How would we use a taxonomy? How does taxonomy relate to metadata?

What are shortcoming and limitations of each taxonomy?

Two initial taxonomies studied:

UN – basically commercial procurement codes

ACM – academic/bibliographic categorization

Proposal: start with UN taxonomic codes and evolve by:

- 1) Categorizing *<the next 1000 S/W artifacts?>*
- 2) Collecting a list of exceptions
- 3) Periodically reviewing the exception list (e.g. every 1000 items) and revising/expanding the set of categories

Questions/comments:

1. Why isn't ACM taxonomy better, since it is computer oriented
(Answer: probably due to bibliographic slant)
2. Ongoing idea: end up with multiple taxonomies on same body of material
(Sharon: the term *categorization* seems preferable to *taxonomy*)
3. There is a good (~20 page) introduction to UN SPSC codes – “a very good read”
4. (Ed T.) Can one item belong to multiple categories within a single taxonomy?
(Sharon: yes, which is part of her preference for avoiding the term taxonomy)

The subcommittee was tasked to draft a writeup on their proposed approach

Software Metadata Subcommittee

Introduced new member: Chris McKay – highly experienced volunteer

Subcommittee proposes to use Dublin Core as foundation. The idea is to define items and relationships. The relationships are captured by Business Rules encoded as metadata.

Example:

Items *Engineer* and *Source-File*

Relationships: *Creates* and *Created-By*

Structure is: *Engineer Creates Source-File; Source-File Created-By Engineer*

Metadata does *not* imply a hierarchy but rather a rich web of arbitrary relationships.

Of course, stylized use of metadata can represent a hierarchy, such as the *Is-A* relationship of taxonomy

Example:

Compiler Is-A Application

Application Is-A Program, etc.

In addition to relationship attributes, metadata represents simple attributes such as *Creation-Date*

More formally, metadata consists of **<name,value>** attributes attached to objects

name = qualified name = part{.part}*
value = literal | objectID

Example 1: encode “A is a part of B”

A has attribute **<name, value>**

where **name** = *Relation.IsPartOf* and **value** = objectID “B”

Example 2: encode “C was created on January 15, 2001”

C has attribute **<name, value>**

where **name** = *CreationDate* and **value** = literal “01-15-2001”

Questions/comments:

1. (Ed F.) Discussion has gone directly to mechanism, rather than starting with problem statement. Perhaps problem being solved is obvious to subcommittee members, but should be discussed to give context for full Committee.

In response, a problem statement was proposed: *Goal is to enable finding any object (or class of objects) in the collection via an appropriate search of the metadata.*

2. (Bernard) Need to control complexity, while providing sufficient expressive power to do the job. This is a crucial tradeoff. Need to understand usage, especially the spectrum between *curation* and *description*. Someone suggested the following definitions:

Basic cataloging (= *description?*) is simple recording of the *what is* of the collection

Curation is a more complete process of providing *context* for collected objects

3. The focus of the metadata is on describing *physical objects*, not their *conceptual content*. This may explain why the UN taxonomy seems to work better than ACM taxonomy.
4. So far, types are just a controlled vocabulary of textual values. Could be embedded as abstract objects, but practical value of doing so is not clear.
5. (C. McKay) There is always tension between an *open-ended* vocabulary of relationships to maximize flexibility, vs a *controlled* vocabulary to maximize stability. Need appropriate balance; in particular, too much flexibility is seductive but can quickly lead to chaos.
6. What is the relationship of our taxonomy and/or metadata to “inter-library loans” and other forms of collaboration? One answer (Sharon): Supporting this kind of activity is a key motivation for adopting the Dublin Core.

Brief update on Webtools & Support

Web Tool: new server exists, tool not yet available for information sharing via server.

Distribution List: Committee alias creation in progress, expect completion within a day or two.

Web Log: Archive of messages and other shared documents being planned. Details (Wiki or etc.) TBD.

Cataloging of Existing Software Collection

This is our initial testbed before all other projects.

The task comprises:

1. Organizing physical objects for preservation
2. Inventory of objects
3. Creation of metadata
4. Entering metadata into database

To do this we will need:

- Volunteer training for participants
- Reference Materials for cataloging
- Database creation based on closure on metadata issues
- Sufficient computers (should be no problem)
- Staff availability for project oversight
- Closure on requirements and goals beyond just a minimal inventory
- Clear understanding of relevant Best Practices (e.g for reading old magnetic tapes etc.)
- Clarification of what can be borrowed rather than reinvented

A task group was formed to write up a specific proposal for beginning this project. The group consists of: Lee, Bernard, Kirsten and Allison.

A Closing Plea

Dick Gabriel reminded the group of the importance of prioritizing the old/endangered/significant/rare materials and making sure that they get preserved before time runs out.

Agenda Items for March 24, 2004 Meeting

Review of Action Items
Continuing Review of "First Ten" List
More on Webtools & Support
Report on Community of Practice
Start work on the "Top 100" list (per February discussion)
(Further reports from Taxonomy & Metadata Subcommittees will resume in April)

Upcoming Meetings

Wednesday March 24 from 2:30 pm to 4:30 pm. [Note different time]
Wednesday April 21 from 1:00 pm to 3:00 pm.
Wednesday May 19 from 1:00 pm to 3:00 pm.
Wednesday June 23 from 2:30 pm to 4:30 pm. [Note different time]
Wednesday July 21 from 1:00 pm to 3:00 pm.
Wednesday August 18 from 1:00 pm to 3:00 pm.
No September meeting