Summary Report of Committee Meeting

Tuesday April 14, 1998
Georgetown University, Washington, DC

Present: Pamela Ellis, David Green, Lorna Hughes, Mike Neuman, John Unsworth.

A subset of the Database Group met in the Reiss Computer Lab at Georgetown University, hosted by Michael Neuman.

SGML v. Database & Dublin Core

John Unsworth demonstrated a prototype that included his reconstruction of the CETH prototype and material using an SGML entry form and SGML editing.

The material from the CETH prototype was, of course, by now quite old and, as an illustration of the necessity in any project for regular, thorough updating, virtually all the links in the material were dead. However, more to the point, this demonstration revived the discussion of whether to use SGML or a database structure for our project.

SGML was described as being ultimately more flexible: in construction, in searching and in longevity. However, the interoperability of most contemporary databases (using the ODBC standard), the ease of output from one to another and the ease of data input using web-based forms made a database structure favored by the majority of the group. Given the recommendation to go with a database structure, the group concluded that Dublin Core would be a good base to use, on to which necessary additional fields be added.

Audience: What This Is: What It Is Not

There was some discussion again about primary audience. Although there was some voicing of the importance for this information to be usable by a general audience of non-scholars, the majority of the group re-affirmed that this database project should principally serve humanities computing practitioners and funders and that its primary function was to quickly and deeply convey the nature of high-quality, exemplary, current humanities computing projects.

It was suggested we develop different interfaces for different users (a general audience; funders; scholars; practitioners).

It was seen to be helpful to define this project against other related projects. Thus, this was not planned to be a comprehensive list of humanities computing resources. In the UK, the Arts & Humanities Data Service was clearly about the compilation and redistribution of digital resources. There is no equivalent to this service in the US; however this form of information delivery is probably inappropriate for the US, for which future integrated digital library structures will emerge. As noted before, ARL is compiling a database of digital library projects at its member institutions. David Chesnutt is working on a national database of documentary editions. Keeping abreast of other projects (and referencing them) will be important in keeping this project clearly defined.

This project is also not to be confused with guides or gateways to Web resources, such as the American Arts & Letters Network, the Voice of the Shuttle, HUMBUL or other Web-based projects.

Characteristics of our database is that it would provide information on the approach, structure, software and tools used in producing a resource; it would be ongoing and records would need to be regularly updated.

Although we would be interested in peer reviews of the projects (and we might encourage Chorus and the Chronicle of Higher Education, for example, to regularly review projects in the database and include the reviews), we are more interested in the technical details of the projects that would indicate the quality of thought and work behind them.

Role of funders & Corporate Funding

It was suggested that funders be subscribed to a private e-mail list of funders and practitioners. To be eligible one would either need to be an interested funder or a practitioner who had submitted a record to the database. One could ask funders to require the submission to the database of digital humanities projects they fund.

The issue of interesting commercial sponsorship was raised: SoftQuad and Informix were mentioned as two potential candidates

Types of Projects to be included: different records for different projects

Discussion revealed that this group was interested in four classes of humanities "projects" that would require slightly different record structures. It was suggested these projects fall into the following classes:

- Content Creation (archives; editions; serials)
- Teaching projects
- Software Tools
- Indexing, finding aids, discovery tools

Under tools (and perhaps elsewhere) we thought it a good idea to declare what tools are needed by practitioners, to stimulate thinking.

Lorna, Pamela, John and Mike each agreed to take one of these and create a prototype record, using Dublin Core fields as a base and adding on fields necessary for adequately describing projects.

Criteria For Inclusion

We need to articulate the standards and values that need to be met for inclusion in this database. David Chesnutt's comment on E-Docs listserv about the minimum requirement for documents (" 1. What is the source text for the document?; 2. How closely does the electronic text follow the source?; 3. How was the document prepared? 4. Who was responsible for creating the electronic document?") is a useful start. We should also look at the DLI-2 application form and NEH for clear statements of criteria to be met before being admitted to the database. The re-usability of material and tools would also be important.

One clear criterion would be a committment to update information at least once a year.

Timeline

Participants at this meeting will present their prototype records and opening set of criteria before the whole group for discussion and an attempt to iron out agreement on management structure and business plan by the end of September via the list and meetings planned at ACH/Hungary and DRH/Glasgow.