

## ITPB Subcommittee on Research and Education Data Management

### Revised DRAFT Report to ITPB

April 24, 2008

- Christine Borgman, Information Studies, Chair
- Marilyn Raphael, Geography
- Alan Robinson, Medical School
- Gary Strong, University Librarian
- Sam Morabito, Business enterprise
- Margo Reveil, ATS
- Elizabeth (Libbie) Stephenson, Social Science Data Archives

### Research and Education Data Management ITPB 2007-08 Priority Area

#### Work to date

- Priority area from July 2007 ITPB retreat
- Problem Scope presented at November 1 ITPB mtg
- Subcommittee formed November 1
- Committee met November 26, January 14, and April 14, 2008
- Alan Robinson reported at Dec 14 ITPB mtg
- Recommendations to ITPB on Research Data: January 29

#### Scope of topic

- The campus research and educational data management - i.e., data management as infrastructure

The committee divided the topic into two related parts, research data and educational data. We report them here in two sections, with general conclusions and recommendations at the end. These are pressing problems for which UCLA should take a leadership position – not a defensive position.

#### I. Research Data

##### A. Initial framing of the strategic issues for UCLA:

1. Provide assistance for researchers to create data management plans
2. Develop an institutional response to the directions of grant agencies regarding data
3. Establish policy and standards for setting up databases
4. Determine who “owns the problem”

## B. Recommendations and priorities for research data

We divided the problem as follows:

- Develop an institutional response to the directions of grant agencies regarding data
  - Help for researchers to create data management plans
  - Policy and standards for setting up databases
- Who owns the problem
  - Identify responsibilities for education, curation
  - Identify variability in responsibilities

Institutional responses: Data management plans

Researchers need assistance to create data management plans

Known constraints

- Funding agencies (e.g., NSF, NIH) are now requiring data management plans for some or all grants, they have not yet established clear guidelines for data management plans
  - Some disciplines/fields have established guidelines, e.g. survey data management in the social sciences, genome data, seismic data
  - Many fields have minimal guidelines or common metadata standards
  - Few national / international bodies have established best practice guidelines for data management
- Notions of “data” vary widely by discipline and by research problem
- Retention practices vary widely by discipline
- Funding agency requirements for data retention may be at odds with IRB policies for data destruction

Actions for UCLA to take

- Identify current practices
  - Survey best practices on campus
  - Survey best practices elsewhere
- Establish / enhance services
  - Website, actively maintained
  - Education and referral services

Staffing requirements in support of these actions:

- Assistance to subcommittee to assemble survey / environmental scan
- Compilation of results and recommendations
- Construction and maintenance of website
- Coordination of education and referral services

## Institutional responses: databases

- Policy and standards for setting up databases
  - Best practices
  - Responsibility

### Who owns the problem

- Identify responsibilities for education, curation
  - Within school, department, research group?
  - Campus level?
    - VC for research
    - Library
  - Funding agency?
  - Discipline / research area?
- Identify variability in responsibilities
  - Content (e.g., by type of material)
  - Origins (e.g., locally generated vs. purchased)
  - External criteria (e.g., funding agencies)
  - Practice (e.g., disciplinary variation)
- Identify areas for collaboration and shared responsibility across UCLA, among UC campuses, UC-wide

## II. Education Data Management

The Subcommittee met on April 3 to address issues specific to educational data. By this we mean both educational modules (work products of UCLA faculty, instructors, and teaching assistants) and records about instruction (e.g., transactions of online courses, chats, content posted to Moodle and other learning environments). We generally exclude records covered by FERPA (grades, student applications, etc.) as these tend to be in protected systems and are not intended for open access or sharing. Our concern here is with pedagogy.

### A. Develop an institutional response to educational/pedagogical data management

While some policy work on management of educational data exists at the national and international levels, campus-level strategy has received little attention. National activity includes the National Science Foundation Task Force on Cyberlearning (*Borgman, Chair; report due June, 2008*), Open Education Resources Program (<http://www.hewlett.org/Programs/Education/OER/>); MIT's Open Courseware Initiative (<http://ocw.mit.edu/OcwWeb/web/home/home/index.htm>), and the National Science Digital Library (<http://nsdl.org/>). Issues surfacing at the national level are equally important at the campus level.

We identified the following set of issues for UCLA:

1. Studying and assessing student progress and learning presents challenges for privacy and confidentiality. State and federal oversight/regulations may increase. Use of course management data for pedagogy and tailoring must be balanced with concerns for ethics, privacy, and regulation.

2. The use of mobile and multi-functional devices by students has increased. Students from higher socio-economic backgrounds tend to have access to better technology outside their classrooms.

3. Platforms for creating, using, sharing and publishing course materials need to adapt to a fluid technological environment capable of being accessed by students and faculty at all levels of technological sophistication and socio-economic background.

4. The Open Textbook movement suggests a move away from “publishing” of course materials and towards open courseware/materials. Ownership of faculty created course materials is already addressed by UC Copyright Policy (i.e., faculty own their course materials unless “extraordinary resources” were invested by the university, in which case rights are negotiated).

<http://www.ucop.edu/ucophome/coordrev/policy/9-25-03copyright.html>

5. Examine UC policies for ownership of research data for alignment with current policies of funding agencies and community best practices. The 1992 policy on copyright

<http://www.ucop.edu/ucophome/coordrev/policy/8-19-92att.html>

states (under Purpose and Scope) “This Policy addresses ownership of copyright; **it does not address ownership or access to the underlying research results or data**, as covered in [Academic Personnel Manual Section 020, University Regulation 4.](#)” APM 20, regulation 4 was drafted in **1958**.

5. Departments and disciplines generally handle faculty-generated course material at the department level at UCLA. This includes storing multiple versions of materials for the same course; shared among colleagues and possibly beyond the department and institution. UCLA has multiple courseware management systems; some are maintained by department staff and some by faculty members themselves. Accordingly, management of educational data has been local, except for registrar records.

6. As the Common Cooperative Learning Environment (CCLE) is deployed on campus, educational data management issues must be addressed.

Recommendations and Priorities:

Address education/teaching materials as sharable community resources

- Stewardship of educational materials should reflect role of the University in the 21st C.

Develop an institutional response to directions of higher education regarding educational materials developed for and about students

- Incorporate campus policies for oversight and management in line with state and federal regulations (FERPA)

Set multiple strategies for common goals; campus strategy does not require monolithic policy

- Evaluate Content Management Systems (e.g., IRODS at SDSC; Hive) for campus or UC wide implementation as a means for faculty to self archive materials in multiple formats with associated metadata.
- Evaluate use of UC eScholarship and other open and public repositories for preservation of data
- Assess faculty preferences and practices/workflows

Identify responsibilities for management of faculty generated teaching materials and for student records

- Clarify goals for preserving multiple types of learning objects in multiple formats
- Establish best practices and responsibilities
- Set retention policy(ies) for variety of materials and/or formats
- Clarify copyright and IP policy with respect to course materials and student records

Clarify assumptions regarding technology to achieve:

- Ability to provide access to students with diverse, mobile devices, through distributed wireless and campus facilities
- Ability to support faculty development of learning objects and course delivery within GIS, medical, engineering, architecture, statistical, other multi-media and textual applications
- Ability to support long term curation, revision and preservation of educational data including creation and manipulation of metadata

### III. Conclusions and recommendations

- UCLA should take a proactive, leadership role in the management of research and educational data.
- We should be aligned with state, national, and international policies.

Specific actions include:

- Identify current practice
  - Survey best practices on campus
  - Survey best practices elsewhere
- Establish / enhance services
  - Website, actively maintained
  - Education and referral services

These actions are being the scope of an ad hoc subcommittee. We recommend the following immediate actions for the campus:

- Assign staffing resources
  - Best practices survey
  - Website development and maintenance
- Determine campus responsibilities
  - Education and referral on data management plans
  - Creation and curation of data resources

Continue this subcommittee for another year to pursue these recommendations, provided that some staff support is available to do so.

The long term campus goal is to identify, establish, and maintain best practices in data management and curation.