

## SOPAG's Digital Preservation and Archiving Committee - Charge - 4/9/01

The long-term retention of digital library materials represents an urgent problem for the University of California. Faculty are hesitant to embrace publishing in "electronic only" formats, as they require more assurance that their scholarship will endure through time. In addition, UC Libraries need to create an environment where these digital materials become part of their permanent collections, thus ensuring they are available for use by future generations of scholars. Finally, the University must be concerned with protecting its significant and growing investment in these digital assets.

The root of the digital preservation conundrum is that the problem is not solvable. Libraries can never *guarantee* that they will be able to migrate 100% of their digital materials forward over time, just as they cannot guarantee the same for print collections. Therefore, the general approach to addressing the digital preservation must be the same as with print. That is, libraries must work to reduce the risk of losing materials over time, to some acceptable level. That said, the risk factors for digital materials are completely different than with print. Print materials can go brittle over time, be stolen, or lost in a physical disaster, such as a fire or flood. Digital materials are at the most risk when technology changes and they need to be migrated forward.

The scope of your charge covers digital materials created by UC libraries to CDL digital standards, which are to become part of their permanent collections (e.g., MARC records, EADs, or photographs, manuscripts, etc., created as CDL standard objects). While it is certainly possible that an archival repository could some day be open to other materials, (e.g., faculty created databases, web pages, etc.) this will be explored at a later time.

You are charged to:

In particular, investigate work on collaborative solutions.

- EADs, Electronic books (digitized and text based), journals, maps, manuscripts, photographs, etc.?
- As a working definition for digitized materials, assume that an archival object captures all the relevant information from the original. Identify and address questions such as, "How much descriptive, structural and administrative metadata must be stored with the object?" Should we archive objects that have not been created to the UC library standard, as they still represent an investment by the University? If so, how do we decide what to keep and how much metadata is needed?
- For example, migration to new technologies, physical disasters (e.g., disks being destroyed), protecting the integrity of an object (e.g., keeping the components of an object and their relationship to each other intact - the 300+ files that make up a digitized book with 12 chapters and 300 pages), links in the object pointing outside the archive to other materials that may not be archived, etc.?
- Should the archival repository have levels of service? For example, should we have one service that stores "bits" only? That is, the organization depositing the files is responsible for understanding and migrating the contents of the files archived. Should we have another level of service for "classed objects," where the archive can take some responsibility for migration, as the contents and encoding of the file is well understood? What type of search indexes should an archival repository support? How are objects within the archive updated?
- What policies and operational best practices should be set before we create an UC archival repository? Should the same be required contractually of outside organizations that offer to archive our materials? If not, what is different?

- *Review the literature and consult with other ongoing digital preservation explorations to ensure we leverage our work on that which has gone before.*
- *Identify the classes of digital library objects that can be created to CDL standards that are most likely to be archived, within the scope of your charge;*
- *Determine what constitutes an archival object;*
- *Identify the risks related to the potential loss of these digital library materials over time and suggest means to mitigate these problems;*
- *Define the functionality of an **archival repository** that would help to mitigate the risks of losing digital materials over time and how it would differ from an **access repository**, if at all;*
- *Draft written recommended policies and guidelines for the creation and operation of an archival repository;*

SOPAG expects your charge to be completed by July 1st, 2001. Thank you in advance for sharing your expertise in this very important project!

### **Membership - SOPAG Digital Preservation and Archiving Committee**

Howard Besser, UCLA  
Curtis Fornadley, UCLA  
Anne Gilliland-Swetland, UCLA  
Sal Guerena, UCSB  
Bernie Hurley, UCB (Chair)  
Richard Marciano, UC San Diego Supercomputer Center  
Reagan Moore, UC San Diego Supercomputer Center  
Barclay Ogden, UCB  
David Walker, CDL  
Brad Westbrook, UCSD