

# Digital Preservation Program



## *Preserving the University of California's research, teaching, and learning*

[http://www.cdlib.org/programs/digital\\_preservation.html](http://www.cdlib.org/programs/digital_preservation.html)

### **The Goal**

Ensure long-term access to digital information that supports and results from research, teaching, and learning at UC.

### **Material at risk**

- “Born digital” primary materials (content produced by scholars in the course of their research)
- Web-published content (content available exclusively online)
- Digitally reformatted content (digital copies of items held by UC libraries and museums)
- The published record (journal articles, monographs, technical reports, conference papers)

### **Strategy**

Develop a backend infrastructure—technology, policies, standards, and procedures—that the UC campus libraries can use to meet their local preservation needs. Tools will support the identification, capture, description, organization, presentation, and persistent management of digital content.

### **A Preservation Solution**

Develop a robust and reliable preservation repository that can be replicated and extended for use by the campuses. Design and develop tools to support independent campus preservation efforts, including web harvesting, as well as guidelines and best practices to support their use. Much more than a failover or backup service, the digital preservation program allows UC to preserve content.

The UC Libraries Digital Preservation Repository (DPR) <http://www.cdlib.org/inside/projects/preservation/dpr/> is the foundation University of California libraries digital preservation program. The digital preservation repository is a set of services that support the long-term retention of digital objects for the benefit of the UC libraries and their users. The services and associated storage are based at the California Digital Library. There are two ways to interact with the Digital Preservation Repository:

### **THE NEED FOR DIGITAL PRESERVATION**

- Number of academic/scholarly journals published online: 15,757
- Percent of web-based references in scientific articles from three major journals inaccessible within two years of publication: 21 percent
- Annual cost of UC libraries' online journals systemwide collections: \$20 million
- The number of terabytes of information that are on the surface (not in databases) of the World Wide Web: 170
- Number of times the Library of Congress's print collections would have to be replicated to equal the volume of World Wide Web: 17
- Percent of U.S. federal government publications produced exclusively online in 2003: 65 percent
- Estimated percent of U.S. federal government publications available exclusively online by 2008: 90 percent

## Web Interface

Using the web interface users can experience the DPR's full services. It is the simplest option and is ideal for users with limited programming support.

## Java Toolkit

Using the Java Toolkit libraries can plug the DPR into their own Java programs and tailor it to meet their needs. This method is ideally suited for campuses with programming support

## Priority collections for inclusion in the preservation repository

**1.** Data managed by the CDL and made available through systemwide services: The CDL acquires and/or manages access to an ever-growing volume of digital information that is made available without restriction and without access charges to individual users. With many of these services and collections, such as the eScholarship Repository and the Online Archive of California, the CDL actively manages the digital files, even as they may reside (and be accessible from) elsewhere.

**2.** Web and third-party collections: The UC libraries may identify a common or pervasive need to capture, manage, and preserve third-party digital information that is essential for UC users, is publicly accessible, and significantly at risk of loss, such as web-based government or political information. The preservation program is developing tools that will enable the UC libraries to identify, capture, describe, organize, and ingest web-based materials, and to provide some access to the collections.

**3.** Campus-based digital collections: The infrastructure developed by the preservation program will be available to campus libraries that will use the infrastructure to preserve materials in which they have a distinctive archival interest. In addition to the online materials in the shared digital collection, materials in this category include reformatted materials and their associated finding aids, and third-party databases that are commercially available and acquired or licensed by campus libraries.

The UC Libraries Digital Preservation Program is also investigating a preservation strategy for online journals. Priority will be given to journals that are of scholarly value, are at risk, represent a significant investment, and have a commitment to advancing the sustainability of scholarly communication.

The CDL's preservation partners include the 10 UC campus libraries, and digital library and preservation researchers around the world. Grant funding has been received from The Andrew W. Mellon Foundation and the Institute of Museum and Library Services.

## Contact

Program Director Patricia Cruse  
Patricia.Cruse@ucop.edu