

**INSTITUTE OF MUSEUM AND LIBRARY SERVICES FINAL  
REPORT FROM THE CALIFORNIA DIGITAL LIBRARY:  
Preserving Digital Materials: Creating an Open Archival  
Information System (OAIS) Conformant Preservation Repository  
for Multi-Institutional Use**

March 2005

1	Introduction .....	2
2	The DPR in Context.....	3
2.1	Introducing the DPR .....	4
2.2	Content Considerations .....	4
2.3	DPR Nuts and Bolts.....	6
3	Guidelines and Procedures.....	10
3.1	Use of the DPR.....	10
3.2	Governance .....	12
4	Communications .....	13
4.1	Communications Within UC.....	13
4.2	Communications Beyond UC.....	16
4.3	Assessment as Communication .....	17
5	Project Deliverables and Outcomes-Based Assessment.....	17
5.1	Establish Preservation Repository with Associated Operational and Governance Structures Using Open Archival Information System (OAIS) Reference Model 17	
5.2	Migrate CDL Digital Objects to METS .....	18
5.3	Successful Delivery of Objects from the Repository.....	18
5.4	Research Questions Regarding Repository Operations and Policy .....	19
5.5	Increased Community Awareness of Digital Preservation Challenges and Digital Preservation Program.....	21
5.6	Document and Share Challenges, Decisions, Costs, Methods, Results .....	21
5.7	Positive Role of a Preservation Repository in Service Quality .....	22
5.8	Create Preservation Program at UC Based on Ongoing Repository Support ..	22
6	Lessons Learned and Issues Raised.....	23
6.1	Empowering Libraries: Preservation as a Utility Service .....	23
6.2	Viability of Self-Service Archiving Model .....	24
6.3	Bringing Consistency to an Otherwise Anarchic Archive: The Standards-Based Approach .....	24
6.4	Content Migration and Geographic Replication .....	25
6.5	Technical and Organizational Re-engineering.....	25
6.6	Implementing a Phased Approach.....	27
6.7	OAIS Reference Model.....	28
6.8	Preservation as Institutional Priority .....	28
6.9	Trust.....	28
7	Next Steps .....	30
7.1	Extending the DPR's Capacity.....	30
7.2	Extending the DPR's Client Base .....	31
7.3	Resolving Key Issues .....	32
7.4	Assessing Extensibility of Standards-Based Approach .....	32
8	Conclusion .....	33
9	Acknowledgements.....	34

# 1 Introduction

With this final grant report we are pleased to announce the successful development and initial launch of the University of California Libraries Digital Preservation Repository (DPR). Hosted at the California Digital Library (CDL), the DPR includes the technical and service infrastructure as well as the policy and protocols required to support its use. This report documents the development of the DPR, its operation, and the lessons learned in its construction, rollout, and continuous assessment. It also identifies next steps as the CDL explores opportunities for making the DPR available to libraries and other cultural organizations that reside outside the University of California.

.....

Digital information is a vital part of our nation's heritage. It is essential in the support of teaching, learning, and research. And it is fundamentally at risk. Research, public, and special collection libraries have a crucial role to play in providing enduring access to this information. Indeed, their historic missions entail a responsibility for safeguarding it. Yet libraries have been hampered in taking up this role in a digital age because they cannot individually build the considerable technical infrastructure that is required.

The purpose of the DPR and the IMLS grant-funded work supporting its initial development is to test a hypothesis: by providing the back-end infrastructural components of digital preservation services as a utility to libraries, libraries will actively engage with the front-end activities, notably the selection, capture, and ongoing management of selected digital information.

The CDL is ideally positioned to assess this proposition. Founded in 1997, CDL acts as a technology service provider to the 10 University libraries of the University of California. By using CDL technologies and services, UC libraries cost-effectively build high-quality and highly distinctive online information services. Although CDL had no prior experience with digital preservation, it is a trusted provider of technology services. Further, the addition of a digital preservation repository seemed a natural extension of its service portfolio.

To test this hypothesis, the CDL built a digital preservation service following the Open Archival Information System (OAIS) Reference Model. OAIS is an international standard reference model that describes the functions and organization of preservation repositories. The DPR is not a centralized service cast in the traditional archive's light. Rather, it is a suite of tools that libraries can use flexibly and adaptively to address their own digital preservation problems however they define them.

The DPR, in this regard, is built to be blind to how and for what data it is used by client libraries. The flexibility is intended to meet diverse library needs. Some will be exclusively concerned to preserve the digital files they create, for example, by digitally reformatting selected holdings in special and archival collections. Others will want selectively to capture web sites that contribute to historically significant collections. Still others will want to offer preservation services to their users, for example, to faculty and research staff who produce digital information as a result of their teaching and scholarship. The DPR's flexibility, then, merely reflects the fact that digital preservation

requires a diversity of approaches to meet the very different information management demands that heritage institutions will have. Furthermore, it appropriately reflects the fact that the field of digital preservation is new and constantly changing. Guidelines and systems — in fact, an institution's entire approach to digital preservation — must be sufficiently flexible to accommodate organizational and technological change.

When the CDL set out to build the DPR, it assumed, perhaps naively, that this was just another complex development effort. Little did we realize in the first flush of enthusiasm that the DPR's development would require the comprehensive re-engineering of the CDL's technology infrastructure, fundamental organizational restructuring, and a major University-wide strategic planning initiative aimed at realizing an appropriate systemwide University policy environment.

For these reasons, we feel that the story of the DPR's development is well worth telling; in part because it introduces a service that we expect to scale to serve other client communities beyond UC, but also because it includes vital clues that may help others who follow in our path to build digital preservation services for libraries.

## **2 The DPR in Context**

The UC system is ideally situated to design, test, and evaluate the viability of a digital preservation service that is offered to libraries as a utility. The 10 campus libraries are already used to relying upon utility technology services supplied to them by the CDL. There is also a history of partnership and collaboration among the campus libraries that enable them formally to design and assess new utilities as they are developed.

In addition, the libraries' preservation needs are themselves diverse. Diversity reflects size, institutional location, and the varied demands that different research, teaching, and curricular needs place on the libraries' service portfolios. In essence, every library in the UC system has some need of a digital preservation service, but the nature of that need varies enormously. The University libraries at Berkeley, San Francisco, and San Diego, for example, have very large collections of digitally reformatted materials that they consider at risk. Their use of the DPR focuses on the capture and persistent management of these natively developed materials. At other libraries, the focus remains on digital information developed by faculty in the course of their research or teaching. Still elsewhere, libraries are turning aggressively to corners of the World Wide Web, seeking programmatically to capture highly volatile publications to complement and extend existing collection strengths (e.g., in government information and area or cultural studies). And libraries across the system are profoundly concerned to ensure persistent access to the online journals and databases that represent so large an annual investment. Given their diversity, the UC libraries represent in microcosm the academic library community more generally. In this regard, it is an ideal testbed environment to develop and assess preservation strategies that aim to empower and enable libraries generally to take up their preservation role in a cost-effective fashion.

## 2.1 Introducing the DPR

The UC libraries' preservation program was established in 2002 when the University Librarians agreed to co-invest to support two permanent staff positions, and the CDL committed its own funding to build out essential technology and service components. Though initially envisaged as a central service that would be owned and operated by the CDL on behalf of the campuses, thinking about the DPR evolved quickly and it was recast in a utility mold. This fundamental change of approach reflects a highly iterative design and development process that the CDL used to ensure that the DPR responded optimally to the needs perceived for it by the UC libraries and that the DPR maximally leveraged existing technology capacity at the CDL and on the campuses.

## 2.2 Content Considerations

At present, use of the DPR emphasizes "built content"; that is, the digital information created and managed by client libraries (in some cases acting as surrogates for third-party data producers). The materials come in a variety of formats, such as digitally encoded texts, digital images, and data. Some are digitally reformatted and based upon selected analog materials (e.g., manuscripts and images). Other materials have no analog counterparts and are "born digital." Most represent considerable financial, intellectual, scholarly, and/or curatorial investment, and some are significantly at risk of loss due to normal deterioration of magnetic media or their existence without print analog.

The rationale for starting with built content is that the UC library community exercises maximum control over its production, dissemination, and use. However, the DPR is designed so it may be used for a far broader range of information, including digital content to which access is acquired from third parties under license (e.g., electronic journals and reference databases) and web-based information that may be captured, for example, by web-crawling.

The rationale underpinning the program's interest in licensed content is straightforward. UC spends upwards of \$27 million a year licensing or acquiring digital scholarly information. In a growing number of instances, the information is exclusively available online, which puts the annual investment in it very much at risk. UC licenses have, as a matter of policy, a persistent access clause through which content vendors commit to making historically subscribed materials available to UC even if a subscription is allowed, for whatever reason, to lapse. The clauses, however, have no implementation paths associated with them. Should a vendor exercise a persistent access clause by submitting a file of historically subscribed content to the CDL, without a preservation program in place, the CDL would be incapable of managing the file persistently.

At the same time, the University libraries are loathe to take responsibility for the persistent management of the 9,000 electronic journal titles to which they collectively subscribe. Some titles are or will be appropriately catered to by emerging third-party archives that may be supported on a subscription basis (the Portico venture, for

example). Others are managed by vendors (e.g., Reed Elsevier, American Physical Society) that are positioned to make reasonable claims to persistence through their use of trusted archives. Accordingly, while the DPR's development gives UC libraries the capacity to manage licensed content, it does not give them the criteria to determine which licensed content to manage.<sup>1</sup>

The UC libraries are equally interested in the capture and persistent management of information that is freely and readily accessible via the open Web. This information is very substantially at risk of loss (the life of the average web page is less than 25 days) and an increasingly essential part of the scholarly and cultural record. For example, the U.S. federal government expects that by 2008, fully 95 percent of its published information will be available exclusively online. The digital preservation program provides the UC libraries with a viable means of capturing and preserving this content, thereby continuing to develop collections in areas where material is migrating largely to an online environment.

As with its work on built content and licensed content, the digital preservation program will develop and maintain a utility infrastructure and a suite of tools that campus libraries can use to capture and persistently manage collections of web-based materials that most closely suit their local needs. The tool suite will include technologies, essential documentation, and administrative and rights management procedures.<sup>2</sup>

---

<sup>1</sup> Such criteria are only beginning to evolve and are presently available as a checklist of questions for consideration by the depositor library:

- Is the journal at risk of loss because of publisher practice or the lack of stability of the publisher?
- Is another trustworthy institution, such as another university library, preserving the journal electronically?
- Is the publisher making positive contributions to changing the nature of scholarly communication?
- How often is the journal used by UC faculty?
- How prominent is the journal in its field? How does it compare to other journals in its field?
- How many journals on this topic exist?
- Is the publisher willing to cooperate?
- How much functionality and look-and-feel should be preserved?

<sup>2</sup> The program's work on web archiving is supported in equal measures by the CDL and by the Library of Congress's National Digital Information Infrastructure and Preservation Program (NDIIPP). It will take place in stages according to a work plan that has been developed by those two bodies.

In 2004, the CDL was awarded a \$2.4 million NDIIPP grant to develop web archiving tools that will be used by libraries to capture, curate, and preserve collections of web-based government and political information (see NDIIPP press release, [http://www.cdlib.org/news/press\\_releases/award\\_announcement\\_final\\_20040930.doc](http://www.cdlib.org/news/press_releases/award_announcement_final_20040930.doc)).

## 2.3 DPR Nuts and Bolts

The DPR is a set of current operational services that supports the long-term retention of digital objects.<sup>3</sup> At present, the services are uniquely available to the eleven University of California Libraries (including the CDL), their affiliates (e.g., campus museums), and their users.

The DPR was built with an eye to extensibility. Although the first phase of implementation is within UC, the infrastructure is currently being extended beyond the UC community, with greater adoption expected in the period ahead.

The DPR provides for the controlled, orderly deposit and persistent and reliable management of digital objects. It also ensures that objects once deposited in the DPR are disseminated only to authorized users. As a utility service, the DPR is designed to interact with staff at libraries and other information organizations rather than with individual end users (e.g., faculty, students, library patrons). It is, in effect, a suite of tools and services that libraries and other information organizations can use to construct locally appropriate digital preservation services.<sup>4</sup>

The management and administration of the repository services and associated storage are based at the CDL. Authorized users are based at client libraries (presently the UC libraries). They have the ability to deposit digital objects and to request their dissemination. The CDL commits to keep the objects in perpetuity, but multiple versions of an object can be deposited and individual versions of an object either kept or replaced as the user decides. Authorized users are entirely free to select what digital objects to deposit, how to record information about those objects, and the terms and conditions of access that apply to objects' dissemination and use.

The DPR was developed conforming to the Open Archival Information System (OAIS) Reference Model.<sup>5</sup> As a reference model, OAIS describes the responsibilities and components of a digital repository as well as its organization and governance. It was developed by NASA's Consultative Committee for Space Data Systems and has become an internationally recognizable reference model for archival systems. OAIS outlines many core functions — how digital objects can be prepared, submitted, stored, maintained, and retrieved as needed — without specifying technologies or archiving techniques. The existence and adoption of OAIS allows for increased awareness and understanding of digital preservation across all types of institutions; creates common terminology and concepts that help to describe and compare data models and archival architectures; expands consensus on digital preservation issues; and offers a framework to guide the identification and development of standards.

---

<sup>3</sup> The DPR System Design document is at [http://www.cdlib.org/inside/projects/preservation/dpr/RepoDesign\\_Final\\_Dev.pdf](http://www.cdlib.org/inside/projects/preservation/dpr/RepoDesign_Final_Dev.pdf).

<sup>4</sup> Note: In this report, the terms “user,” “participant,” and “depositor” are used interchangeably to represent DPR constituents. In the first phase of the repository, the user community is the UC system, but this is quickly being extended to a wider base of users.

<sup>5</sup> OAIS Reference Model, <http://ssdoo.gsfc.nasa.gov/nost/wwwclassic/documents/pdf/CCSDS-650.0-B-1.pdf>.

Access to the DPR is restricted to authorized users. Access is simple (based on a unique user ID) and controlled: the DPR is not designed to be a fully featured end-user access system. Thus, for example, a library that deposits in the DPR data that underpin a complex digital library exhibit that enables deep search and retrieval, as well as a variety of features that make use a highly interactive event, will explicitly not be able to use the DPR as an access service that replicates the functionality of the original. To avoid the impression that the DPR is about fully featured access, we say that objects will be "disseminated" to explicitly identified users. During the registration process, which takes place before objects are submitted, the client library decides who has access and at what level.

To submit an object to the DPR, an authorized user "wraps" the objects with metadata and an inventory of the files. The wrapper format, which is based on XML, is called METS (Metadata Encoding and Transmission Standard). The METS format is maintained by the Library of Congress. METS is increasingly gaining popularity within the digital library community, and is already planned for use in both proposed and operational digital preservation repositories.<sup>6</sup>

To meet a diversity of needs, authorized users can interact with the DPR in three different ways. The HTML browser interface is the easiest way.<sup>7</sup> Using it, the DPR appears to behave like any web-based service such as Amazon.com or eBay. Transactions are governed by individuals via the Web. Using the Java API layer or the SOAP (Simple Object Access Protocol) layer, more technically savvy client libraries can integrate the repository directly into their own systems environment and tailor it closely to meet local needs.

Just as it is important to know what the DPR is, it is necessary to know what it is not. Crucially, the DPR is not a disaster recovery system. In fact, the DPR differs from backup systems in several important ways. First, organizations typically back up onto recyclable tapes, and files are rarely retained for more than one year. In addition, backups are typically created by more or less indiscriminately recording updated computer files, which may be in many formats; these formats may or may not last over time. In the DPR, objects and their versions will never be deleted unless explicitly removed by the submitting organization. Also, the DPR is designed to selectively record coherent, platform-independent information about the objects that are deposited by the producer (hence the use of the METS wrapper).

Stand-alone, platform-independent objects are an important attribute of the DPR, and crucial for understanding why objects deposited in a preservation repository can still be used decades later, whereas content retrieved from a backup tape probably will not be reusable after any significant duration of time. Even if one were able to recover a 10-year-old file from backup tapes (and, with many systems used today, this is not a sure thing), it might be unusable without specific files, database records, and computer platforms upon which it depends. In the DPR, file and record dependencies are minimized by packaging related files in one object. Platform dependencies are reduced by requiring that deposited objects conform to the standards-based, non-proprietary METS standard, which the CDL has adopted as its Digital Object Standard.<sup>8</sup>

---

<sup>6</sup> For more METS information, see <http://www.loc.gov/mets> (Library of Congress METS site).

<sup>7</sup> [http://www.cdlib.org/inside/projects/preservation/dpr/ui\\_users\\_guide.pdf](http://www.cdlib.org/inside/projects/preservation/dpr/ui_users_guide.pdf)

<sup>8</sup> CDL Digital Object Standard web page, <http://www.cdlib.org/inside/diglib/guidelines/>.

Finally, unlike ordinary computer files, objects saved in the DPR all have globally unique persistent identifiers, which provide a long-term, stable association between a string of characters and an object. In the case of an object hosted in the DPR, one can query the CDL directly, at any time in the future, using the persistent identifier. With the persistent identifier, users can also access an object or obtain a basic description of the object. Once assigned, a persistent identifier is flagged so that it will never be re-assigned.

### **2.3.1 Service Components**

In developing service components for the digital preservation repository, the CDL used the OAIS to identify the following functions that had to be supported:

- ❑ Pre-ingest: develop procedures for negotiating the submission and preparation of materials for deposit
- ❑ Ingest: how objects are taken in and validated
- ❑ Infrastructure: the core functions of the DPR for managing data, security, and authenticating users (including the Java API layer that allows clients to customize the DPR and integrate it into their own information environments)
- ❑ Access: how users retrieve objects (including the relationship between clients' access systems and the DPR access system)
- ❑ Administration: how users generate reports about their content; logins and passwords; identification and authorization of users
- ❑ System administration: how the CDL builds a staging and production environment; how it installs and configures software, including the storage system
- ❑ Database administration: how the CDL installs and configures necessary servers and develops backup and restore procedures
- ❑ Quality assurance: how the CDL develops testing tools and uses them; how it informs DPR users about errors that result from a transaction
- ❑ User interface: how the CDL designs and develops a user interface for those accessing the DPR via the web; how it conducts usability studies to assess the user interface and makes changes recommended by users.

### **2.3.2 Design Principles**

Design of the DPR was undertaken with an eye to its extensibility, so that it could serve a broader range of clients than those represented by the UC libraries and manage a broader range of content than the digitally reformatted "built content." The DPR was built using open source software, and the code will be made public to allow for its adoption and enhancement. The preservation repository was designed to have the ability to plug into client-based services (access services, content management systems, etc.) and into services maintained at the CDL.

To achieve this level of systems integration, the DPR was built according to a common information architecture (what the CDL refers to as its "Common Framework") that comprehends the relationship and the interaction between the full variety of services, abstractly conceived as may exist at any one time in the information environment.

Another design principle entails the standard representation of objects within the DPR. The standards-based approach to digital preservation is well known and enables the more cost-effective management of digital repositories.<sup>9</sup> By standardizing the representation of digital objects, the repository obscures the heterogeneity that characterizes its holdings (e.g., with regard to how objects are created, encoded, compressed, etc.). It can consequently manage holdings using common procedures (e.g., for storage, migration, discovery). An alternative approach would entail ingest of digital objects as they are presented and the development of different management and dissemination procedures for objects of different types. Given the virtually infinite array of data types that a digital preservation repository will encounter when offered as a utility service, the non-standards-based approach doesn't scale.

There is a downside to the standards-based approach, however; it entails a degree of data normalization at the point of ingest. Simply, data developed differently to meet different needs have to be transformed so they comply with the digital object standard. Although the work can be automated to some extent, it still needs to be done either by staff at the repository or by the client depositor. With the DPR service, client depositors are encouraged to ensure that deposited objects conform to the CDL's digital object standard. Although clients bear this responsibility, the CDL does provide a variety of support services so the DPR can cater to client depositors whose technical capacities and understandings vary widely.

A related design principle is the DPR's "self-service" model. Clients interact directly with the DPR. Interactions between CDL staff and users occur mainly at the point of submission negotiations, which take place before objects are deposited. The DPR was designed so that clients can customize it to meet their own needs and incorporate it directly into local work- and data flows. For example, a client library running its own well-developed digitization service might, working with the DPR's published APIs, integrate the DPR with their local content management system. In this way, the client library could ensure that newly digitized material is automatically deposited into the DPR as it is created or as it is added to the local content management system.

With the self-service model, clients share responsibility with the CDL for the efficacy of their digital preservation efforts. The CDL is responsible for reliably and persistently managing the technical infrastructure for the deposited objects and for disseminating those objects in renderable form to authorized users. If content is omitted erroneously from the DPR, or if the content is ill-formed at deposit (for example, if it lacks sufficient descriptive information to enable its future use), only the client can be held responsible. While the self-service model entails a high degree of client accountability, it also enables client independence and flexibility. Clients are entirely free to choose what information to preserve, how to preserve it, and how and by whom deposited information may be distributed.

The self-service model stands in stark contrast to a more traditional archiving model in which data producers entrust a third party with the selection, ingest, persistent management, and dissemination of its content. In this traditional model, the content owner abdicates a certain measure of responsibility for and control over its content. In

---

<sup>9</sup> Beagrie and Greenstein, "A Strategic Policy Framework for Creating and Preserving Digital Collections," <http://ahds.ac.uk/strategic.pdf>

the self-service model, the content owner maintains that responsibility and control. The advantages of the self-service model in a digital age are threefold:

- The self-service model enables and empowers numerous organizations to take on a measure of responsibility for ensuring the longevity of digital information. Given the vast quantity of digital information that is being produced and its highly distributed ownership, only this self-service approach seems remotely capable of ensuring the broadly redundant effort that will be necessary to capture and maintain even a fractional proportion of our digital heritage.
- It ensures redundant effort (more organizations become involved in preserving at least some digital information) and redundant practice (preservation is conducted differently by different organizations). Given the volume of digital information and the relatively immature state of our knowledge about how best to manage it over the long haul, both are valuable.
- It enlists clients as partners in co-development. As clients develop new ways of using the DPR, these options become available to and enable others to emulate and use. Thus, in a real life example, a UC campus library is considering using the DPR to manage electronic theses and dissertations (ETDs). Working with the DPR and its published APIs, the individual campus library is developing a suite of tools with which it will capture ETDs, produce them as standard conformant digital objects, and transfer them into the DPR. The tools as they are developed will serve the client library well. They will also add to the suite of tools available to other clients who may wish to use the DPR for their own ETDs. In this regard, the self-service model enables the CDL to create a community of practitioners around digital preservation who figuratively and literally can share their evolving expertise in a manner that ensures each benefits tangibly from the work of others.

### **3 Guidelines and Procedures**

Well-written policies, procedures, and documentation enable clients successfully to use the DPR to meet their local preservation needs. In a self-service model, these elements are particularly important. They allow for clients' independent and flexible use of the digital preservation utility. The policies, procedures, and documentation developed to date have evolved through much discussion and consultation, and will continue to evolve as the DPR is put to new uses and offered to a broader range of clients. The description that follows pertains to the materials and processes that have been developed in support of our initial work with UC libraries and their "built content."

#### **3.1 Use of the DPR**

Documentation describes how clients intersect with the DPR via its HTML browser interface, via its Java APIs, or as a web service.<sup>10</sup> User guides are available online

---

<sup>10</sup> DPR user interface documentation:  
[http://www.cdlib.org/inside/projects/preservation/dpr/ui\\_users\\_guide.pdf](http://www.cdlib.org/inside/projects/preservation/dpr/ui_users_guide.pdf)

(<http://www.cdlib.org/inside/projects/preservation/dpr/toolkit/>; [http://www.cdlib.org/inside/projects/preservation/dpr/ui\\_users\\_guide.pdf](http://www.cdlib.org/inside/projects/preservation/dpr/ui_users_guide.pdf)) and include screen shots, instructions, problem-solving solutions, and a help section. Sample programs are included in the Java guide. Even greater emphasis is placed on preparing client libraries to launch preservation programs, to register as DPR client libraries, and to develop effective data submission and transfer routines.

When a client library takes an active interest in using the DPR, it begins specific discussions with the DPR Project Manager at the CDL. At that stage, the client is directed to fill out forms that enable CDL staff to do two things: register the client as a *bona fide* user of the DPR and gain a sense of how the client wishes programmatically to use the DPR. The former (registration) is a one-time process in which authentication information on individual users is collected from the client and entered into the system (with appropriate logins and passwords). The latter (gaining a sense of client use) helps the CDL identify issues that the client will need to resolve and practices that the client will want to consider. It also helps the CDL identify new development needs that client uses may require and client-side development efforts that may benefit the CDL and the broader community of clients.

Once a client is registered, it prepares a submission agreement. The document, among other things, gives the CDL the right to archive the client's materials, and, where appropriate, to transform them (e.g., in the case of a data migration) for the purposes of preservation. Clients need only fill out a single submission agreement, which applies to the lifetime of their association with the DPR. A submission inventory must also be completed for each deposit.<sup>11</sup> The submission inventory includes information about the deposited content, including its file format, the date of submission, etc. The submission inventory process also helps the client identify how it will conform to the CDL's digital object standard, how data will be transferred to the DPR, and other details. In effect, the finalized submission inventory serves as a lasting and formal record of how key data preparation and transfer issues are addressed and resolved.

Rights management is another key issue that the client library addresses in submission negotiations.<sup>12</sup> It is also a complex one which finds clients turning to the CDL for advice. The CDL retains a vested interest in providing this advice because it helps the CDL reduce its own risk. Although limited — in submission agreements, client libraries hold the CDL harmless for any breach of copyright that might be entailed in their use of the DPR — the CDL maintains some residual risk in operating the DPR. In order to advise clients consistently, the CDL is in the process of developing rights management guidelines to help client libraries recognize, assert, and express the rights that are bound up in the materials they wish to deposit.

Client libraries are also turning to the CDL for advice about what content they should deposit in the DPR. This was rather unexpected by the CDL, which had developed the DPR as a self-service utility in response to the diverse preservation needs that client libraries had expressed in the earliest consultation phases.<sup>13</sup> Initially, then, the CDL set

---

<sup>11</sup> The client library completes the pre-ingest worksheet, which begins the process of submission negotiations between client library and CDL. The result of those negotiations is a finalized submission inventory, which the CDL completes, representing those negotiations.

<sup>12</sup> The pre-ingest worksheet contains a detailed section on rights.

<sup>13</sup> 2001 Systemwide Operations and Planning Advisory Group (SOPAG) final report of the Digital

very minimal content requirements that reflect the DPR's operation as a centrally funded service of the University of California.<sup>14</sup> As the DPR evolves and extends its client and funding bases, these requirements will undoubtedly change. At present, they are as follows: material for deposit in the DPR should be of persistent value to the broader UC community and should support and/or result from research, teaching, and/or learning at UC. In addition, the client library must have the right to authorize deposit of the digital objects in the DPR. In this regard, they must be able to demonstrate either that the content is in the public domain, that the copyright to the content is held by the client library, or that the client library has obtained permission from the copyright holder to deposit the content in the DPR. Obviously, the content also needs to conform to the CDL Digital Object Standard.

In planning their use of the DPR, however, client libraries repeatedly turned to CDL (and to one another) for guidance about what to deposit and how to prioritize their potential preservation needs, which they identified through surveys and discussions. In response, the CDL facilitated a discussion amongst client libraries and identified a list of issues that are commonly being addressed by clients in determining their preservation needs and priorities. As the DPR is extended, these guidelines will be adapted to meet other needs as they are identified. At present the guidelines are constituted as follows.

- ❑ Content is at significant risk of loss, through predictable or unpredictable actions (e.g., an unreliable publisher or reduction in funding)
- ❑ Disappearance of content would constitute an unacceptable level of loss to the UC community
- ❑ Content is of common interest to the UC libraries, or of demonstrated interest to one or more of the UC campus libraries
- ❑ UC systemwide services depend on content
- ❑ Material is considered authoritative
- ❑ Material has significant scholarly value (e.g., a high citation rate)
- ❑ Evidence of high demand, current and/or potential
- ❑ UC played a fundamental role in the creation of the materials
- ❑ High level of UC investment (e.g., licensing fees or infrastructure/project investment)
- ❑ Preservation is promised as part of an external funding agreement
- ❑ Added functionality in the digital format is unavailable in analog form
- ❑ Content is not made available by another trustworthy institution
- ❑ Content is not being preserved by another trustworthy institution
- ❑ Deposit of content facilitates the development of preservation technologies
- ❑ Reasonable cost to preserve content (e.g., negotiations with producer or securing rights)
- ❑ Technical and/or funding constraints to preservation are not prohibitive

## 3.2 Governance

Because the DPR is built as a utility service, all parties involved have a responsibility for its effective and successful use. Thus, the commitment to preserve an object is a joint

---

Preservation And Archive Committee (DPAC), "Preserving Digital Materials," <http://libraries.universityofcalifornia.edu/sopag/dpac/DPACFinalReport.pdf> .

<sup>14</sup> See <http://www.cdlib.org/inside/projects/preservation/dpr/>.

responsibility of the CDL and the client library that deposits the object in the DPR. The CDL is responsible for maintaining the infrastructure and for the policies, practices, and guidelines that govern its use; that is, for the controlled, orderly deposit and appropriate dissemination of objects, and for the reliable, long-term storage and maintenance of those objects to agreed-upon specifications. When credible certification processes become available, the CDL will also subject the DPR to external audit.

Clients take responsibility for understanding the service and using it effectively. They make decisions about what content should be deposited in the DPR, and about how that content is formatted, documented, and ultimately used. They are responsible for following guidelines supplied by the CDL and for ensuring that deposited content conforms to specified standards. Additionally, participants have a responsibility for maintaining their content by adding, replacing, and deleting versions as necessary. They are also ultimately responsible for the content's accessibility from future hardware and software generations.

The DPR's governance reflects this shared responsibility as well as the service's current reliance on central funding from the University of California. The CDL maintains the DPR and is the final arbiter over investment decisions. It makes these decisions in consultation with the client libraries, who share responsibility for the DPR's effective use. The consultation process is shepherded through a highly evolved committee structure that ensures the closest possible collaboration and functional coordination across the UC system's 11 university libraries.<sup>15</sup>

## **4 Communications**

The development and launch of the UC libraries' preservation program has required a comprehensive communications effort, which has focused on three closely related objectives. Communications within UC have created a policy environment appropriate to the program's sustainable development as well as a client community that is appropriately informed to utilize the program effectively. Communications beyond UC have ensured that the program benefits from and in turn informs leading-edge development efforts and the consensus around standards and best practices that is beginning to result from them. Finally, the program has benefited from continuous assessment, which plays a vital role in informing the University and broader communities about progress and plans while ensuring the program's responsiveness to perceived client needs. These three aspects of the program's communication efforts are reviewed in slightly greater detail below.

### **4.1 Communications Within UC**

Ongoing support of the digital preservation program relied on two substantive policy-level changes at the University of California: in the scope of the library program and in the way in which libraries are funded. With regard to program scope, the University has always recognized the support that libraries offer academic programs through their collections, user services, and provision of critical civic spaces. Like most universities

---

<sup>15</sup> More about the UC Libraries systemwide groups and activities:  
[http://libraries.universityofcalifornia.edu/about/groups\\_activities.html](http://libraries.universityofcalifornia.edu/about/groups_activities.html).

and colleges, however, the University of California has fallen short of claiming a persistent stewardship role over libraries' holdings — print or digital. Further, the UC libraries have historically been funded independently as campus-based entities, each of which is responsible for developing their own collections and services. Over the past quarter century, library collaboration has been encouraged and has proliferated in a variety of areas.

The preservation program departed from past collaborative practices, though, in at least two ways. First, it required at least a tacit admission on the part of the University that as an institution of higher education the University bears some responsibility for ensuring the longevity of crucial digital information that results from or supports research and teaching. It also required mechanisms appropriate to funding and governing technology utilities that are deployed widely across the system but funded centrally, rather than on the basis of a top-slice tax or a recharge made against distributed campus funding. Although libraries had for some time been aware of these issues, and were even well placed to address them effectively, they did not, in 2002, reside in a policy environment that enabled realization of the preservation program.

The communications and outreach component of the digital preservation program aimed first and foremost to create that environment. It entailed support for a two-year strategic planning process convened by the Provost's Systemwide Library and Scholarly Information Advisory Committee and led by the University Librarians.<sup>16</sup> The process aimed at refreshing the libraries' strategic plan and raising both awareness of and support for essential utility technologies, including a digital preservation program. The plan, released and endorsed by the University's faculty and its administration in 2004 after extensive consultation, addresses these issues specifically. It positions the libraries to sustain the preservation program and, perhaps more importantly, commits the University to its role as steward over selected digital information.<sup>17</sup>

This latter development is significant and perhaps most effectively realized in a recent UC press release in which the University's Provosts and Senior Vice President claim the steward's role as being a central one. Commenting on the launch of an institutional repository through which UC faculty are able to distribute their publications openly via the Internet, Provost M.R.C. Greenwood claims the steward's role as a central one. "The program...is only partly about access", Greenwood said. "It also demonstrates how seriously UC takes its commitment to acting as steward over the vast well of scholarly and cultural information that is produced by faculty, staff, and students, and acquired or created by its libraries and museums. These materials form a significant part of the scholarly and cultural record.... They contribute directly to the state's economic progress, educational advancement and cultural well-being."<sup>18</sup>

This level of awareness and support creates a rock-solid foundation for the digital preservation program, which did not exist in 2002. Perhaps as critically, the preservation program, and the stewardship role that it heralds for the University, appears as an

---

<sup>16</sup> More about SLASIAAC: <http://www.slp.ucop.edu/consultation/slasiac/>.

<sup>17</sup> Systemwide Strategic Directions for Libraries and Scholarly Information report: [http://libraries.universityofcalifornia.edu/planning/library\\_strategy.pdf](http://libraries.universityofcalifornia.edu/planning/library_strategy.pdf).

<sup>18</sup> "UC Launches Postprints Service to Provide Greater Access to UC Scholarship," February 23, 2005, [http://www.cdlib.org/news/press\\_releases/postprints\\_final\\_20050223.doc](http://www.cdlib.org/news/press_releases/postprints_final_20050223.doc).

important thrust in the budget statement that the Board of Regents for the University of California submits to the Governor's office in support of their request for state funding.<sup>19</sup>

The communication effort required to achieve this substantive shift in policy and funding patterns was intensive. It involved the University libraries having to listen carefully to concerns expressed across various quarters of the University and translating these concerns into a preservation program that promised to respond to them. Particularly challenging was demonstrating to the University's faculty and senior leadership how a preservation program was essential to the University's fulfillment of its research, teaching, and service missions. To make the case, the UC libraries appeared regularly before campus- and systemwide senate committees (on educational policy, research, planning, and budget, as well as on libraries), and worked closely with the University administration to build the evolving case for preservation. In the process, the libraries realized that their compelling interest in preservation (protecting very substantial subscription investments in digital information) only animated the faculty and the leadership to a limited extent.

A more compelling case was grounded in the growing sense of risk that faculty feel for their own digital output, including their publications but perhaps more pressingly their research data, web sites, and so on. The faculty were also motivated by the collection development opportunities that would be missed if the University did not sustain a preservation program. That is, they recognized that important areas of research and teaching would not be adequately supported unless and until the University libraries programmatically capture web-based information that increasingly is the exclusive source of current content (e.g., government information, cultural and area studies, etc.). The University administration finally was animated by the threat posed by possible digital decay to an information organization that increasingly produces and relies upon online information.

A further aim of the preservation program's communications effort was to introduce the program to client libraries and their staff in a manner that encourages and supports effective use of the DPR. In July 2004 a meeting of pilot DPR users from the UC campuses and DPR staff from the CDL was convened. Materials prepared for this meeting were exemplary. They included an overview of the DPR as well as in-depth analysis of technical and programmatic topics. The DPR FAQ (<http://www.cdlib.org/inside/projects/preservation/dpr/faq.pdf>) served as a cornerstone of this body of information. Additionally, a preservation flier was also created for wide distribution.<sup>20</sup> Brief updates on the preservation program's progress are also carried regularly in the biweekly CDL newsletter, *CDLINFO*, which is distributed to library staff throughout the UC system.

The corpus of information on the preservation program is available digitally, of course, and some considerable portion also exists in print form. As the program evolves and as it extends its client base to include libraries and other information organizations outside the UC system, the materials will undoubtedly require modification. At present, they address the parochial needs of the University of California and its libraries. At the same time, these materials embody the compelling arguments and incentives that have

---

<sup>19</sup> Further information on the UCOP budget: <http://budget.ucop.edu/budgetnews.html>.

<sup>20</sup> Digital preservation program flier:  
[http://www.cdlib.org/inside/instruct/preservation\\_spring2004.pdf](http://www.cdlib.org/inside/instruct/preservation_spring2004.pdf).

effectively motivated substantial investment in and support of the program while also proving effective in guiding operational use.

## 4.2 Communications Beyond UC

Internally, our efforts have focused on creating a policy environment capable of sustaining the preservation program and a clientele able to use it effectively. In this regard, we have focused in an insular way on the UC community. Our work on the repository's design (both functional and technical), meantime, has involved extensive consultation with the broader library, computer science, publishing, and information service communities. Digital preservation is a rapidly evolving practice, the efficacy of which will ultimately rely upon deployment of community-agreed standards. In this regard, UC's efforts exist on a node in an international network of activity. Communication with other nodes on that network remains vital to our progress and to the network's general development.

Two collaborative engagements have been particularly formative.

- A partnership with New York University and the University of North Texas, which is funded by the Library of Congress's National Digital Information Infrastructure Preservation Program, to build and practically evaluate components that will enable client libraries to use the DPR selectively to capture, organize, and manage web-based digital information. The project will be undertaken in close collaboration with three technical partners: the San Diego Supercomputer Center, Stanford University Computer Science Department, and Sun Microsystems, Inc. Additionally, the project includes a number of client libraries who will utilize and help to evaluate the DPR's evolving web archiving capacity.<sup>21</sup>
- Close involvement with the further development and promotion of the Metadata Encoding & Transmission Standard (METS), which serves as the CDL's digital object standard. The CDL's reliance on METS has helped to refine and improve the standard. It has also encouraged its widespread adoption amongst libraries. METS is now one of the two or three standards (MPEG21 and ARC are others) that are most commonly used by digital library preservation repositories.

Additionally, CDL has presented on its preservation work at a variety of venues. Meetings of the Digital Library Federation (DLF) provide a biannual forum for exchanging information; CDL staff have presented information about the progress of the DPR at numerous DLF meetings. Other meetings and conferences, including the annual European Conference on Digital Libraries, an ERPANET symposium on economic models for digital preservation, meetings of the American Library Association and the Coalition for Networked Information, gatherings of librarians from across the UC system, and a meeting of awardees for National Digital Information Infrastructure Preservation Program grants from the Library of Congress, serve as examples of arenas where the UC Libraries Digital Preservation Program has been discussed. These conferences and

---

<sup>21</sup> More information on the CDL's "Web at Risk" NDIIPP project:  
<http://www.cdlib.org/inside/projects/preservation/webatrisk>.

informal conversations have provided valuable insight into other institutions' approaches to digital preservation and opportunities for collaboration.

### **4.3 Assessment as Communication**

Continuous assessment of the DPR has been another important objective of the preservation program's communications efforts. Communications activities indicated above were designed to elicit feedback from numerous stakeholding constituencies as well as to inform them about the preservation program's progress and plans. In addition, the preservation program team undertook formal assessment within client libraries to find out how they perceive and scope their preservation problems and about the tool suites they require to address them. Results helped inform the design of the DPR and the development of supporting policies, procedures, and implementation guidelines. Formal usability testing was conducted to evaluate and refine the DPR's user interface and supporting documentation.

A cornerstone event of the assessment process was the July 2004 DPR users meeting, which involved preservation representatives from client libraries. At that meeting the DPR was rolled out in some detail for extensive review and comment. The result was twofold. First, the meeting surfaced a number of detailed but important modifications to the DPR. More crucially, it launched a DPR user group, which looks set to meet regularly, in addition to regular listserv communication, as a mean of ensuring continual formative evaluation.

## **5 Project Deliverables and Outcomes-Based Assessment**

In the process of fulfilling this grant, specific deliverables have been created. These deliverables are described below with their success measured through the outcomes-based assessment model.

### **5.1 Establish Preservation Repository with Associated Operational and Governance Structures Using Open Archival Information System (OAIS) Reference Model**

- **Digital object submissions from all UC campuses and Online Archive of California (OAC) partners**

The creation of an OAIS conformant UC Libraries Digital Preservation Repository (DPR) achieves this objective. The repository evolved from a concept to a design plan to an operational service with accompanying governance, policies, and user support services and documentation. The technical infrastructure is described in Section 2.3, the governance structure, usage procedures, and support services and documentation in Section 3.

Clients can utilize the DPR to store and preserve digital objects over the longer term. Use has been piloted during a beta operational phase by the CDL (which has deposited many objects sourced from a variety of online information services, including the OAC). At the completion of the beta or pilot phase, the service will be made available to the 10

campus libraries of the University of California. Campus plans for utilizing the DPR (deposit plans) were collected and analyzed through a survey conducted by the CDL in order to do appropriate capacity planning for the repository.

## **5.2 Migrate CDL Digital Objects to METS**

The use of METS (the Metadata Encoding and Transmission Standard) is an important technical criterion for the DPR that is described more fully in Section 2.3 of the report. At present, we have migrated over 50,000 digital objects to METS as a means of developing migration tools and protocols, and of defining, documenting, and evaluating the digital object standard and its use in the DPR.

The work has been influential in other ways as well, notably in helping to refine the METS standard, which is being adopted very widely by the digital library community. It has also helped us to anticipate the need and begin designing tools that will substantially automate the process for generating standard-conformant objects. This will significantly enhance the DPR's utility because it will lower the largest single barrier to use — transforming digital objects from their native format to something that the DPR is capable of ingesting and managing.

## **5.3 Successful Delivery of Objects from the Repository**

The DPR has been developed with an administrative interface that enables distribution of objects. The interface has been used successfully by the CDL and by client libraries to retrieve objects that have been deposited. The successful use has also been evaluated and verified by the assessment team. The administration functions of the DPR are described more fully in Section 2.3. It is also covered in the DPR design document ([http://www.cdlib.org/inside/projects/preservation/dpr/RepoDesign\\_Final\\_Dev.pdf](http://www.cdlib.org/inside/projects/preservation/dpr/RepoDesign_Final_Dev.pdf)).

## 5.4 Research Questions Regarding Repository Operations and Policy

- **Certification**
- **Division of submitter and repository responsibilities**
- **Alternatives for storage and format migration**
- **Establishing access services that avoid redundant storage of deposited content**

These have been evaluated in extensive consultation with and practical evaluation by client libraries, and in continuous discourse with the broader digital preservation community.

**Certification.** Discussed in Section 6.9, certification was researched and deemed to be the kind of endeavor crucial to be undertaken by the community rather than by an individual library. The rationale is straightforward. Certification requires a range of auditing procedures that can be agreed upon by data depositors (those who utilize a digital preservation repository to manage their objects persistently), data users (those who rely on a digital preservation repository for the persistent management of digital content on which they somehow depend as end users), and digital archives (entities that are responsible for managing digital content persistently and mediating occasionally competing demands of depositors and users). It also requires credible third parties capable of implementing community-agreed auditing procedures. Accordingly, it is inappropriate for a single digital repository to “develop” and “implement” certification procedures. Consequently, the CDL undertook to work with the broader digital preservation community rather than to develop certification procedures as a sidestream activity and attempt to impose these more generally (or even to use them in isolation). It has, as a consequence, worked with the RLG/OCLC task force in developing and evaluating attributes of a trusted digital repository. It is also participating as a partner in the Library of Congress’s National Digital Information Infrastructure Preservation Program, which will be tackling this problem head on, in part through CDL efforts.

**Responsibilities.** Responsibilities of the CDL and of the client libraries that utilize the DPR are determined through negotiations that precede submission. Both parties agree to specifications for the deposit in a process that results in two formal documents: the submission agreement and submission inventory. The process has been developed in close consultation with client libraries and reflects the self-service aspect of the preservation utility that the CDL has built. It is explained at greater length in Section 3 of this report.

As explained elsewhere in this report, our understanding of the division of responsibilities between CDL as archive and client libraries (as depositors) has evolved somewhat in the course of developing the DPR. We began with a centralizing service model that reflected traditional archiving practice in both print and digital realms. According to this model, the archive effectively takes on responsibilities for identifying, building, and managing collections persistently. Depositors have a role to play, obviously, in ensuring the flow of information into the archive but are otherwise relatively passive agents.

In the DPR's self-service model, the archive has responsibility for maintaining an infrastructure and a process for ensuring the orderly deposit, long-term storage, and appropriate distribution of client-deposited digital objects. Selection decisions as well as decisions pertaining to the creation, documentation, and distribution of digital objects are left entirely to the depositing client library. This rather unique service model emerged through an iterative process through the CDL's consultation with its client libraries. Results of a user survey were particularly significant.

Inevitably, the CDL has also learned that a self-service model requires richer client support than is perhaps required in a more centralized service model. It is essential, for example, that clients are clearly aware of the responsibilities that they share with the CDL for the longevity of their deposits. Both process and documentation are vital; these are available in the form of supporting documentation that includes an FAQ that describes the DPR and how and why it may be used, an FAQ that describes the process that client libraries go through when depositing objects to the DPR, a form that enables client libraries to describe the objects — format, rights, and other details — they want to deposit in the DPR, and submission agreements and inventories, the legal documents between depositors and the CDL. The planning process for objects to be deposited into the DPR has also been documented in a so-called "pre-ingest workflow," which has proved helpful in clarifying roles and responsibilities both for CDL preservation staff and for the client libraries.

**Storage and format migration.** A range of storage issues were examined during the course of this project, including exploring potential partnerships with the San Diego Supercomputer Center and other institutions. The market for storage is still maturing. New product offerings continue to emerge, and product costs (when measured in dollars per bytes stored) continue to decline. As we investigated different solutions, we determined to wait for the market to stabilize more before committing wholly to a single storage option. This decision has not, of course, undermined our ability to build out and operationalize the service. Extant storage is likely to remain adequate for some time to come, and the modular architecture of the DPR enables us to swap out specific components (of which storage is certainly one) without disrupting or having to re-engineer others. CDL has adopted a similar approach to format migration.

**Access Services.** In researching access services for a digital preservation repository, we heard again and again from our client libraries how they wanted to maintain control of their objects and the manner in which they were distributed and made accessible. To satisfy client need — and, in fact, to build client trust in the DPR, we developed a service that carefully separated access and preservation functions. The latter is the responsibility of the CDL, the former of depositing client libraries. This distinction emerged as a crucial design consideration and a key characteristic both of the DPR and the service model that it adopts.

**Redundancy.** Initially we assumed that data stored in the repository would be accessible to end users from the DPR. Accordingly, we anticipated that redundant data storage would act, in effect, as content backup. Distinguishing access from preservation functions, however, changed our thinking about the level of appropriate redundancy. By positioning the DPR as a preservation service and yielding access responsibility to client libraries (acting as depositors), we introduced a further level of data redundancy. At least for data that need to remain accessible to end users (not all archived information needs to be available immediately to end users), an access copy needs to reside at the client library (or its surrogate) in whatever format is appropriate for the designated access

service. Another benefit of this approach is that it satisfies the very serious concern that client libraries and their patrons (faculty) expressed about redundancy, briefly, urging that the preservation program make data more rather than less redundant.

## **5.5 Increased Community Awareness of Digital Preservation Challenges and Digital Preservation Program**

There has been an exponential increase in awareness about digital preservation issues at the University of California and in the broader library, academic, and heritage communities over the two-year term of this grant. The CDL has played key roles raising awareness in both spheres, as explained in Section 4 of this report. Our outreach and awareness raising efforts have taken on a variety of forms ranging from staged events to focus groups, surveys, and web- and print publications.

Within UC a very considerable effort was required to create the policy environment necessary to support a preservation program, which, in effect, required the University to commit to a new stewardship role over its digital information assets. While this role entailed considerable consultation with the University's senior administration and its faculty, a further and also substantial effort was required to build a community of client libraries sufficiently conversant with preservation issues and DPR protocols to enable them to use the preservation service effectively.

Outside UC, the CDL has worked very extensively with other universities, libraries, and key players in the broader digital preservation arena to surface practices and protocols likely to support the development of an extensive and highly distributed network of interoperable preservation repositories. Interoperability is essential at at least two levels: because any repository will need replication by others in order to safeguard its contents, and because our users are likely, at some level, to want the distributed nature of preserved content to be wholly or partially obscured from their purview.

Finally, we have discovered through our work how assessment (of the DPR's design, service model, support services, etc) has played an important part in building a community of users within UC, as well as a more informed network of digital preservation stakeholders beyond the University's boundaries.

## **5.6 Document and Share Challenges, Decisions, Costs, Methods, Results**

Throughout the DPR project, we have been collecting and writing information on the range of issues that impact the preservation program, including our successes, challenges, decisions, choices, procedures, policies, and costs. We've been particularly interested in collecting information on preservation challenges across the University.

DPR surveys have been conducted to this effect. Other surveys and solicitations of input on preservation issues have been conducted by the University Librarians and through the Librarians Association of the University of California (LAUC).

Much of the information gathering and sharing has been done in the context of in-person meetings (many described in the preceding section) and materials prepared in association with those meetings. At UC, issues related to digital preservation have been discussed among DPR users, University Librarians, the University Archive Council, SOPAG, DPR technical leads, and the DPR steering committee. Outside of UC, there have been many further opportunities to share information on our digital preservation program — at DLF, NDIIPP, and other venues discussed in Section 4.2.

The Preservation Program description

([http://www.cdlib.org/inside/projects/preservation/UC\\_Libraries\\_preservation\\_overview.doc](http://www.cdlib.org/inside/projects/preservation/UC_Libraries_preservation_overview.doc)) codifies both challenges and approaches. The system design of the DPR ([http://www.cdlib.org/inside/projects/preservation/dpr/RepoDesign\\_Final\\_Dev.pdf](http://www.cdlib.org/inside/projects/preservation/dpr/RepoDesign_Final_Dev.pdf)) reflects many of our decisions, methods, and results. Materials used in the submission process, FAQs, guidelines, checklists, publicity, and other forms of collateral information on the DPR (see <http://www.cdlib.org/inside/projects/preservation/dpr/>) are also important for documenting and sharing information for use by the DPR team, within UC, and especially with the wider digital library community. Finally, this report to the IMLS represents the best summary to date of the lessons we have learned and the challenges we continue to face in developing the preservation program.

## **5.7 Positive Role of a Preservation Repository in Service Quality**

As discussed throughout the report, especially in Section 6.9, the issue of trust is an essential part of the success of the DPR. Without trust in the DPR, which includes excellence in service quality, users will not participate by depositing content into the DPR.

We have taken many steps to ensure that the DPR's service quality is viewed in a positive light by its user base. Usability testing and solicitation of feedback from users is one element. This has been achieved through formal usability testing, sharing of submission materials with users and usability testing on those documents, regular check-in calls for DPR users and CDL DPR staff, the 2004 DPR users meeting, conversations with the digital preservation reps on all the campuses, and consultation with a variety of UC governance structures.

## **5.8 Create Preservation Program at UC Based on Ongoing Repository Support**

Building an economically sustainable preservation program has been one of our principal aims from the outset. This goal was achieved with the policy and funding re-organization that is described in Section 6.5. The UC Libraries' Preservation Program is a permanent one, with a funding model appropriate to its maintenance. Here, as elsewhere, the utility, self-service model has been enormously beneficial since it distributes preservation costs as well as responsibilities between the CDL (as archive manager) and the campus libraries (as depositing clients). While the CDL maintains

responsibility for maintaining DPR staff, infrastructure, and client support services, client libraries bear the costs of data selection, collection development, and, crucially, access.

The question to which the program now turns is whether the program can extend its client base to libraries and other information organizations outside UC, and how, in so doing, it can continue to support itself. What, in other words, are the marginal additional costs involved for the CDL in supporting a preservation service that is available to the broader library community; how can those marginal additional costs be met in a manner that does not require the University of California unreasonably to subsidize non-UC beneficiaries of such a broadened service. This discussion is taken up at greater length in Section 7.

## **6 Lessons Learned and Issues Raised**

The activities described throughout this report all constitute “lessons learned.” This section highlights issues of particular significance as well as those that are likely to be encountered by other proximate initiatives.

### **6.1 Empowering Libraries: Preservation as a Utility Service**

As mentioned earlier in this report, UC presents an ideal microcosm for experimenting with the development and deployment of a utility preservation infrastructure that enables and empowers libraries to take up their historic preservation roles with digital information. In building out this utility, we have emphasized flexibility, client choice, and a self-service operational model. In this way, we have attempted to enable client libraries to define their distinctive preservation needs and to meet them in a manner that suits their local purposes and integrates optimally with local systems and workflows. Overall, the approach promises to build redundancy of practice into the digital preservation process, which, we feel, works to the long-term advantage with regard to the reliable persistent management of our nation’s digital cultural record.

The approach, we now recognize, has a further advantage insofar as it allows client libraries with very different local capacities to begin to play preservation roles. Our libraries with well-developed technology staff and infrastructure naturally gravitate toward the DPR’s APIs, which allow them to integrate the preservation utility into their local information environment. The HTML browser interface, on the other hand, satisfies the needs of less technologically sophisticated client libraries. All are equally enabled.

Additionally, by phasing the DPR’s rollout, we have effectively enlisted the “can-do” client libraries in a co-development effort that benefits the less technologically enabled. Their active participation in the DPR’s initial use has done two things in particular. First, it has helped in a practical way to review and refine the various aspects of the DPR — to get the kinks out — so that the DPR is more readily and more easily available to the less enabled client libraries. Secondly, the can-do client libraries have helped to develop ingest routines where they did not already exist to manage the data flows in which they were particularly interested. These ingest tools add to the toolkit now available to future users.

By offering a utility service into a distributed and highly varied environment of client libraries, then, the CDL has managed to empower a larger group of libraries to undertake digital preservation in an efficient and highly localized manner. The CDL has also created a community of co-developers whose efforts prove to be mutually beneficial and serve to extend the DPR's capacity.

## **6.2 Viability of Self-Service Archiving Model**

Previously, archiving was conceptualized as follows: "give us your material, and we will put it somewhere safe." As a consequence, archives often bore the full cost of and responsibility for the collection and persistent management of their holdings. The model works well so long as there are a sufficient number of archives putting themselves forward. It survived in a paper-based world largely on the backs of libraries (academic, corporate, and public) and historical societies. Two things happened in the transition to a digital age that broke the back of the old model. First, the volume of archivable information expanded geometrically as the Internet made everyone their own publisher. Second, as demand for possible archiving services escalated, supply of services dwindled to almost nothing. The costs and uncertainty inherent in digital preservation drove traditional providers out of the marketplace.

Offering preservation as a service utility to libraries enables them to take up their historic roles with digital information. But it also requires the elimination of the gatekeeping function that the archive historically exercised over its collection. By offering preservation services to libraries, the DPR had to plan for libraries to use the services differently. Rather than surface centrally the mechanisms necessary to deal with the full range of digital content that libraries might offer to deposit, the DPR preferred a self-service route — one by which client libraries take full responsibility for selecting, capturing, and depositing collection content into the archive. The model is infinitely scaleable, but it also requires that the archive share responsibility with its depositors for the successful persistent management of its holdings

## **6.3 Bringing Consistency to an Otherwise Anarchic Archive: The Standards-Based Approach**

As already indicated, in a more centralized preservation service the archive acts as a gatekeeper and determines what content is stored and how it is managed. In the utility, self-service model, the archive relinquishes a substantial amount of that control to the client library acting as depositor. And yet, the archive must be able reliably to manage its vastly heterogeneous collections and to migrate them through changing technical regimes. To do this efficiently, it requires that a degree of consistency be imposed across its otherwise anarchic collections. To balance depositors' needs for an archive capable of managing diversity against its own needs for consistency, the DPR has adopted a standards-based approach to data ingest.

Client libraries determine what information to save in the archive; the archive is blind to how that information is formatted, documented, compressed, and presented through appropriate access services. The archive insists, however, that deposits are represented in a standard way — in the CDL's case, that they are bound in an XML wrapper that is

described by the evolving library standard known as METS. Because deposits are likely to comprise digital information that is produced originally as METS files, the CDL has developed an ingest process through which it is able to work with client libraries helping them conform to the CDL's digital object standard. In future, it looks to the development of tools that enable client libraries to produce standard conformant objects more quickly and easily.

## **6.4 Content Migration and Geographic Replication**

When the CDL began to develop the DPR, we planned to offer two services in the initial release of the DPR: 1) format migration and 2) geographic replication. The aspiration proved to be too ambitious, particularly as the DPR grew to take on the self-service approach and a vastly more heterogeneous collection than was originally anticipated.

The CDL is currently investigating migration and replication strategies appropriate for a digital preservation repository that manages deeply heterogeneous and only minimally consistent digital information. That investigation is set to continue over the next 36 months. In fact, it is a major research and development component that the CDL hopes to undertake in the context of the work funded by the Library of Congress's NDIIPP program. Having said that, initial forays into this field suggest that while progress in these areas is likely to be made, no single reliable solutions or solution sets are likely to emerge anytime soon. We would encourage, instead, implementation and formal and preferably comparative evaluation of a variety of different approaches.

## **6.5 Technical and Organizational Re-engineering**

The DPR's development required comprehensive organizational restructuring of the CDL as a utility information service provider as well as a fundamental re-engineering of its underlying technology infrastructure. The logic is sufficiently clear. When the preservation program was launched in 2002, the CDL was organized around a small number of technology-based services, each of which benefited from its own technology platform and its own project team. Although project teams differed to some extent in their composition, they typically included a project manager, a technical lead, a user services coordinator, and, where appropriate, a content development specialist. In this highly "siloized" configuration, projects tended to develop their own different technology and service solutions, even when focusing on generically equivalent problems. Although the model enabled the CDL very effectively to operate and manage a small number of technology based services, it did not adequately scale and was stressed to the point of fracture with the introduction of a project as large and as complicated as the DPR's development.

The DPR's start-up costs also figured prominently in the rationale underpinning the restructuring. Although the preservation program benefited significantly from permanent additional funding (roughly the cost associated with two additional FTE staff members), the initial programming effort that emerged from the detailed design phase involved fully 26 FTEs over a 9-month period. Further, although the DPR was designed as a self-service utility, the negotiation of submission procedures with client libraries required a dedicated consulting and ingest effort that simply did not exist in the CDL and that was not provided for in any preservation budget line.

The net result was a substantial reorganization of the CDL that focused simultaneously in three areas:

- ❑ The development of a highly modularized, standards-based, and, crucially, common technology infrastructure capable of supporting the widest possible range of application services.
- ❑ The development of a highly matrixed organization divided not into project teams but into functional divisions whose staff and capacities could be combined as appropriate to meet project and program needs.
- ❑ The development of planning and decision-making processes appropriate to an increasingly and very highly matrixed organization.

The logic underpinning a common technology infrastructure (we call it “the Common Framework”) is simply presented. Rather than having several content management systems (one for each application service), the CDL sought to develop a single content management system capable of handling digital objects independent of the access systems that deliver them to end users with the required range of functionality. Where access services are concerned, the aim is to develop re-usable functionality (search and retrieval, relevance ranking, user annotation, etc.) that can be combined in different ways and with different content in the construction of new application services. Similarly where data capture (collection building) is concerned, we seek tool suites that generically can be used to support categorically similar strategies (web crawling, metadata harvesting, data ingest). The aim is the development and maintenance of technologies that can be used and re-used in different combinations, rather than ones that are bespoke — that is, appropriate — only to a single application service.

The organizational realignments have a similar logic. Rather than equipping individual, content-rich application services (union cataloging, Online Archive of California, the DPR, etc.) with their own data capture staff, we would aim to establish a single department of the CDL called “ingest,” and leave “ingest” to develop and maintain the data capture capacity as may be required across the entire CDL. As a consequence of this approach, CDL was transformed from an organization whose staff was pigeon-holed by the application services they were associated with to one whose staff are assigned to functional units such as “data ingest (data capture),” architecture and infrastructure (ensures a hardware and software environment suitable to our needs), technology programming (a pool of skilled programmers able to work on any aspect of the technology infrastructure), advanced technology (an R&D unit that helps us keep abreast with and adapt appropriate technology innovations), support services (interact with clients generally), assessment (conducts evaluation and assessment as appropriate across the unit), and web design and production.<sup>22</sup>

Interestingly, the most difficult part of the transformation entailed the development of a planning and decision-making apparatus appropriate to a highly matrixed organization. In an old model, decision-making was traditional and hierarchical. So long as project managers communicated their needs to the CDL Director, decisions could be made about the organization’s priorities and resources could be allocated appropriately. In the current configuration nothing is quite so simple.

---

<sup>22</sup> An overview of the CDL organization and staff responsibilities is available at: <http://www.cdlib.org/glance/organization.html>.

The delivery of particular application services (or projects) in the new CDL does not require allocation of funding direct to the project. Rather, it requires aligned effort across a variety of functional units (e.g., ingest, programming, support services, web design and development, assessment). In this organizational model, communication is paramount. So too is a shared understanding of the organization's evolving priorities. The latter is essential to an effective sequencing of development effort. Planning and communication resources and protocols are already extensive and continue to develop; they are available upon request.

## **6.6 Implementing a Phased Approach**

When discussions about the DPR first began at UC, electronic journals were considered a priority. However, with further planning, analysis, and surveys of other efforts, we realized that ejournals were not the place to begin but rather the arena into which ultimately to extend our preservation efforts. In addition, our user needs analysis uncovered a very significant concern amongst client libraries: to have the capacity they needed selectively to capture and persistently manage whole areas of the open web where this content complemented and promised to extend particular collections.

The government information collection requirement is exemplary in this regard. At present, more than 65 percent of all published government information is published exclusively to the web. That proportion is expected to grow very rapidly toward 100 percent. As a consequence, the UC libraries, which include 8 very extensive government information collections, are faced with either ceasing to collect government information or developing capacity to collect that information systematically from the open web. The same is true of library collections that are built in large part to support area and cultural studies, where web-based materials are becoming increasingly vital.

Thus, from early in its development, the UC libraries preservation program was confronted with having to support three potentially very different preservation needs: for the content developed or managed on behalf of third parties, for content licensed from scholarly publishers and other vendors, and for content captured selectively from the web.

To satisfy these three potentially very different needs, the program adopted a phased approach through which it could build and evaluate in successive development cycles a common suite of technology, protocols, and support services. Given this phased approach, it was deemed appropriate to begin with the data content — built content — over which the UC libraries exercised the greatest degree of control, and then iteratively to test the preservation capacity against different kinds of data.

Looking forward, we are now developing a tool suite appropriate for client libraries that wish selectively to capture, organize, and persistently manage materials drawn directly from the open web. The work defines our partnership in the Library of Congress's NDIIPP program. Work with licensed content, meanwhile, is still very much in a design and consultation phase. Interestingly, it is unlikely, in our view, to require very radical changes or even additions to the DPR as it is developed for built and web-based materials. It will, however, require a more sophisticated understanding than we have at

present about what licensed materials to actually preserve. It will also require a decision-making apparatus that is not currently available to make selection decisions.

## 6.7 OAIS Reference Model

The CDL's plan had always been to build the DPR to conform to the OAIS reference model. As the project progressed, we came to appreciate even more the essential nature of this standard. OAIS allows us to collaborate with other institutions within the larger digital library and digital preservation communities, providing a *lingua franca* even where other characteristics diverge. Furthermore, OAIS serves as a kind of internationally recognized gold standard for preservation repositories, showing that, even without the availability of a formal certification process, the DPR meets certain important criteria and thus should be considered trustworthy. Additionally, because OAIS is a common place from which to begin, there are relevant OAIS repository resources on which on we could draw for planning, building, and policymaking, enabling us to learn from the experiences of other institutions.

OAIS was developed for the space science community rather than for the library community, and thus some tweaking of the model to suit libraries has been necessary. This customization has turned into a collective effort within the digital library community, providing a common starting point for collaboration and discussion.

## 6.8 Preservation as Institutional Priority

When we began this project, digital preservation was mostly a concern of a small number of UC library staff. Today it is considered a mission-critical part of the University's stewardship role and a priority for its libraries. It has also emerged recently as a priority concern for the University's strategic direction.<sup>23</sup> This prioritization was codified in making digital preservation one of the five pillars of the UC libraries' strategic plan. It was important to communicate the need for preservation in a compelling way so that it could be understood and adopted as a priority throughout the University (see Section 4.1). A key message was that digital preservation would allow an institution to continue building first-class collections and to save money. Print collections are fairly well understood by the University administration. The CDL demonstrated that we could make more effective use of print collections by tackling the digital realm. As proof of our success, digital preservation now exists as a separate line in the UC Regents' budget.

## 6.9 Trust

A critical success factor for any preservation repository is that it is trusted. Without trust, the repository will attract neither depositors nor users.<sup>24</sup> From the outset, we found ourselves having to address the challenge of building trust for a service that has no track record.

---

<sup>23</sup> See the Systemwide Strategic Directions for Libraries and Scholarly Information report: [http://libraries.universityofcalifornia.edu/planning/library\\_strategy.pdf](http://libraries.universityofcalifornia.edu/planning/library_strategy.pdf).

<sup>24</sup> "Trusted Digital Repositories: Attributes and Responsibilities: An RLG-OCLC Report," May 2002, <http://www.rlg.org/longterm/repositories.pdf>.

So what is trust? As one UC campus user memorably defined it, in the context of the DPR trust entails "no screw-ups." Less crassly, client libraries that deposit in the DPR do not want to worry about their content. At least, they want to know that the CDL is living up to its responsibility to reliably manage that content through changing technical regimes. To develop a sense of trust, the CDL has adopted a number of strategies, communicating these appropriately to its constituents. In particular, it has:

- ❑ built the DPR to conform to an international standard: the OAIS Reference Model — as a consequence, the CDL can convey the fact that it has developed something that is instantly recognizable in an international community as a digital preservation repository;
- ❑ ensured that the DPR has all the key attributes of a library-based digital preservation repository defined by respected library associations — RLG and OCLC, in particular;<sup>25</sup>
- ❑ fostered the evolution of a policy environment that supports and thus builds institutional trust in, or at least credibility for, the repository;
- ❑ leveraged the existing reputation of the CDL, which is known for the maintenance of robust, high-quality, and widely trusted online information services;
- ❑ surrounded the preservation program with a consultative structure that ensures that investment decisions benefit from input received from key stakeholding communities, including libraries, faculty, and the University administration; and
- ❑ publicly announced plans to subject the DPR to a formally recognized certification process, including third-party audit, as soon as a credible process emerges (this is likely to happen as an outgrowth of the Library of Congress's work building a national digital information infrastructure for preservation).<sup>26</sup>

Inescapably, however, these strategies seem at best to be temporary. Perhaps inevitably we expect that trust will evolve (or not) and be related directly to the absence of mishap and potentially to the first well-documented successful data migration.

---

<sup>25</sup> *Ibid.*

<sup>26</sup> More information on the Library of Congress's National Digital Information Infrastructure and Preservation Program, including the NDIIPP awards, is available at [www.digitalpreservation.gov](http://www.digitalpreservation.gov).

## 7 Next Steps

The next stages of our project can be categorized in four areas:

- Extending the DPR's capacity
- Extending the DPR's client base to other libraries
- Resolving key issues of certification, data migration, and data replication
- Assessing the extensibility of our standards-based approach to ingest of objects into the DPR

### 7.1 Extending the DPR's Capacity

At present, the DPR has strictly limited capacity, particularly with respect to the variety of ingest tools and routines that are currently on hand. Specifically, our work to date has equipped us adequately to manage digitally reformatted content that is typically created by libraries, notably encoded texts, digital images, finding aids, and catalog records. Working with the data that has been acquired by the CDL's eScholarship program and with campus libraries (interested in depositing electronic theses and dissertations), we are also anticipating gaining very considerable experience with digital monographs and similar objects. We have only broken ground, however, and recognize that we will need to manage GIS, digital audio and video, data, and, increasingly, mixed media materials. Large and small collections of web pages taken directly from the Internet are also in our sights.

Our strategy for developing tools and competence with an ever broadening variety of information is very much the one with which we launched out — working closely with client libraries who need to develop preservation strategies and tactics for handling these materials. Accordingly, we anticipate that the DPR's capacity will be extended incrementally over the next 12-24 months. As part of that development process, we will choose our partners carefully, ensuring that we do not overextend our limited staff resources by taking on too many new challenges.

One particular challenge that we anticipate encountering soon is the one associated with the selective capture, organization, and persistent management of materials drawn directly from the open web. The effort focuses on our work as a partner in the Library of Congress's NDIIPP program and entails the development of appropriate ingest routines (as indicated above). It also extends the DPR's functionality to include data capture routines, particularly those appropriate to the selective acquisition of web-based information.

## 7.2 Extending the DPR's Client Base

The DPR was developed initially for the UC libraries. From the outset, however, we sought to build tools and processes that would enable us to offer the service to a wider client base. Although we believe we have achieved extensibility from a technical point of view, we are certain we do not yet have the policies, procedures, or business and governance models that will credibly enable us to provide preservation services to other libraries. Accordingly, we look forward to exploring whether, to what extent, and how we might offer the DPR as a utility that other clients could use to address their preservation challenges, and also the business and operational processes that would enable this in a sustainable way.

Three lines of development are already planned and look particularly promising:

- We will be developing a web-based environment through which to offer a number of utility services to UC and other libraries. All support libraries in the development, delivery, and persistent management of a variety of high-quality, highly localized, online information services, including the DPR. Each of the utilities is offered according to the same self-service model that was developed for the DPR. They will be available to the UC libraries and, initially, to a limited group of non-UC libraries as well, notably those funded by the California State Library to build digital collections. It is hoped that this web site will evolve as the place where any number of client libraries can intersect with and use the utilities offered on a self-service basis by the CDL. It is also envisaged that through this work the existing range of DPR services will become available to and used by a growing number of California-based academic and public libraries.
- Over the next 12 months we will be offering the DPR as a utility to three client libraries (New York University, University of North Texas, Arizona State Library) that reside outside California. The work, conducted under the terms of the Library of Congress's NDIIPP program, will involve these clients in using the DPR and its new generation of web-crawling tools to capture, organize, and persistently manage web-based materials that they select.
- Through both of these activities, we will have an opportunity to review and revise our policies, procedures, and governance structures to ensure they are appropriate for a broader client community than that represented by the UC libraries. We will also work with our extended client base to formally model business practices that can sustain the DPR as a utility that cost-effectively enables and empowers libraries to take up their historic preservation role, irrespective of whether they are located within the UC system.

### **7.3 Resolving Key Issues**

Obviously, a number of critical challenges remain unaddressed, and these are likely to occupy the CDL for at least the next couple of years. Our partnership with the Library of Congress is particularly promising, especially with respect to data migration and replication. In our work with NDIIPP, we will partner with the San Diego Supercomputer Center, Stanford University Computer Science Department, and Sun Microsystems, as well as with two partner libraries (New York University and University of North Texas), to explore mechanisms for replicating digital repository content through and across a network of interlinked repositories. A detailed plan is available for this work, which has only recently commenced.

Data migration is another area where we will be working under the auspices of NDIIPP with the partners enumerated above. In effect, we will be developing mechanisms that will enable us formally to test and evaluate the efficacy and cost of various different data migration strategies that might be applied to the data content residing within the DPR.

The efficacy of a repository's data migration strategies is one aspect of its service that requires formal validation by a trusted third party. Others include the integrity and accuracy of ingest routines, the management of magnetic media, precision of file location, and the administrative routines that ensure that data are only distributed to *bona fide* users.

Certification is required, but none of this can easily be developed and deployed without input from other digital repositories and the communities of data depositors and users that surround them. General development and validation of such procedures are likely to surface demand sufficient to sustain the third-party auditors that will be required to implement them. Progress is being made at the conceptual level (notably through the work by RLG and OCLC). We look forward to participating in the practical initiative that will be shepherded by the Library of Congress.

### **7.4 Assessing Extensibility of Standards-Based Approach**

Finally, planning for sustainability is an important endeavor as well. In order to better understand costs associated with digital preservation at the level of both provider (CDL) and user (the UC campuses and other institutions), for both the short- and long-term, we need to examine a range of costs and potential revenue streams for content types, maintenance and user support, replication, migration, and other services we may offer in the future. Business planning will focus on investigation and discussion until more quantitative data is gathered. At that point, our efforts will turn to developing a sustainability plan, working in concert with users as they examine sustainability issues. This level of investigation and discussion represents an important inquiry with relevance to digital library services beyond just the DPR, especially as the utility service model is used more extensively at the CDL and elsewhere in the digital library community.

Our initial efforts have focused on the UC community and other large institutional digital libraries. Now we need to reach out beyond this constituency to public libraries and

other less technologically enabled library organizations. The Local History Digital Resources Project is one vehicle through which we will be extending the DPR infrastructure to other institutions.<sup>27</sup>

## 8 Conclusion

In conclusion, it is with great excitement about the future that we look forward to the many possibilities that exist ahead of us. In creating the DPR, we came to a fuller understanding of the challenges and opportunities inherent in digital preservation. Our next phase includes extending the infrastructure of the DPR more widely for the benefit of the greater digital library community. In the past three years, awareness about digital preservation has grown immensely within some communities, including UC, but further knowledge of issues and solutions needs to extend still wider to the public libraries, museums, and other memory institutions charged with the vital task of preserving the information that comprises our heritage.

There is still much ground to cover, but the CDL is well poised to continue on to future phases of this endeavor. We know what questions to ask and appreciate the diversity of approaches that would benefit from exploration. Through our accomplishments in building the DPR, we have amassed a host of ideas for next steps and methods of further extending the core base of the DPR infrastructure, thus enabling an even wider library community to take part in the vital and historic process of protecting our nation's assets for the long term.

---

<sup>27</sup> More information on the Local History Digital Resources Project, <http://www.cdlib.org/inside/projects/oac/lsta/#projdesc>.

## 9 Acknowledgements

Input from California Digital Library staff, the University Librarians of the University of California, digital preservation representatives and DPR users from the campus libraries throughout the University of California, and fellow members of the digital preservation community has been instrumental to the University of California Libraries Digital Preservation Repository and to the learnings and achievements detailed in this report.

In particular, members of the CDL DPR technical leads team deserve special mention: Peter Brantley, Lynne Cameron, Mary Heath, John Kunze, Margaret Low, Stu Sugarman, and Steve Toub. Special thanks and credit at the CDL also go to David Loy, Anita Auyang, Ravi Gudipati, Sean O'Hara, Raymund Ramos, Michael Thwaites, Paul Fogel, Randy Lai, Shifra Raffel, Mark Reyes, Brian Tingle, Eric Satzman, Claudia Woo, Michael Russell, Felicia Poe, Jane Lee, Jennifer Colvin, Michele Manuel, Cate Hutton, Garland Giles, and Ellen Dunn. Special appreciation to our pilot users on the UC campuses: Kathleen Cameron, Eric Choi, Michael Conkin, Stephen Davison, Curtis Fornadley, Chris Frymann, Lynne Grigsby-Standfill, Bernie Hurley, Terry Ryan, Heidi Schmidt, Stephen Schwartz, and Dawn Talbot. Thanks also to Martha Crowley at IMLS for her help.

This report was written by Suzanne Samuel, Project Manager, Digital Preservation; Patricia Cruse, Director, Digital Preservation; and Daniel Greenstein, Associate Vice Provost, Scholarly Information and University Librarian, Systemwide Library Planning and California Digital Library.