Preservation Is Not a Place

Stephen Abrams, Patricia Cruse, John Kunze

California Digital Library, University of California

December 2008

Abstract

The Digital Preservation Program of the California Digital Library (CDL) is engaged in a process of reinvention involving significant transformations of its outlook, effort, and infrastructure. This includes a re-articulation of its mission in terms of digital curation, rather than preservation; encouraging a programmatic, rather than a project-oriented approach to curation activities; and a renewed emphasis on services, rather than systems. This last shift was motivated by a desire to deprecate the centrality of the repository as *place*. Having the repository as the locus for curation activity has resulted in the deployment of a somewhat cumbersome monolithic system that falls short of desired goals for responsiveness to rapidly changing user needs and operational and administrative sustainability. The Program is pursuing a path towards a new curation environment based on the principle of devolving curation function to a set of small, simple, loosely coupled services. In considering this new infrastructure, the Program is relying upon a highly deliberative process starting from first principles drawn from library and archival science. This is followed by a stagewise progression of identifying core preservable values, devising strategies promoting those values, defining abstract services embodying those strategies, and, finally, developing systems that instantiate those services. This paper presents a snapshot of the Program's transformative efforts in its early phase.

Introduction

Information technology has become integral to the pedagogic mission of modern universities. The scholarly community routinely produces and utilizes a wide variety of digital assets in the course of teaching, learning, and research. These assets represent the intellectual capital of their institutions; they have inherent and enduring value and need to be preserved for use by future scholars. The Digital Preservation Program of the California Digital Library (CDL) has a broad mandate to ensure the long-term usability of digital assets within the 10-campus University of California (UC) system. To better position itself to meet this task, the Program is engaged in a process of reinvention involving significant transformations of its outlook, effort, and infrastructure.

Firstly, the Program now defines its mission in broader terms of digital *curation*, rather than preservation. While preservation and access were previously considered disparate functions, they are now properly seen as complementary: preservation aimed at providing access *over time*, while access depends upon preservation at a *point in time*. Curation better expresses the new programmatic emphasis on activities over the full digital lifecycle (Higgins, 2008). Secondly, the Program's developmental focus is now on content *services* rather than systems. Technical systems are inherently ephemeral, their useful lifespan being constantly encroached upon by disruptive technological change. Rather than pursuing the somewhat illusory goal of long-lived systems, curation goals are better served by concentrating on long-lived content, sustained by an evolving repertoire of nimble, commodified services. This change in emphasis is best exemplified by deprecating the centrality of the curation repository as *place*. Rather than relying on a conceptually monolithic system as a locus, curation outcomes will be the product of loosely coupled, independent, distributed services. To maximize the flexible applicability of these services throughout the digital lifecycle, and especially in the early, upstream stages, these services should be capable of operating on content in situ – a field research station, a laboratory workbench, a desktop workstation – without a necessary precondition of being transferred to a central locus for curation processing.

In effecting these transformations, the Program has engaged in a thorough review of its University stakeholders, progammatic policies, strategies, practices, and infrastructure. One important consideration of the new approach is the deferral of technical decision making until curation intentions and objectives are clearly understood and unambiguously defined. During the planning process for the new curation environment it has been useful to start from first principles drawn from library and archival science to ensure a sound conceptual basis for the work. This paper presents a snapshot of the Program's transformative efforts at an early phase.

Foundational Principles

Library and archival science have a deep, rich history. Their historically established principles still have currency when applied to the digital realm. Consider, for example, Ranganathan's five laws of library science (Ranganathan, 1931). The first three laws ("*Books are for use*"; "*Every reader his book*"; "*Every book his reader*") are fundamentally concerned with use. By loose analogy we initially assert that digital assets are preserved in order to be used. Furthermore, use entails that these assets be

both discoverable and utile, that is, users can find assets of interest and the information content encapsulated into those assets are meaningfully exposed to their users.

The fourth law ("*Save the time of the user*") is fundamentally concerned with service. All digital assets inherently require technological intermediation in order to be useful. Again by analogy, we assert that user services for curated assets must be available, responsive, and comprehensive, that is, they can be used at the time and place of user's choosing and they conform, moreover, to the user's expectations in terms of performance and functionality.

The fifth law ("*The library is a growing organism*") is fundamentally concerned with change. Due to the nature of intermediation, digital assets are inherently fragile with respect to technological change. Again by analogy, we assert that curation services must evolve over time in a sustainable manner in order to continue to provide value and mitigate threats to that value.

The concept of archival diplomatics stresses the importance of provenance, the understanding of an asset's source and relationship to the information content it encapsulates (Pearce-Moses, 2005; Ross, 2007). One of the distinguishing characteristics of digital content over analog forms is its ease of undetectable mutability. By analogy we finally assert that curated assets must not only be accesible and usable, but also authentic, that is, they are what they purport to be.

These assertions form the basis for the Program's curation imperative: providing highly available, responsive, comprehensive, and sustainable services for access to use, reuse, and enrichment of, authentic digital assets over time. Or more succinctly, "maintaining and adding value to a trusted body of digital information for current and future use" (Digital Curation Centre [DCC], 2007). It is important to note that these goals merely restate a set of stewardship responsibilities and activities that scientific and cultural memory institutions have aways carried out as their core mission

Curation Objects

The primary unit of curation management is the digital object, an encapsulation in digital form of an abstract intellectual or aesthetic work. The fundamantal components of a digital object with regard to curation considerations are content and description (see Figure 1). [1] Content is the primary intrinsic value carried and transmitted by the object and possesses abstract semantic meaning that is exposed through pragmatic behavior operating on tangible syntactic form. Description is extrinsic information about the object and expresses its significant syntactic, semantic, and pragmatic characteristics, and the evolution of these characteristics over time. [2] This tripartite definition of object content (*meaning*, *behavior*, *form*) follows from a semiotic conceptualization of preservation activity. Semiotics explains the transference of meaning across time and space in terms of signs, or meaning-bearing symbols. As

---

[1] Note that the object decomposition model (*content*, *description*) is a reasonable analogue to the OAIS information package model ISO 14721:2003 (International Organization for Standardization [ISO], 2003):

*OAIS*ObjectInformation objectContentData objectDescriptionRepresentation information

[2] The use of the terms *description* and *descriptive* in a metadata context is often intended to indicate a purely intellectual description as opposed to a technical, structural, behavioral one, and so on. This is not the sense of the term here. We use *description* in its most generic sense: any extrinsic information > *about* something, in this case, a digital object.

formulated by Morris, syntax is the relationship between signs; semantics is the relationship between signs and the things they represent or signify; and pragmatics is the relationship between signs and their interpreters (Morris, 1938; Rochberg-Halton, & McMurtrey, 1983). [3]

The semiotic perspective highlights the importance of behavioral considerations in curation activities. The use of digital assets presupposes their rendering, that is, conversion into human-sensible analog form, whether visual, aural, tactile, etc. Thus the preservation of behavior is an essential component of effective curation. While behavior is implied in the concept of representation information, that information generally stresses structural and semantic description.
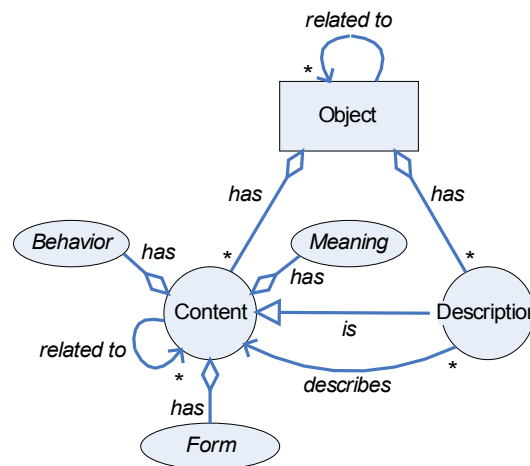

Figure 1. Curation object modeling.

Content components can be associated with other components through an arbitrary network of typed relationships, for example, *new-version-of*, *color-profile-for*, *page-image-of*, etc. Similarly, objects themselves can be associated through typed relationships, e.g. *new-edition-of*, etc.

Within the curation infrastructure, the internal form of a digital object is an aggregation of one or more formatted files representing intrinsic meaning and extrinsic description. These files are managed as opaque containers in an unstructured storage abstraction; conventionally, a file system (Linden, Martin, Masters & Parker, 2005). For operational convenience, some subset of description, or metadata, may be managed in a structured (or fielded) storage abstraction; conventionally, a relational or XML/RDF database. Regardless of this duplicative management, the metadata of record is that found in an object's files. This is an important consideration supporting a business continuity requirement that all automated services and systems can be fully reinstantiated from a file system expression.

Curation Strategies

Digital curation is a highly elastic term applicable to a continuum of intentions, activities, and outcomes, each with its attendant level of effort and efficacy. [4] In trying

---

[3] Note that the semiotic conception is a reasonable analogue of the National Archives of Australia (NAA) performance model (Heslop, Davis & Wilson, 2002):

Morris *NAA* Content Semantics Performance Behavior Pragmatics Process Form Syntax Source
[4] See for example, the variety of definitions of digital preservation (Association for Library Collections

to devise appropriate and attainable goals for the Program in an environment of ever increasing demands for services on an expanding variety and volume of content, it is necessary to achieve a careful balance of desirable function and available resources. To achieve this balance, the Program has found it useful to articulate curation intentions first in terms of desirable object- and service-centric values (see Table 1). Specific curation activities can then be articulated in terms of strategies designed to foster those values (see Tables 2 and 3).[5]

Value	Justification	Strategy
Identity	To distinguish an object from others	Persistant naming, actionable resolution
Viability	To recover an object from its storage medium	Redundancy, heterogeneity, media refresh
Fixity	To ensure that an object is unchanged from its accepted state	Redundancy, error-correcting codes, message digests
Authenticity	To ensure that an object is what it purports to be	Provenance, cryptographically-secure signatures
Ontology	To understand an object's significant nature	Syntactic, semantic, and pragmatic characterization
Visibility	To enable users to find objects of interest	Public discovery systems and registries, exposure for web harvest
Utility	To expose an object's underlying information content	Behavior-rich delivery
Portability	To faciliate content sharing and succession planning	Self-contained, self-documenting objects, packaging standards
Appraisment	To understand the consequences of the passage of time	Analysis and assessment
Timeliness	To know when a preservation value is threatened	Technology watch, stakeholder engagement

Table 1. Object-centric curation values and strategies.

---

and Technical Services [ALCTS], 2007; Lavoie & Dempsey, 2004).
[5] An earlier version of these curation values and strategies was presented in Abrams and Kunze (2007).

Value	Justification	Strategy
Availability	To provide access at a time of user choosing	Redundancy, automated failover
Responsivity	To provide appropriate throughput	Redundancy, load balancing
Security	To enfore appropriate use of services and content	Cryptographically–secure identity and role management
Interoperability	To facilate creative object reuse	Standard interfaces
Extensibility	To enable graceful evolution	Granularity, orthogonality, virtualization
Trustworthiness	To promote users' sense of predictability and reliability	Transparency, audit, certification
Sustainability	To ensure ongoing access and use	Commodity components, institutional commitment, financial cost-recovery, professional development

Table 2. Service-centric curation values and strategies.

Once the curation values are well-defined, specific strategies (or groups of related strategies) can be formulated as abstract services and then implemented through concrete human activities or automated systems. By deferring technical decision making until the underlying intentions and outcomes are clearly understood, the Program hopes to arrive at curation solutions that are fully functional yet minimally resource-intensive to develop and deploy.

Curation Services

The cultural, scientific, and economic value of a digital object is predicated by both its intended and actual use. However, in the context of curation over any significant time period, the users and uses of a digital asset cannot be definitively known a priori. The Program will therefore accept custodial responsibility for digital objects from UC-affiliated agents regardless of provenance, structure, format, or characterization by metadata. However, the level of assurance of ongoing usability applicable to a given object is subject to limitations imposed by these formal factors (or their absence), the general state of preservation understanding, and other CDL priorities.

While it is certainly necessary to have a robust, secure, and sustainable technical environment in which to manage digital objects, their preservation and use are also dependent upon significant human competencies, analysis, and decision making, both on the part of CDL curation managers and UC collection managers and curators. The analytical and consultative services provided through human effort, albeit often with significant technological augmentation, include:

- Best practice recommendations for the creation and acquisition of curation-amenable objects, including:
  - Selection of formats best suited for representing content.
  - Development of appropriate technical specifications and workflows for object creation; and selection of "best edition" from multiple versions of object content for use in creating derivatives.
- Consideration of appraisal and selection factors leading towards a decision to curate content, including:
  - Assessment of intellectual, aesthetic, economic, and artifactual value

- o Rarity (or ubiquity), ease (or difficulty) of re-acquisition, and degree of alternative access.
- Preservation planning activities focused on ensuring ongoing usability of managed digital objects, including:.
  - o Understanding object significant properties, and formal characteristics – format, structural relationships, behavior – that expose those properties.
  - o Surveying the technological environment that mediates object use.
  - o Consideration of changing behavioral expectations for that use.
  - o Determination of the prospective resilience of managed objects based on their formal characteristics evaluated in light of environmental monitoring and user expectation.
  - o Development of action plans, with associated trigger events, to ameliorate identified preservation risks.
- The complex of preservation intervention activities involved in executing action plans following trigger events, including quality assurance testing subsequent to the execution of the plans to ensure the efficacy of the intervention and the invariance of the resulting object's significant properties.
- Service brokerage to select appropriate curation service providers and mediation of service-level agreement negotiation.

Curation Infrastructure

The locus for automated curation services is conventionally defined as a *repository*, most often considered in terms of the OAIS reference model. However, an undue insistence on the centrality of the repository often leads to large, cumbersome systems that are expensive to deploy and support. A more nimble and sustainable approach follows from recasting curation as a content-centric, rather than system-centic activity (Gladney, 2007). In other words, preservation is not a *place* into which content is put for safe-keeping, but rather, it is a *process* in which content evolves proactively and reactively through the application of strategy-embodying services.

Consistent with best practice for component-oriented architecture, curation services should be independent and single-purpose (Factor, 2008; Hamilton, 2007; Liegl, 2007). These services should expose themselves through well-defined abstract interfaces that constitute the service contract. Service requests can be made, and responses received, through concrete language bindings (for procedural interaction) or protocol bindings (for network interaction) that expose the interfaces (see Figure 2). Individual services should be instantiated as self-contained, easily deployable, and redeployable, units.

Curation Micro-Services

The individual automated curation services (known as "micro-services" in view of their fine granularity) are intended to execute in the context of a computational cloud and operate on objects managed in a storage cloud.[6] The cloud abstraction removes specific environmental dependencies and permits the flexible deployment of the services for purposes of adaptive load balancing and failover, and faciliates

---

[6] This concept is strongly informed by the iRODS system, although no specific decision has yet been made on how best to achieve it (Moore, 2008).

deployment in the native environments in which digital content is created, acquired, and used. The definition of the services follows the general REST-like principles of granularity, orthogonality, uniform interfaces, and stateless session (Fielding, & Taylor, 2002). Atomistic services can be combined through pipe-like layering to provide more complex behaviors. Creating complexity through the composition of simple components should simplify overall implementation and facilitate service responsiveness to changing technical and functional demands.[7]
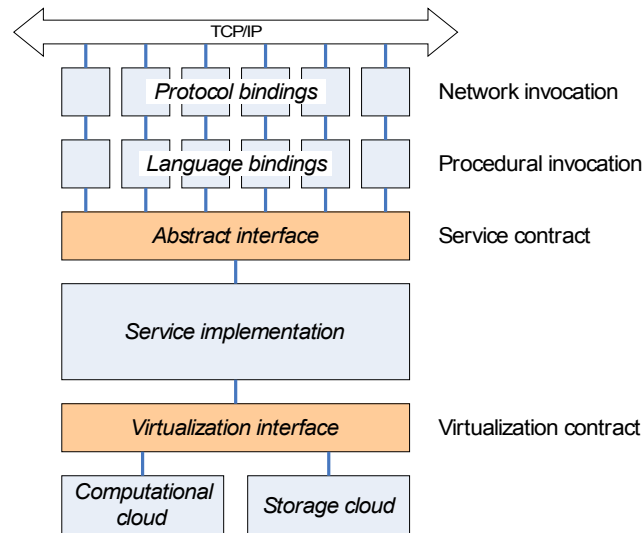
Figure 2. Preservation service stack.

Identity Service

An identifier is a persistent association between a unique character string and typed referents or name/value pairs. In the curation context there are three important reference types related to managed digital objects:

- Information used to retrieve a representation of a digital object, for example, an actionable URI.
- Information descriptive of that object, for example, kernel metadata (Kunze & Turner, 2007).
- Information descriptive of the identifier itself, for example, a statement of persistence obligation by the service provider.

The Identifier services supports three processes:

- Minting, the generation of new identifiers conforming to the syntactic rules of a specific namespace.
- Binding, the association of an identifier string with a typed referent.
- Resolution, the retrieval of a referent for a given identifier string.

The current implementation intention of the Program is to use Archival Resource Key (ARK) identifiers and the Nice Opaque Identifier (NOID) minting, binding, and resolution management system as the basis for the Identity service (California Digital Library, 2006, 2008).

---

[7] As the work described in this paper is still in its earliest phase, a number of these micro-services are not yet associated with specific implementation plans.

Storage Service

The Storage service provides unstructured storage in which to manage the file components of digital objects. Although the individual files are semantically opaque to the service, object coherence – the fact that a given file is associated with a specific object – will be maintained by the service. If necessary, objects can be reinstantiated from information managed solely by the storage service.

The current implementation intention of the Program is to use Pairtrees as the underlying storage abstraction (Kunze, Haye, Hetzner, Reyes & Snavely, 2008). A Pairtree is a file system hierachy into which objects files are deterministically placed based on a bigram, or character pair, decomposition of the object identifier. For example, version three of the file *abcd* of object *123456* would be found at:

```
pairtree_root/
          12/
            34/
              56/
                v003/
                    data/
                        abcd
```

The bigram decomposition was chosen to provide reasonable subdirectory fan-out (hierarchical breadth vs. depth) to optimize read and write performance for the widest variety of file systems.

Characterization Service

Characterization is information that describes a digital object's format-specific character or significant nature (Brown, 2007). Characterization processing has four important aspects:

- Identification, the determination of an object's purported format on the basis of suggestive extrinsic hints and intrinsic signature
- Validation, the determination of an object's conformance to the normative requirements of its format
- Feature extraction, the reporting of an object's intrinsic properties that can be used as a surrogate for the object itself in the context of much curation analysis and decision making
- Assessment, the determination of an object's acceptibility for a specific purpose based on local policy rules

The Characterization service can be utilized in the following contexts:
- Client-side Submission Information Package (SIP) packaging
- Server-side Ingest processing
- Post-transformation quality assurance testing

The current implementation intention of the Program is to use JHOVE2 as the basis for the service (Abrams, Owens, & Cramer, 2008).

Catalog Service

The Catalog service maintains descriptive information about digital objects and their files in support of optimized queries for preservation decision making. In general, this will be a subset of the full information encapsulated within the object

itself as it is instantiated in the storage service. Descriptive information can be supplied explicitly as part of object acquisition or implicitly through the use of the Characterization or Annotation services.

Annotation Service

The Annotation service is used to enable user-driven enrichment of managed digital objects.

Fixity Service

The Fixity service provides a means to verify the bit-level integrity of individual files managed by the Storage service using typed message digests.

Replication Service

Storage resilience is promoted through two primary strategies: use of enterprise, rather than commodity hardware; or redundancy. Recent research strongly suggests that global redundancy using commodity components obtains the highest level of assurance at the lowest cost (Rosenthal, 2008). Note that redundancy is most effective when it is most uncorrelated (Pinheiro, Weber & Barroso, 2007).

The Replication service will provide methods to set object replication policies regarding the number and location of the replicas. It is assumed that various Storage service instantiations will provide global heterogeneity and decorrelation.

Transformation Service

The Transformation service provides a means to perform the transcoding of digital object representations from existing available forms to newly required forms. The service can be utilized in the following contexts:
- Ingest canonicalization to conform to internal standards for AIPs
- Preservation localization to remove external dependencies such as schemas or font definitions
- Preservation desiccation to produce deliberately lossy, although inherently more preservation-friendly, derivatives as preservation copies of last resort, for example, an ASCII representation of a PDF document
- Preservation migration to mitigate format obsolescence
- Access derivation, for example, the creation of downsampled compressed copies of master images

Ingest Service

The Ingest service provides a means to bring new material into the curation environment for management. It makes use of the Characterization, Transformation, Identity, Storage, and Catalog services.

Access Service

The Access service provides a means of requesting object representations. It makes use of the Catalog, Identity, Storage, and Transformation services.

Conclusions

The CDL Digital Preservation Program is engaged in a significant transformation of its fundamental outlook and effort. Curation more properly connotes the full range of lifecycle activities central to the mission of the Program. Due to the inherent fragility of digital assets with respect to technological change, curation activities must be proactive, rather than reactive. This necessitates an open-ended programmatic, rather than time-bounded project-oriented approach. Since the automated systems utilized in a curation context are inevitably ephemeral, curation goals are better served by concentrating on long-lived content sustained by a constantly evolving repertoire of nimble, commodified services. This reinvention will better enable the Program to remain responsive to the ever-expanding needs of its University stakeholders, and to collaborate more effectively with the broader curation community.

Inspired by the principle of Ockham's Razor, the conceptual framework developed in this paper started by considering the question, How simple can a curation environment be and still be effective? The proposed solution can be summarized in terms of three rather simple aphorisms:

- Lots of copies keeps stuff safe[8]
- Lots of services keeps stuff useful
- Lots of uses keeps stuff valuable

At the technical level, rendundancy ("lots of copies") is the key principle for ensuring the safety of curated assets and the availability of services built around those assets. At the programmatic services level, responsiveness to the needs of users ("lots of services") is the key principle for ensuring the widespread integration of curated assets into the research, teaching, and learning activities of the University.[9] At the level of scholarly discourse, the multiplier effect of the creative use and re-use of curated assets ("lots of uses") is the key principle for enriching that discourse.[10]

The deliberative multistage design process (*value → strategy → service → system*) employed by the Program is intended to guard against over-engineered solutions. By insisting on clear, well-defined intention and outcome definitions upfront, implementation efforts can focus on specific necessary function. Of course, the determination of "necessary" will evolve over time, but the Program believes that the governing principle of its new environment – providing curation function through granular, interoperable, virtualized components – will allow that environment to evolve freely and responsively.

References

Abrams, S., & Kunze, J. (2007). Preservation is not a location. *Preservation and Archiving Special Interest Group (PASIG) Conference, San Francisco*.

Abrams, S., Owens, E., & Cramer, T. (2008). "What? So what?" The next-generation JHOVE2 architecture for format-aware characterization. *Proceedings of the Fifth International Conference on Preservation of Digital Objects*. London: British Library.

---

[8] This phrase was originally coined by the LOCKSS Project (Reich & Rosenthal, 2001).
[9] See, for example, Crane (2008).
[10] See, for example, Blyth & Chachra (2005) and Van de Sompel, Lagoze, Bekaert, Lui, Payette, & Warner (2006).

Association for Library Collections and Technical Services. (2007). *Definitions of digital preservation*. Retrieved November 5, 2008, from
http://www.ala.org/ala/alcts/newslinks/digipres/index.cfm

Blyth, E., & Chachra, V. (2005). The value proposition in institutional repositories. *EDUCAUSE Review Magazine 40*(5). Retrieved November 5, 2008, from
http://www.educause.edu/EDUCAUSE+Review/EDUCAUSEReviewMagazineVolume40/TheValuePropositioninInstituti/158011

Brown, A. (2007). Developing practical approaches to active preservation. *International Journal of Digital Curation, 2(1)*, pp. 3-11.

California Digital Library (2006). *NOID(1): Batch Identifier Infrastrucuure*, Retrieved November 5, 2008, from http://www.cdlib.org/inside/diglib/ark/noid.pdf

California Digital Library (2008). *ARK (Archival Resource Key)*. Retrieved November 12, 2008, from http://www.cdlib.org/inside/diglib/ark/

Crane, G. (2008). On scholarly access to a million books. *Using Digital Collections: Open Content Alliance Annual Meeting*, San Francisco, October 27-28, 2008.

Digital Curation Centre (2007). *What is the DCC: What is Digital Curation?* Retrieved November 5, 2008, from http://www.dcc.ac.uk/about/what/

Factor, S. (2008). What makes a design "Googley"? *The Official Google Blog*, April 23, 2008. Retrieved November 5, 2008, from
http://googleblog.blogspot.com/2008/04/what-makes-design-googley.html

Fielding, R., & Taylor, R. (2002). Principled design of the modern web architecture. *ACM Transactions on Internet Technology 2(2)*: pp. 115-150.

Gladney, H. (2007). *Economics and engineering for preserving digital content*, December 5, 2007. Retrieved November 5, 2008, from
http://eprints.erpanet.org/141/01/LDP_Engineering.pdf

Hamilton, J. (2007). On designing and deploying internet-scale services. *21st Large Installation System Administration Conference (LISA '07).*

Heslop, H., Davis, S., & Wilson, A. (2002). *An Approach to the Preservation of Digital Records*. Retrieved November 17, 2008, from
http://www.naa.gov.au/Images/An-approach-Green-Paper_tcm2-888.pdf

Higgins, S. (2008). The DCC curation lifecycle model," *International Journal of Digital Curation 1(3)*: pp. 134-140.

International Organization for Standardization. (2003). ISO 14721: *Space data and information transfer systems – Open archival information systems – Reference model*.

Kunze, J., Haye, M., Hetzner, E., Reyes, M., & Snavely, C. (2008). *Pairtrees for Object Storage (V0.1)*. Internet draft. Retrieved November 12, 2008, from http://www.ietf.org/internet-drafts/draft-kunze-pairtree-00.txt

Kunze, J., & Turner, A. (2007). *Kernel Metadata and Electronic Resource Citations (ERCs)*. Retrieved November 12, 2008, from http://www.cdlib.org/inside/diglib/ark/ercspec.html

Lavoie, B., & Dempsey, L. (2004). Thirteen ways of looking at … digital preservation. *D-Lib Magazine* 10(7/8). Retrieved Nov. 5, 2008, from http://www.dlib.org/dlib/july04/lavoie/07lavoie.html

Liegel, P. (2007). The strategic impact of service oriented architectures. *Proceedings of the 14th Annual IEEE International Conference and Workshops on the Engineering of Computer-Based Systems (ECBS '07)*, pp. 475-484.

Linden, J., Martin, S., Masters, R., & Parker, R. (2005). *The large-scale archival storage of digital objects*. Digital Preservation Coalition Technology Watch Series Report 04-03.

Moore, R. (2008). Towards a theory of digital preservation. *International Journal of Digital Curation 3*(1): pp. 63-75. Retrieved June 22, 2009, from http://www.ijdc.net/index.php/ijdc/article/viewFile/63/42

Morris, C. W. (1938). *Foundations of the Theory of Signs*. Chicago: University of Chicago Press.

Pearce-Moses, R. (2005). Diplomatics. *A Glossary of Archival and Records Terminology*. Retrieved May 26, 2009, from http://www.archivists.org/glossary/term_details.asp?DefinitionKey=21

Pinheiro, E., Weber, W.-D., & Barroso, L. A. (2007). Failure trends in a large disk drive population. *Proceedings of the 5th USENIX Conference on File and Storage Technologies. San Jose*.

Ranganathan, S. R. (1931). *The Five Laws of Library Science*. Madras.

Reich, V, & Rosenthal, D. S. H. (2001, June). LOCKSS: A permanent web publishing and access system. *D-Lib Magazine* 7(6). Retrieved Nov. 5, 2008, from http://www.dlib.org/dlib/june01/reich/06reich.html

Rochberg-Halton, E., & McMurtrey, K. (1983). The foundations of modern semiotic: Charles Peirce and Charles Morris. *American Journal of Semiotics, 2*(1-2), pp. 129-157.

Rosenthal, D. S. H. (2008). Bit storage: A solved problem? *Proceedings of the Fifth International Conference on Preservation of Digital Objects. British Library, London.*

Ross, S. (2007). Digital preservation, archival science, and methodological foundations for digital libraries. *11th European Conference on Research and Advanced Technology for Digital Libraries. Budapest.*

Van de Sompel, H., Lagoze, C., Bekaert, J., Lui, X., Payette, S., & Warner, S., (2006). An interoperable fabric for scholary value chains. *D-Lib Magazine 12*(10). Retrieved November 5, 2008, from http://www.dlib.org/dlib/october06/vandesompel/10vandesompel.html