Durable Digital Objects Rather Than Digital Preservation

and Professional Implications

H.M. Gladney
HMG Consulting

Abstract: Long-term digital preservation is not the best available objective. Instead, what information producers and consumers almost surely want is a universe of durable digital objects—documents and programs that are as accessible and useful a century from now as they are today.

Given the will, we could implement and deploy a practical and pleasing durability infrastructure within two years. Tools for daily work can embed packaging for durability without much burdening their users. Moving responsibility for durability from archival employees to information producers also avoids burdening repositories with keeping up with Internet scale. An engineering prescription is available.

Research libraries' and archives' slow advance towards practical preservation of digital content is remarkable to outsiders. Why is their progress stalled? Ineffective collaboration across disciplinary boundaries has surely been a major impediment. We speculate about cultural reasons for this situation and warn about possible marginalization of research librarianship as a profession. †

Introduction

> "[No] concise and well-developed [digital preservation] strategy that represents the views of a broad community has yet emerged. Since 1989 at least twelve have been published. … as a community we need to re-think how we are approaching research … [and] need to engage … researchers in this process, and especially those with a strong computing science and engineering background." Ross[1]

Last year's ECDL keynote address, characterized by this quotation, suggests the import of many articles written by librarians, archivists, and information scientists.[2] The statement is remarkable because computer scientists and software engineers believe they know solutions to all the pertinent technical problems[3] and how to realize these solutions in practical and pleasing document management tools.[4]

The difference of opinions is puzzling. The current paper speculates, "Why have archivists and research librarians reached an impasse while ignoring applicable software engineering?"

---

† All cited Web pages were viewed in January 2008 or more recently.

1 S. Ross, *Digital Preservation, Archival Science and Methodological Foundations for Digital Libraries*, 11th European Conference on Digital Libraries (ECDL), Budapest, 2007.

2 Readers can judge the assessments made in this paper by inspecting conference proceedings, such as that available at http://www.kb.nl/hrd/congressen/toolstrends/programme-en.html, plans such as that of M. Bellinger et al., *OCLC's digital preservation program for the next generation library*, Advances in Librarianship 27, 25-48, 2004, workshop reports such as that cited in footnote 21, and other writings cited in these.

3 Solutions to technical challenges evident in 2004 can be seen in two papers: H.M. Gladney, *Trustworthy 100-Year Digital Objects: Evidence After Every Witness is Dead,* ACM Trans. Office Information Systems 22(3), 406-436, 2004; H.M. Gladney and R.A. Lorie, *Trustworthy 100-Year Digital Objects: Durable Encoding for When It's Too Late to Ask*, ACM Trans. Office Information Systems 23(3), 299-324, 2005.

Other work in this direction has appeared: J.A. Smith and M.L. Nelson, *Creating Preservation-Ready Web Resources,* D-Lib Magazine 14(1/2), 2008.

4 H.M. Gladney, *Economics and Engineering for Preserving Digital Content,* http://eprints.erpanet.org/139/.

> "[T]otal incomprehension gives, much more pervasively than we realize …an unscientific flavor to the whole 'traditional' culture, and that unscientific flavor is often … on the point of turning anti-scientific. … If the scientists have the future in their bones, then the traditional culture responds by wishing the future did not exist. It is the traditional culture … which manages the western world.
>
> "This polarization is sheer loss to us all. To us as people, and to our society. It is at the same time practical and intellectual and creative loss." C.P. Snow[5]

We speculate that, half a century after C.P. Snow identified the cultural gap between Oxbridge liberal arts and science communities as an impediment to progress, its effects continue unabated.[6]

Durable Objects and Trusted Repositories

Cold butter is a good medium for beautiful sculpture. Ice works too. Nevertheless, when a sculptor wants to please future generations, he crafts in stone or pours molten bronze into short-lived molds.

In ancient custom, statues were mounted in open spaces accessible to many people. But custodians learned that scultures' longevity was enhanced by moving them into churches, museums, and palaces. They provided access with replicas mounted on the original pedestals, and sometimes poured additional copies for a far-flung public.[7]

Almost everybody, including this author, has been gulled by the phrase "digital preservation." History teaches us that it might be more effective to make objects (material and digital) durable when they are first created or, more precisely, when they are first shared.

Preservation should be an objective only for the (all too frequent) cases in which durability has not been built into artifacts early in their creation/dissemination. Repositories are merely tools for housing and disseminating the best human artifacts. Straining to make cultural institutions do what they are ill-suited to accomplish makes little sense. Instead it wastes skills and resources that could be better employed.

A lesson is evident. Prescribing how clerical procedures might achieve digital preservation by creating "Trusted Digital Repositories" is not the best available objective. Instead we should focus on structure and content that create usefully "Durable Digital Objects."

Social and Economic Context

> "[P]reservation of digital materials [is] a labour-intensive artisan or craft activity. … there is widespread agreement that the handicraft approach will not scale to support the longevity of digital content in diverse and large digital libraries." Ross[1]

The Information Revolution is relatively new, beginning about 50 years ago and still evolving rapidly. A century ago, recorded data held in governmental institutions and in the private sector were several orders of magnitude smaller than they are today.[8] Publishing written works was labor-intensive. Audio-visual recording was little more than a laboratory curiosity. Scientific records were notebook scribbles. Written information to manage individuals' personal health and welfare was rare. Radio, television, and the Internet did not exist.

The research community has grown 100-fold since 1930. Large increase in university faculties, coupled with the publish-or-perish syndrome, has created a periodical subscription crisis for research libraries. Our reading is burdened by scholarly articles that convey little new. Similar difficulties are evident in the popular press and business trade press.

The professional community ready to invest in preservation (alluded to below as the "digital preservation community (DPC)") numbers between 500 and 5000 archivists and academics worldwide. The

---

5 Loc. cit. footnote 6, pp. 9-10.

6 C.P. Snow, *The Two Cultures,* Cambridge U.P., 1959.

7 Safety measures are illustrated by Ghiberti's *Doors of Paradise,* crafted for the Baptistry of Florence's Duomo, but hidden and replicated in 1943. When both the original panels and the replicas survived the war, the replicas were acquired for San Francisco's Grace Cathedral, where they please viewers almost as much as do the remounted originals. See http://www.gracecathedral.org/enrichment/crypt/cry_19960703.shtml.

8 J.F. Gantz et al., *The Expanding Universe: A Forecast of Worldwide Information Growth Through 2010*, IDC White Paper, 2007, http://www.emc.com/about/destination/digital_universe/.

information technology community providing tools to create and share digital content is much larger—between 200,000 and 2,000,000 computer programmers. How many people add to the information pool? Perhaps as many as 50,000,000; if technology to record personal activities[9] appeals to the public, this number will increase ten-fold.[10]

The DPC is too small to keep up with software engineers' activities and ordinary citizens' software uptake. The pace of information creation greatly exceeds archival institutions' ability to select for preservation. The facts suggest that no attempt archivists make to archive a significant fraction of newly created digital content can succeed.

> "Maybe we need to empower the individual, or, even, to understand that individuals will come to assume more and more responsibility for preserving our digital heritage—rather than records professionals' constant search for the magic solution for all systems in all institutional and individual applications. [I]nspiration comes from Leonardo da Vinci … 'He was an endless doodler … who tucked several notebooks of varying sizes into his waist belt to record his thoughts…'" Shneiderman[11]

What Might Preservation Clients Expect?

What might one of our descendants want of information stored today? He would be satisfied if, for any document preserved in a world-wide repository network, he could:

1. Retrieve a copy of the bit-string that represents the content;
2. Read or otherwise use the content as its producers intended;
3. Decide whether the information received is trustworthy;
4. Exploit embedded references (links) to retrieve contextual information; and
5. Exercise all this functionality conveniently.

In addition to professional authors, some citizens will want to preserve information without asking anybody's permission to do so. They will want convenient tools and infrastructure to:

6. Package any content to be preservation-ready;
7. Submit such readied content to repositories that promise to save it.

What technology will repository institutions want? In addition to digital library technology, they will want support for:

8. Continuing to use deployed content management software without disruption;
9. Sharing content and metadata; and
10. Avoiding preservation burdens beyond their resources.

What Research Is Needed?

Ross articulates a long-term digital preservation (abbreviated LDP below) research agenda that "responds to the lack of progress … in the delivery of preservation solutions, methods and techniques over the past twenty years."[1] The challenges he identifies as needing research attention are represented in condensed form in Table 1, together with a computer scientist's reaction to each.

We take "digital preservation" to mean "mitigation of deleterious effects of decay, obsolescence, and human error or misbehavior that might impair the value of digital document copies long after originals were created." This choice partitions "archiving" into digital library services for a perfect world—a world in which information access does not degrade with time—and mitigation of degradation.

With this definition, "archiving" is patently a different (and broader) topic than "digital preservation." Archiving includes content selection, service delivery by archives, social acceptance, technology uptake by prospective users, library community education, and institutional management. The current paper

---

9 G. Bell and J. Gemmell, *A Digital Life,* Scientific American 296(3), 58-65, 2007.

10 Our statistics are crude. Any might be incorrect by a factor of three. Such uncertainty hardly affects what the numbers imply—that no plausible increase in repository institution resources will reduce the challenges.

11 B. Shneiderman, *Leonardo's laptop: Human needs and the new computing technologies*, MIT Press, 2002.

avoids such topics—issues of digital library infrastructure to please "perfect world" clients. If the reasons for doing so are not immediately obvious, they will become so.

Table 1: Views of preservation research

Research challenge according to Ross	A software engineer's perspective
Restoration: when digital objects have broken, how can we ensure the correctness of restored versions?	Restoration is not urgent. Readers wanting at-risk content can foot R&D expenses when they know they want access.
Conservation of intact digital objects: how can we ensure copy integrity and authenticity?	How to do this is known.[3] Open questions include whether cheaper methods exist.
Collection and repository management: what archival methods will satisfy collection users' quality expectations?	More than 100 digital library packages exist. Any institution can tailor one for its preferences. LDP support will require at most modest adaptation.
Risk management: how can archivists quantify content risks and benefits to allocated resources?	Crude risk estimates will be good enough. What is worth saving is opinion inappropriate for research publication.
Preserving digital object interpretability: how can archivists save any digital object to be fully useful in centuries to come?	How to do this is known.[3] Open questions include whether and when file-type-specific methods would be preferred over the generic solution.
Collection cohesion and interoperability: how can archives integrate collections for sharing among independent repositories?	Context is subjective choice to be made by authors and subject experts. As it is required for today's services, this need not be considered an LDP topic.
Preservation automation: what automation can effectively manage the digital object flood?	Methodology for handling tomorrow's data floods is a task worthy of the best engineering skills
Preserving digital object context: what are the legal and social expectations for contextual information and how can these be realized?	Existing metadata initiatives are on the right track.[12] How to reliably package unique object context is known.[3]
Storage technologies and methodology: how can we deploy an archival repository network?	How to manage immense storage is understood and realized in available software offerings.

Models of Information Communication

Subtle information packaging weaknesses might not be discovered until readers access content saved many years earlier. Such risks can be reduced by careful analysis of communication models.[13] Figure 1 models digital object transmission. Figure 2 models digital repository software and users. Figure 3 abstracts archival object structure.

Information Flow

A very simple model (Figure 1) teaches critical aspects of document communication. The **0? 1** and **9? 10** transmission steps involve subjectivity. Every attempt to communicate is confounded by accidental information; signaling what is essential and what is accidental is partially feasible, but difficult and rarely done. Social convention asserts that the primary authority for essential/accidental distinctions is the information provider. The secondary authority is the Figure 1 information consumer, who judges what information is interesting in the light of his objectives. For instance, books' marginal notes might be considered essential by paleographers and almost no-one else.

---

[12] Northwestern Univ. Digital Library Committee, *Inventory of Metadata Standards and Practices,* at http://staffweb.library.northwestern.edu/dl/metadata/standardsinventory/.

[13] More elaborate versions of these models are available in H.M. Gladney, *Preserving Digital Information*, Springer Verlag, 2007; ISBN 978-3-540-37886-0.

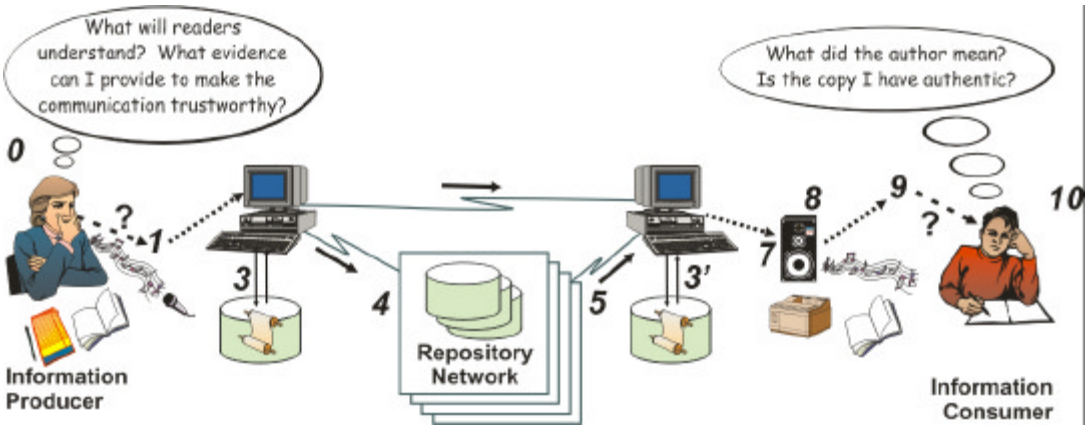

Figure 1: Information flow model

There might be much machinery and many potential transmission paths between each producer's information copy **3** and the version **3'** that arrives in the consumer's PC. The most reliable way of ensuring the integrity of the consumer's version, and its independence of the transmission path, is to arrange that the copy **3'** is identical, bit by bit, with the copy **3**. The customers of archiving service don't care how the intervening machinery works, provided that it is inexpensive, fast, and unobtrusive!

Archival Models

> "[T]he OAIS model has several shortcomings … as a basis for developing a software system. It is therefore necessary to … fill the gap between the OAIS model and software development." Egger[14]

Figure 2 corresponds to the OAIS functional entities[15] depiction (Figure 4), but emphasizes software partitioning and layering rather than archival procedures that can and should be the private business of any service institution.

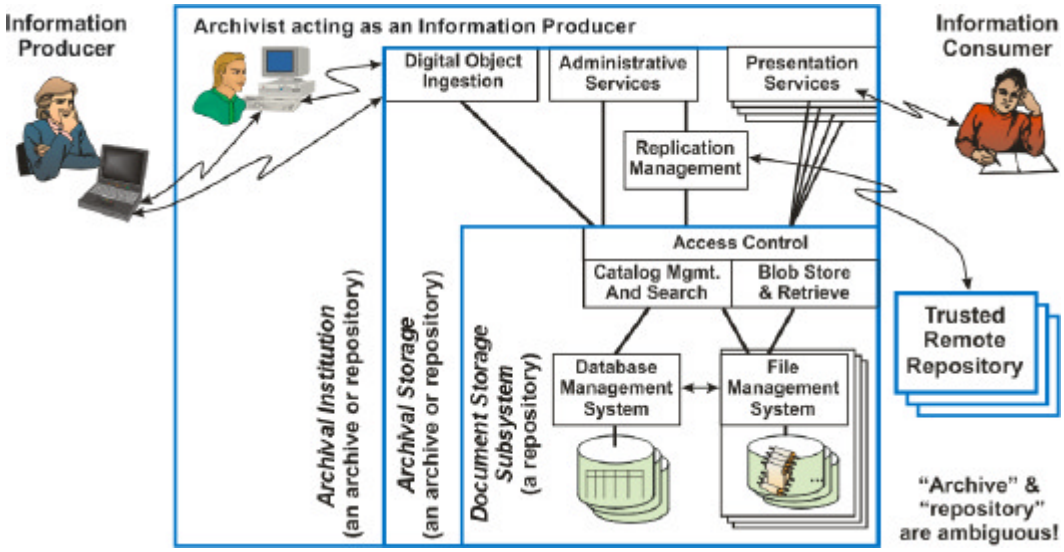

Figure 2: Nested repositories and human interfaces

---

[14] A. Egger, *Shortcomings of the Reference Model for an Open Archival Information System (OAIS)*, TCDL Bulletin 2(2), 2006, http://www.ieee-tcdl.org/Bulletin/v2n2/egger/egger.html.

[15] CCSDS 650.0-R-2, *Reference Model for an Open Archival Information System* (OAIS), 2001, Fig. 4-1.

Bit-string replication services suggested by the figure are sufficiently well understood that they need no attention in current preservation literature.

Figure 2 suggests that all the rest of preservation support can and should be implemented as PC programs. From the perspective of information consumers, archivists' modifications of preserved documents should be limited to and documented as if they were authorized by the original information producers, e.g., perhaps along the lines of editors' help for authors. Any other changes to document originals would be potential authenticity defects.

A consequence is that everything else needing attention concerns the design, content, and use of preserved information and tools for document preparation and inspection.

A Digital Object Model

Figure 3 suggests packaging for any digital object or collection. Any conforming object can be made trustworthy with cryptographic signatures. Such signatures need evidentiary infrastructure support that could be maintained by networked repositories.[16] The figure reminds us to include metadata both for collections and for collection contents and integrity evidence for links (references).

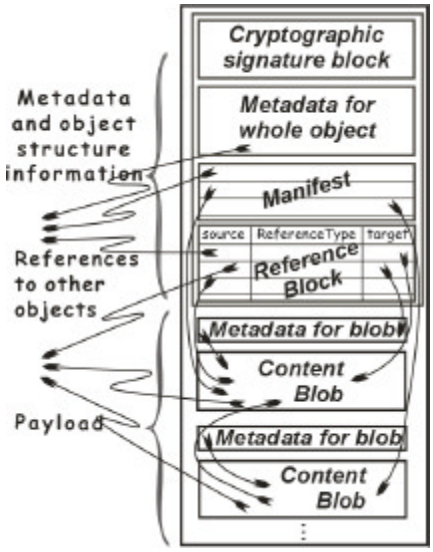

Figure 3: A digital object model

Approaches to Digital Preservation

Two preservation approaches have received attention—one focusing on archival procedures and one focusing on information clump structure. The former has received more attention, but is technically flawed and economically impractical. The latter is sound, focused on what will please end users, and could be deployed to be convenient and cost effective.

Trusted Digital Repositories

> "Long-term preservation of digital information on a scale adequate for the demands of future research and scholarship will require a deep infrastructure capable of supporting a distributed system of digital archives.
> …
> "A process of certification for digital archives is needed to create an overall climate of trust about the prospects of preserving digital information." Task Force on Archiving[17]

---

[16] G. Caronni, *Walking the Web of Trust*, Proc. 9th Workshop on Enabling Technologies. 2000.

[17] Findings in J. Garrett et al., *Preserving Digital Information: Report of the Task Force on Archiving Digital Information,*1996, http://www.clir.org/pubs/abstract/pub63.html.

Librarians and archivists, much influenced by the Figure 4 model of archival institutions, have diligently explored how their traditional methods[18] might be adapted for LDP and how to achieve public certification of institutional practices.[19] The Trusted Digital Repository approach depends on a premise that its proponents never question—the premise that LDP can be achieved by modestly adapting traditional archival methodology. Their recommended procedures might be necessary. However, they are insufficient to meet the expectations of information clients.

> *"… so steht der Benutzer/die Benutzerin der Zukunft gleichwohl vor einem Problem. Er oder sie ist ohne weitere Unterstützung nicht in der Lage, den archivierten Datenstrom zu interpretieren, da die erforderlichen technischen Nutzungsumgebungen längst nicht mehr verfügbar sind."* Schwens and Liegmann[20]

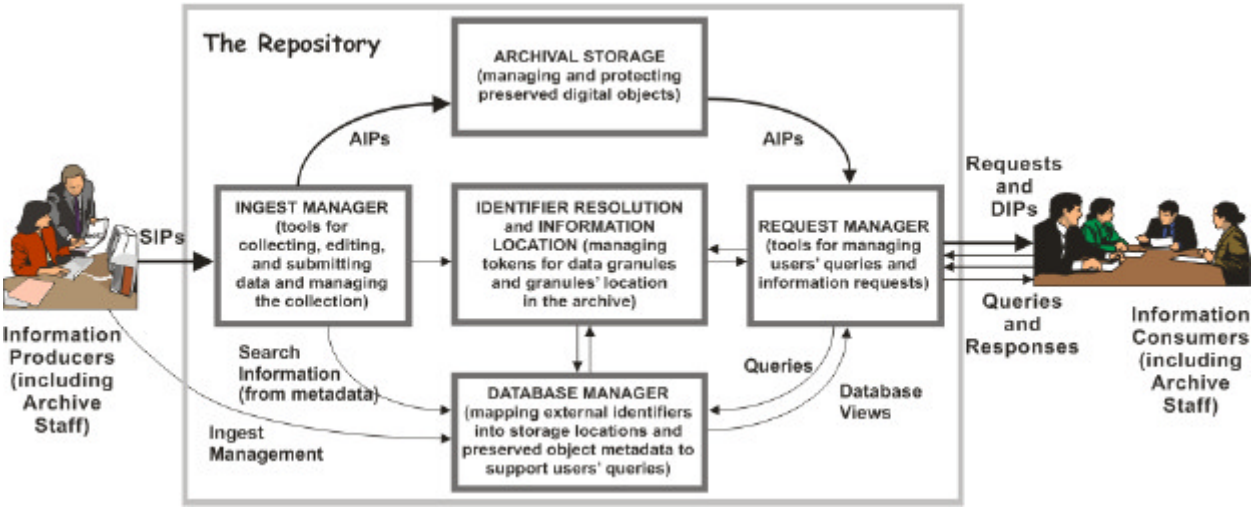

Figure 4: OAIS repository model

Furthermore, it is unlikely that procedures depending partly on human agents will be flawlessly executed for decades or longer. "Trusted digital repositories" certified by credible auditors are implausible as a guarantee of reliable preservation.[21]

- They depend on the doubtful premise that exposing an archive's procedures can persuade its clients that its content deliveries will be authentic.
- Nobody has described an affordable, readily comprehensible, and foolproof procedure whereby an end user can judge the reliability of a repository's deliveries.[22]

---

18 N. Beagrie, *E-Infrastructure for Research: Final Report from the OSI Preservation and Curation Working Group*, Jan. 2007. §3, its "theoretically ideal situation", includes:

> "Long-term threats to preservation and curation of digital information arising from organisational and administrative disruption, funding instability, or lack of clarity surrounding handover of curatorial responsibility will have been addressed. This will have been achieved through development of a network of repositories and services, replication and collaboration between them, longer-term funding frameworks, and definition of different types of repository, roles, and responsibilities over the lifecycle of research information.
>
> "We will have a complex network of trusted digital repositories and policies in place across sectors and disciplines."

See also R.H. McDonald and T.O. Walters, *Sustainability Models for Digital Preservation Federations*, and other papers of a 2007 digital curation symposium at http://www.ils.unc.edu/digccurr2007/program.html.

19 Center for Research Libraries, *Trustworthy Repositories Audit & Certification (TRAC): Criteria and Checklist, 2007*, http://www.crl.edu/PDF/trac.pdf.

20 U. Schwens and H. Liegmann, *Langzeitarchivierung digitaler ressourcen* in *Grundlagen der praktischen Information und Dokumentation,* K.G. Saur, 2004, pp. 567–570.

21 See *Issues of Trust* in M. Seadle and E. Greifeneder, *In archiving we trust: Results from a workshop at Humboldt University in Berlin,* First Monday 13(1), 2008.

- No audit of a digital repository can prove that a sensitive record is uncompromised by improper changes made by its employees[23] or hackers[24] many years before it is accessed.[25] What about bribery to persuade repository clerks to alter critical records? What about rogue governments?
- Repository procedures do not protect network transmissions against improper modifications.
- The trusted digital repository literature does not pretend to deal with document readability.

The custodians of cultural and scholarly collections might argue the improbability of willful falsification of their holdings. None has done so. Nor have any suggested that the archival procedures they propose are intended only for low-risk collections. The public will want universal procedures good enough for financial, juridical, political, and diplomatic records.

Durable Digital Objects

How to achieve reliable future access becomes obvious if one shifts attention from an archetypical archive to an archetypical saved object. As with diplomatic paper, trustworthiness can be ensured by fortifying each document

so that improper modification is readily discovered. This can be achieved by binding a document's parts firmly to one another and sealing the whole with a signature that is difficult to forge.[16]

Several years ago we published a long-term digital preservation solution based on a "trustworthy digital object" (TDO) conforming to Figure 3. TDOs are compatible with interchange conventions described by other authors[26] and will be compatible with metadata standards still being discussed.[27] The scheme makes the authenticity any TDO reliably testable and ensures that eventual users will be able to render or otherwise use its contents. Implementation is readily partitioned into almost independent components:

I. Content servers that store packaged works, and that provide search and access services.

II. Replication mechanisms that protect against the loss of the last remaining copy of any work.

III. Schema for packaging a work together with metadata that includes provenance assertion and reliable linking of related works, ontologies, rendering software, and package pieces with one another.

IV. Standard bibliographic metadata and topic-specific ontologies defined, standardized, and maintained by professional communities.

V. A bit-string encoding scheme to represent each content blob in language insensitive to irrelevant and ephemeral aspects of computing environments.

To prepare the TDO that represents a work, an editor causes conversion of each content bit-string into a durably intelligible representation and collects the results, together with standardized metadata, to become the TDO payload. In addition to its payload, each TDO has a protection block into which a human editor records metadata and relationships among its parts and with other objects. The final construction step, executed at a human agent's command, is to seal all these pieces within a single bit-string with a *message authentication code*. In a valid TDO representing some version of an object:

---

22 Ross (loc. cit.) suggests that an archive user can "determine whether a digital object is what it purports to be … only if institutions have adequately and transparently documented the processes of digital entity ingest, management and delivery." How many archive users will have the knowledge, ability, and time to do this?

23 An RSA Security survey (http://newsletter.varbusiness.com/cgi-bin4/DM/y/eBF2G0GkrVF0Elt0FhfX0ED) found many security problems originating in employee carelessness or ignorance.

24 Tools for Internet exploits are widely publicized. For instance, see *SANS Top-20 Internet Security Attack Targets*, 2006, https://www2.sans.org/top20/.

25 H.R. Tibbo in a 15th Oct. 2007 posting to the *MOIMS-Repository Audit and Certification* blog (moims-rac@mailman.ccsds.org), writes, "What is the purpose of [the TRAC] standard? … Even the highest level of certification will not ensure digital longevity and authenticity, any more than best practices in analog repositories will ensure that no objects go missing or that none are defaced in some way."

26 Open Archives Initiative, *Compound Information Objects: the OAI-ORE Perspective,* 2007, http://www.openarchives.org/ore/documents/CompoundObjects-200705.html.

27 R. Guenther, *Using Metadata Standards in Digital Libraries: Implementing METS, MODS, PREMIS and MIX*, ALA Annual Meeting, 2007.

- The bit-string set that represents the version is XML-packaged with registered schema.
- These bit-strings and metadata are encoded to be platform-independent and durably intelligible.
- The metadata include identifiers for the version and for the set of versions of the work.
- The package includes or links reliably to all metadata needed for interpretation and as evidence.
- All these contents are packaged as a single bit-string sealed using cryptographic certificates based on public key message authentication.
- Each cryptographic certificate is authenticated by a recursive certificate chain.

By these means any information whatsoever can be represented for later reliable use. TDO structure and infrastructure will allow consumers' authenticity testing to whatever level they choose. And representing every content blob using either durable standards or Lorie's Universal Virtual Machine[3] will ensure permanent interpretability.

A Critique of DPC Literature

> "Subject-based and institutional digital repositories are increasingly being hailed as the preferred means for safeguarding the future accessibility of digital information." Nestor Newsletter[28]

Apparently archivists see their institutions as the key to digital preservation. Does any other community see it this way?

Scientists and engineers work, think, and write differently than do graduates of university liberal arts faculties. I could not help but notice this as I read digital preservation papers written by archivists. Compared to what referees of top scientific and engineering periodicals (such as the publications of ACM, IEEE, Am. Phys. Soc., …) demand, this literature exhibits striking weaknesses:

- The cumulative number of words discussing the nature of the challenge, often without saying much that has not been said many times before, far exceeds the number of words describing solutions.
- Papers rarely identify what novel content they offer, in contrast to what they include to provide context.
- DPC literature exhibits surprisingly little attention to know-how originating outside the small DPC community. Its authors seem not to have noticed applicable commercial tools, such as almost automatic replicated storage hardware.[29]
- DPC literature includes surprisingly little analysis of end user needs and preferences.
- DPC papers rarely partition the issues and solution proposals into topics that can be handled with only modest interactions. For instance, how to avoid losing bit-strings has little interaction with how to create representations that will be intelligible century from now.
- There is little evidence that archivists foster effective partnerships with software engineers, even though techniques for digitally represented information originate mostly in that community.

Scope: In recent years, DPC literature has narrowed to attention only to cultural and academic works, paying little attention to content of business and other practical interest.[30] It is unsurprising that almost nobody outside this small community much cares to support it.

Holism vs. atomism: Arguably, the DPC difficulties with digital preservation comes from its failure to apply scientific methodology. The depth and pervasiveness of different thinking styles were strongly suggested to me by a 1996 debate on *Documents in the Digital Culture.* The organizers had structured the panel as a *Two Cultures* debate, with four social scientists and liberal arts representatives facing four scientists and engineers.

Each social scientist began along the lines, "My scientific colleague talked about …, a topic for which we must consider the relationship with …, which itself cannot be understood without [the following broad

---

[28] Nestor Newsletter 13/2007, available at http://nestor.sub.uni-goettingen.de/newsletter/index.php?lang=en.

[29] This is well described in trade periodicals, such as *Information Week.*

[30] Contrast D. Bearman, *Reality and Chimeras in the Preservation of Electronic Records*, D-Lib Magazine 5(4), 1999. See also http://www.archimuse.com/papers/200711-bearman-brazil/brazil.e-records.2007.pdf

context]." The more the speaker progressed from a narrow topic to a very broad spectrum, the more discomfort we saw among the scientists, who fought down urges to interrupt the speaker's thrust.

The style of each scientist was along the lines, "The previous speaker dealt with …, a topic too broad for me to say anything specific about. I'll deal with [thus and such] a small piece." Implicit in this was confidence that other contributors would address and integrate other pieces, possibly not until years later, and that the whole would amount to a large step forward. The further the speaker progressed towards solving his problem segment, the more discomfort we saw among the social scientists.

Proper research topics: Different communities seem to have different notions of worthwhile research. If a computer scientist can describe how to satisfy a service requirement, he would say it is not a proper research topic. In contrast, the U.S. NDIIPP plan[31] reflects a common view that a research topic exists for any information management need unsupported by available software.

In IBM Research corridors in the 1980s, the boundary between research and practical engineering was called "SMOP"—"a simple (or small) matter of programming." This did not necessarily mean that the task being discussed was either uncomplicated or inexpensive. Instead it meant that computer scientists knew answers to its difficult questions, allowing most of the work to be passed to software developers.

Consensus and criticism: Experiencing 1990s digital library workshops peopled mostly by research librarians gave me the impression that librarians value consensus[32] much more highly than do scientists, and are uncomfortable with the kind of criticism that scientists value.[33] In fact, I was quietly taken aside and asked not to voice disagreements with librarians' proposals—proposals that seemed to me unworkable for reasons similar to some expressed above and in a review of the NDIIPP program.[34]

Too much consensus[35] is illustrated by the many published assertions that their authors' institutions conform to the OAIS description.[15] Since OAIS is primarily a definition of what it means to be a digital archive—a definition constructed by inspecting institutions similar to those now insisting they conform, these articles convey almost nothing new. Evaluation of its working methods in terms of the OAIS standard might be a useful internal exercise for a repository institution, but is helpful to almost nobody else—certainly not to information producers and consumers who are the primary archival clients. To engineers, publication about bureaucratic internals is surprising; instead, technologists are likely to feel that "how I run my shop is no outsider's business, as long as it provides excellent service."

This is not to deprecate the value of consensus for appropriate issues. An example of essential consensus is that needed to achieve digital interchange standards.[26]

Bureacratic vs. "free market" processes: One cannot read DPC publications without noticing that this community seems comfortable with bureaucratic procedures, and distinctly uncomfortable with "free market" ways of making things happen.[36] Seadle[21] illustrates this in its comments on DPC attitudes toward interoperability standards. The DPC is correct to doubt the longevity of standards it might foster itself; it is simply too small a community to establish wide-spread use of its consensus procedures. Furthermore, for a standard to become durably entrenched, it must satisfy a pervasive day-to-day need. A special standard for archiving would be problematical. An effective approach for DPC would include it participating in and influencing broadly based standards initiatives.

---

31 Library of Congress, *Plan for the National Digital Information Infrastructure and Preservation Program*, 2002.

32 For instance, notice the allusion to consensus in the opening quotation from ECDL 2007 (footnote 1).

33 See the discussion of "falsifiability" in K.R. Popper, *The Logic of Scientific Discovery*, Anchor Press, 1959.

34 H.M. Gladney, *Digital Preservation in a National Context: Questions and Views of an NDIIPP Outsider*, D-Lib Magazine 13(1/2), 2007.

35 D.B. Marcum, *Too Much Consensus*, CLIR Issues 18, 2000.

36 This bias is evident in European Commission funding views. See Carlos Oliveira, *Digital Curation and Preservation: Funders' Perspectives,* 2007 at http://www.ils.unc.edu/digccurr2007/slides/oliveira_slides_3-1.pdf.

Marginalization of professional archivists: The above problems seem symptomatic of wider risks for the community represented by DPC authors: creeping marginalization of their professions.

Twenty years ago my first tool for finding information was research library catalogs; today it is a Google search of on-line resources. Although these include digital versions of university library catalogs, most useful "hits" seem to come from other sources. Today, several projects are rapidly creating digital versions of massive book collections. Within limits dictated by copyright law, their output will replace my visiting nearby libraries. Of course, a few research librarians are helping achieve this convenience. However, it will reduce our future need for help from their colleagues. Simultaneously, work on search semantics will reduce how long it takes me to find the material I want.

These changes, and others that I cannot anticipate, are inevitable. The economic circumstances have eliminated entire professions, such as stenography. The only at-risk professions that survive are those whose members invent replacement services for which they are uniquely qualified. The apparent passivity of the library community is noteworthy.[37]

Discussion

Digital longevity thinking occurs in unrelated patches. Most of the world pays little attention to deterioration of its digital assets. Computer scientists seem uninterested in durability issues, perhaps because they regard the solution components as obvious. A few engineers work on commercial records whose retention periods are usually less than 30 years. These engineers pay little attention to cultural and scholarly literature because its custodians do not present themselves as customers. And, as suggested above, the DPC thinks in terms of repositories extending the services of their institutions. Little evidence suggests that any of these communities pays attention to the work of any other.

The biggest social barriers to progress seem to be market barriers. Private sector enterprises that might benefit from LDP are preoccupied with short-term concerns. Research libraries, archives, and other cultural institutions do not present themselves to technology providers in the guise of customers, but as supplicants for free technology and charitable support. Free market economies, of course, provide them little alternative. However, they have not much considered indirect opportunities such as reducing the real expense of LDP by thorough automation, shifting workload from central services to their clients, and burying LDP expense by integrating it into day-to-day content handling.

Document handling is technically complicated because human beings are sensitive to communication nuance. People expect this sensitivity to be reflected in semi-automatic tools. They also expect much more of digital objects than they do of works on paper—more reliable authenticity, rapid linking to context, rapid creation of derivative works, and so on, all delivered so as to minimize how much detail they need to understand. Some DPC members might be uncomfortable with the technical depth required to understand that the TDO solution is correct and complete. Unfortunately for them, much of the solution to any digital information problem is applied computer science.

The digital preservation community (DPC) has been seeking practical methods of ensuring durable usefulness of digital documents—methods that will handle immense numbers of objects and deliveries. Its search space has been surprisingly narrow—processes within archival institution perimeters and for persuading clients that archival deliveries are trustworthy. Many facts—the number of digital objects, the number of authors, the speed of information creation and dissemination, the expectations of citizens, the cost trends of technology, relative skills of different communities, and so on—suggest shifting as much as possible of the responsibility from repository institutions to those who are served—information producers and information consumers. Repository procedures alone cannot accomplish what's needed. No plausible investment in repository institutions would change this fact.

It seems that archivists are mostly interested in methods for repository management, a topic of scant interest to information producers and consumers. Instead, consumers' interests focus on finding documents that they trust and can use more or less as their producers intended.

---

[37] A cruel, but apt, joke asks, "What are the three kinds of people in the world?" The answer is, "People who make things happen. People who watch things happen. And people who, when they wake up, ask, 'What happened?'"

Conclusions

> Everything under the sun has been said before. However, since nobody listened … – attributed to André Gide

Ross summarizes preservation interests of archivists with "after more than twenty years of research … the actual theories, methods and technologies that can either foster or ensure digital longevity remain startlingly limited."[1] Many years work without developing its own solution should persuade the DPC to search outside its institutional perimeters.

A solution exists, but is ignored by DPC authors—a perplexing situation. This paper suggests that inattention across the boundary between cultural custodians and software engineers has clouded the distinction between what is not known and what has merely not been implemented to please archivists—an example of cultural dissonance Snow described half a century ago. That a community dedicated to ensuring ready availability of research information seems itself to ignore pertinent publications is more than slightly ironic. Technologists have an unflattering description of such behavior; they call it the "not invented here" syndrome. They mention it so often that they use an acronym: "NIH".

Archivists and research librarians would more rapidly achieve their preservation objectives by ending their near-isolation from outsiders' lines of thinking and by participating towards more broadly applicable solutions than those suitable only for scholarly and cultural publications. In fact, such changes might be critical towards their avoiding creeping marginalization of their professions. Absent their vigorous efforts to reinvent their working methods and roles in the world, their professions' futures seem inevitable and other than they might want.

We need to avoid what this paper has argued are procedures that are impractical because they are susceptible to human weaknesses and because they cannot keep up with the worldwide information growth. A change of rhetoric will help us focus on what is truly wanted and also feasible. **Instead of seeking digital preservation, we should learn to create durable digital objects.**

Acknowledgements

John Bennett and John Swinden reviewed draft manuscripts and made suggestions incorporated in the current version.