

Digital Preservation in a National Context

Questions and Views of an Outsider

[H.M. Gladney](#)

Independent Consultant

<hgladney@gmail.com>

(This Opinion piece presents the opinions of the author. It does not necessarily reflect the views of D-Lib Magazine, its publisher, the Corporation for National Research Initiatives, or its sponsor.)

INTRODUCTION

Repository institution representatives have expressed urgency for preserving authentic digital works. The needs they express will expand to include those of businesses wanting safeguards against fraud, attorneys arguing cases based on the probative value of digital documents, and our own dependencies on personal medical records.

"..., the problem of digital preservation is urgent; can be addressed only by a distributed, decentralized, and networked infrastructure; and collaboration among all stakeholders to share the responsibility for the fate of digital culture can and should be catalyzed by the Library of Congress ..." [1]

Among international initiatives focused on preserving digitally represented information, the U.S. NDIIPP [2] is the largest concerted effort. Recent articles [3, 4, 5] inform readers about NDIIPP history, remind us that the funded program has reached its mid-point, and suggest that it is time for evaluation that might lead to course changes.

"meetings established ... areas of consensus: on the need for the national preservation initiative, ... for a distributed or decentralized solution, ... for more research into the technologies for digital preservation, and recognition that technology is an important part of the solution but not the most important." [1]

Most authors contributing to digital preservation literature belong to the cultural heritage community, a loosely knit group drawn mostly from librarians, archivists, and museum curators educated in the liberal arts and social sciences. Among these the American participants are a closely knit group funded by the NDIIPP and related U.S. Government agencies. This community values consensus highly [6] and has achieved remarkable homogeneity of scope and approaches.

In contrast, the author is a scientist and engineer, and an outsider to the communities mentioned. This leads to distinctions reflected in the content and style of what follows, such as its "divide and conquer" approach – called "partitioning" throughout this article – important for achieving technical progress.

Scope and Style to Discuss Digital Preservation

The current article offers a critical analysis of NDIIPP technical progress. As such, it emphasizes apparent shortfalls because there would be little point in repeating success stories already published. It further emphasizes technical requirements that are not yet satisfactorily resolved in NDIIPP-funded [7] and related literature. It anchors its topics by quoting objectives found in the planning documents. Recent publications suggest that today's NDIIPP thinking is still well represented by its 2002 planning documents [8].

This article focuses on preservation technology, partly because a careful discussion of the entire NDIIPP plan would be impossible in a short article. It does not address issues of repository organization and management, intellectual property rights, or education and training for digital curators, except indirectly when effective application of technology can mitigate non-technical problems. The aspects not discussed all depend on subjective choices – matters of taste and judgment that are more difficult to resolve than technical matters, especially when their effective resolution requires community consensus.

Until archiving is partitioned into almost independent components, tasks that can be defined by objectively described rules are hopelessly confounded with subjective tasks that require human creativity, judgment, and taste. Many engineers favor "divide and conquer", designing no mechanism or process until it is segregated from the overall task sufficiently so that it can be treated with minimal, well-defined interfaces with other components.

The preservation approach reported here resolves confusions by careful attention to the meanings of critical words and by modeling human communication mediated by machinery. The analysis is necessarily painstaking – perhaps even tedious.

Selection of what is to be saved, including choices among versions of any work, are outside this article's scope. Audio/video performances pose a challenge not evident for other data – existing fixed versions might not be suitable for long-term preservation. This is not a preservation challenge in the sense of mitigating the effects of obsolete technology. However, disentangling subjective and objective factors is needed to generate archival versions. Choosing performance records is therefore discussed briefly. For other genres, the article discusses information preservation action only for already-selected works.

The article emphasizes matters of principle, as contrasted to software design, implementation, and deployment. This style, while typical of the computer science literature, is different from that of most of the preservation literature. The latter seems to treat solutions that have not yet been widely deployed and integrated into everyday library operations as R&D topics [9].

Preservation is only a small part of what NDIIPP calls "archiving". In contrast to the situation for preservation methodology, aspects of digital repository technology needed to capture, organize, and provide access to digital documents are no longer research topics, being demonstrated by many satisfactory digital library services.

In this paper, preservation is defined to be "only the additional activities needed to protect information from effects of media deterioration, technology obsolescence, and fading human recall."

What about Digital Repositories?

The NDIIPP technical architecture description [10] is centered on the Open Archival Information Systems (OAIS) idea of what it means to be a repository. Following the OAIS reference model, it identifies essential archiving processes, but says surprisingly little either about their precise functional details or how to accomplish them.

The Figure 1 model, a functional elaboration of the OAIS repository depiction [11], is helpful towards partitioning essential mechanisms into relatively independent components whose functionality can then be analyzed to any level of detail needed. Its structure is mostly forced by practical requirements. For instance, replication management must use access control to enforce rules about what can be backed up in administratively independent repositories.

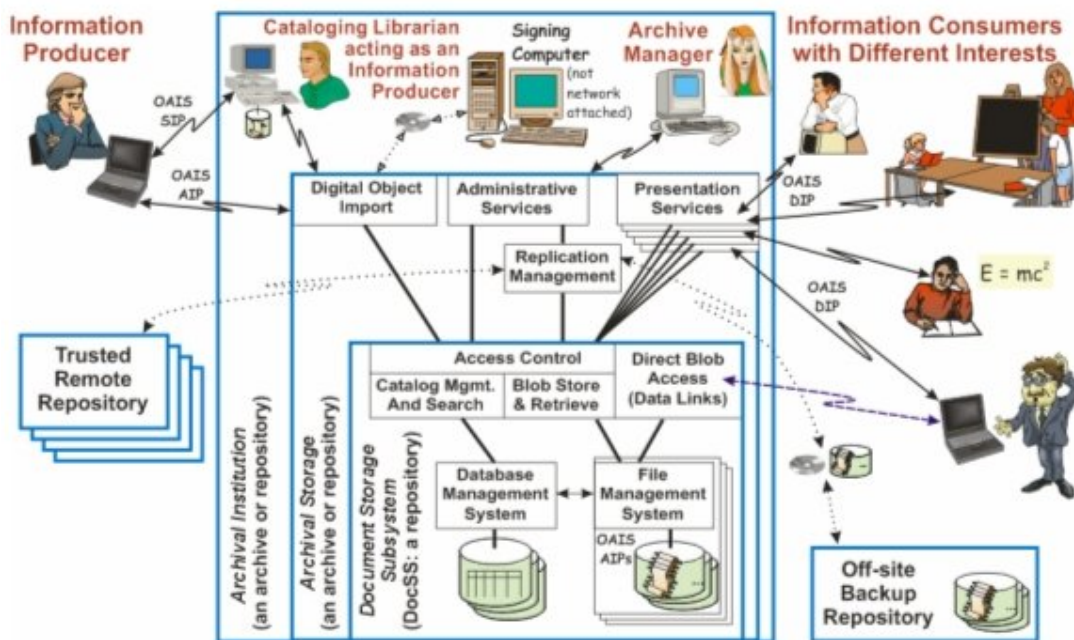


Figure 1: Nested digital repositories and human roles in their management and exploitation

As the figure suggests, the words 'archive' and 'repository' are ambiguous; various authors use the terms differently. The figure's largest box depicts the digital library services of an Archival Institution, which is what OAIS and NDIIPP seem to mean by "a repository" or "an archive". Their documentation mostly discusses what is inside this largest box without being inside the figure's next largest box – Archival Storage.

The core Document Storage Subsystem level is today available in commodity products. Its interface to higher level software is described by a standard [12] that is just beginning to be implemented in content management offerings. That it is still an area for significant innovation is evidenced by IBM's latest DB2 release, which integrates XML objects into the Database Management System. This might interest libraries because it can enable high performance data mining of an archive's content.

The software that differentiates among today's hundred-odd digital library offerings implements local tailoring within Figure 1's Archival Storage box. With few exceptions, user access to the Document Storage Subsystem is mediated by this box. It is a shared resource

available similarly to every user, repository administrators as well as end users, except as limited by access control rules it itself enforces.

The structure sketched in the prior two paragraphs suggests that nearly all digital preservation software must execute in client machines depicted by the personal computers in the figure. The exception is bit-string replication. Embedding any other preservation support in the document subsystem level would be unattractive because it would have to be added to every digital library offering used by repositories, with similar behavior in every implementation. While the situation is not so clear-cut at the archival storage level, adding it there would still require disruptive software change. In short, business circumstances and practical difficulties make it unattractive to add preservation support to deployed digital library offerings.

Scope of a National Plan?

The NDIIPP plan calls for action to "identify barriers to and opportunities for building a distributed preservation infrastructure." I have not found as much evidence of progress as I had expected for requirements the NDIIPP plan identifies. Nor do I find the NDIIPP plan displays the breadth citizens might expect of a "National" initiative.

"This plan is to encourage shared responsibility for digital content and to seek national solutions for:

- the continued collection, selection, and organization of the most historically significant cultural materials and the most important information resources, regardless of evolving formats,
- the long-term storage, preservation, and authenticity of those materials, and
- persistent, rights-protected access for the public to the digital heritage of the American people." [\[1\]](#)

What kinds of information should National preservation activities target? What constituencies should a National program serve? NDIIPP literature emphasizes information genres that research libraries have long collected, extended by audio/visual media from entertainment and news. Compared to its careful consideration of the concerns of academics and of curatorial staff, the NDIIPP plan pays surprisingly little attention to helping individual document creators or to relieving obsolescence impacts on the general public. Nor does it consider the possibility of fraud by modification of individual records, a risk important for business, personal, and military records.

What about infrastructure information, such as engineering records for public utilities and transportation? Curiously, NDIIPP technical plans make no reference to business records, even though their typical flow is sufficiently different from that for academic papers to require different handling. Being essential to the social fabric, they would need attention even if NDIIPP scope were limited to public agencies and educational institutions. An example is the office records used to manage the Library of Congress.

What about health and medical records? As a nation, we should finally enable a dream of physicians, the so-called "longitudinal patient record", which would collect every medical

report for a citizen's lifetime [13]. A promising early application domain is the patient record collection of the U.S. Veterans Affairs hospitals [14].

"The formats explored – e-books, e-journals, digital music, digital television, digital video, and Web sites – are complex and present enormous technical challenges ... These new formats redefine genres: even such things as books and journals that seem so stable in the print regime are redefined online.

"[Representations for] ... documents "on the fly" ... A prime example is the Geographic Information System (GIS), which "maps" a response to a query by matching data to spatial coordinates. ... within a decade or two, GIS will supersede the mass production of maps on paper. ... What ... infrastructure ... and user services must be in place to ensure future access to such geo-spatial materials?" [1]

What data formats should a comprehensive preservation program support? A university generates almost every well-known information type, and also many novel types within its research programs. An institutional archive must accommodate every kind of information generated by its constituency, and must be readily extensible to information representations that become important in the future, even though their specifics are unknown today. In particular, technical research generates computer programs among which some surely should be preserved.

"[Stakeholders' interests] ... at best overlap. Just as often they are in conflict ... because they do not share common understandings of the value of that information – for whom, for how long, for what purpose." [3]

What minimal cooperation is required for effective and economical digital preservation while maximizing the freedom of action within each repository organization?

"The information and technology industry ... should be a contributing partner in addressing digital access and preservation issues inherent in the new digital information environment" [15]

"At this time, there is virtually no coordination of preservation efforts between commercial archives, such as those of the record companies, and institutional archives." [1]

"The Library continually sought out interviews and meetings ... with leaders in industry, technology, librarianship and archiving, and scholarship. The interviewees were asked to identify their enterprises' needs for preservation; to identify what roles their organizations could ... play in a distributed preservation infrastructure; ... These conversations ... have deeply informed the planning process," [1]

Do current partnerships across business sectors accomplish what Congress called for? If not, what is needed to persuade private sector participation? NDIIPP authors allude to industry representatives' comments in the kinds of meeting described as evidence of support for a vigorous preservation program. However the social pressure to attend such meetings (if only for business intelligence) is great. Once there, politeness is a great incentive for concurring with expressions of the importance of digital preservation, if only to avoid

unresolvable arguments over social values. In short, kind words are inexpensive, so that oral concurrence without supporting action is unconvincing.

If commercial entities are indeed concerned with long-term preservation, why is there little evidence of such preservation work in the private sector [\[16\]](#)?

DIFFERENT TECHNICAL EMPHASIS

"The digital realm is one of change and uncertainty, and it is likely to remain so for the foreseeable future. Even the most astute businesspeople cannot forecast anything comfortably because change is so rapid that it is too difficult to develop viable business models." [\[1\]](#)

Curators will surely see some information disappear, perhaps including some that they had hoped they made permanently durable. Preservation failures can make themselves known, but successes cannot. Presuming that they have diligently and correctly applied the best known methodology, how can curators achieve peace of mind?

As posed, the question is unlikely to bother many scientists or engineers. Curators need to learn to live with not knowing for sure that they have succeeded. "Not knowing for sure" is unavoidable. However, worrying about it is neither warranted nor healthy.

Scientists do not normally claim sure truth for any law or deduction, but instead hold that any assertion is conditional on no counter-examples being found. Engineers do not claim perfect reliability, but instead design for "good enough" to meet the requirements in view of cost estimates.

Preserved information can be represented with schemes that are most unlikely to change or to be lost. Data processing standards effectively freeze information formats by making noncompliance unattractive for commercial or altruistic reasons. Permanently intelligible information may be built starting with a few simple standard languages and a simple virtual computer definition.

What Would a Preservation Solution Provide?

Digital preservation is a special case of communication – the case in which no information consumer will be able to question information producers on whose work (s)he depends. [Figure 2](#) can remind everyone that information might travel from its producer to its consumer by any of several paths. Participants will want the details of information copies to be independent of transmission channels. Any recipient will also want access to all closely associated information, including links to whatever is needed to render the holding or to validate its authenticity.

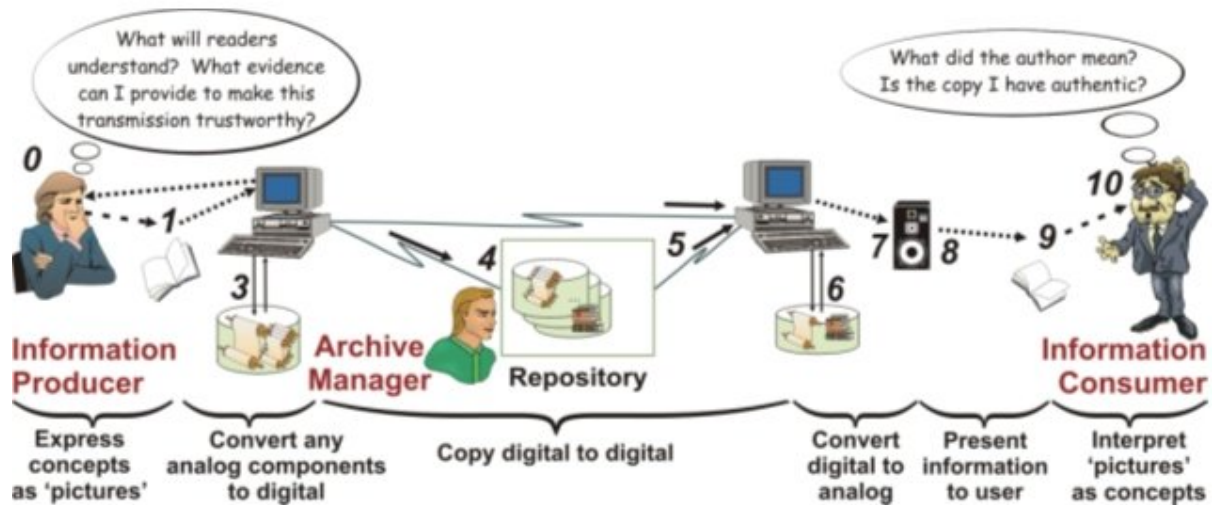


Figure 2: Information interchange and repositories, suggesting preservation to be a form of communication, and identifying challenges faced by human users

What might someone a century from now want of information stored today? That person might be a critic interpreting our writings, a businessman guarding against fraud, an attorney litigating a property dispute, a mechanic maintaining a 40-year old airframe, or a physician consulting your charts of 30 years earlier. For some applications, information consumers will want evidence that information they depend on is authentic [18]. They will be frustrated by received information that they cannot use as they believe was originally intended and possible.

Figure 2 helps us discuss preservation objectives and constraints. In addition to services that digital library offerings and metadata schema already provide, a preservation solution would incorporate methods for:

- Ensuring that each saved bit-string survives as long as somebody might want it;
- Ensuring that readers can find and use any preserved object as its producers intended;
- Providing evidence with which readers can judge information authenticity;
- Integrating preservation support seamlessly with current information services; and
- Hiding technical complexity from end users.

What about a Technical Solution?

"Maintaining the integrity of original content and intent ... is particularly acute with digital morphing capabilities to change images in ways that cannot be detected." [1]

NDIIPP could speed progress towards practical digital preservation by focusing on representation of what I call a "trustworthy digital object" (TDO, as in Figure 3) and on standard network information interchange protocols. The following TDO methodology sketch summarizes technical descriptions in readily accessible periodicals; these details are not repeated here [19].

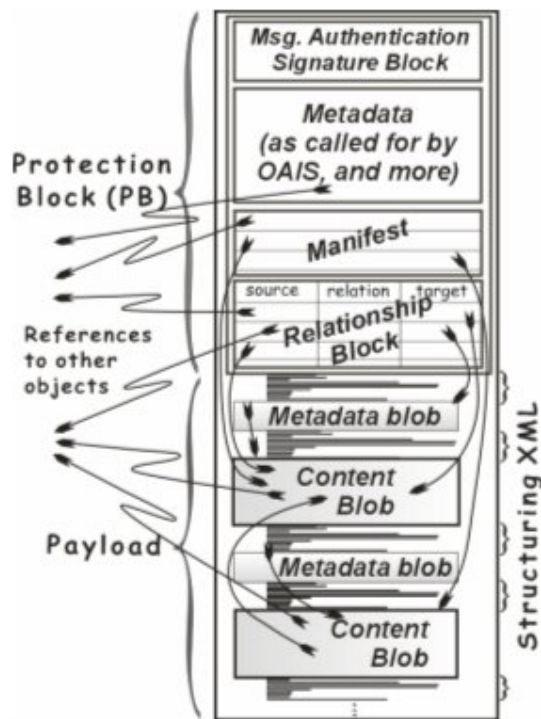


Figure 3: Trustworthy Digital Object (TDO). The structuring XML conforms to the MPEG-21 standard.

- Bit-string retention can be made reliable by replication in independently managed repositories [20].
- Interpretability forever can be provided by combining information representation standards [21] with the use of Turing-complete virtual machine representation of programs [22].
- Authenticity evidence can be saved in message authentication signature blocks. The included cryptographic public keys can themselves be authenticated in a recursive hierarchy whose roots are a small number of widely published public keys [23].
- Hiding technical complexity is a conventional software engineering task for which discussion would be less useful than trial and error with prototypes and competition among software providers.
- All information objects depend on other objects for context, interpretability, and security. Dependencies can and must be represented by links (a.k.a. pointers, a.k.a. references) as suggested by [Figure 3](#).

Among these solution components, the only essential one is reliable saving of bit-strings. Every other preservation activity is merely for user convenience and cost minimization that includes incurring costs today instead of burdening our descendants with much greater costs. Wasteful redundancy among autonomous repositories could be avoided with the assistance of union catalogs that identify preserved objects [24].

Except for bit-string replication, much of the function needed is digital document editing and rendering qualitatively similar to that in PC office suites such as OpenOffice™. It can be implemented either as extensions to such a suite or as independent software packages [25].

What about Dynamic Data?

"People in both the technical and the arts communities urged attention to "ephemera" ... Respondents in film, television, and sound noted that the distinction between publication and ephemera is blurred." [1]

Difficulties with dynamic data are a recurrent theme in the digital preservation literature [26]. Performance information adds two challenges to what other genres pose: the subjective challenge of version selection, and the technical challenge of capturing work not yet prepared to be saved.

Consider preserving a symphony concert broadcast. This seems more difficult than preserving a book. The signals that start as light and sound in a concert hall are repeatedly converted on the way to your television monitor, with tailoring at many stages. Somebody must decide which of many versions to save, and electronics need to copy this version onto tape; this is a mostly technical challenge. Somebody also will have positioned cameras and chosen which of several feeds to transmit at every moment; such decisions are mostly aesthetic. What makes a book seem easier to preserve than a concert is that its corresponding decisions usually have been completed long before a curator decides to make the book part of the historical record. Viewers value the immediacy of television sufficiently to compress its generation, editing, transmission, and viewing processes to be concurrent, with each performance increment occurring in small fractions of a second, rather than sequentially over many months. If someone wants to save a performance version, this must be done at the same time scales.

The decisions alluded to are social choices which are logically prior to actions for long-term preservation. Every kind of inscribed information – digital files, analog recordings, and works printed on paper – experiences dynamic episodes during production, but is static most of the time. The challenge has not been qualitatively changed by new technology. To be eligible for copyright protection, a work must be fixed – written in a stable representation. When a computing system shares a copy, also this is usually a stable version. The candidates for long-term preservation are necessarily such fixed versions.

A "dynamic object" is not a different kind than a "static object". "Dynamic" merely refers to periods when the object described is changing significantly. Any performance can be described by a sequence – a function $P(t)$ of time t – which can readily be fixed. A repeat performance $R(t)$ of an earlier performance $P(t)$ is called authentic if it is a faithful copy except for a constant time-shift, t_{start} , i.e., if $R(t) = P(t - t_{\text{start}})$.

In effect, since *preserving dynamic digital objects introduces no new technical challenge* beyond those for stable documents, no new best practices are needed for preserving them.

How To Make Progress?

"What most distinguishes the digital preservation context from the analog one now in place for ... heritage institutions is the sheer scale of it. It comprehends vastly larger amounts of information, created in a greater variety of formats, and distributed in new venues to a larger and more heterogeneous user base." [1]

How can the scale and diversity of what might merit preservation be handled? How can NDIIPP leverage national resources? The massive scale of Internet resources is fueled by affluence that affords many citizens more discretionary time than ever before, and by private sector competition. Archival libraries cannot manage information growth and client expectations without emulating private sector practices selectively.

Only four tactics can achieve productive scale – task automation [27], avoiding unnecessary work, splitting tasks among independent processors (machines and people), and using "waste cycles" for work that need not be synchronous with human requests. To implement them, the effective focus would be on tools for millions of government officials, businessmen, and citizens, instead of methodology for a few hundred curators. Semi-automatic tools are attractive because many tasks, such as metadata generation [28], intertwine subjective and objective steps.

Technical progress depends on principles and methodology that software engineers take for granted, but that I do not find in the NDIIPP planning documents. I/T methodology starts by partitioning processes and data formats, followed by careful attention to the subjective/objective distinction. Compare our figures with their NDIIPP counterparts. [Figure 1](#) partitions repository processes. [Figure 2](#) partitions the flow of a document. [Figure 3](#) partitions a digital object. Each figure suggests only high-level partitioning; each piece shown is typically partitioned recursively as design and implementation proceed.

An effective partitioning identifies itself as such. The behavior of each partition is readily described entirely in terms of its interfaces. Such descriptions can be precise, complete, and compact. Any partition's internals can be changed without affecting other partitions. Widely used interfaces become formalized as standards which are implicit contracts between independent engineers.

After engineers choose a set of partitions, they ask how to manage each, perhaps doing so in collaboration with potential participants in its implementation and use. They ask: what skills are needed to design it? To implement it? What kind of organization would be its natural "owner", some government department or private sector elements? Which design and implementation tasks are already addressed in available components?

"Companies and archives might wish to pursue collaborative preservation projects ... Such collaborative projects would not be easy to undertake. Record companies today feel bruised by the rampant swapping of music files ... and may be reluctant to authorize the use of master files outside their domains." [1]

Paying attention to the economic and business circumstances of each major partition is important, but is not apparent in NDIIPP planning. For instance, production software is not typically created by research libraries, but rather by private industry or in the open-source community. However, the cultural heritage community does not seem to have adopted an I/T customer role, which might contribute to NDIIPP's difficulty in engaging I/T providers effectively.

No repository user cares how it works internally if it conforms to published, standard interfaces. If document representations and network protocols are standardized, and if repositories publish directories, then each repository can autonomously choose its internal methods.

Partitioning functionality makes it possible to decouple business issues from technical matters. Technology need not dictate policy, but can be used to eliminate or simplify non-technical challenges. Instead of calling for outside audits of "trusted digital repositories" [29], archive managers might welcome freedom from external meddling.

WHY WILL THE RECOMMENDED PRACTICES WORK?

"There is no clear ... set of solutions to meet the challenges of digital preservation. The unpredictability of technological development ...; the business climate ...; and the global political environment ... contribute to the challenge of plotting a course in the face of a wide range of possible futures." [\[1\]](#)

"Each new technological innovation is likely to have unintended and negative impacts on the techniques available for archiving." [\[1\]](#)

A few cultural heritage articles deplore "here and now" emphasis that precludes attention to long-term preservation. TDO methodology provides a "clear set of solutions" for most, if not all, of the technical challenges of digital preservation, even for sensitive business information. Absent a clear demonstration of a systematic failure that cannot readily be repaired, I can only presume the work summarized in the current article is correct.

TDO design specifies document editing and document viewing similar to that of office automation software suites [\[25\]](#). TDOs are XML documents conforming to MPEG-21 and other widely used standards. Using TDOs will cause negligible disruption to repository institutions. Information encoded as a correct TDO will be easily interpretable forever and also testable for its authenticity. Such durability is insensitive to technology changes. These facts make TDO implementation compatible with most repository software without any change to that software. I invite criticisms that suggest my confidence to be unwarranted. The NDIIPP worry that "technology innovation is likely to have ... negative impacts on ... archiving" is effectively bypassed.

For heritage institutions, the strategic choice is among waiting for some years (risking being left behind in a competitive world), learning enough about the technologies to balance early adoption with its risks and costs, or some combination of these.

The first is a "do nothing" strategy that calls for no further comment in the current article. To execute the second, NDIIPP needs to engage the engineering community more effectively than seems to have been the case so far.

The explosive growth of information content has been fostered by practices bent on pleasing people that the private sector considers to be customers. Libraries cannot keep up without identifying which private sector practices are effective for creating tools that please customers and emulating these practices. Part of this would be to recognize that surveys won't help without being coupled with pilot software in a few end user's hands, modifying this software until it truly pleases its users [\[30\]](#).

It would be difficult to overemphasize the value of task partitioning. Its productive consequences include that apparent non-technical problems prove not to be obstacles after all. For instance, any repository institution can autonomously change its software without damaging with its preserved content and without coordinating changes with other repositories, provided that it communicates externally only via published standard interfaces. Other tasks are much simplified. For instance, providing document authenticity evidence replaces auditing to measure repository trustworthiness, a method discussed by many authors, but incapable of protection against felonious document change.

Other tasks seem to be outside the scope of what NDIIPP should fund. For instance, cultural institutions do not and cannot, for the most part, create production-level software. Furthermore, each unanswered challenge can be categorized as technical or non-technical, a question of principle or merely an implementation need, a matter of software development or one of personnel training, independently decidable or requiring negotiations, and so on.

CONCLUSIONS

The U.S. National Digital Information Infrastructure Preservation Program has reached the mid-point of its ten-year funded plan – a time to pause and consider changes of direction.

Today's NDIIPP program is focused on cultural content similar to traditional research library holdings, with little attention to preserving practical information critical to most social infrastructure and of interest to most citizens. I think that taxpayers would expect much more of a program labeled "National", were they to notice NDIIPP and understand its current scope.

During many years of industrial software R&D experience, I saw that business executives consistently demand substantial progress towards agreed objectives within a year of the beginning of any project's funding. From this perspective, NDIIPP technical progress seems very slow. Manifestly successful engineering methodology has hardly been considered, much less tried in prototypes, field tests, or deployments. The best available technical skills and I/T resources have not been effectively tapped.

Compared to their discussion of digital repositories, NDIIPP technical activities have paid little attention to the structure and representation of preservation-worthy digital information – an opportunity not only to solve technical challenges, but also to reduce non-technical obstacles.

What is sad about this situation are the missed opportunities. No technical obstacle impedes the production of convenient and reliable tools for preserving digital information, or deploying infrastructure within which information producers, information consumers, and archive managers could effectively and conveniently exploit those tools. Although some of the software needed is not yet available, providing it would be easy and inexpensive.

Acknowledgements

Many citizens read drafts. In particular John Bennett, Tom Gladney, and John Swinden discussed their detailed change recommendations with me at length. Nevertheless, the responsibility for controversial opinions is mine alone.

Notes and References

1. Library of Congress, *Plan for the National Digital Information Infrastructure and Preservation Program*, October 2002, <<http://www.digitalpreservation.gov/about/planning.html>>. See also the NDIIPP Website, <<http://www.digitalpreservation.gov/>> especially its page about *Partnerships*, <<http://www.digitalpreservation.gov/partners/creative.html>>.

2. Laura E. Campbell, *Envisioning Future Challenges in Networked Information*, JISC/CNI Conference, July 2006, <<http://www.ukoln.ac.uk/events/jisc-cni-2006/presentations/l-campbell.ppt>>.

3. Abby Smith, Distributed Preservation in a National Context, *D-Lib Magazine* 12(6), June 2006, <<http://dx.doi.org/10.1045/june2006-smith>>.

4. Guy Lamolinara, *What if NDIIPP knew what NDIIPP knows?* NDIIPP Website, 2006, <http://www.digitalpreservation.gov/partners/project_mtg010906.html> and <<http://www.digitalpreservation.gov/>>. Also ARL briefing, October 2005, <http://www.arl.org/arl/proceedings/147/LeFurgy_files/LeFurgy.ppt>.

5. Deanna B. Marcum, *The Future of Preservation*, Paris, March 8, 2006, <<http://www.loc.gov/library/reports/paris-speech-preservation.pdf>>.

6. Deanna B. Marcum, Too Much Consensus, *CLIR Issues* 18, Nov./Dec. 2000, <<http://www.clir.org/pubs/issues/issues18.html>>.

7. Library of Congress, *Digital Preservation Technical Infrastructure*, <<http://www.digitalpreservation.gov/technical/aiht.html>> ; see also *D-Lib Magazine* 11(12), <<http://dx.doi.org/10.1045/december2005-contents>>.

8. Loc. cit., footnotes [2](#), [3](#), [4](#), [5](#), and [7](#).

9. What is a worthwhile research topic? The computer science literature avoids discussing what some people call "a simple matter of engineering", leaving such matters to open-source and commercial component specifications and to less formal collegial and trade literature. If a computer scientist can describe how to satisfy a requirement, (s)he would say it is not a valid research topic. In contrast, NDIIPP reflects what seems to be a cultural heritage community criterion that a research topic exists for any need unsupported by available software. NDIIPP planning literature emphasizes establishing collaborative networks. The current article adopts an engineer's view that such networks already exist, and that using them effectively is a matter of training rather than a research or technical issue.

10. *Update to the NDIIPP Architecture: Version 0.2*, 2004, §7, <http://www.digitalpreservation.gov/technical/NDIIPP_v02.pdf>.

11. As suggested by Figure 4.1 of the *OAIS Reference Model*, <<http://public.ccsds.org/publications/archive/650x0b1.pdf>>.

12. JSR 170: *Content Repository for Java™ technology API*, <<http://jcp.org/en/jsr/detail?id=170>>.

13. See the *Michigan Electronic Medical Record Initiative*, <<http://www.memri.us/>>.

14. According to *BusinessWeek*, July 17, 2006, pp.50-6, "The VA has by far the most advanced electronic medical-records system in the U.S. Every VA facility is linked to the national database, and every patient protocol and interaction is recorded."

15. PL 106-554 and Conference Report 106th Congress (H. Report 106-1033), H. R. 4577, Chapter 9.

16. In 2000, IBM Research told me that there was insufficient business case for my joining Raymond Lorie in addressing digital preservation, thereby contributing to my decision to leave that organization. As far as I know, IBM Research has ceased digital preservation research.

Almost coincidentally with the publication submission of the current article, an IBM Israel group announced its participation in a European Sixth Framework programme to preserve cultural and scientific resources. See

<http://www.research.ibm.com/haifa/info/news_ibm_caspar.html>. Is this an exceptional action, or a signal of emerging private sector participation in digital preservation support?

17. This question, differently phrased, came from an NDIIPP participant, suggesting that it represents common unease within the cultural heritage community.

18. H.M. Gladney and J.L. Bennett, What Do We Mean by Authentic? *D-Lib Magazine* 9(7), July 2003, <<http://dx.doi.org/10.1045/july2003-gladney>>.

19. H.M. Gladney, Principles for Digital Preservation, *Comm. ACM* 49(2), 111-116, February 2006 and papers it cites. The key ideas have been available since 2003 in the *Digital Document Quarterly*, <<http://home.pacbell.net/hgladney/ddq.htm>>. Similar object structure has been prototyped in the Los Alamos aDORe service: Jeroen Bekaert and Herbert Van de Sompel, A Standards-based Solution for the Accurate Transfer of Digital Assets, *D-Lib Magazine* 11(6), June 2005, <<http://dx.doi.org/10.1045/june2005-bekaert>>.

20. Vicky Reich and David Rosenthal, LOCKSS: A Permanent Web Publishing and Access System, *D-Lib Magazine* 7(6), June 2001, <<http://dx.doi.org/10.1045/june2001-reich>>.

21. A selection from today's widely used EDP standards suffices. Unicode, XML, and BNF figure prominently among the technologies to be exploited.

A limitation is that documents represented by proprietary formats cannot be preserved without violating copyright privileges. The most prominently difficult case, Microsoft Office® representations, is mitigated by a Microsoft commitment for 2007 support of the Open Document Format standard, <<http://www.oasis-open.org/specs/index.php#opendocumentv1.0>>.

22. The Koninklijke Bibliotheek has deployed an electronic publication deposit system based on an IBM prototype of its Universal Virtual Computer (UVC) mechanism, which is described in the work cited in footnote 19.

23. H.M. Gladney, Trustworthy 100-Year Digital Objects: Evidence After Every Witness is Dead, *ACM Trans. Office Information Systems* 22(3), 406-436, July 2004.

24. The OCLC WorldCat service illustrates the idea, <<http://www.worldcat.org/>>.

25. For a design sketch, see the final Appendix of H.M. Gladney, *Preserving Digital Information*, Springer Verlag, 2007 (in press) <<http://www.springer.com/3-540-37886-3>>.

26. Luciana Duranti, The Long-term Preservation of the Dynamic and Interactive Records of the Arts, Sciences and E-Government: InterPARES 2, *Documents Numerique* 8(1), 1-14, 2004.

27. Miguel Ferreira et al., A Foundation for Automatic Digital Preservation, *Ariadne*, 48, July 2006, <<http://www.ariadne.ac.uk/issue48/ferreira-et-al/>>.

28. Steve Mitchell, Machine-assisted Metadata Generation and New Resource Discovery: Software and Services, *First Monday* 11(8), August 2006, <http://www.firstmonday.org/issues/issue11_8/mitchell/>.

29. Ronald Jantz and Michael J. Giarlo. Digital Preservation: Architecture and Technology for Trusted Digital Repositories, *D-Lib Magazine* 11(6), June 2005, <<http://dx.doi.org/10.1045/june2005-jantz>>.

30. Consider the NDIIPP study at the University of Michigan, *Incentives for Data Producers to Create Archive-Ready Data Sets*, <<http://www.si.umich.edu/incentives/>>. I doubt that such surveys will produce useful and novel insights.

(On February 1, 2007, in Reference 4, the author of "What If NDIIPP Knew What NDIIPP Knows" has been corrected to read Guy Lamolinara. On February 7, 2007, in line links to references 28 and 29 were added.)