

2010

Digital Preservation and Authentic Legal Information

G. Patrick Flanagan

Follow this and additional works at: https://digitalcommons.law.uw.edu/law-lib_borgeson

Recommended Citation

G. Patrick Flanagan, *Digital Preservation and Authentic Legal Information* (2010),
https://digitalcommons.law.uw.edu/law-lib_borgeson/20

This Article is brought to you for free and open access by the Law Librarianship Program at UW Law Digital Commons. It has been accepted for inclusion in Borgeson Papers by an authorized administrator of UW Law Digital Commons. For more information, please contact jafrank@uw.edu.

DIGITAL PRESERVATION AND AUTHENTIC LEGAL INFORMATION

G. Patrick Flanagan^{*}

Information has never been stable. That may be a truism, but it bears pondering. It could serve as a corrective to the belief that the speed-up in technological change has catapulted us into a new age, in which information has spun completely out of control. I would argue that the new information technology should force us to rethink the notion of information itself.¹

INTRODUCTION

Writing and researching about the permanence of digital documents is a quizzical, self-referential activity. I set out to uncover approaches to the problems facing the longevity of authentic legal information. How did I do this? Primarily, I accessed and read electronic documents. My exercise here might very well suffer the same issues raised in the information science and legal literature. Faulty, inconsistent, and potentially inauthentic electronic databases may unduly – however subtly – shade my analysis. For what I’m doing here – a student’s attempt to add to an academic discourse – I’m pretty unconcerned.

The subject of my focus, however, is an entirely different matter. The law has consequences large and small. Legal information, whether in the form of publications or records, impacts our lives in concrete ways. Criminal codes convict. Election rules elect. Construction standards keep roofs from falling in. The law is *important*. Further, although “the law” is a complex political, social, and interpersonal construct, it is ultimately expressed through words.

Digital files distributed through the internet present problems to information generally, and to the law specifically. If we take a skeptical view – as I think we should –

^{*} ©2010 Gerald Patrick Flanagan, II. Submitted in partial fulfillment of the M.L.I.S. Law Librarianship degree at the University of Washington.

¹ ROBERT DARNTON, *THE CASE FOR BOOKS: PAST, PRESENT, AND FUTURE*, at 29 (2009).

how do I know that words I read are the words that were written? This paper is a meditation on that question and search for an answer.

Part I describes the issues of authenticity as they're discussed in the English-language library and information science literature. Finally, Part II describes the importance of authenticity to the law. Part III describes the potential dangers if we fail to adequately address preserving authentic legal information. At each step along the way, we'll see the need for libraries and archives to take an active role in the creation of documents and foster a trustworthy informational environment.

PART I – INFORMATION THEORY AND DIGITAL OBJECTS

A. DIGITAL CONVERGENCE

Libraries provide access to materials. Archives safeguard materials. Museums display materials. Digital technology blurs those traditional roles. As Paul Marty notes, “the topic of the “digital convergence” of libraries, archives, and museums has a lengthy history.”² “Lengthy,” however, only in the context of internet-based distribution of digital materials. Compared to the history of the written word – but a speck in time.

As digital technologies unfold, much of the information literature seeks to translate tried and true theory onto these radically new materials – and rightly so. There's no need to throw everything out the window. But we see traditional roles eroding. Suddenly (in historical terms), traditional roles of access, preservation, and display are interwoven. As I ponder preserving digital materials, I – like many information scholars – soon find myself pressured under the weight of historicity. A large body of scholarship exists concerning the issues of preserving information.

Marc Truitt provides welcome relief when he suggests that,

[p]erhaps we're thinking too big when we speak of
“forever.” Maybe we need to begin by conceptualizing and

² Paul F. Marty, *An Introduction to Digital Convergence: Libraries, Archives, and Museums in the Information Age*, 80 LIBR. Q. 1, 2 (2010) (citing W. B. Rayward, ELECTRONIC INFORMATION AND THE FUNCTIONAL INTEGRATION OF LIBRARIES, MUSEUMS, AND ARCHIVES IN HISTORY AND ELECTRONIC ARTEFACTS 207 (E. Higgs ed., Oxford Univ. Press 1998)).

implementing on a more manageable scale. Or, to adopt a phrase that seemed to become the informal mantra of both this year's PASIG and the immediately preceding iPres meeting, "To get to forever you have to get to five years first."³

Concentrating on attainable goals does indeed make the task seem easier. And perhaps that's a key to success – putting together a string of smaller attainable steps. If we concentrate too hard on the big, "forever" questions too hard we might miss something.

Authenticity of information is something we can't afford to miss. As I discuss below, authenticity is particularly important for legal information. "The user faces a lot of claims of authenticity from people with many different interests, not all of them public spirited."⁴ Nonetheless, developing and maintaining verified and trustworthy records are important in every information context. Archival arts and sciences have long struggled with how to develop systems of trustworthy documents. Heather MacNeil aptly describes the problems presented by digital information:

With non-digital forms of records, continuous custody has been considered sufficient grounds for asserting their authenticity. [...] Authenticity is particularly at risk when records are transmitted across space (that is, when they are sent between persons, systems, or applications) or time (that is, when they are stored offline, or when the hardware or software used to process, communicate, or maintain them is upgraded or replaced). Therefore, in the case of records maintained in electronic systems, the traditional presumption of authenticity must be supported by evidence that a record is what it claims to be and has not been inappropriately modified or corrupted.⁵

Essentially, MacNeil describes the problem of no longer having a physical object embodying information. Authenticity for physical objects can – under most

³ Marc Truitt, *Editorial: Reflections on What We Mean by "Forever"*, INFO. TECH. & LIBR., Dec. 2009 at 150.

⁴ John D. Gregory, *Electronic Legal Records: Pretty Good Authentication?*, 1997 "The Official Version:" A National Summit to Solve the Problems of Authenticating, Preserving and Citing Legal Information in Digital Form, Proceedings – November 20-22, 1997 available at <http://web.archive.org/web/20041011203751/www.callacbd.ca/1997summit/auth-johngregory.html> (last visited May 12, 2010).

⁵ Heather MacNeil, *Providing Grounds for Trust II: The Findings of the Authenticity Task Force of InterPARES*, 54 ARCHIVARIA 24, 28 (Fall 2002).

circumstances – be determined by physical inspection. Or, at least, evidence for a physical object’s authenticity resides in the physical world (Where was it kept? Do the pages appear altered? Are the materials appropriate for what it claims to be?) As George Barnum puts it,

For the most part, if a printed document is in some way altered after its publication, other than the production of another edition or impression, it can be detected. One of the greatest challenges in preserving access to electronic publications is the transitory nature of networked information: the here today and easily changed or gone later today aspect. This inherent instability presents challenges of various kinds: instability of location or address, the ease with which information can simply disappear, and establishment and verification of “officialness” which exists alongside the actual authenticity or integrity of the digital object.⁶

Digital objects obviously have physical underpinnings with issues of custody, access, and alteration but inspecting them won’t yield similar results without technological intervention and interpretation. Evidence for a digital object’s authenticity must come from somewhere outside the object.

B. AUTHENTICITY IS A SOCIAL CONSTRUCT

As MacNeil notes, “research and reflection on the preservation of authentic digital materials has tended to focus on the identification and elaboration of procedural or technological criteria for assessing and protecting the trustworthiness of those resources.”⁷ Moreover, the software industry spends a great deal of effort on developing the tools of trustworthiness. Encryption technologies seek to establish trust in

⁶ George Barnum, *Availability, Access, Authenticity and Persistence: Creating the Environment for Permanent Public Access to Electronic and Government Information*, 19 GOV’T INFO. Q. 37, 39 (2002).

⁷ Heather MacNeil & Bonnie Mak, *Constructions of Authenticity*, 55 LIBR. TRENDS 26, 27 (Summer 2007) (citing INTERPARES, THE LONG-TERM PRESERVATION OF AUTHENTIC ELECTRONIC RECORDS: FINDINGS OF THE INTERPARES PROJECT, (L. Duranti ed., 2005); L. Duranti & Heather MacNeil, *The Protection of the Integrity of Electronic Records: An Overview of the UBC-MAS Research Project*, 42 ARCHIVARIA 46 (1996); H.M. Gladney & J.L. Bennett, *What Do We Mean by Authentic? What’s the Real McCoy?* 9 D-Lib Magazine (July/August 2003) available at <http://www.dlib.org/dlib/july03/gladney/07gladney.html> (last visited May 12, 2010)).

communication. Digital rights management (DRM) schemes seek to establish trust in intellectual property markets.

But people don't trust technology per se. More precisely, any trust in a specific technology is implicitly trust in its creator. As Jacobs, Jacobs, and Yeo put it:

Ultimately the problem of authenticity is a social problem, not a technological one. By using technical tools (e.g., the public key infrastructure-PKI), creators of documents can provide a way for users to verify, through a third party, that a document is what it purports to be and has not been altered. While that is good, it does not solve the problem, but shifts the trust from the party that delivered the document to a third party. The user must still trust that third party.⁸

For an example, we can turn to the State of Alabama's efforts at creating digital access and preservation protocols. Rickey Best explained the project in a 2009 article of *Southeastern Librarian*. When they shopped for archival tools they ultimately chose OCLC's Digital Archive, in part, because "the stability of OCLC as an organization gave confidence that digital images archived there would be safe."⁹ Functional issues played a role, but resolving the issues of preservation relied on the social standing of OCLC. Organizational stability is vital because digital preservation is an ongoing, active process. A conservator of digital objects that can't "keep the lights on" jeopardizes the collection. Furthermore, stable organizational structure gives confidence that controls are in place to minimize tinkering with archival tools.

MacNeil further explores this reliance on social structure for authenticity, by drawing an example from evidence law. As she explains:

[T]he notion of an authentic record in evidence law (in the sense of a trustworthy statement of facts), like the notion of an authentic art work or literary text, is shaped to a considerable degree within a specific social and institutional framework; its authenticity does not inhere in the record itself but is actively constructed in accordance

⁸ James A. Jacobs, James R. Jacobs, & Shinjoung Yeo, *Government Information in the Digital Age: The Once and Future Federal Depository Library Program*, 31 J. ACAD. LIBRARIANSHIP 198, 202 (2005).

⁹ Rickey Best, *Preserving and Providing Access to Digital State Publications in Alabama: A Case of Cooperation*, 57 SOUTHEASTERN LIBR. 1, 5 (Fall 2009) (citing Theodore R. SCHELLENBERG, THE MANAGEMENT OF ARCHIVES, 41-45 (1965)).

with the theoretical and methodological assumptions operative within that framework.¹⁰

Evidence law, generally, relies on the framework of an adversarial legal process. Through that process, litigants develop authenticity through testimonial assertion. The judge and fact finder draw conclusions about authenticity informed by the litigants' framing of the evidence. These conclusions exist in a social framework – both the microcosm of the litigation itself and the community at large. Similarly, as we evaluate the digital objects in a repository, we must place the archive in a social framework to evaluate authenticity of the items collected.

Developing a model for that social framework must be principled precisely because digital objects are unstable. Like any other, money plays a role in the social framework for the preservation for digital materials. As Richard Johnson observes,

Digital materials must be the object of appropriate preservation. Preservation activities require the development of standards and best practices as well as models for sustainable funding to guarantee long term commitment to these materials.¹¹

Whatever trust we give to a digital archive based on their position in the social framework, must necessarily incorporate their ability to maintain the ongoing costs associated with digital preservation. “The rush to develop the technological processes necessary to preserve authentic electronic records appears to have come at the expense of first addressing cost and policy.”¹²

¹⁰ MacNeil, *supra* note 7, at 43.

¹¹ Richard K. Johnson, *In Google's Broad Wake: Taking Responsibility for Shaping the Global Digital Library*, 7 ARL 1 (2007).

¹² Shelby Sanett, *Toward Developing a Framework of Cost Elements for Preserving Authentic Electronic Records into Perpetuity*, 63 C. & RES. LIBR. 388, 402 (2002).

C. LIBRARIANS AND ARCHIVISTS EFFECT DIGITAL OBJECTS

Librarians strive for impartiality and fairness. Consider one of S. R. Ranganathan's five rules of librarianship: "to every reader his book."¹³ Every reader should have access to the writings he requires and – by implication – the library should promote that access without criticism. If we extrapolate that rule across the vagaries of time, the library—or digital archive—promotes the stability of access for when "the reader" seeks his book. However, "[t]he digital environment resists the imposition of traditional structures of stability because it dramatically accelerates the process of change. It is precisely this dynamic characteristic of digital technology that has been the source of anxiety for librarians and archivists."¹⁴ If we give in to that anxiety, we might reflexively resist efforts toward digital preservation because librarians hold tight to notions of impartiality.

Better, as MacNeil suggests, librarianship should acknowledge that

[l]ibrarians and archivists are not neutral preservers of digital resources, but active agents in the reconstitution of these resources over time. The decisions that they make about preservation determine how the materials will be accessed, read, and understood by users. For that reason, their decisions should be made known: Which intentions and which meanings have been privileged and preserved, and for what reason?¹⁵

And hopefully, this attitude becomes less problematic if we continue to ponder Darnton's truism: "information has never been stable."¹⁶ Paper pages rot. Sources get misquoted. A misshelved volume is as good as lost. The challenge is to incorporate digital materials into the fold and accept the greater responsibility they demand.

Any models for digital preservation must acknowledge the greater role that those outside the library must play. "It is clear that the preservation of authentic electronic records is a responsibility shared by record creators and preservers and, in many cases,

¹³ See also, Morris Cohen, *Toward a Philosophy of Law Librarianship*, 64 L. LIBR. J. 1 (1971).

¹⁴ MacNeil, *supra* note 7, at 44.

¹⁵ *Id.*, at 47.

¹⁶ Darnton, *supra* note 1.

the creator and preserver will be the same person[.]”¹⁷ For legal information—a particularly government information—the creator might be an organization with complex rules and standards informing the process. In creating digital preservation models and systems, libraries must reach out to creators and embrace an active role in the creation of materials as well as preservation.

PART II – THE IMPORTANCE OF AUTHENTICITY TO THE LAW

With rare exceptions, the aphorism holds true: ignorance of the law is no defense. Worse, relying on bad law or inaccurate statements about the law can have serious consequences. The citizenry must have access to the law and have some way to rely on it as authentic. The digital environment presents sources of errors and corruption unlike non-digital sources. Incorporating these digital sources to the overarching legal framework is necessary. The United States’ (and others’) tradition of an informed citizenry forms the basis for the legitimate rule of law.

The notion of one centralized and easily located source of legal and government information is seductive. Simplifying the process of finding the text of the law and legal materials would certainly give us a sense of satisfaction with legal research. That task, however, presents authenticity problems. The bibliographic universe of the legal materials is immense. Even if we limit ourselves to primary materials, the task of compiling them and making them uniformly accessible is daunting. Even so, some feel that they are up to the task.¹⁸

Although a centralized archive would seem to solve problems of authenticity, it creates a single point of failure. If part of the archive fails, becomes corrupted, or lacks desirable systemic elements, then those errors effect the whole information landscape. The internet makes it easy to propagate errors and unreliable information. However, a distributed network of trusted sources creates a systems of checks-and-balances and creates confidence in the authenticity of information.

¹⁷ Heather MacNeil, *Providing Grounds for Trust II: The Findings of the Authenticity Task Force of InterPARES*, 54 *ARCHIVARIA* 24, 39 (Fall 2002).

¹⁸ See e.g., Public.Resource.Org, *Law.Gov: America’s Operating System, Open Source* (Feb. 19, 2010), <http://resource.org/law.gov/index.html>.

A. THE WORD OF THE LAW

Claire Germain gets right to the heart of the importance of authenticity in the law when she observes:

“Because in every country of the world, in an environment where online sources have replaced official print legal information, citizens need to trust the ‘official word of the law’ in the same way that they trust print information. Since the digital medium is vulnerable to errors in management and control, corruption, and tampering, it is of utmost importance to make the digital information both official and authentic.”¹⁹

In a way, she appeals broadly to textualism. When analyzing a statute, the textualist would argue that we ought to strictly adhere to the plain meaning of the text of the statute. Going beyond that (so the argument goes) invites us to color our interpretation with ideas that weren’t promulgated by the legislature. But in either event – whether we agree or disagree with the textualist – intellectual honesty requires us to construct our analysis on *some* text. We might consult legislative histories. We might consult agency materials. We might consult news reports – or any number of texts that we could consider legal information. Whatever the source, we must trust those written words we consult.

Furthermore, nearly any source of information can find its way into the law – especially in common law systems like the United States. As a judge incorporates a source into an opinion, that source carries weight under the rules of precedence. With digital information we must be concerned with its authenticity and permanence. “Even when cited sources remain on the Internet unchanged, [...] some question whether appellate courts have overstepped their roles when they use Internet sources to bolster their understanding or interpretation of the case facts at issue.”²⁰

¹⁹ Claire M. Germain, *Legal Information Management in a Global and digital Age: Revolution and Tradition*, 35 INT’L J. LEGAL INFO. 134, 154 (2007).

²⁰ Coleen M. Barger, *On the Internet, Nobody Knows You’re a Judge: Appellate Courts’ Use of Internet Materials*, 4 J. APP. PRAC. & PROCESS 417, 435 (2002).

As to authenticity, Barger makes the argument that courts are under a special duty to scrutinize the reliability and authenticity of sources they consult. She notes:

It is one thing for law review articles to rely too heavily on questionable sources. It is quite another for courts to do so. The case law handed down by appellate courts, for the published opinions at least, is the primary authority that others will rely upon tomorrow. Even dicta and non-majority opinions can provide the inspiration for someone's good faith argument to change the law at a later date. Case law authority is built on the foundations laid down by judicial authors. Those foundations deserve to be solid and visible to those who will later learn from and add to that body of law.²¹

Thus, the vagaries of information as it appears on the internet get incorporated into the law. Moreover, the information is not necessarily static. As Barger further observes, a website's content could change or move. The URL that a court cites to could change. "[O]ne could argue that the reader of the opinion, using her own research skills, simply ought to conduct a little Internet research to find the misplaced site. Perhaps, but that argument ignores the basic function of a citation-to permit readers to easily locate the precise source referenced by an author."²²

To their credit, the federal judiciary has begun to address issues of permanence when consulting digital materials, validating Barger's position. The Judicial Conference has issued a series of "suggested practices" to assist courts in the use of Internet materials in opinions.

The guidelines suggest that, if a webpage is cited, chambers staff preserve the citation by downloading a copy of the site's page and filing it as an attachment to the judicial opinion in the Judiciary's Case Management/Electronic Case Files System. The attachment, like the opinion, would be retrievable on a non-fee basis through the Public Access to Court Electronic Records system.²³

²¹ *Id.* at 447.

²² *Id.* at 439.

²³ Administrative Office of the U.S. Courts Office of Public Affairs, Third Branch Newsletter, July 2009, http://www.uscourts.gov/ttb/2009-07/article09.cfm?WT.cg_n=TTB&WT.cg_s=July09_article09_newsroom

At least two of the circuit courts, the fifth and the ninth, have already gone further and made those materials accessible to the public through their law libraries.²⁴

B. AN INFORMED CITIZENRY

Much like criminal defendants have a right to know the charges leveled against them, the citizenry ought to have appropriate access to the law and legal materials. The movement that ultimately led to the Freedom of Information Act in the 1960's illustrates that point. Among other concerns leading to its enactment, "[t]he Washington law firms and the organized bar wanted a requirement that agency "secret law" policies must be codified and published, so as to facilitate the representation of more sophisticated clients in agency adjudicatory proceedings."²⁵ But notions of fairness and freedom dictate against any "secret law" – even for the unsophisticated or unrepresented.

Widespread access to authentic versions of legal materials has long been part of the United States cultural and legal fabric. As Jacobs, Jacobs, and Yeo describe:

In the United States, there are deeply rooted values that a democracy requires an informed citizenry, that government must be accountable to its citizens, and that citizens therefore must have full, free, easy access to information about the activities of their government. These values have led to the creation of the Joint Committee on Printing (JCP), Government Printing Office (GPO), and the Federal Depository Library Program (FDLP) to facilitate this process.²⁶

²⁴ See, Ninth Circuit Library / Websites Cited In Ninth Circuit Opinions, 2009, <http://www.lb9.uscourts.gov/webcites/2009.php>; Fifth Circuit, Archived URLs of the Fifth Circuit, <http://www.lb5.uscourts.gov/Resources/ArchivedURLs/>. It does not appear, however, that the Fifth Circuit has continued the practice beyond 2007.

²⁵ James T. O'Reilly, "Access to Records" Versus "Access to Evil:" Should Disclosure Laws Consider Motives as a Barrier to Records Release?" 12 KAN. J.L. & PUB. POL'Y 559, 561 (2002).

²⁶ James A. Jacobs, James R. Jacobs, & Shinjoung Yeo, *Government Information in the Digital Age: The Once and Future Federal Depository Library Program*, 31 J. ACAD. LIBRARIANSHIP 198, 198 (2005)(citing James Madison Letter to W.T. Barry, August 4, 1822, in JAMES MADISON, THE WRITINGS OF JAMES MADISON COMPRISING HIS PUBLIC PAPERS AND HIS PRIVATE CORRESPONDENCE, INCLUDING NUMEROUS LETTERS AND DOCUMENTS NOW FOR THE FIRST TIME PRINTED (Gaillard Hunt ed., G.P. Putnam's Sons, 1900) and U.S. GOVERNMENT PRINTING OFFICE, KEEPING AMERICA INFORMED: FEDERAL DEPOSITORY LIBRARY PROGRAM (GPO, 2000); Harold C. Relyea, *Public Printing Reform and the 105th Congress*, 16 GOV'T INFO. Q. 129 (1999).

Through the FDLP the federal government has long provided access to authentic materials. Many – but increasingly fewer – FDLP libraries have undertaken to preserve these materials. Thus, a distributed archive of reliable and authentic information has been cobbled together.

Paradoxically, the internet – itself a highly distributed information network – is changing this arrangement. Writing in 2005, Jacobs notes that faced with increased budgetary pressure and different demands from consumers of information,

it is clear that GPO plans to change not only the role of depository libraries, but its own role as well. GPO plans the "creation of a fully digital database of all past, present and future government documents. This, combined with omission of any intention of depositing digital publications with depository libraries effectively describes a vision in which depository libraries are replaced with a single monolithic database of government documents."²⁷

Now, in 2010, we see the first robust collections appearing on GPO's Federal Digital System (FDsys). Many FDLP libraries are slashing their print government collections. As these inherently and demonstrably authentic print resources disappear, we must place more trust in the digital versions.

For an example of one problem created by having a single digital source of legal information, Jacobs highlights an effort to withhold materials from the FDLP:

While documents that are deposited with FDLP libraries can be withdrawn, it is cumbersome and can be controversial to do and much easier to quietly remove a single digital copy from a government controlled Web server. The attempted withholding of the volume *Foreign Relations of the United States, 1964-1968, vol. XVI: Cyprus; Greece; Turkey* demonstrates how even the intention of depositing publications with libraries can stave off a political decision to withdraw, withhold, or destroy government information.²⁸

²⁷ *Id.* at 200 (citing BRUCE R. JAMES, KEEPING AMERICA INFORMED IN THE 21ST CENTURY: A FIRST LOOK AT THE GPO STRATEGIC PLANNING PROCESS-A WORK IN PROGRESS, 3 (GPO, 2004).

²⁸ Jacobs, *supra* note 26, at 201 (citing George Lardner Jr., *History of US-Greek Ties Blocked CIA Opposes Disclosure of Proposed Covert Actions in '60s*, WASH. POST, Aug. 17, 2001, at A21).

The fact that multiple physical copies have been distributed through the FDLP provides a check on the government. One could argue that this redundancy check is only a side-effect of the program – not inherent to its original mission.

There are potentially many technological ways to accomplish this redundancy. When discussing the usefulness of mirror sites to increase access and authenticity, Claire Germain notes that

“there is much interest and usefulness in having a “neutral” site, such as a library Web site, which will insure the integrity of the database of information in the same way that libraries have insured access to official versions of governmental texts in the print world with the various editions of state and federal codes and volumes of court decisions that cannot be tampered with.”²⁹

The ease of replicating digital files makes it feasible—although certainly not cost-free— for many “neutral” sites to provide redundancy.

For another example of automating the redundancy process, a joint project of Sun Microsystems and Stanford created project called Lots of Copies Keeps Stuff Safe (LOCKSS). Libraries or other institutions maintain independent nodes of redundant repositories. This network of “LOCKSS boxes” continually audit each others’ content and repair damage or incompleteness. “The more organizations preserve given content, the stronger the guarantee that they will all have continued access to it.”³⁰ By having multiple copies in disparate places, we have stronger confidence in its permanence and authenticity.

²⁹ Claire M. Germain, *Web Mirror Sites: Creating the Research Library of the Future, and More...*, 21 LEGAL REFERENCE SERVICES Q. 87, 99 (2002).

³⁰ Victoria Reich & David Rosenthal, *Distributed Digital Preservation: Private LOCKSS Networks as Business, Social, and Technical Frameworks*, 57 LIBR. TRENDS 461, 463 (2009).

PART III – POTENTIAL DANGERS

A. TECHNOLOGICAL OBSOLESCENCE AND VULNERABILITIES

In a 1998 conference of the Canadian Association of Law Libraries, they recognized the need for preservation. They recommended, “[i]n terms of media and format, the optimum strategy for preservation was considered to be a combination of optical imaging and full-text, preferably with the text being stored in SGML.”³¹ Now, in 2010, we recognize the limitations of the technological recommendations they made. It may be obvious now, but any standards that we recommend must account for such changes.

Again, writing over ten years ago, Germain observes that “[t]he PDF (Portable Document Format) Adobe Acrobat format is now the standard for federal government publications, even though it is criticized by some because of its proprietary nature. It reproduces the physical appearance of a page much better than technologies such as HTML.”³² The PDF has enjoyed a lengthy useful status – in digital technology timeframes – but has started to evolve. Adobe, in conjunction with the International Standards Organization has developed an archival file format (PDF/A-1) based on the PDF.³³ But, even with wide adoption (or perhaps because of), such standard formats use technologies that are subject to attack through the form of viruses and data corruption.³⁴

B. UNINTENDED CONSEQUENCES

But the dangers to preservation and authenticity need not be nefarious or based in the technology. Legal information creators may simply not consider the preservation

³¹ C. Anne Crocker, *The Official Version: Authenticating, Preserving and Citing Legal Information in Digital Form*, 26 INT’L J. L. INFO. 23, 32 (1998).

³² Claire M. Germain, *Content and Quality of Legal Information and Data on the Internet with a Special Focus on the United States*, 27 INT’L. J. LEGAL INFO. 289, 294 (1999).

³³ See, National Digital Information Infrastructure & Preservation Program, PDF/A-1, PDF for Long-term Preservation, Use of PDF 1.4 (Mar. 16, 2010) <http://www.digitalpreservation.gov/formats/fdd/fdd000125.shtml>

³⁴ See, Kim Zetter, Google Hackers Targeted Source Code of More Than 30 Companies, WIRED.COM, Jan. 13, 2010, at <http://www.wired.com/threatlevel/2010/01/google-hack-attack/> (“hackers used a zero-day vulnerability in Adobe Reader to deliver malware to many of the companies”).

ramifications of their actions. For example, Best describes the fallout in Alabama for legal information in the political setting:

"Upon the inauguration of each new governor, many digital records are removed in preparation for transition to the new administration. At the same time, information about specific documents posted on state agency websites disappears as well because agencies are lax in complying with requirements to send web-pages to the Alabama Department of Archives and History for long-term preservation and archiving. Regretfully, citizens are as unlikely to find many of the publications as they a misshelved book."³⁵

The government's laudable efforts to provide current information undercut preservation initiatives.

The several states provide a microcosm for understanding legislative initiatives at digital preservation. Faced with many dozens of government authors with separate sources of information, Marcia Oddi notes that what results "is a growing, separate, uncodified body of law, inaccessible for all practical purposes except to the cognoscenti."³⁶ The American Association of Law Libraries (AALL) conducted a study in 2003 to survey how the states have addressed these issues. In their report, the AALL defined digital preservation as "permanent public access."³⁷ They observed that

"states that have sought to address permanent public access issues have tended to look to public records laws, and records retention and disposition schedules created under them, to try to solve permanent public access issues. [...]The preservation processes do not effectively address publications, inasmuch as publications fall outside the definition of records, or the responsible agencies simply fail to treat publications as records."³⁸

Thus identifying the limitations of those initiatives.³⁹

³⁵ Best, *supra* note 9, at 2.

³⁶ Marcia J. Oddi, *Can You Rely on the Indiana Code?*, RES GESTAE, May 2008, at 20.

³⁷ RICHARD J. MATTHEWS, ANNE E. BURNETT, MARY ALICE BAISH, CHARLENE C. CAIN, SUSAN L. DOW, & DAVID L. MCFADDEN, STATE-BY-STATE REPORT ON PERMANENT PUBLIC ACCESS TO ELECTRONIC GOVERNMENT INFORMATION, (AALL 2003) *available at* <http://www.aallnet.org/aallwash/PPAreport.asp>.

³⁸ *Id.* at 14.

³⁹ The report contains a note that the online version of their report is **permanently available** online and provides a URL. Ironically, the URL no longer works.

The National Conference of Commissioners on Uniform State Laws (NCCUSL) has attempted to address problems of preservation by promoting a model law on digital preservation to state legislatures. Like the AALL, NCCUSL adopts the definition of “permanent public access.” Their model legislation provides that

[i]f a document is made available exclusively electronically, it must remain available electronically permanently, either in its original location or in an archived location. The official publisher must ensure that all amended, changed, or superseded documents shall remain available on conditions of access similar to those in effect for then-current documents.⁴⁰

The models of digital preservation would thus have the weight of legislative mandate. The problems associated with preservation at the agency level would be solved by a centralized publisher that would assume the responsibility for permanent public access.

I would argue, however, that this centralization is equally problematic. By centralizing the archive, we create a single source of failure – a single source of corruption. As discussed above, the authenticity of the information in the archive derives greatly from the trust the institution garners in the social framework. For the law, those in government might well be swayed by political incentive to do damage to the record of legal information. As Jacobs describes:

“Failure” in a single-source information culture includes technological failure, accidents, intentional altering or destruction or removal of information, changing budget priorities that are unable to keep up with a rapidly growing amount of information, changing political priorities, and other unforeseen technical, economic, social, and political problems.⁴¹

Although a centralized source would certainly be convenient and very likely trustworthy, it can’t be the whole story of a solution.

We can also glean conclusions from scholarly publications. Legal scholarship, after all, certainly plays a role in the world of legal information. As I’ve noted above, the

⁴⁰ National Conference of Commissioners on Uniform State Laws, *Authentication and Preservation of State Electronic Legal Materials Act*, March 5 – 7, 2010 Committee Meeting Draft available at <http://www.law.upenn.edu/bll/archives/ulc/apselm/2010jan8%20meeting%20draft.pdf>

⁴¹ Jacobs, *supra* note 26, at 202.

creator must address issues of preservation in the creation. But scholarly creators might not have the incentive or time to give preservation much thought. Richard Danner notes that

[a]lthough scholars are concerned about the continuing availability of electronic journal and other forms of scholarship, it is likely that they themselves will be less interested in the preservation issues raised by a burgeoning electronic grey literature than they have been in making use of electronic publishing and networked communications for disseminating and publicizing their works in both formal and informal forms.”⁴²

Accelerating the publication process through digital means and relying on centralized sources for distribution has the unintended consequence of hampering authentic materials appearing the archives. Again, it is the social framework of the scholar—with pressures to publish widely—that most influences the preservation of authentic information.

Moreover, as we continue automating the process of digital preservation, we must be ever more vigilant in maintaining the underlying social frameworks that inform the creation and maintenance of digital repositories. Matthew Fagan describes the issues associated with automated caching of websites by the Internet Archive and their use in litigation. The Internet Archive passively crawls the web and stores copies websites and provides free access to them through their Wayback Machine service. He notes the dangers that

A malicious attacker could place inaccurate information into the system at any number of points; caching services may copy the fake web pages, or have their originals replaced with fakes directly. Attorneys should heed George Orwell's warning: “If all others accepted the lie which the Party imposed-if all records told the same tale-then the lie passed into history and became truth.”⁴³

⁴² Richard Danner, *Issues in the Preservation of Born-digital Scholarly Communications in Law*, 59 L. LIBR. J. 591, 603 (2004).

⁴³ Matthew Fagan, “Can You Do a Wayback on That?” *The Legal Community’s Use of Cached Web Pages In and Out of Trial*, 13 B.U. J. SCI. & TECH. L. 46, 64 (2007) (quoting GEORGE ORWELL, 1984 29 (Plume Books 1983)(1948).

C. LOSS OF INDEXING AND FINDING AIDS

Digital tools allow greater ability to find materials. Searching the full text of documents gives a sense that we are being more thorough and accurate in our retrieval. But we should not conflate access with understanding. As Germain notes:

The reliance on Internet search engines leads to the loss of a lot of sophisticated indexing tools, such as subject and digest keyword indexing, the elaborate system created by West since the end of the 19th century. The Internet makes legal information much more accessible to the public. But, it is not clear that the greater accessibility makes the law more understandable because it may lack a context. People can misinterpret the text of the law, unless there are disclaimers.⁴⁴

By providing a context with indexing tools and linking ideas in the law by subject, we increase understanding.

Those indexing tools, whether created by publishers or bibliographers, should be maintained. The AALL calls for the government to maintain them. In their state-by-state report they declare that

“AALL believes strongly that, in the online environment, the government is responsible for creating useful finding tools to locate electronic government information; for ensuring its authenticity when the decision is made to no longer produce the information in an official tangible version; and for ensuring that valuable electronic government information will be not only retained and preserved, but will also remain available for *permanent public access*.”⁴⁵

But, just as government actors might have political, budgetary, or social incentives to shirk their preservation mandates, they might also fail to create appropriate finding tools. Legal information would thus lose appropriate context for understanding. Sophisticated users—like librarians and lawyers—might have the technological and intellectual tools to navigate the legal information, but the law should be physically *and intellectually* accessible to everyone

⁴⁴ Claire M. Germain, *Content and Quality of Legal Information and Data on the Internet with a Special Focus on the United States*, 27 INT’L. J. LEGAL INFO. 289, 295 (1999) (citing Robert Berring, *Legal Research and Legal Concepts: Where Form Molds Substance*, 75 CAL. L. REV. 5 (1987)).

⁴⁵ Matthews, *supra* note 37 at 8.

CONCLUSION

The Library and Information Science scholarly literature continues to struggle with applying long-held theories of permanence and authenticity to digital materials. Technological approaches to preservation exist in an overarching social framework. Standards and models for preservation must account for the potential pressures and instability presented in the social framework. For legal information, political posturing and institutional apathy can undermine preservation efforts. Libraries and archives should play a role as neutral preservers of authentic information. The ephemeral nature of digital objects, however, requires information professionals to play an active role in the creation process because preservation issues arise the moment a digital object is created. The rule of law and democratic notions of fairness require permanent access authentic legal information.