LONG-TERM DIGITAL OBJECTS

TRUSTWORTHY EVEN AFTER EVERY WITNESS IS DEAD

H.M. Gladney

HMG Consulting
Saratoga, CA 95070
hgladney@pacbell.net

24 October 2003

In ancient times, wax seals impressed with signet rings were affixed to documents as evidence of their authenticity. A digital counterpart is a message authentication code fixed firmly to each important document. We sketch an architecture that accomplishes this by using encapsulation of content with provenance metadata, cryptographic sealing, and webs of trust rooted in respected institutions.

A prototype will integrate software that is mostly available, will conform to many standards, and will be interoperable over most computing bases. It will avoid administrative overhead inherent in other methods.

INTRODUCTION

Digital interchange is a growing activity that is beginning to be accompanied by attention to preserving content for periods that exceed practical technology lifetimes and sometimes human lifetimes. What might someone a century from now want of information created today? For some applications, **people will want *evidence* that the information they depend on is *authentic***. [Bearman 2] [Gladney 2]

Consider what some user might want of a century-old digital object, and of any object it depends on:

(1) A copy of the bit-string representing the object should have survived and be accessible.

(2) The content should be comprehensible for human reading or correctly executable if it is a procedure.

(3) Associated metadata should identify the purpose, historical context, and limitations of the content.

(4) Consumers should be enabled to determine whether content received is trustworthy. I.e., their should be evidence that the object is what it purports to be and that it has not been altered.

(5) It should be possible to resolve included references reliably. For instance, if the object identifies a program or a certificate needed for its safe use, the user should be able to obtain a copy of this together with evidence that it is the correct object and valid.

The current article focuses on *authenticity* for documents to make them useful as *evidence,* in line with archivists' criteria refined by years of critical thinking. It condenses a more detailed manuscript [Gladney 3] by omitting details that would obscure the main design features.

Terms of Reference and Assumptions

This work addresses any data that might pass between digital machines; the kind of information and its purpose are irrelevant. We assume a computing environment in which a producer—a human being who endorses an object as being in some sense authoritative—conveys information to consumers by the Internet or on storage volumes. (Figure 1) This transmission might be asynchronous, with consumers finding deposited records only many years after deposit.

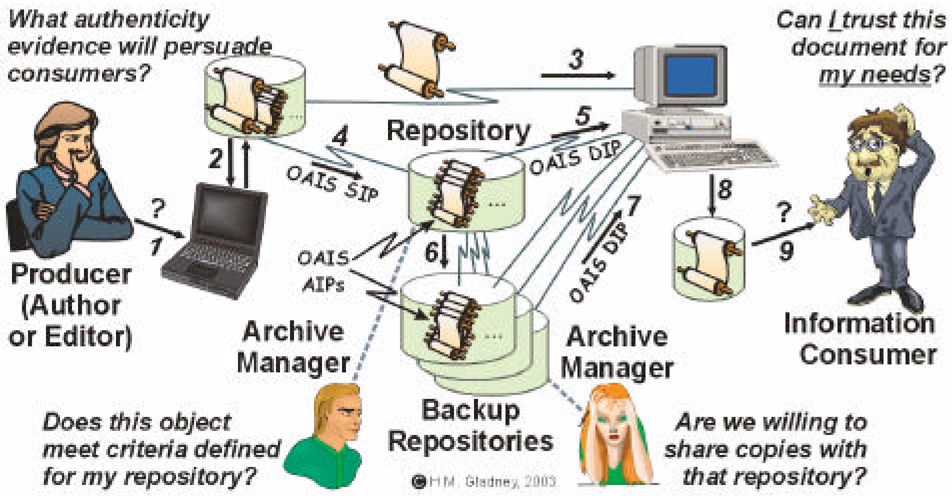

Figure 1: Information interchange and repositories, illustrating alternative pathways by which a digital object might move from its producer to its eventual consumer.

Nouns such as "producer", "author", "artist", and "composer" are effectively synonyms. Similarly, the reader should construe "read", "print", and "report" broadly. Allusions to people denote their roles in transactions, rather than their positions. For instance, most archives will have an employee who acts sometimes as a producer, sometimes as a consumer, and sometimes as an archive manager.

The words *archive* and *repository* have multiple meanings. [OAIS] avoids ambiguity by distinguishing between an *archival institution* and *archival storage*. (Figure 4) We assume that the reader has elementary familiarity with digital security technology. [Schneier]

What Will Users Want?

Many institutions already have digital repositories, and will want to extend their services to durable content. Naturally, they will want to accomplish this without disruption, and might refuse any preservation solution that requires them to migrate from their installed software to a replacement.

Since nobody has surveyed end users' trust needs, we speculate in the context of Figure 1. Content producers will want to please consumers. Archive managers will want to please both producers and consumers. Archive managers are likely to have sufficient contact with producers to resolve information format and protocol issues, but will have contact with only a small fraction of their consumer clients.

Some producers might serve content to consumers directly (transmission #3 in Figure 1). They might deposit copies into public archives (transmission #4) only years after it was produced. They will not be willing to prepare each document for early delivery and, years later, differently for a repository.

Information consumers will prefer that the formats and content of documents received do not depend whether they come directly from publishers (transmission #3), from primary repositories (transmission #5), from backup repositories (transmission #7), or from other consumers.

When a repository shares a holding with another repository—whatever the reason for sharing might be—the second repository will want the delivery to include all the information closely associated with that holding as part of the sharing. It will further want a ready test that everything needed is included.

Consumers will mostly decide whether to trust preserved information without conversations with producers or archivists. They will see only information delivered into their local machines.

Every change to recorded information introduces risk that some critical meaning will be distorted. That the person who makes a change might be the agent of an archival institution is irrelevant, as are the possibilities that this editor means well, or that the change might be "merely" the addition of a metadata attribute or a reference to another document.

In the light these circumstances, we assert the following technical objectives:

(1) The preservation method should work for **all kinds of information and representations.**

(2) **All closely related metadata and historical evidence** should be accessible to whoever properly receives a copy of a digital object.

(3) For interoperability, core **elements of preservation schema need to be common world-wide**.

(4) **Every process step that can be automated should be automated**, but no intervention of human judgment should be precluded.

(5) The preservation method should be **upwards compatible from digital library software** in wide use, and should not interfere with existing library procedures or information finding mechanisms.

(6) Preserved object **bit-string representations should be automatically replicated** in repositories in which their primary custodians have confidence.

(7) The trustworthiness mechanism should **not interfere with other security mechanisms.**

(8) The trustworthiness mechanism should be **secure end-to-end.**

(9) Consumers should be able to **test preserved object trustworthiness** themselves.

(10) Producers should be able to **provide whatever authenticity evidence they choose**.

The design below treats only syntactic transformations because machines can do nothing more. In Figure 1, the steps 2 through 8 include at most syntactic changes. In contrast, step 1 and step 9 involve decisions about representing and interpreting meaning—semantic decisions that we do not treat.

PRIOR WORK

Trust, and *evidence* for deciding what digital information is *authentic* are considered in [Bearman 2], [CLIR 107], [Conway], [Cullen], [Lynch], [MacNeil], and [RLG]. For risks and mitigation mechanisms see [Caronni], [Ellison], [Gerck], and [Schneier]. Current understanding is suggested by:

> In their seminal report of 1996, Waters and Garrett lucidly defined the long-term challenges of preserving digital content. Ironically, the questions and issues that swirl about this new and perplexing activity can be relatively simply characterized as a problem of integrity of the artifact. … It is the mutable nature of digital information objects that represents one of the principal obstacles to the creation of archives for their long-term storage and preservation. [Okerson, p.53]

To be evidence, information must be fixed, as is required for copyright protection. The literature is replete with metadata articles, such as [Bearman], [Hunter], [Lupovici], [METS], and [NLNZ].

Embedding metadata has a long history. The idea that each information object should embed its own metadata became explicit in "object oriented programming" in the late 1970's. [Kahn] recommended it for digital libraries, but the idea was not used for digital documents until recently.

STRUCTURE OF EACH TRUSTWORTHY DIGITAL OBJECT (TDO)

Input *DO*s (*digital objects*) can be any expressions of information that somebody wants to preserve. A producer packages and seals each DO set with a message authentication code (MAC). Each such resultant TDO must include its own authenticity evidence as suggested by Figure 2. This includes a Protection Block that contains assertions about the TDO history, (§0) and a safe representation of every critical reference (link) needed.

Many digital objects will have prior versions. We signal this relationship by a shared *digital resource identifier* (DRI). Any TDO might itself be a payload portion; i.e., TDOs can be nested. Every such nesting of a prior TDO, or of a reliable reference to a prior TDO, within a new TDO is **an effective audit trail element**.

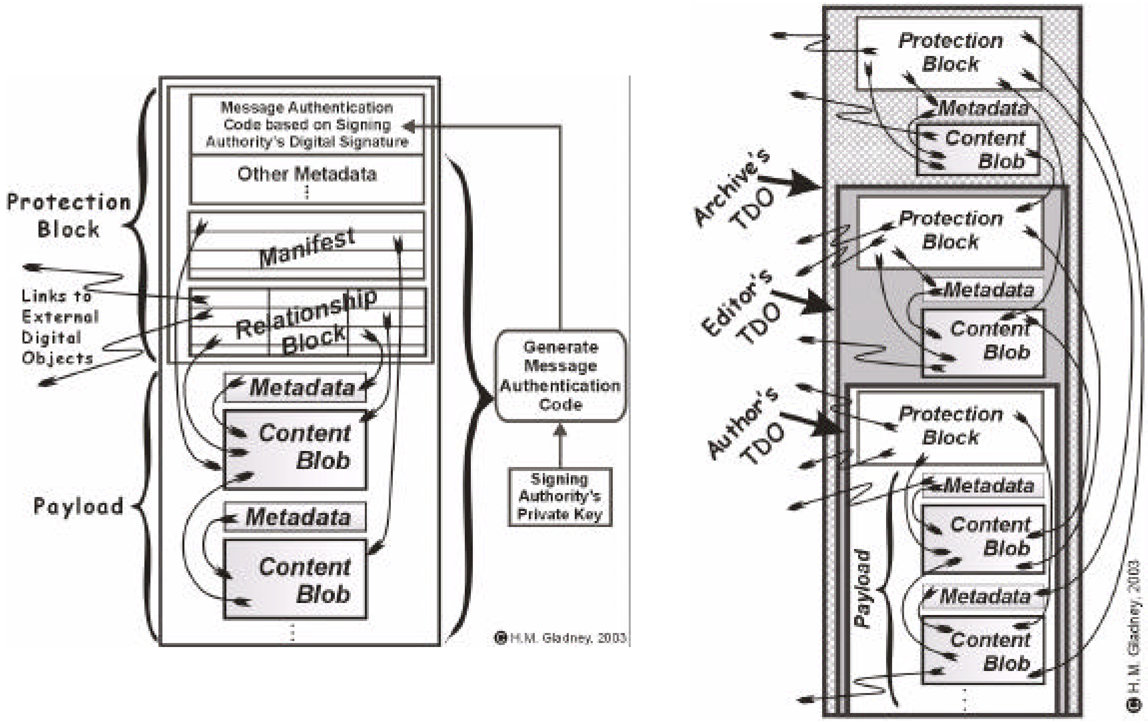

Figure 2: Preservation data packaging (a TDO) and an example of TDO nesting, depicting on the left a TDO created from a DO (the payload) by adding a message authentication code and, on the right, how the publication version of a book might nest TDOs. The diagrams also depict internal and external links.

In summary, the TDO design specifies encapsulating and managing information for preservation so that:

(1) The blob set that represents a work is XML-packaged with registered schema.

(2) Integrity is assured by public key message authentication [Eastlake]. (§0)

(3) The package includes its own identifier and metadata needed as provenance evidence. (§3.2)

(4) Links to other works are secured by the message authentication codes of the linked entities.

(5) Information loss is minimized by replication in independent repositories. (§0)

(6) The cryptographic key management uses the Web-of-Trust model (§0) grounded in keys that widely trusted institutions publish and periodically replace with new keys. (§0)

Object Packaging and Sealing

An author might package his book draft as a TDO. (Figure 2 right) The editors might package their revised version as a TDO re-using the digital resource identifier (DRI) chosen by the author. The publisher might share this version with a copyright depository library and request encapsulation with standard cataloguing metadata endorsed with the library's seal. The library would re-use the author's DRI in creating the TDO that it returns (§0); this would be a good version for the publisher to disseminate.

When the time comes for a new edition, the participants might repeat these steps, reusing the original's DRI. Each participant might include new information. For instance, if the book is about computation, the author might include new sample programs, the publisher might include new links to Web software sources, and the library might include new bibliographic links. Reusing the DRI will help intellectual property managers and others who care about the work's history.

From an archivist's perspective, the significant events in a work's history are its transfers from each producer to a successor in a chain of custody. Only then do two people assuredly see the same version. To create an audit trail, each successive producer would include in the TDO he creates a copy of the TDO that he received, or a reliable link to that prior version.

Suppose that the publisher's first and second versions are publicly accessible. Then a reader who finds a version might seek all works with the same DRI, receiving both versions. From the library's time-stamped signatures she might believe that both were valid. From protection information described in §0 she would learn more of the work's history. Furthermore, she could compare versions for change details.

Protection Block Content and Structure

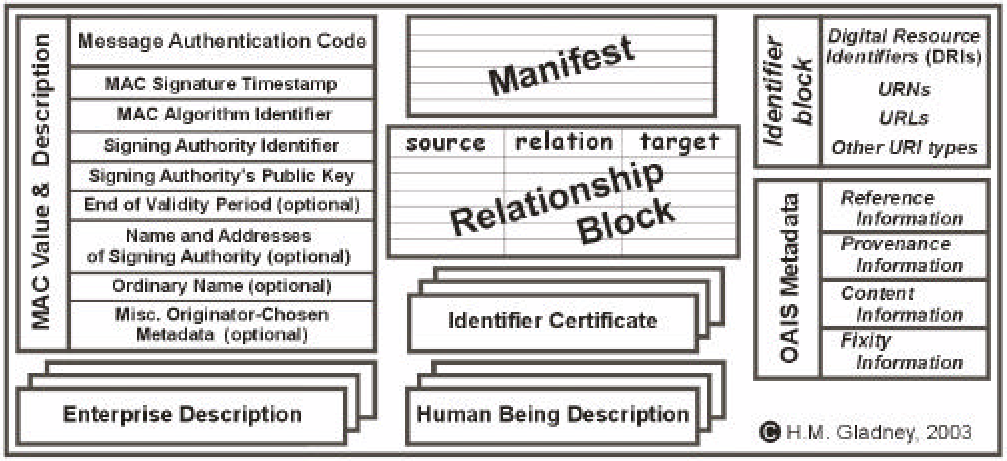

Figure 3: Contents of a protection block (PB), depicting abstract syntax, without suggesting any concrete syntax or structuring beyond what is discussed in [Gladney 3].

Each TDO is a MAC-sealed package of a protection block and a payload. (Figure 2) The *protection block* has a new scheme that might include references into the payload and into external objects. Its fields are not scalars, but instead extensible value sets that might include links and substructures. A structure proposed by [CNRI] is adequate. A *protection block* **(PB)** consists of (Figure 3):

- A block containing one or more identifiers;
- A MAC description block containing information related to signing [Eastlake], perhaps extending what is specified for X.509 [Gerck]; the essential information is a timestamp, a signature algorithm identifier, a signing authority identifier, and the signing authority's public key value. Each item can be anything, including text specifying what TDO properties are certified and liability disclaimers.
- A ***manifest*** describing the payload blocks; the *n*th manifest element describes the *n*th payload block.
- An optional ***relationship block*** with any number of 3-cell rows. Each row identifies two blobs in the TDO or external objects, or bookmarks into such objects, and a relationship.
- Optional descriptions of enterprises and human individuals;
- Any number of identifier certificates; and
- An optional [METS] block, e.g., for preservation description information. [OAIS, §4.16]

Almost every TDO will need external references: to standards, schemas, and authenticity evidence. TDO design provides for internal and external links. (Figure 2). Each link can have a prefix and a suffix. The *prefix part* would identify a digital object. The *suffix part* would indicate either a place or a data extent in that resource. External resources can be digital resources or physical resources such as land parcels.

A work itself, each edition, and each instance might merit distinct identification. A *Digital Resource Identifier* (DRI) **identifies all the objects that producers choose to indicate as being versions of a work.** Its syntax conforms to the URI specification [RFC 2396].

An internal TDO link is secure against undetectable tampering. To make an external TDO link proof against tampering, it is accompanied by the MAC of its referent.

INFRASTRUCTURE FOR TRUSTWORTHY DIGITAL OBJECTS

Safe preservation depends on infrastructure that can be mostly deployed digital library and communication software. §0 suggests how archival storage needs to fit into an archival institution. §0 discusses repository structure forced by tension between global interoperability and local customization.

TDO integrity assurance depends on the secrecy of its producer's private cryptographic key. Consumers' ability to test TDO authenticity depends on certainty that the producer's corresponding public key belongs to the purported producer. §0-§0 sketch a convenient satisfying design.

Services of Archival Institutions

Library patrons' expectations are likely to constrain TDO editing performed by cataloguing librarians. An archival institution might publish such constraints as part of its public commitments. For instance, it might assert that it will not lose any information received in producers' packages that it republishes as TDOs and that it will not add anything beyond certain kinds of metadata.

The current work focuses on records that pass across the boundary of Figure 4's archival storage, in contrast to [OAIS], which focuses on the archival institution's boundary with the rest of the world. The Digital Object Import service should accept a submission only if it meets the institution's published TDO criteria. These properties would typically include specifics for each object's metadata, and also submission by an agent the institution knows and trusts for such submissions. To ensure audit trail completeness, any TDO change made by an agent of the archival institution (depicted as a "Cataloguing Librarian" in the figure) must be treated similarly to submissions from outsiders.

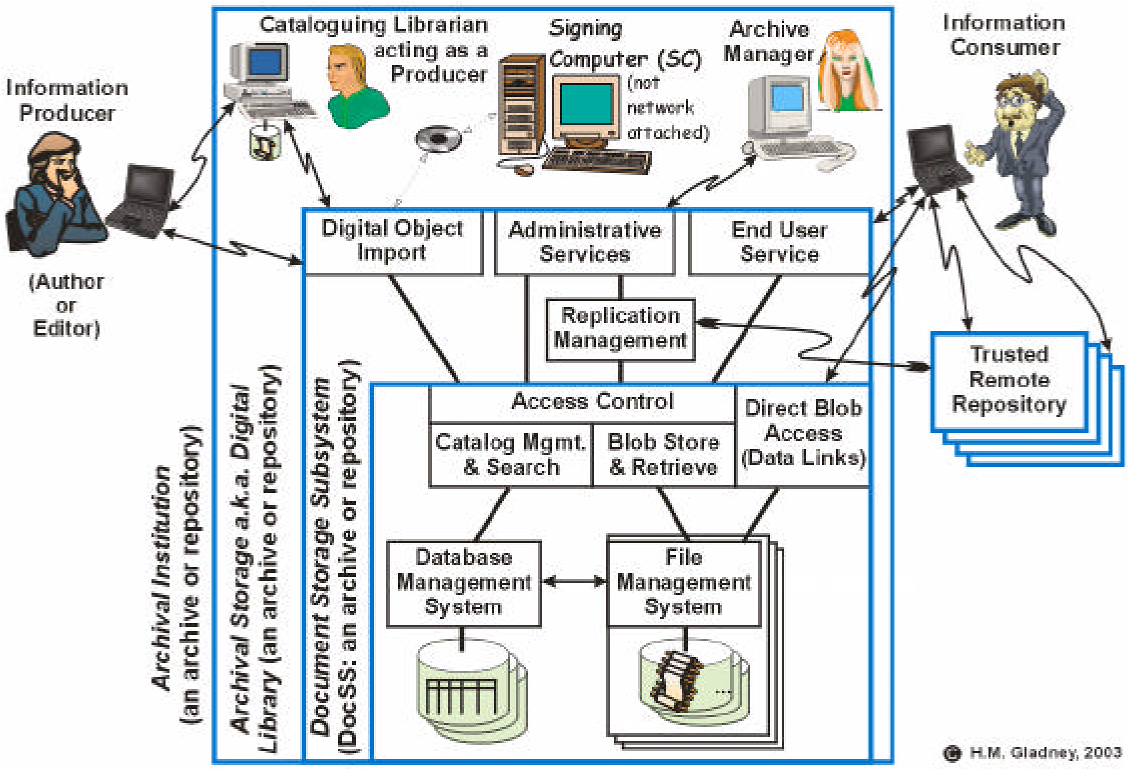

Figure 4: Archival institution, archival repository and nested document storage subsystem in a repository network, showing human roles in the use and management of long-term digital repositories. The role, use, and business controls of the *Signing Computer* are discussed in §0

Figure 1 illustrates that a TDO might travel to its consumer by any of several paths; which path will actually be used cannot be anticipated by producers. Consumers will want the content and format of objects they receive to be independent of its route. This forces what [OAIS] calls a *Distribution Information Package* (DIP) to be identical to the corresponding *Submission Information Package* (SIP).

Repository Structure and Repository Networks

As well as serving its primary patrons, an archival institution is likely to serve external communities. For this it must provide popular interfaces to catalog querying and object retrieval; exposing DocSS interfaces (Figure 4) is therefore forced by interoperability objectives. However, TDO service interfaces can still be tailored for local users at the Archival Storage layer. Such layering has been available in IBM Digital Library since 1993. [Gladney]

Information producers within an archival institution are not distinguished from external producers by the Digital Library (DL) interface in its enforcement of institutional TDO validity criteria. However, the Import module might include workflow functions that distinguish between internal and external producers. For instance, it might route unacceptable submissions to a Cataloguing Librarian (Figure 4) at the same time as it returns exception information to producers.

Replication Management in Figure 4 provides bit-string safety. [Reich]. Like the Import and End User modules, it operates according to rules specified by an Archive Manager. Interposing DocSS access control between it and the repository core provides administrative autonomy of cooperating repositories.

Business controls are managed through an Administrative Services module in the DL layer. This includes specification of who may modify the collection or TDO acceptability criteria.

Exchanging Public Keys: a "Web of Trust"

The credibility of TDO content will depend on the security of producers' identity certificates, and on facts that people—perhaps the public at large—trust. Evidence that will impress consumers is certification by an institution, such as a national library, that is widely trusted to handle documents correctly and faithfully.

Such trust can be encouraged by an institution's public assertions that has little to gain and much to lose by misrepresenting what it publishes. These might include publication of its own public key, and an offer to issue certificates to certain classes of organization. A national library might advertise something like, "The [X Library] offers certification of the public key of any [some class of institution] that provides [certain information about itself] by a visit of its accredited representative to X premises. The X public key from [beginning date] to [end date] is [key value]."

The widely publicized public key infrastructure (PKI) depends on *certificate authorities* (CAs) as the roots for certificate hierarchies. Each certificate asserts the relationship of some computer user or enterprise to a particular public key. [Gerck] identifies a flaw allowing fraudulent certificates from a CA. Since **impersonation may be the most important exposure to content authenticity certification**, we cannot accept this kind of risk.

Whereas the participants in commercial transactions are often not acquainted before their transactions, the participants in a producers' chain are often personally acquainted. (§0) This enables the "Web of Trust" method. [Caronni] Without imposing undue burden, we can require each such collaborating pair to exchange their own public keys personally. Later the partner at risk can test incoming information from the other partner. A producer should certify as correct only TDO content that he is confident was endorsed by a known producer that the incoming TDO itself identifies, or that he has himself generated.

For instance, books will mostly be submitted to a trustworthy institution (TI) by publishers in the same geopolitical area as the TI, e.g., American publishers asking the Library of Congress to serve as a TI. Publishers can deliver their public keys in professional meetings of librarians, giving the TI the reliable knowledge it needs to test their submissions. (§0) A TI agent who certifies a digital object acts as a producer who must be at least as diligent as any other producer in judging the authenticity of information that he will certify. A TI agent needs to be aware that **flawed certifications will jeopardize the reputation for integrity and quality that creates and maintains the trusted status of his/her employer.**

A TI can enlarge the community that might trust its offerings by persuading other TIs to certify its public keys. Each such TI would participate in creating a Web of Trust by publishing the public key certificates it has signed; this action would endorse the public-key-to-identity mapping of its sister TIs. The benefit to each endorsing TI would be reciprocal endorsement.

Object Signing by a Trustworthy Institution (TI)

After a TI receives content from its producer, it will test this input and what it knows about the producer to determine whether these satisfy its own published criteria for document certification. If it is satisfied, it will create a new DO by copying, editing, and augmenting the input metadata into a new protection block that meets its own specifications. When this editing is complete, it will copy the resulting DO into a signing computer SC (Figure 4) that is used for creating institutional MACs to convert DOs into TDOs, being exclusively devoted to this task as a security measure for protecting the private key of the TI.

Whenever it contains a sufficient DO input batch, its operator will **detach SC from all networks** and then load it with TI's private (secret) key. The operator will then start a utility program that (1) tests that each input meets all TI-required quality criteria, (2) fills in missing PB portions into each input, including metadata for the pending MAC, and (3) computes and adds the MAC, thereby finishing the TDO construction. (Figure 2) Finally, the operator must remove all copies of the private key from SC before reattaching it to networks to deliver the newly certified TDOs.

To assure users about the age of TDOs it has sealed, and to protect its private key further, the TI might choose a new public/private key pair periodically—annually for instance—and should then **destroy <u>all</u> copies of the expired private key**. It should further publish the history of its public key values. This mimics an 19th century Japanese practice, in which the censors of ukiyo-e ("pictures of a floating world") changed their seals approximately annually and published these keys. [Illing]

Consumers' Tests of Authenticity and Provenance

A cautious consumer will not judge a received TDO to be authentic unless he believes certain things:

- That the enclosed MAC demonstrates that the TDO has not been altered after it was certified;
- That the enclosed identifications of the most recent MAC signatory and date are authentic;
- That the producer of each TDO version had appropriate authority for changing prior versions;
- That the final signatory's TDO-signing procedure is sound and includes judging the authenticity of information it includes in any TDO it creates; and
- That this TI (trustworthy institution) signing procedure has been correctly executed.

As evidence, the consumer will have the published public keys of some endorsing TI (§0), which itself might be endorsed by other TIs' cross-certifications (§0), and the certified public keys and known identities and changes made by prior producers in a TDO history. (§0). The consumer might additionally be able to judge the TDO payload (Figure 2) as corroborative evidence and might also use context provided by other documents that he knows professionally.

DISCUSSION AND CONCLUSIONS

This work is part of a large project—creating software for preserving any kind of digital object. Digital preservation poses four technical challenges:

- Ensuring that our descendants can understand and use whatever information we save. [Gladney 1]
- Ensuring that anybody can decide whether saved data is sufficiently trustworthy.
- Replacing human effort by automatic procedures.
- Helping each information producer create metadata.

The current article reports a design that solves the second challenge by sealing digital objects—a **return to ancient practice in which rulers certified their decrees with wax signet ring impressions.** By embedding critical information in each digital object, we achieve end-to-end security and minimize bureaucratic overhead. By connecting identity certificates in a "Web of Trust", we ground evaluation of digital object authenticity in the public keys of a small number of institutions that some user community trusts <u>for the kind of content being certified</u>.

Security technology cannot eliminate the need for administrative controls or independent audits, but can reduce their complexity and cost by shifting the locus of risk to a relatively small part of the infrastructure. TDO architecture shifts trust management from protecting digital repositories to protecting private keys—a relatively simple task suited to administrative and physical control.

What information a consumer will find sufficient to decide the authenticity of a TDO is a judgment that will depend on how he plans to use the content. An information producer's reasonable response is to guess what will be wanted by prospective consumers, and to provide that as part of the TDO metadata.

Ideally, we would compare TDO design to alternatives; however nobody seems to have designed one. The closest approach, [RLG], suggests a direction for inquiry, but provides insufficient depth for detailed evaluation. Repository-centric proposals for trustworthy preservation have intrinsic weaknesses:

- They depend on an unexpressed premise—that exposing an archive's procedures can persuade clients that its content deliveries will be authentic.
- Certifying archival institution procedures is inherently more expensive and less reliable than having each stored object embed its own audit trail. For instance, periodic audits of an archive cannot prove that its contents were not improperlyb altered years before the current audit.

The TDO design enables authenticity testing that is independent of the distance in time, space, or organizational affiliation from the objects' producers. Specific design elements address specific needs, viz.,

- MAC-sealing an object together with its metadata keeps content and descriptors reliably connected, simplifying information sharing and helping make it easy to load a digital library.

- Having each object carry its own authenticity evidence avoids any need for digital repository changes, allowing institutions to avoid costly disruption.
- Including identifiers and links in each TDO enables flexible object structuring and finding.
- The open design makes for compatibility with other initiatives and emerging standards.
- A broad notion for "resource" ("the British fleet" is a resource) helps each digital object record its own history, simplifying scholarly research, document production, and library administration.

TDO architecture makes possible authenticity assessment for digital objects. **Digital authenticity assessment will, when the ideas reported are refined, become better than its counterpart for works on paper: more reliable, cheaper, faster, and requiring less end-user expertise.**

Acknowledgments

This work originates in collaboration with Raymond Lorie and discussions with former IBM Research colleagues. John Bennett and John Swinden read manuscript drafts carefully and provided scores of specific comments and illuminating discussions.

Bibliography

[Caronni] Germano Caronni, *Walking the WWW of Trust*, Proc. 9th Workshop on Enabling Technologies, IEEE Comp. Soc. Press, 2000.

[Conway] Paul Conway, *Preservation in the Digital World*, CLIR pub92, March 1996.

[CNRI] Corporation for National Research Initiatives, *Handle System: A general-purpose global name service enabling secure name resolution over the Internet,* http://www.handle.net/, 2001.

[CLIR 107] Council on Library and Information Resources and the Library of Congress, *The State of Digital preservation: An International Perspective*, pub107, April 2002. ISBN 1-887334-92-0

[Cullen] C.T. Cullen, *Authentication of Digital Objects: Lessons from a Historian's Research*, in *Authenticity in a Digital Environment*, CLIR pub92, 2000; ISBN 1-887334-77-7.

[Eastlake] Donald E. Eastlake and Kitty Niles, *Secure XML: The New Syntax for Signatures and Encryption,* Addison Wesley, 2002. ISBN 0-201-75605-6

[Ellison] C. Ellison and B. Schneier, *Ten Risks of PKI: What You're Not Being Told About Public Key Infrastructure*, Computer Security Journal 16(1), 1-7, 2000.

[Garrett] J. Garrett et al., *Preserving Digital Information: Report of the Task Force on Archiving of Digital Information*, commissioned by The Commission on Preservation and Access and The Research Libraries Group, August 1995.

[Gerck] Ed Gerck, *Overview of Certification Systems: X.509, PKIX, CA, PGP, SKIP,* 2000. http://www.mcg.org.br/certover.pdf

[Gladney 1] H.M. Gladney and R. Lorie, *Trustworthy 100-Year Digital Objects: Durable Encoding for When It's Too Late to Ask*, June 2003, available at http://daedalus.lib.gla.ac.uk/archive/00000007/.

[Gladney 2] H.M. Gladney and J.L. Bennett, *What Do We Mean By Authentic? What's the Real McCoy?* D-Lib Magazine 9(7/8), July 2003.

[Gladney 3] H.M. Gladney, *Trustworthy 100-Year Digital Documents: Evidence Even After Every Witness is Dead,* May 2003. Preprint at http://daedalus.lib.gla.ac.uk/archive/00000008/.

[Hunter] J. Hunter, *MetaNet - A Metadata Term Thesaurus to Enable Semantic Interoperability Between Metadata Domains*, Journal of Digital Information 1(8), 2001.

[Illing] Richard Illing, *The Art of Japanese Prints,* Calmann & Cooper, London, 1980. See p.170.

[Kahn] R. Kahn and R. Wilensky, *A Framework for Distributed Digital Object Services*, May, 1995. http://www.cnri.reston.va.us/cstr/arch/k-w.html

[Lupovici] Catherine Lupovici and Julien Masanès, *Metadata for the Long Term Preservation of Electronic Publications*, NedLib Report Series 2, 2000. ISBN 9062 59 1469

[Lynch] Clifford Lynch, *Authenticity and Integrity in the Digital Environment: An Exploratory Analysis of the Central Role of Trust*, in *Authenticity in the Digital Environment*, CLIR Report pub92, 2000; ISBN 1-887334-77-7. *When Documents Deceive: Trust and Provenance as New Factors for Information Retrieval in a Tangled WWW*, J. Am. Soc. Info. Sci. 52(1), 12-17, 2001.

[MacNeil] *Trusting Records in a Post-Modern World,* Archivaria 51, 36-47, Spring 2001.

[METS] *Metadata Encoding and Transmission Standard* (METS), 2001-2. http://www.loc.gov/standards/mets/

[NLNZ] National Library of New Zealand, *Metadata Standards Framework – Preservation Metadata*, November 2002. http://www.natlib.govt.nz/files/4initiatives_metaschema.pdf

[OAIS] CCSDS 650.0-R-2, *Reference Model for an Open Archival Information System (OAIS)*, Red Book, July 2001. The ISO version is at http://www.ccsds.org/RP9905/RP9905.html.

[Okerson] A. Okerson et al., *YEA: The Yale Electronic Archive, One Year of Progress: Report on the Digital Preservation Planning Project,* 2002. http://www.library.yale.edu/~okerson/yea/

[Reich] Vicky Reich and David S.H. Rosenthal, *LOCKSS*: *A Permanent WWW Publishing and Access System*, D-Lib Magazine, June 2001.

[RFC 2396] T. Berners-Lee et al., *Uniform Resource Identifiers (URI): Generic Syntax,* IETF RFC 2396, 1998.

[RLG] Neil Beagrie, Meg Bellinger, Robin Dale, Marianne Doerr, Margaret Hedstrom, Maggie Jones, Anne Kenney, Catherine Lupovici, Kelly Russell, Colin Webb, and Deborah Woodyard, *Trusted Digital Repositories: Attributes and Responsibilities,* RLG-OCLC Report, May 2002. http://www.rlg.org/longterm/repositories.pdf

[Schneier] B. Schneier, *Applied Cryptography: Protocols, Algorithms, and Source Code in C*, John Wiley & Sons, New York, NY, 1996