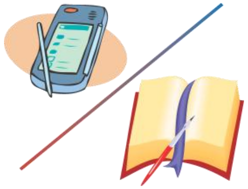


An Institutional Framework for Creating Authentic Digital Objects

4th International Digital Curation Conference

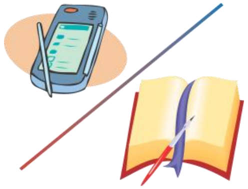
1- 3 December 2008
Hilton Edinburgh Grosvenor Hotel
Edinburgh, Scotland

Ronald C. Jantz
Rutgers University Libraries

21st Century Libraries and Services

Martell (2000) eloquently implores us to " to create a range of services unthinkable in the twentieth century, but mandatory in the twenty-first century, if we are to provide society with the value added services it will need from its professionals."

Martell, C. (2000). The disembodied librarian in the digital age. College and Research Libraries, (61), (1), 10 – 28.



Working Assumptions

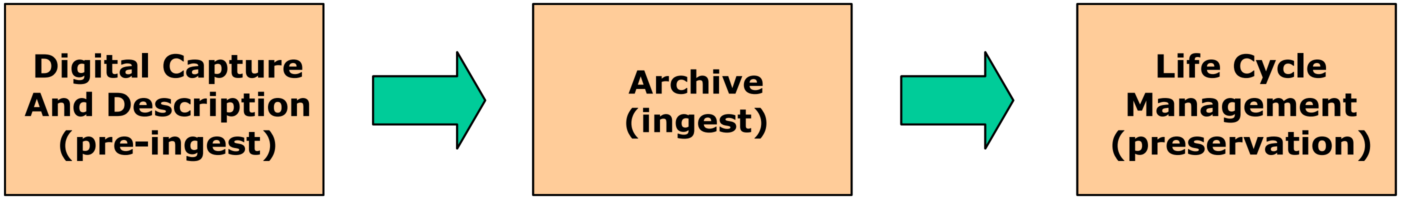
- Digital scholarship requires authentic digital objects
- Some many years after archival, researchers who reuse a digital object will want to know with some assurance:
  - Who was the “author” of the digital content?
  - When was the object created? By whom?
  - When last modified?
- A new service - with appropriate transformations, libraries and archival institutions can create authentic digital objects

The Challenge To Digital Scholarship

Two examples where the archivist could have added value:

- Faulty methods – the Cold Fusion physics debacle of the 1980s
  - See http://www.guardian.co.uk/education/2005/mar/24/research.highereducation2
- Fraudulent science – "biggest scientific fraud in history"
  - ". . .last January an investigation . . . concluded that not a single human embryonic stem cell had been cloned" - Cynthia Fox, Fortune Magazine, December 22 2006.

A 3-Stage Model for Digital Preservation

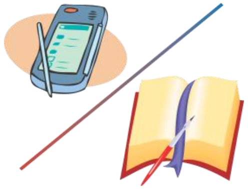

Digital Capture And Description (pre-ingest)
Archive (ingest)
Life Cycle Management (preservation)

- Digital capture – capture/create descriptive data and the archival master, in open and non-proprietary format.
- Archive – "fix" and store the digital object in safe place with appropriate markers of authenticity and reliability.
- Life Cycle – preserve and curate for scholarly use.

Essential Attributes of a Digital Object

Digital objects are both surrogate and born-digital with:

- A well-defined object architecture
- A persistent ID for citation integrity
- An audit trail
- Retention of versions
- Integrity – e.g. a periodically verified checksum on the archival master
- Descriptive and administrative metadata
- Digital provenance (part of metadata) recording life cycle events (e.g. migration)



Trustworthy Digital Objects

Two Qualitative Dimensions (from Archival Diplomatics):

- Authenticity - the object is what it claims to be
- Reliability - the object is capable of standing for the facts - it has not been inadvertently or intentionally modified to distort the facts

A Context for Authenticity

- The digital object is an artifact in its own right.
- Claims are made by the archivist*, curator, preservationist – the one who moves the object into the digital archival space.

*For simplicity, the term “archivist” is used to refer to the person or role which involves the storing of the digital object in a safe place, although from an institutional perspective this person might be a curator, cataloger, metadata creator, special collections librarian, subject specialist, archivist or some other designated role.

Claims to Be Made by the Archivist

- The 3-stage process is trustworthy
  - The end-to-end process is certified.
  - Certification process is transparent with appropriate documentation
- The resource is authentic:
  - Archivist makes claims on behalf of the depositor
  - Extensive liaison with the scholar or researcher is required.
- The description, context, and provenance – the metadata – is authentic. Based on:
  - Professional methods of the archivist
  - Establishes authorship, date, original, etc.

Achieving Authenticity
(A Proposal)

- The archivist supports claims by digitally signing* with a secure timestamp (based on public-key cryptography).

- What is needed:
  - The archival institution becomes its own certification authority.
  - A timestamp authority which produces a timestamp from a trusted source, establishing the real time when the object is signed.
  - An archival service for certificates which enables the researcher to obtain an authoritative record of the public key for any signer.

*See research of Haber, Stornetta, Maniatus, Baker, et al – see references, last chart.

A Scenario for a New Service

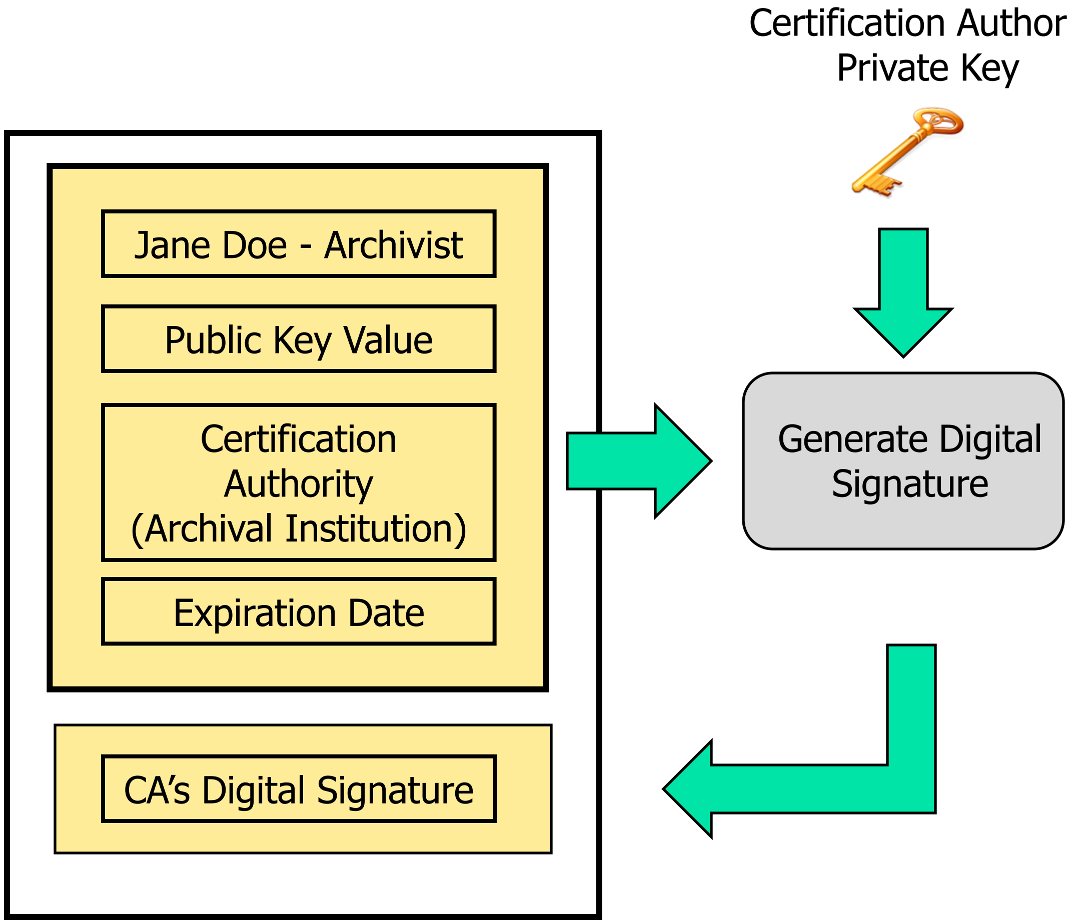
The Institution – The Certification Authority (CA)

- CA creates and archives its own master signing key.
- The archivist registers and creates a public and private key, keeping the private key secure.
- The CA issues an identify certificate for the archivist with name, public key, and expiration date.
- The certificate is archived and is available to researchers.

The Archivist

- In liaison capacity, works with the researcher and signs the digital object with the private key
- Submits archival master to the time stamping service to acquire a secure time stamp
- Archives the object with the signature and time stamp

A Public-Key Certificate*



Certification Authority Private Key

Jane Doe - Archivist

Public Key Value

Certification Authority (Archival Institution)

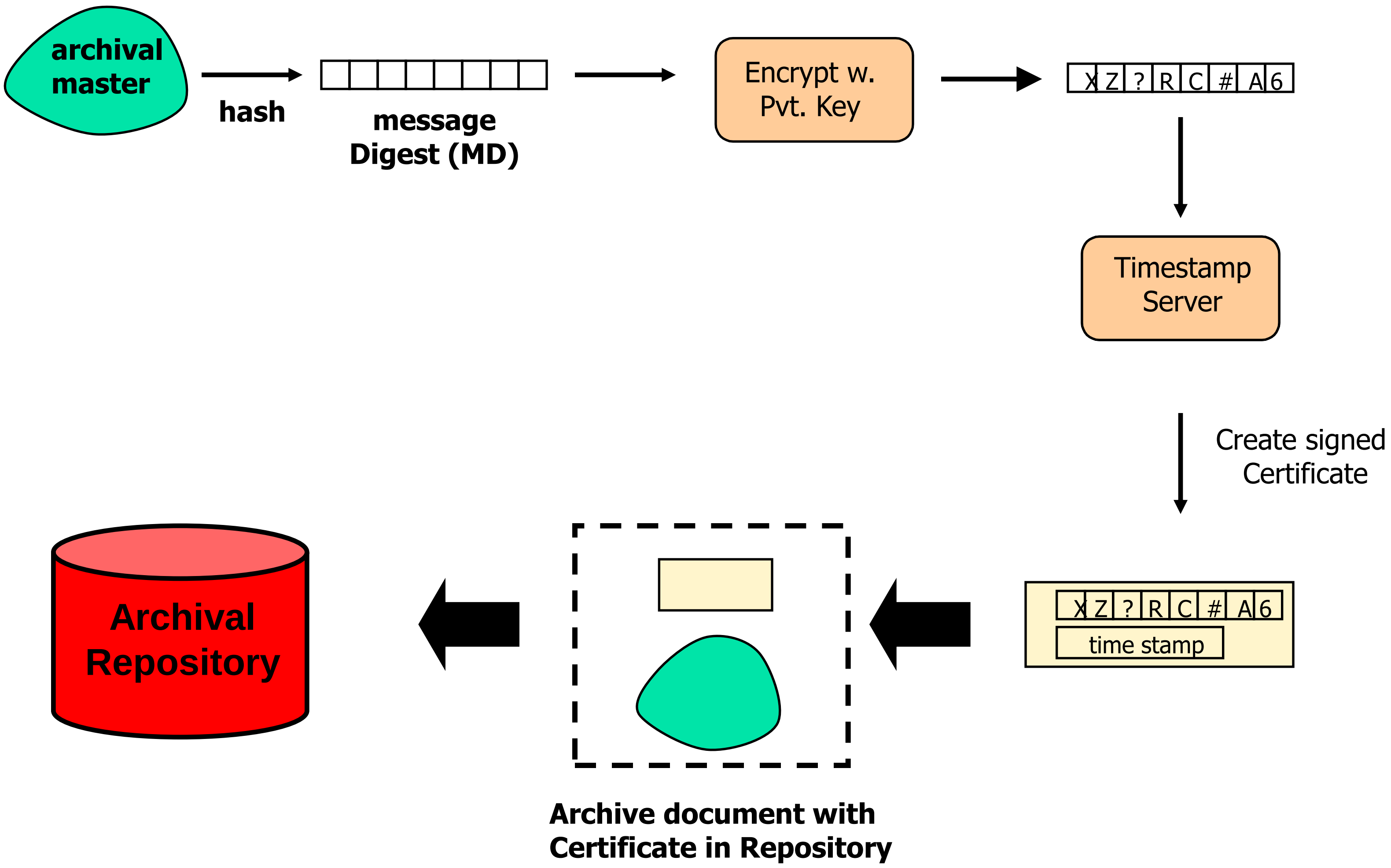
Expiration Date

CA's Digital Signature

Generate Digital Signature

*Ford, W. & Baum, M. *Secure Electronic Commerce: Building the Infrastructure for Digital Signatures and Encryption,* 2nd Editionl.

Signing and Time Stamping Process



archival master

hash

message Digest (MD)

Encrypt w. Pvt. Key

X Z ? R C # A 6

Timestamp Server

Create signed Certificate

X Z ? R C # A 6

time stamp

Archival Repository

Archive document with Certificate in Repository

A Scenario for Re-Use

A Researcher, some months or years hence, verifies the authenticity of the digital object.

- Researcher retrieves authoritative archivist’s certificate from the archive.
- Certificate shows identity of archivist with the public key.
- The public key can be used to decrypt the signature, re-compute the hash and verify that the digital content has not been modified since the time stamp.

The Archival Institution – Steps Forward

- Institution Becomes a Trusted Certification Authority
  - Certifying itself and archivists and managing the signing keys (security, revocation, renewal).
  - Modification of the object (e.g. migration) requires re-signing and attesting to the trustworthiness of the migration process

- Archivist – A New Role
  - Accountable for claims
  - Liaison role representing the depositor.

- Certification of the Institution as trustworthy (e.g. by using TRAC*)
  - A more quantitative approach
  - Repeatable to insure continued trust
  - Official "seal of approval"

**Trustworthy Repositories Audit & Certification: Criteria and Checklist.* http://www.crl.edu/PDF/trac.pdf

In Conclusion
Balancing Risk and Expense

- Assumptions:
  - The archival institution is more persistent than commercial organizations.
  - Authenticity is necessary for scholarly and research purposes.
- The Institution (e.g. academic library) must be certified as trustworthy for certificate signing and generating secure time stamps.
- The roles of librarians and archivists can be transformed to provide a new service for creating authentic digital objects.
- Collaboration with like-minded institutions is necessary to create a community of trust and share the cost of implementation.

Thanks for Listening!

Questions?

References

Ford, W. & Baum, M. *Secure Electronic Commerce: Building the Infrastructure for Digital Signatures and Encryption,* 2nd Edition, Upper Saddle River, NJ: Prentice Hall PTR.

Busey, J. (2004). A proposal for distributed digital time-stamping. Available at: http://ww2.cs.fsu.edu/~busey/samplework/DigitalTimestamping.pdf. Accessed July 9, 2008.

Haber, S. & Kamat, P. (2006). A content integrity service for long-term digital archives. *Proceedings of the 2006 Imaging Science & Technololgy Conference,* Ottawa, Canada, May 23 – 26, 2006.

Haber, S., Kaliski, B., & Stornetta, W. (August/1995). How do digital time-stamps support digital signatures? *Cryptobytes, RSA Laboratoriess 1,* (3), 14 – 15.

Haber, S. & Stornetta, W. (1991). How to time-stamp a digital document. *Journal of Cryptology: The Journal of the International Association for Cryptologic Research, 3,* (2), 99 – 111.

Maniatis, P. & Baker, M. Enabling the archival storage of signed documents. *Proceedings of the FAST 2002 Conference on File and Storage Technologies,* Monterey, California, Janurary 28-30, 2002.

Yamaji, K., Kataoka, T., Sonehara, N., & Namiki, T. (2008). Time stamping preprint server environment using Eprints 3. Poster session. Available at: http://pubs.or08.ecs.soton.ac.uk/77/ *The Third International Conference on Open Repositories, Southampton, UK, April 1 – 4, 2008.*