Digital Preservation 101 for Records Managers

Erin Faulder
Assistant Director for Digital Strategies
RMC

NY ARMA meeting
November 12, 2020

What’s a record?

records, records, records!

Records: recorded information on a fixed storage medium
(Compared to transitory memory or speech)

Records: evidence of activities that are set aside in an official recordkeeping system

Electronic Records: (from a technical perspective) bits that are used to re-constitute a record

Business record

A record created, received, and maintained as evidence and information by an organization, in pursuance of legal obligations or in the transaction of business.

Business records

Have a disposition

Destruction after a time period (e.g. 3 years) or conditions met (e.g. after employee separates from company)

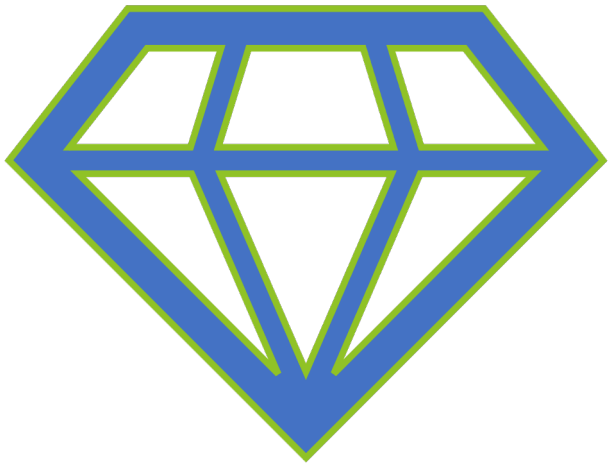
Permanent

Can be Active or Inactive

Identified as the Record copy vs Use or Convenience copy

Electronic record

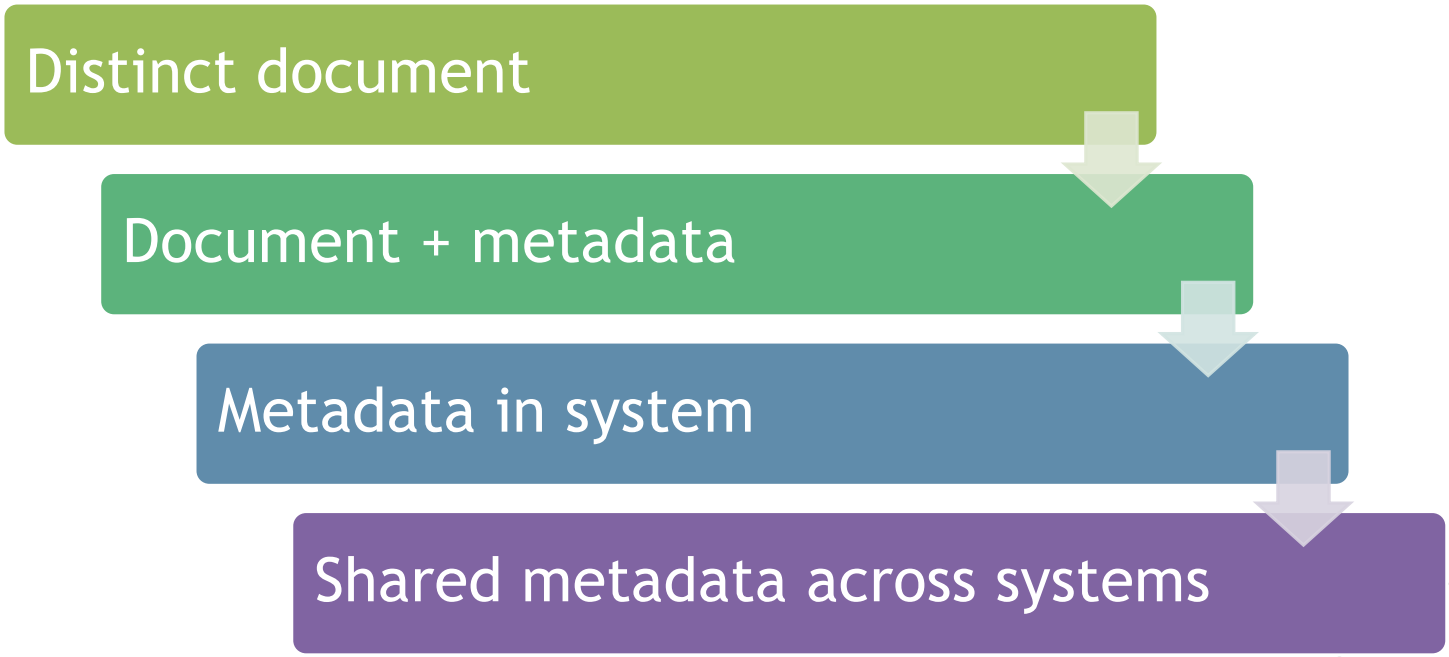
Data or information that has been captured and fixed for storage and manipulation in an automated system and that requires the use of the system to render it intelligible by a person.



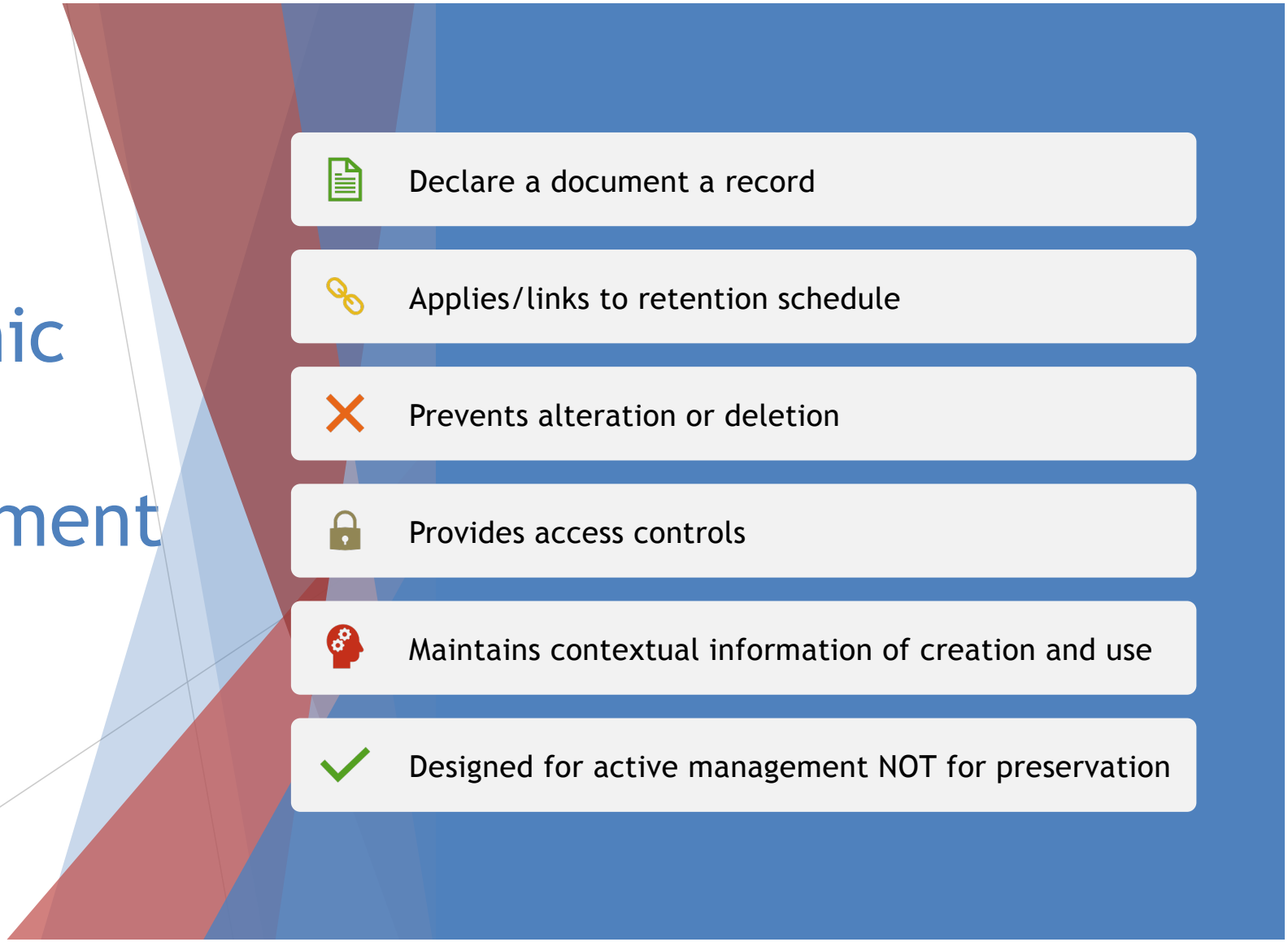
Digital business records proliferate. Copies emailed, drafts saved, versions overwriting one another.

Differentiating between record copy, draft versions, and use or convenience copy is critical for records management and digital preservation.

Electronic record creation and keeping changes



Electronic records management system

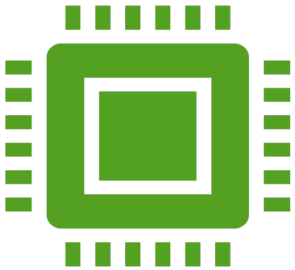


Trusted Digital Repository

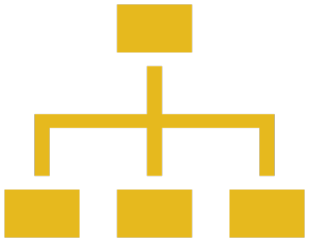
- Compliance with OAIS
- Administrative responsibility
- Organizational viability
- Financial sustainability
- Technological and procedural stability
- System security
- Procedural accountability

Electronic records do not survive by accident

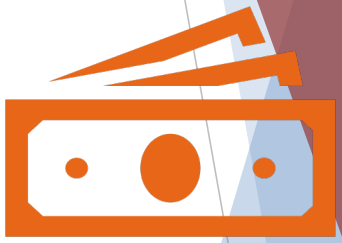
Technological obsolescence



Electronic records change in unpredictable, increasingly complex ways

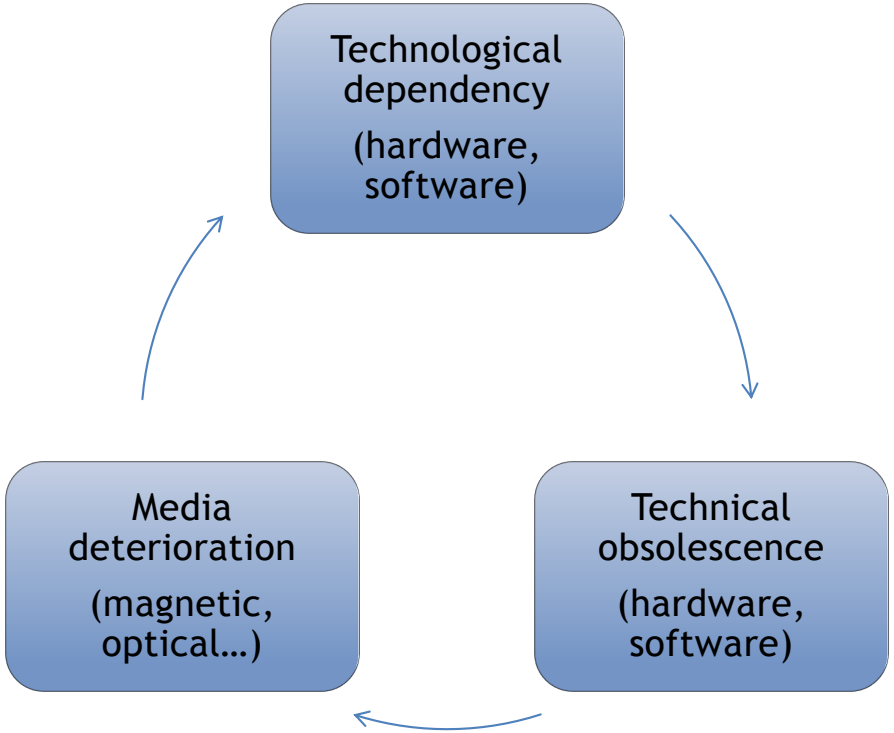


Change is foundational to technologies' business model



Little incentive to support older technology over time

Technical obsolescence cycle



Technological obsolescence timeline

Short-term (0-5 years)

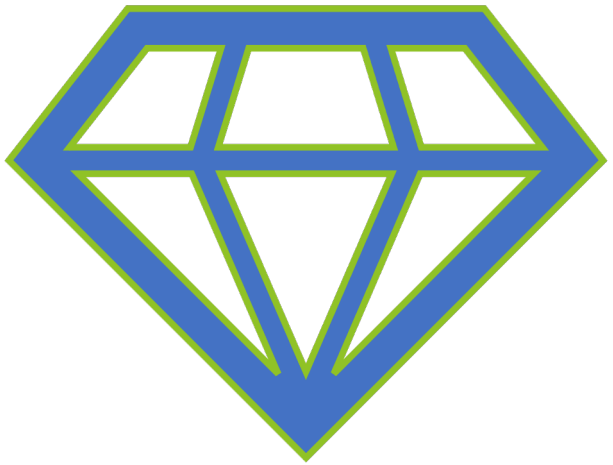
- Active = on-line storage
- “Normal” management issues
- Possible application or version upgrade

Mid-term (5-10 years)

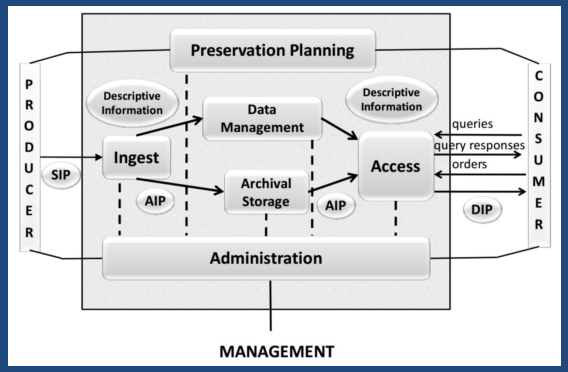
- Less active = near-line or off-line storage
- Multiplying version control issues
- System upgrades
- Hardware, software migrations

Long-term (10+ years)

- Least active = off-line storage
- Migration/conversion likely



Given a record's retention period and disposition, how much will technological obsolescence effect its longevity?



Digital Preservation Concepts

Digital preservation is active management

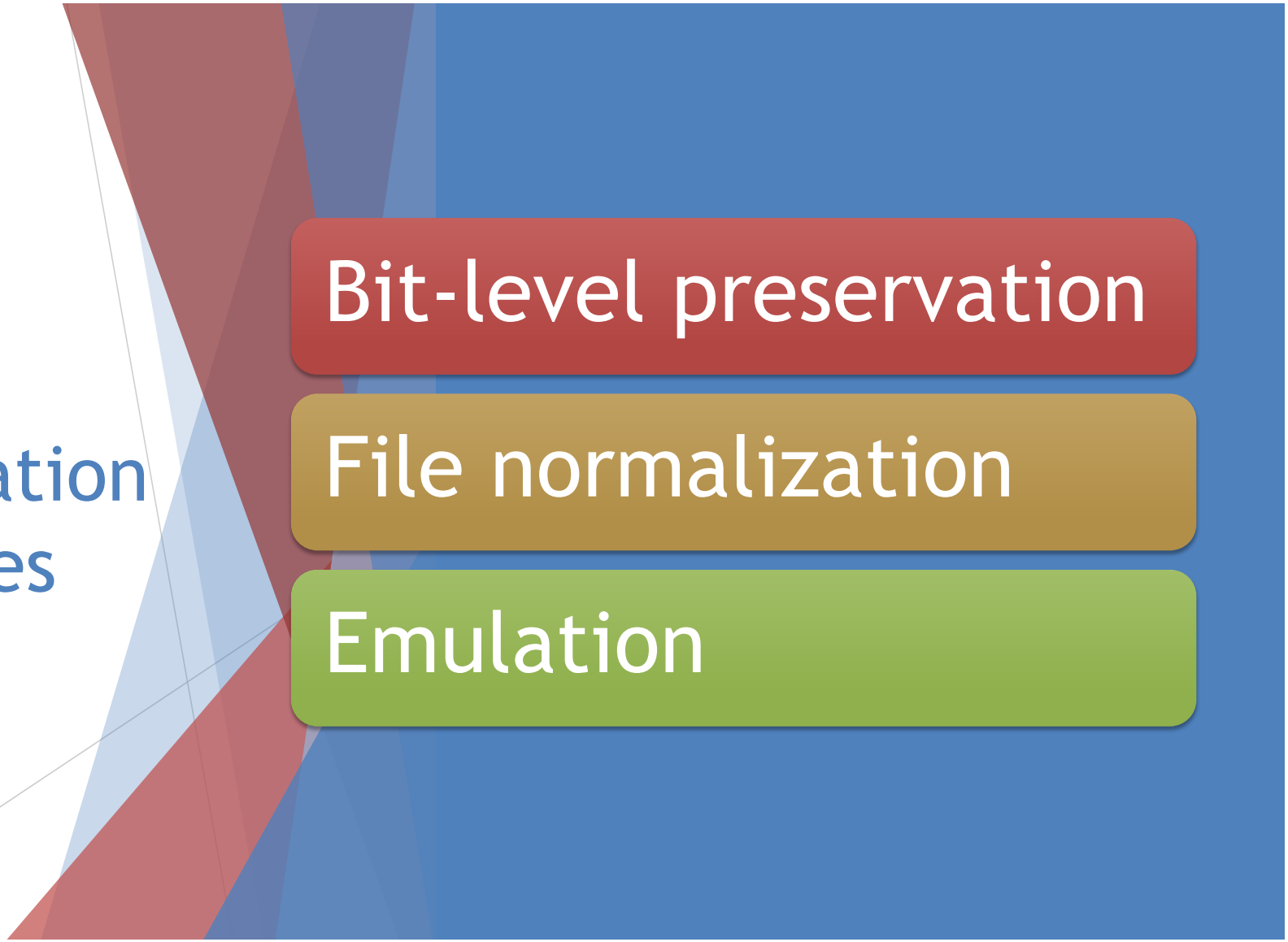
Preservation storage

Best practice: 3 copies, in 2 or more geographically distinct areas on 2 or more storage mediums

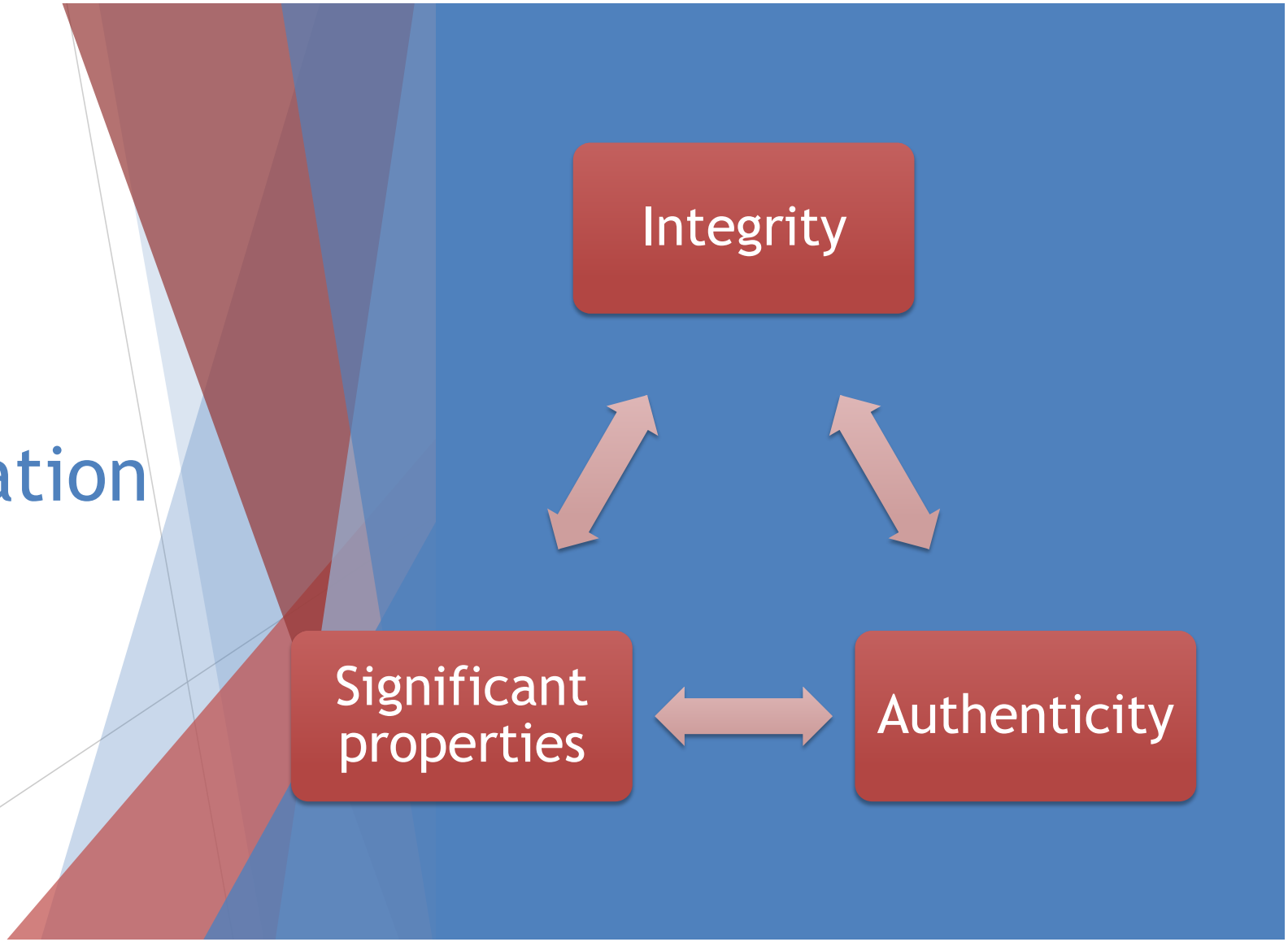
Not for active records

Only preserve ~5% of digital records

Digital preservation strategies



Digital preservation goals



Record integrity

Evidence of record integrity in digital preservation is usually maintained through checksums to ensure the file is whole and unaltered as a result of data loss, corruption, or tampering.

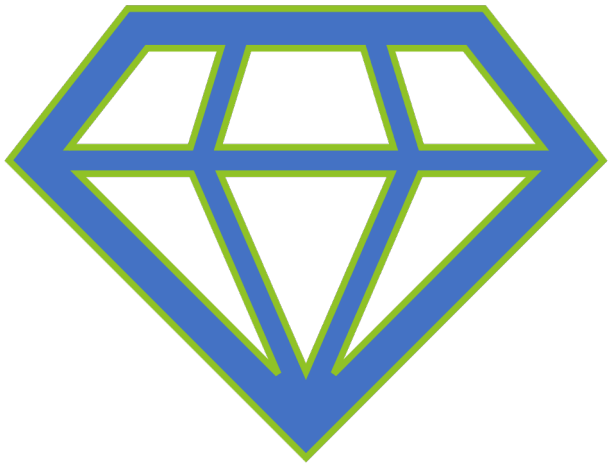
Record authenticity

An *authentic* record is one that can be proven to be what it professes to be, to have been created or sent by the person claiming to have created or sent it, and to have been created or sent at that time.

Evidence of digital authenticity relies on clear internal and external metadata to demonstrate that the preservation copy is the record copy.

Metadata of creation and use

- Administrative
  - Access information
  - Audit trails
  - Retention schedules
- Technical
  - File format information
  - Checksums
  - System information and requirements
- Descriptive
  - Bibliographic
  - Indexes/finding aids



How much evidence of authenticity do you need?

Significant properties

"The characteristics of an Information Object that must be maintained over time to ensure its continued access, use, and meaning, and its capacity to be accepted as evidence of what it purports to record." (2009)
- InSPECT Project Final Report, http://www.significantproperties.org.uk/inspect-framework.html

Content

- The intellectual substance of a document, including text, data, symbols, numerals, images, and sound.

Context

- Creator
- Business function that resulted in the record
- Dates of modification, access, creation (M/A/C properties)

The Who, What, When, Where, Why, How? Metadata

Structure

- Logical attributes of the record
  - Hierarchy, other keys to how the information is organized
  - How aspects of the record relate to itself and other records
- Physical attributes of the record (Rendering)
  - File format, software version, font, line spacing, colors, margins

ERIN FAULDER

Associate Archivist

Ithaca, NY [redacted]

EDUCATION

Simmons College, Boston, MA — 2013
M.S. in Library Science with an Archives Concentration
M.A. in History

Bates College, Lewiston, ME — 2008
B.A. Classical and Medieval Studies, Minor in Latin
Magna cum laude
Phi Beta Kappa

ARCHIVES EXPERIENCE

Cornell University, Ithaca, NY — February 2020 - Present
Division of Rare and Manuscript Collections (RMC), University Library (CUL)

Assistant Director for Digital Strategies

- Develop strategy to strengthen digital program in RMC by working with colleagues in Digital Lifecyle Services and Digital Consultation and Production Services.
- Work with other RMC Assistant Directors to support and coordinate ongoing collecting, arranging, describing, teaching, and access efforts with particular focus on digitized and born digital content.

Structured record

ERIN FAULDER
Associate Archivist
Ithaca, NY [redacted]
EDUCATION
Simmons College, Boston, MA 2013
M.S. in Library Science with an Archives Concentration
M.A. in History
Bates College, Lewiston, ME 2008
B.A. Classical and Medieval Studies, Minor in Latin
Magna cum laude
Phi Beta Kappa

ARCHIVES EXPERIENCE
Cornell University, Ithaca, NY February 2020 – Present
Division of Rare and Manuscript Collections (RMC), University Library (CUL)
Assistant Director for Digital Strategies
* Develop strategy to strengthen digital program in RMC by working with colleagues in Digital Lifecyle Services and Digital Consultation and Production Services.
* Work with other RMC Assistant Directors to support and coordinate ongoing collecting, arranging, describing, teaching, and access efforts with particular focus on digitized and born digital content.
* Work with CUL partners to streamline and automate ingest of content into our OAIS-compliant digital preservation storage (CULAR).
* Standardize digitization project workflows in coordination with curators, digitization staff, preservation unit, metadata services, access staff, and external stakeholders.
* Supervise one full-time staff person and one student worker.
* Develop python scripts to push metadata from external preservation and access systems into ArchivesSpace using the API to enhance discovery and support digital asset management.
* Accession born-digital materials using documented processes, averaging 6TB per year.

Unstructured record

Usability (behavior)

- How content is interactable
  - Hyperlinks in webpage
  - Database
  - GIS data

Records Managers' role in digital preservation

Good records management

...

Decreases records clutter so everyone can focus on the important records.

Provides easier access to the right records to appropriate users.

Reduces liability associated with unmanaged records that are improperly secured and past retention periods.

…supports better preservation

Only permanent records (or possibly longer-term, inactive records) are sent to preservation storage, reducing management and storage costs.

Authentic records are easily discoverable and deliverable to user in a renderable way 50 years after creation.

Reduced liability when permanent records are preserved and scheduled records are properly disposed.

Practical tasks

Filing structure

Group records with same disposition together

All creators agree to structure

File naming conventions

Include ISO (YYYY-MM-DD) date so files sort

Develop shared, meaningful vocabulary for common records

Deleting copies

Avoid digital hording and delete non-record copies as appropriate

What about …?

Email

Create folders or tags that match to disposition

Look at sent mail as well

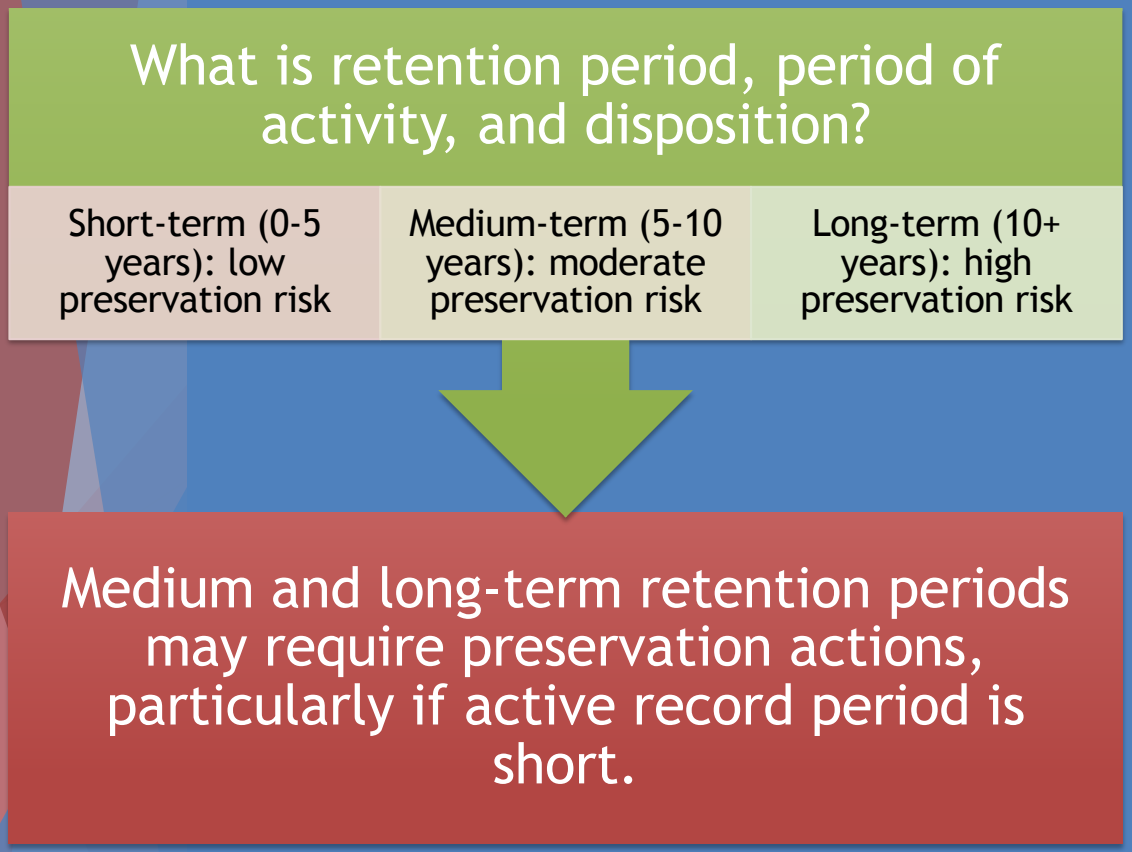
Social Media

What is the record?

Evaluating record for preservation

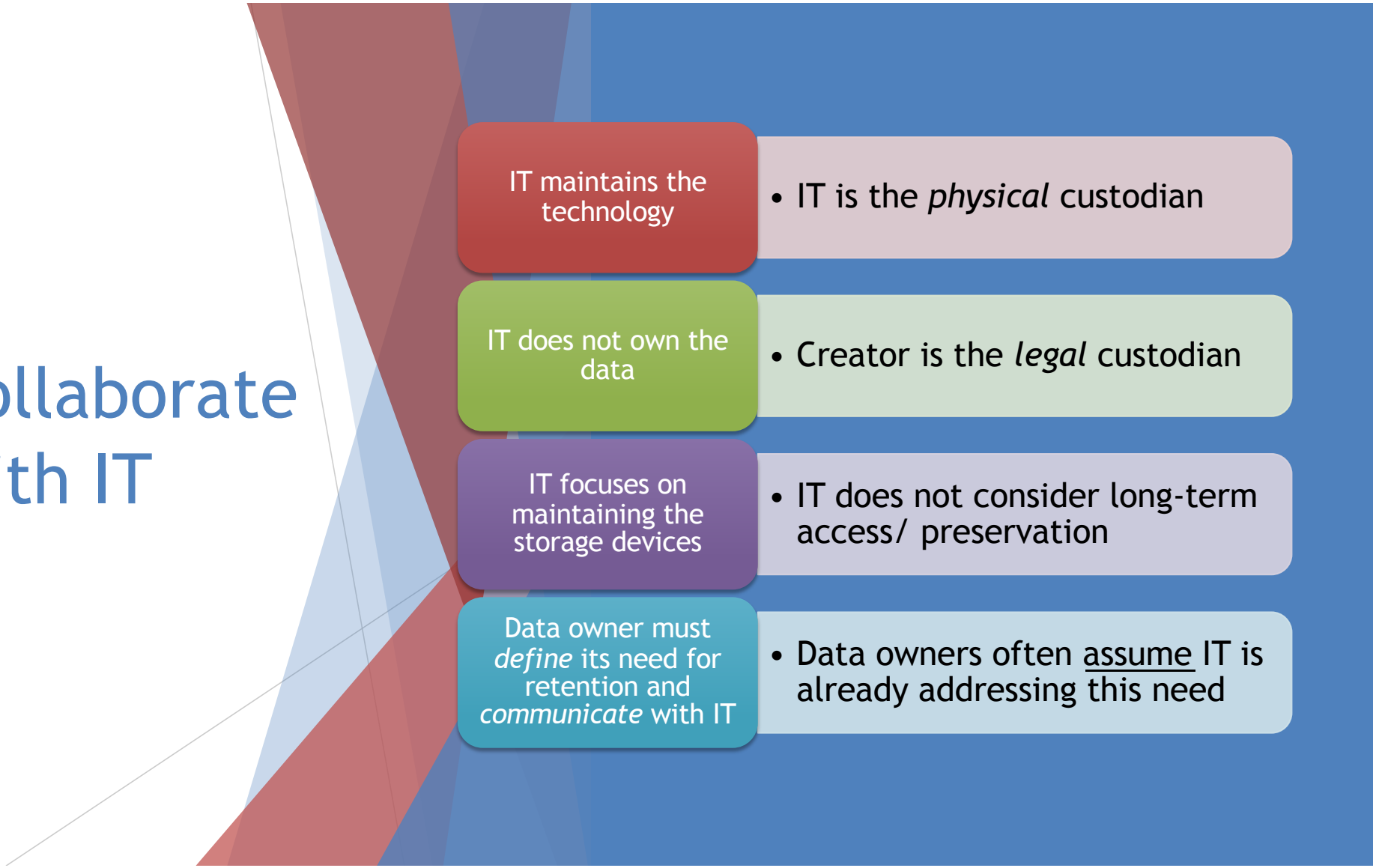
- Consider function of record over the container
  - Are there other forms of the information?
  - Is it the official record copy? Is it a convenience copy?
  - Is it an announcement? Press Release? Correspondence?
- Consider a record's significant properties
  - Content
  - Context
  - Structure
  - Usability
- Consider how the record will be removed from its active management system

Evaluating preservation risks in your records



It’s in everyone’s interest to identify and preserve the good records and delete the useless.

Collaborate with IT



Questions?