



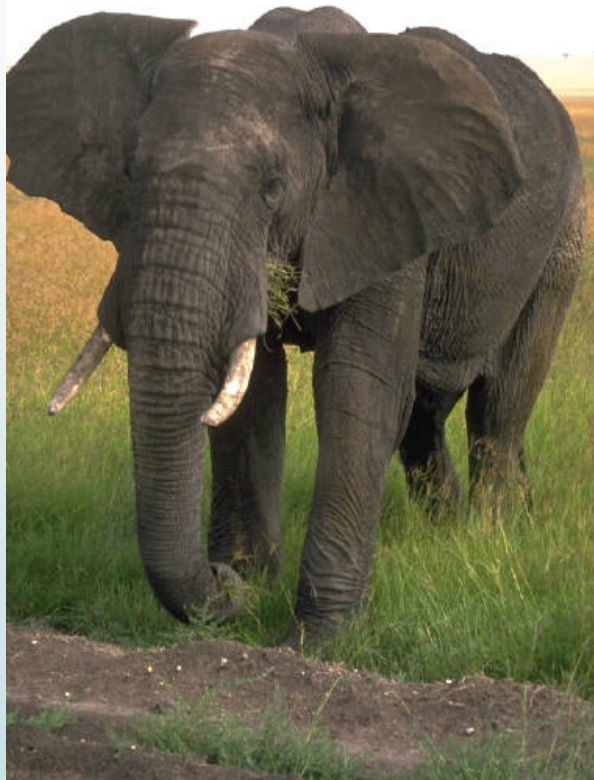
Archiving Digital Content: Creating an Environment for Success

Jackie R. Esposito

Penn State University Archivist

Best Practices Exchange

October 2015





Opportunities for Success

- ▶ Doing things on one perfectly coordinated decentralized platform is a massive act of hubris –*Clifford Lynch, 2009*
- ▶ **ADVOCACY** --
- ▶ Corporate Culture
- ▶ Keep the right people informed
- ▶ Don't assume
- ▶ Policy vs. Custodianship
- ▶ Don't be afraid



Opportunities for Success

- *ESTABLISHING AN E-RECORDS PROGRAM*
- Coordinate w/existing Archival and RM programs
- Identify initial priorities

- Assess technical capabilities
- Determine who can provide support

- Determine training needs
- Determine staffing needs
- Develop a prioritized, phased plan



OPPORTUNITIES

- ▶ *Planning Strategies* –
 - Develop an internal mandate
 - Determine measurable objectives
 - Establish specific timeframes
 - Build a team of players with:
 - Differing viewpoints
 - Common Goals
 - Shared Commitments
 - Understand your restraints
 - Maximize your opportunities



OPPORTUNITIES

- ▶ *General Strategies for Change* –
 - Break problem into manageable portions
 - Minimize complexity
 - Standardize
 - Compromise
 - BE FLEXIBLE

OPPORTUNITIES



► *Storage Infrastructure and Planning –*

Attributes: redundant, distributed, secure

Optionals: self-checking, healing, online

Examples:

External media library

Local/attached storage

Network attached storage

Distributed computing/storage systems

Cloud computing/storage



CORE REQUIREMENTS

- ▶ The key premise underlying the core requirements: repositories of all types and sizes preservation activities must be scaled to the needs and means of the defined community, in this case, designated archival business and operational records.
- ▶ 1. The repository maintains and ensures the integrity, authenticity and usability of digital objects within its scope of responsibility.
- ▶ 2. The repository will operate with an effective and efficient policy framework based on the strictures outlined in the institutions applicable policy structure.
- ▶ 3. The repository will acquire and maintain requisite contractual, legal, historical, fiscal, and organizational rights and fulfill those responsibilities.

CORE REQUIREMENTS



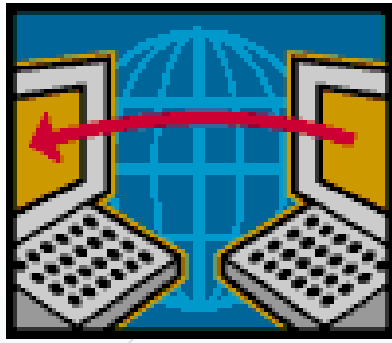
- 4. The repository commits to continuing maintenance of digital objects for identified units.
- 5. The repository will be maintained permanently with sufficient resources to commit to the continuing maintenance of the stored digital objects.
- 6. The repository creates and maintains requisite metadata about actions taken on digital objects during preservation as well as about the relevant production, access support, and usage process contexts before preservation.
- 7. The repository acquires and ingests digital objects based upon stated criteria delineated by the Records Management Advisory Committee that correspond to its commitments and capabilities.

CORE REQUIREMENTS



- ▶ 8. The repository fulfills requisite dissemination requirements.
- ▶ 9. The repository will develop and maintain a strategic program for preservation planning and action including a feasible and trackable disaster plan.
- ▶ 10. The repository has technical infrastructure adequate to continuing maintenance and security of its digital objects.

CORE REQUIREMENTS



- 11. The repository is subject to existing records retention schedules, data security classifications, privacy protocols, and access parameters.
- 12. The repository operates as the digital archives and as such follows the same protocols and policies as the existing archives.



REPOSITORY CORE CRITERIA

An Intellectual Context for IR:

Commitment to digital object maintenance

Organizational Fitness

Effective & Efficient Policies

Integrity, Authenticity & Usability

Dissemination

Adequate Technical Infrastructure

Legal & Regulatory Legitimacy

Acquisition & Ingest Criteria

Audit Trail & Metadata

Preservation Planning & Action

Seamus Ross (UToronto)



STRATEGIC GOALS

- 1) E-Records Management Assessment Criteria:

Are records retained for as long as legally, historically and operationally necessary regardless of format?

Can records be found and accessed by authorized users when needed?

Can records be processed, read and used when found?

Can records be interpreted within the context of business?

Can records be trusted?

Are records destroyed at the appropriate time, in the regular course of business?



STRATEGIC GOALS

2) Evaluate and Prioritize Risk – What is the risk...

- * of no records management or archives program?
- * of inadequate electronic records management?
- *to financial assets?
- *to physical assets?
- *to human assets?
- *to intangible assets?
- *probability of risk
- *severity of risk
- *risk tolerance

STRATEGIC GOALS



► 3) Evaluate Available Resources

Staff

Expertise/Experience

Budget/Capital

Current Technology

Facilities

STRATEGIC GOALS



- ▶ 4) Decide on Your Strategy

 - Start Simple

 - Focus on record series importance

 - Reduce volume, consider growth cycles

 - Reduce risk

- ▶ Work S-M-A-R-T

 - S – Sensible

 - M – Manageable

 - A – Achievable

 - R – Reasonable

 - T – Trackable



STRATEGIC GOALS

- 5) Develop Strategic Action Plan –
 - Based on risk and resource assessment analysis
 - Develop short-term, mid-term and long-term action plans

 - Identify specific tasks/activities/tactics
 - Define roles and responsibilities

 - Plot realistic timelines and milestones
 - Report progress

WORKING GROUPS & HOT TEAMS



- ▶ 1) Identify Participants:
 - Content Stewards
 - Working Group Members
 - Stakeholders
 - Records Liaisons/Coordinators
 - Administrative Supporters
 - Budget Executives



WORKING GROUPS & HOT TEAMS

► 2) Elements of Productive Teams:

Resource Allocation – team members, operating funds, access to technical support, equipment, facilities

Cooperation/Synergies

Communication

Permission to Try/Permission to Fail

Flexibility

WORKING GROUPS & HOT TEAMS



► 3) Create “Just-in-Time” Hot Teams:

Small, well-defined projects

Short turnover time

Clear goals and outcomes

Ready results



DIGITAL CURATION MESSAGE

► 1) Five Principles of E-Records

E-Records are records

Archival principles still apply

Early intervention is key

Doing something trumps doing nothing

Understanding Records Life cycle



DIGITAL CURATION MESSAGE

- 2) E-Records Management
- Distributed creation leads to decentralized recordkeeping
- RM practices are inconsistent
- Identifying records series is complicated
- Multi-disciplinary solutions are necessary



DIGITAL CURATION MESSAGE

3) Digital Curation: Essential Steps

Appraisal – at the point of creation including technical characteristics

Accessioning – acceptable formats, adequate documentation, verification of accuracy

Processing – arrangement, evaluating technical characteristics, conversion to other formats

Description – preliminary inventory during accessioning, develop standard descriptions [bibliographic record, series descriptions, administrative histories], develop user guides

Preservation Challenges – technological dependency, technical obsolescence, media deterioration



DIGITAL CURATION MESSAGE

► 4) File Formats Matter

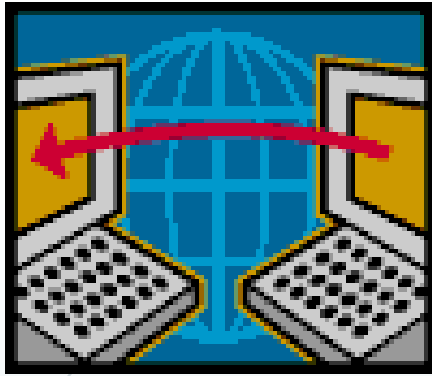
Best formats are open source formats

No format is “permanent”

Most require conversion, migration, emulation

Some are difficult to manage

Control formats at creation if possible



DIGITAL CURATION MESSAGE

► 5) Requirements for Preservation

Perfect translation

Accessibility

Functionality

Uniformity

Authenticity

Low cost

Limited intervention

DIGITAL CURATION MESSAGE



- ▶ 6) Preservation strategies
 - Migration
 - Normalization
 - Emulation
 - Conversion to hardcopy



WORKFLOWS: ACCESSION/INGEST

- Survey for formats
- Check for Privacy issues/Scan for Viruses
- Run Checksums
- Rehouse
- Preliminary Description/Metadata
- Document Access Restrictions
- Assign Location/Secure Digital Storage
- Assign Processing Priority



MANAGING ELECTRONIC CONTENT

► 1) Operational Challenges:

Proliferation of electronic records

Ease of duplication and dissemination

Cost of storage

Limited access and control

Lack of coordinated classification between e-records and established retention schedules

E-records Management Paralysis

Policies without operational guidelines, education and compliance reviews

E-records responsibilities – who is in charge?

Records/Content Creators/Owners vs. Records Custodians



MANAGING ELECTRONIC CONTENT

► 2) Legal and Regulatory Challenges:

Accuracy and Completeness

Discoverability

Records posing greatest risk are the most difficult to manage and control -- Email

Evolution of regulatory requirements

Interpretation of requirements and case law



MANAGING ELECTRONIC CONTENT

► 3) Technology Challenges:

Obsolescence

Software/hardware dependencies

Limited stability of electronic storage media

Migration/reformatting/conversion/emulation

Information processing systems are not recordkeeping systems

Volume vs. value

Implementing a solution with unrealistic expectations

Failure to address migration/conversion costs

Assuming a data backup procedure is the same as an e-records retention strategy



MANAGING ELECTRONIC CONTENT

► 4) Current Computing Environment:

Enterprise-wide or business-process specific systems

Database management systems

Data and document repositories

E-Mail Management

Websites

BYOD – Bring your own device

Social Media

Desktop Applications

Messaging

Cloud Computing

Gaming Systems

MANAGING ELECTRONIC CONTENT



► 5) Models for Electronic Records Programs:

Custodial

Non-custodial

Hybrid



MANAGING ELECTRONIC CONTENT

➤ 6) Methods for Providing Access to E-records

Direct web

FTP download

Public Terminals

Media provided on request

Personal Device Applications

Box Components



CURATION MANAGEMENT PROTOTYPE

► 1) A Digital Repository:

Plays a vital role in the curation of digital materials

Offers a convenient way to store, manage, reuse and curate a variety of digital materials

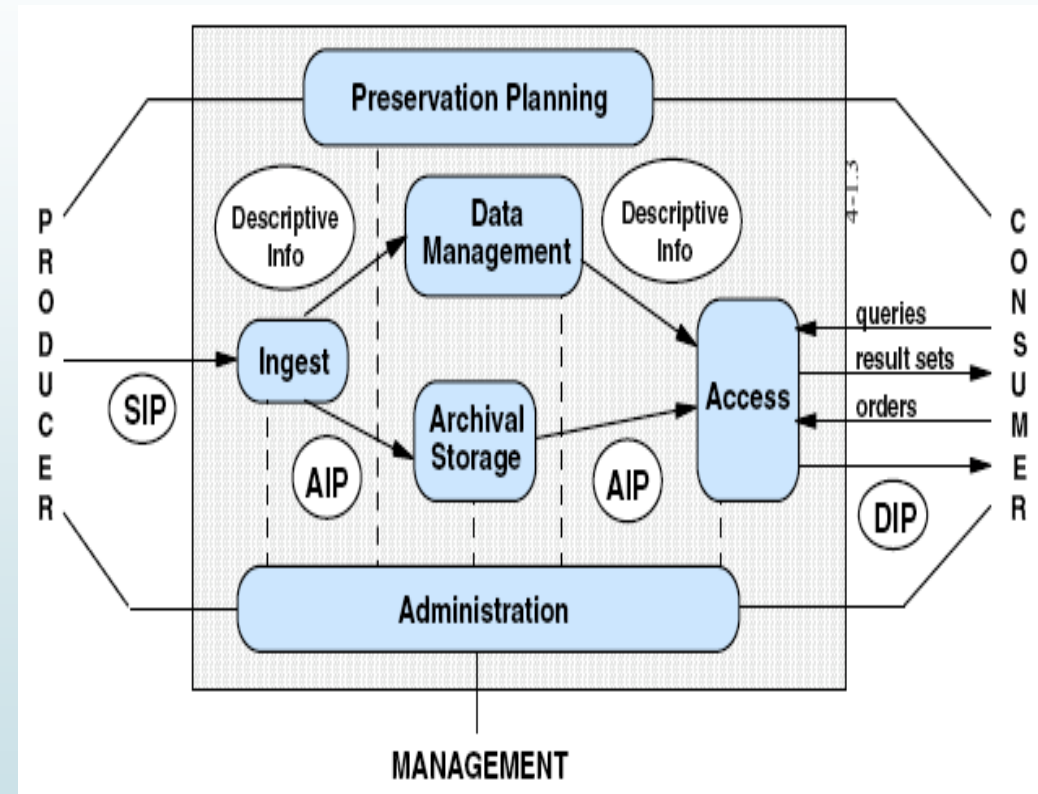
Different digital storage initiatives including institutional repositories, digital archives, digital libraries

Serve many different communities with differing functions and needs

CURATION MANAGEMENT PROTOTYPE



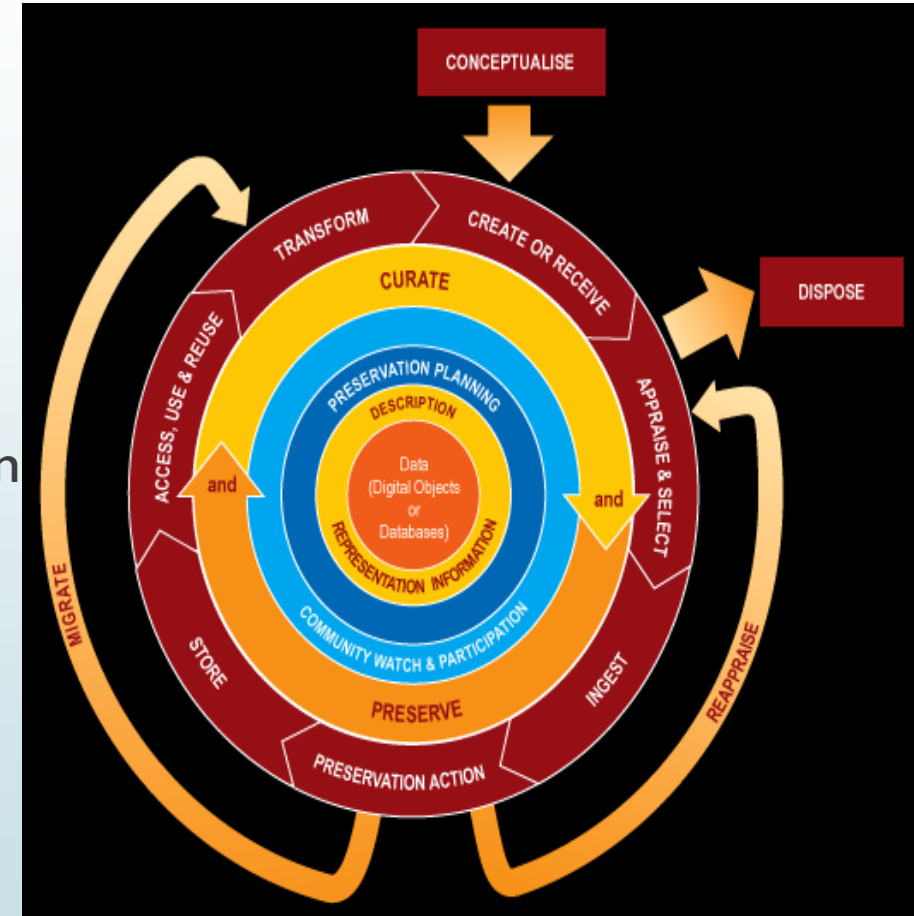
- OAIS: Open Archival Information System Reference Model
- Submission Information Package (SIP)
- Archival Information Package (AIP)
- Distribution/Dissemination Information Package (DIP)





DCC CURATION LIFECYCLE MODEL

- Data Conceptualization
- Data Creation or Receipt
- Data Appraisal and Selection
- Data Ingest or Transfer
- Data Preservation
- Data Categorization or Classification
- Data Storage
- Data Access, Use and Reuse
- Data Transformation
- Data Migration
- Data Reappraisal
- Data Disposal





TRAC: Trustworthy Audit Checklist

- ▶ TRAC: Trustworthy Repository Audit Checklist – Builds on OAIS functions

Audits:

- ▶ Organizational Infrastructure
- ▶ Digital Object Management
- ▶ Technologies, technical infrastructure and security
- ▶ Comparing local capabilities against a set of core criteria
- ▶ Objective measure and/or planning document
- ▶ http://www.crl.edu/sites/default/files/attachments/pages/trac_0.pdf.

GUIDELINES & STANDARDS



- NISO's Framework for Building Good Digital Collections:
 - Organizational planning and implementing digital collections initiatives
 - Funding organization to develop good digital collections
 - Sets of principles for operational guidance
 - Identifies existing resources that support development of sound local practices
 - Encourages community participation
- Composite of Functional Entities: Manage Change of Preservation (InterPARES)

GUIDELINES & STANDARDS



- ▶ Preservation Metadata: Implementation Strategies (PREMIS) Data Model
- ▶ DoD 5015.2 and FISMA Federal Requirements for Department of Defense as of Fall 2011
- ▶ ISO 15489, 19005-1, and 15386 (Dublin Core)



GUIDELINES & STANDARDS

- ▶ ISO16363:2012 –Audit and Certification of Trustworthy Digital Repositories
- ▶ Library of Congress Standards – www.loc.gov/standards/
- ▶ Email Standards – www.dpconline.org/newsroom/latest-news/805-email-tomorrow...
- ▶ HIPAA, FERPA, PCIS



Responsible Management

Survey and Inventory Current Holdings

Locate Existing Holdings

Count and describe all identified media

Prioritize Collections for further treatment

Repeat these steps every time you receive new media



Technical Steps for Readable Media

Clean, dedicated computer

Use a Write blocker

Create a directory on the clean machine for the current project

Copy data from physical media to subdirectory

Generate a copy of the disk directory

Generate and record a checksum

Create a readme file

Copy the project directory to trustworthy archival storage

Return the original physical media to storage

Create an associated finding aid

CURRENT TOOLS



- ▶ Archive-It -- Internet Archive Storage and Access Site for web archiving
- ▶ Archivemataca – http://archivemataca.org/wiki/index.php?title=Main_Page
- ▶ Archivists Toolkit –<http://archiviststoolkit.org/>
- ▶ AIMS/Hydra – <http://www2.lib.virginia.edu/aims/>
- ▶ Duke Data Accessioner –
<http://library.duke.edu/uarchives/about/tools/data-accessioner.html>
- ▶ DRAMBORA -- <http://www.repositoryaudit.eu/>



ESSENTIALS

► Understanding the Issues:

Enterprise System Data (financial, student records)

Employee desktop data (email, files)

Research data sets

Faculty pre-prints

E-theses and Dissertations

Websites

Publications

Personal digital archiving

Social media



ESSENTIALS

► Digital Preservation Considerations:

Worth – the work to design and create a digital product adds value to the information contained in the documents that serves as resources.

Maintenance – the persistence of digital products requires careful attention to the maintenance of content (bits and bytes) functionality (how the bits work in the system).

Longevity – Digital preservation is not absolute, one time only. It depends on the continuing transformative impact of the digital product on the information work of end-users.



ESSENTIALS/KEY COMPONENTS

- ▶ Hardware:

 - Servers, storage and back-up systems

- ▶ Networking and connectivity:

 - Power

 - HVAC

 - Security

- Software:

 - Support back-end, front-end, and “middleware” applications

 - Vendor-purchased, open source, and/or hybrids



ESSENTIALS/KEY COMPONENTS

- Trust

- Reliable, long-term access

- Responsibility for long-term maintenance of digital resources

- Commonly accepted conventions and standards

- Policies, practices, and performance that can be audited and measured (TRAC)

- Relationships – an ever-changing web of personal, organizational, cooperative, conceptual, and contractual understandings, agreements, and consensus points.

ESSENTIALS/KEY COMPONENTS



Controls

Deposit

Access

Security

Disaster/Risk Mitigation

Review

Reproduction

Content

EFFECTIVE TEAMS/ ENGAGED STAKEHOLDERS



Recommended RM Infrastructure:

RMAC: Records Management Advisory Committee –

(sample membership includes: Reps from Legal
Department, Archives, IT, Key Business Units)

Records Manager (and affiliated positions)

Departmental Records Coordinators (or Liaisons)

Archives Presence/Leadership

IT Partnerships



EFFECTIVE TEAMS/ ENGAGED STAKEHOLDERS



- ▶ ROLES: Repository Managers --

Stewards entrusted with content, collections, and resource management (digital archivists)

Maintain content and metadata in forms that are archival, reusable, and interoperable

Enforce conditions of use, rights, ownership

Guide repository communities

Guide outreach, marketing, and collection development

Guide capture best practices to ensure long-term preservation and access



EFFECTIVE TEAMS: ROLES

- ▶ **System Administrators**
 - Manages hardware and infrastructure
 - Responsible for security
 - Coordinates software development and preservation activities at the system level
 - Resource allocation – equipment, staffing, licensing
- ▶ **Content Creators/Depositors**
 - Workflows
 - Standards
 - Best practices
 - System-driven format transformations
- ▶ **End Users -- Everyone**



ENGAGED STAKEHOLDERS

► Stakeholders:

High-level Administrators

Funders and broader community supporters

Advocates

Strategic decision makers

Organizational direction setters

Budget administrators

Records Liaisons



DIGITAL REPOSITORIES

Mission Statement: to provide reliable, long-term access to managed digital resources to its designated community, now and into the future.

Data Classification/Categorization System

Public

Restricted

Classified

Functionality:

Content Deposit

Metadata Creation/Template

Discovery and Access

Usage Controls

Security

Disaster recovery

Migration

Preservation planning



DIGITAL REPOSITORIES/ POLICY AREAS

- Organizational
Governance
Community organization

Participation
Funding support

- Content and Metadata
Deposit and usage
Identification
Metadata

File Format support
Ownership



DIGITAL REPOSITORIES/ POLICY AREAS

- ▶ Access, Rights and Usage
 - Authentication
 - Copyright
- ▶ Preservation and Infrastructure
 - Preservable formats
 - Repository termination
 - Technical infrastructure

POLICIES



- ▶ Collection Development
- ▶ Submission Agreements
- ▶ Records Management
- ▶ Technical Infrastructure:
 - Hardware
 - Discovery System
 - Preservation/Permanence

Service Level Agreements

Use Agreements

Disaster Recovery

Dark Archiving



MESSAGE TO THE MASSES

- ▶ **EDUCATION on Steroids**

 - Training, Training, and More Training

 - In-Class and Online Modules

 - Annual Recertification

 - System template data entry practicum

- ▶ **Community of Users/Liaisons**

 - Members only spaces

 - Listservs

 - Blogs

 - Mailing lists

 - Wikis

 - Social media



MESSAGE TO THE MASSES

► Publications

Press Releases

White papers

Brochures /Flyers

Talking points

Presentations

Annual Reports

FAQs

Sales brochures

► Recruit Advocates

Enlist stakeholders

Satisfied content creators

Best practices

End user case studies



MEASURING SUCCESS

➤ Repositories MUST be both useful and useable

Annual TRAC audits

Web statistics

Surveys

Focus groups

“Early Adopters” feedback

File download count

Anecdotal evidence

➤ Repository Preservation

Content integrity includes Documentation of all policies, strategies, and procedures

Use of persistent identifiers

Recorded provenance and change history for all objects

Verification mechanisms

Attention to security and disaster recovery requirements

Routine audits



MEASURING SUCCESS

- Content maintenance includes

- Robust computing and networking infrastructure

- Storage and synchronization of files at multiple sites

- Continuous monitoring and management of files

- Programs for refreshing, migration and emulation

- Creation and testing of disaster prevention and recovery plans

- Periodic review and updating of policies and procedures



Integrating Archival Expertise

Ten Areas of Archival Expertise –

Ownership – who is the legal owner of digital property?

Donor Relations – factors to address when working with donors

Intellectual Property – rights, rights, and more rights

Appraisal – Significance of material in the context of the institution's collecting mission

Context of Creation and Use – Circumstances under which materials were produced



Integrating Archival Expertise

Authenticity – Factors to determine whether materials are accurate, reliable record

Restrictions on Access and Use – Limitations

Transfer of Ownership – Terms and conditions at the point of transfer

Permanence – Preservation into the foreseeable future

Collection-Level Metadata – descriptive metadata for discovery and access



FIVE STAGES OF DIGITAL CURATION

- 1) Acknowledge – Understand the scope of your local issues
- 2) Act – Initiate digital curation/preservation projects
- 3) Consolidate – Build your program framework
- 4) Institutionalize -- incorporating the larger environment and rationalize programs as part of the operational imperative *****MISSION CRITICAL*****
- 5) Externalize – embrace inter-institutional collaboration and dependency
 - You are NOT alone
 - Learn from trailblazers
 - Do NOT re-invent the wheel
 - Right size your program for your institution, even if that means outsourcing it.



SUMMARY RECOMMENDATIONS

- ▶ Stop waiting and start proactive engagement locally
- ▶ Stake a claim in the production cycle

- ▶ Start retraining and repurposing staff
- ▶ Be a doer, not a broker, whenever possible

- ▶ Consider digital curation collaborations
- ▶ Actualize collaborative engagement

TOMORROW



► NEEDS –

Staff

Standards

Policies

Partnerships

Models

Tools

Workflows

Infrastructure/Tech Support

► WANTS –

A positive attitude toward change

AND

A flexible response structure



SUMMARY

