Digital Repositories

Presenter:

Lori Lindberg MLIS CA

©2016 Society of American Archivists

Digital Archives Specialist (DAS)

Curriculum and Certification Program offered by SAA:

Foundational Courses—*must pass 4*

Tactical and Strategic Courses—*must pass 3*

Tools and Services Courses—*must pass 1*

Transformational Courses—*must pass 1*

Course examinations are administered online.

Getting Started…

Introductions, Logistics, Goals

DAS Core Competencies

This course contributes to the following DAS Core Competencies:

1. Explain the nature of digital records and their lifecycle.

2. Communicate and define terminology, requirements, roles, and responsibilities related to digital archives to a variety of stakeholders.

3. Formulate strategies and tactics for appraising, acquiring, describing, managing, organizing, preserving, and delivering digital archives.

4. Incorporate technologies throughout the archival lifecycle.

5. Strategically plan for the sustainability of digital archives.

6. Employ standards and best practices in the management of digital archives.

7. Design a defined set of services for designated community

Learning Outcomes

At the end of this course, you should be able to:

- explain the basic decisions underlying the development of a digital repository program
- differentiate between the components necessary to implement a viable digital repository service

Learning Outcomes

- evaluate existing and proposed repository initiatives at your local institutions for identified elements of a successful deployment
- reference existing digital repositories and the characteristics they illustrate
- identify areas in which you might build your knowledge base and/or skill sets to meet the needs of a digital repository program

In Other Words:

We will review…

- what YOUR digital repository service needs are
- what choices YOU need to make
- standards, best practices, models, and resources that can guide and support YOUR decision-making process

7

Today's "Nots"

We will not…

- focus on how to digitize
- focus on how to create metadata
- focus on specific hardware/software products; or make product recommendations
- focus too heavily on any one component

8

"…offers a convenient way to store, manage, reuse and curate a variety of digital materials…"

Module 1:
Identifying Repository Needs, Motivations, and Key Decision Points

Maureen Pennock, DCC TechWatch - Eprints

9

Why Repositories? (2005)

"All institutions which maintain long-term collections of paper, recorded sound and recorded images have to make the shift to managing long-term electronic storage."

PrestoSpace

10

Why Repositories? (2011)

"The future of scholarly communication is a Web in the Cloud"

Thorny Staples

11

SAA Code of Ethics

V. Authenticity and Integrity

■ Archivists strive **to preserve** and **protect** the **authenticity** of records in their holdings by **documenting their creation and use** in hard copy and electronic formats.

■ They have a **fundamental obligation to preserve the intellectual and physical integrity** of those records. Archivists may not alter, manipulate, or destroy data or records to conceal facts or distort evidence.

12

SAA Code of Ethics

VI.   Access

- Archivists strive to **promote open and equitable access** to their services and the records in their care without discrimination or preferential treatment, and in accordance with legal requirements, cultural sensitivities, and institutional policies.

- Archivists recognize their responsibility **to promote the use of records as a fundamental purpose** of the keeping of archives…

13

SAA Code of Ethics

VIII.   Security/Protection

- Archivists **protect all documentary materials** for which they are responsible and guard the**m against defacement, physical damage, deterioration, and theft**…

14

Worksheet 1:
What Motivates YOU

When considering local digital repository service drivers and needs, what are your motivations?

15

Worksheet 1: What Motivates YOU?

As the archivist (or other information professional) in charge of a digital repository service, you need a clear understanding of the needs and motivations driving your initiative.

This exercise begins to build YOUR digital repository profile, by considering the following:
- What are YOUR motivations for developing a digital repository?
- Which ones are MOST important to your institution, your mission, and your values? Why?
- Which ones do you consider to be LESSER motivations? Why?

Motivator	Value: High	Value: Medium	Value: Low
Convenience: Digital objects in central storage are easy to access and manage			
Management: A system that is not dependent on operating systems or file systems			
Access and Use: Interoperability and integration in other contexts			
Authenticity: Verifiable and certifiable			
Security: Access controls, rights and usage management			
Preservation: Durability and authenticity over the very long term			
Other:			

Module 2:
The Repository Landscape;
Standards and Best Practices;
Content and Metadata Choices

16

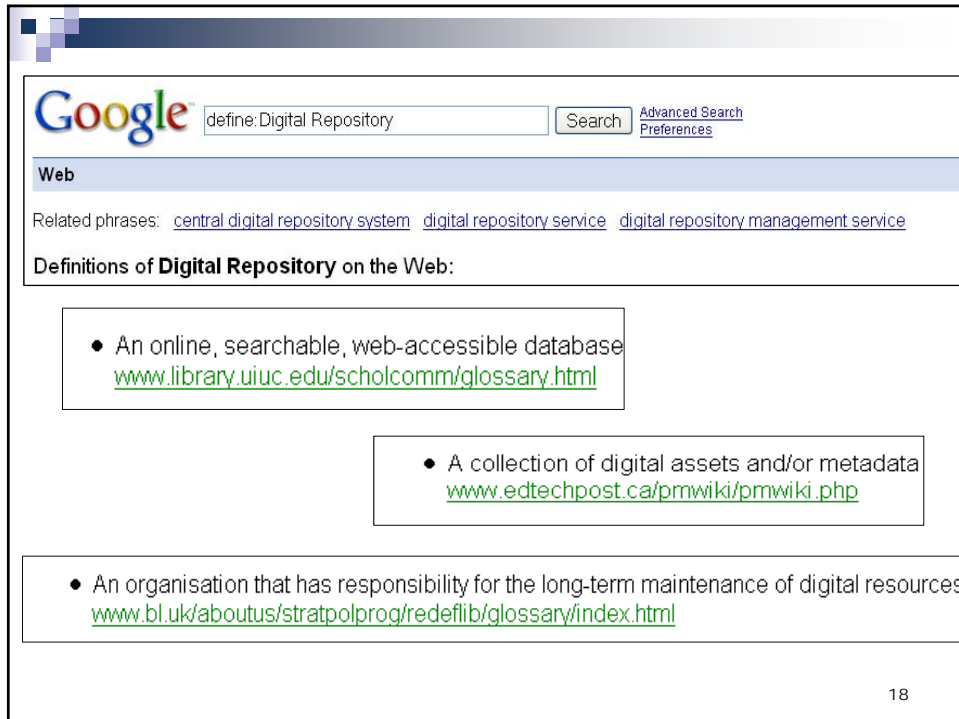Digital Repositories

DEFINITIONS

17

Google define:Digital Repository   [Search]   Advanced Search
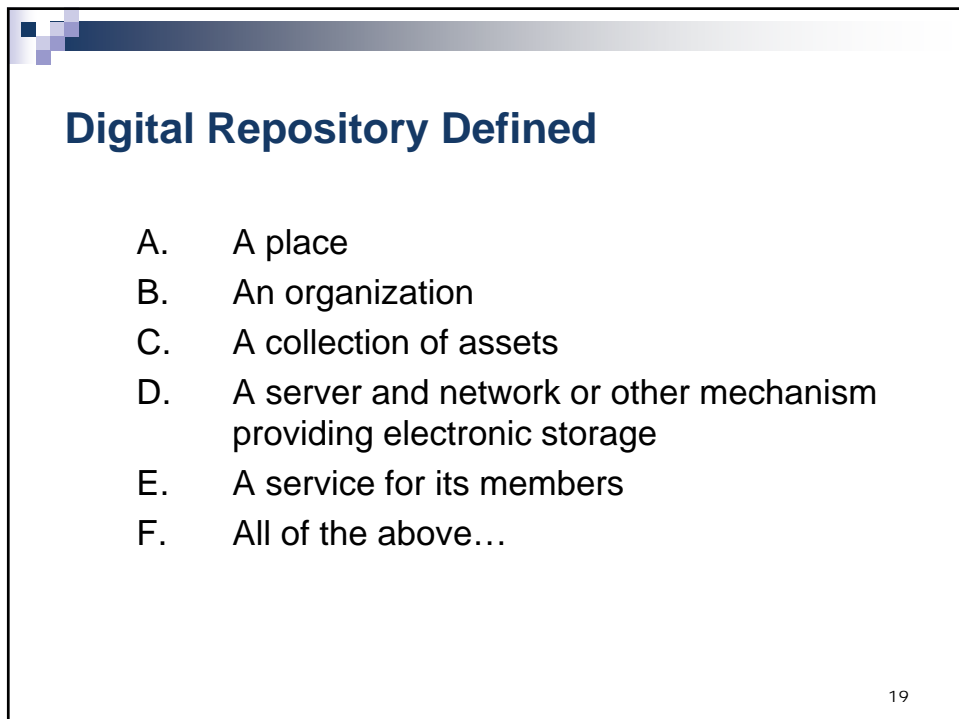   Preferences

Web

Related phrases: central digital repository system   digital repository service   digital repository management service

Definitions of **Digital Repository** on the Web:

- An online, searchable, web-accessible database
  www.library.uiuc.edu/scholcomm/glossary.html

- A collection of digital assets and/or metadata
  www.edtechpost.ca/pmwiki/pmwiki.php

- An organisation that has responsibility for the long-term maintenance of digital resources
  www.bl.uk/aboutus/stratpolprog/redeflib/glossary/index.html

18

Digital Repository Defined

A. A place
B. An organization
C. A collection of assets
D. A server and network or other mechanism providing electronic storage
E. A service for its members
F. All of the above…

19

Digital Repository Defined

A. A place
B. An organization
C. A collection of assets
D. A server and network or other mechanism providing electronic storage
E. ~~A service for its members~~
F. All of the above…

20

Digital Repository Defined

- 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
- can take many forms and carry out many different functions, to service many different communities
- can mean a number of different digital storage initiatives, which are often also referred to as
    - □ institutional repositories
    - □ digital archives
    - □ digital libraries

Maureen Pennock, DCC TechWatch - Eprints

21

Let's Consider…

"Preservation is the creation of digital products worth maintaining over time…"

"Preservation is the creation

- IS.
    - Preservation is a reality and not merely a metaphor for or symbol of access.
- CREATION.
    - The time to be concerned about the long-term persistence of digital products is when a system is designed and before digital conversion has begun.
- PRODUCTS.
    - A digital product has its own identity and exists within a market economy.
    - It is not necessary to sell or license a digital product for the product to have an identity within a community of end-users.

"Preservation is the creation of digital products <u>worth</u> <u>maintaining</u> <u>over time</u>

- WORTH.
  - □ The work to design and create a digital product adds value to the information contained in the documents that serve as sources.
- MAINTAINING.
  - □ The persistence of digital products requires careful attention to the maintenance of content (the bits and bytes) functionality (how the bits work in a system).
- OVER TIME.
  - □ Preservation in the digital world is not absolute, but depends instead on the continuing transformative impact of the digital product on the information work of end-users.

Paul Conway, Rationale for Digitization and Preservation

24

Data Curation Lifecycle
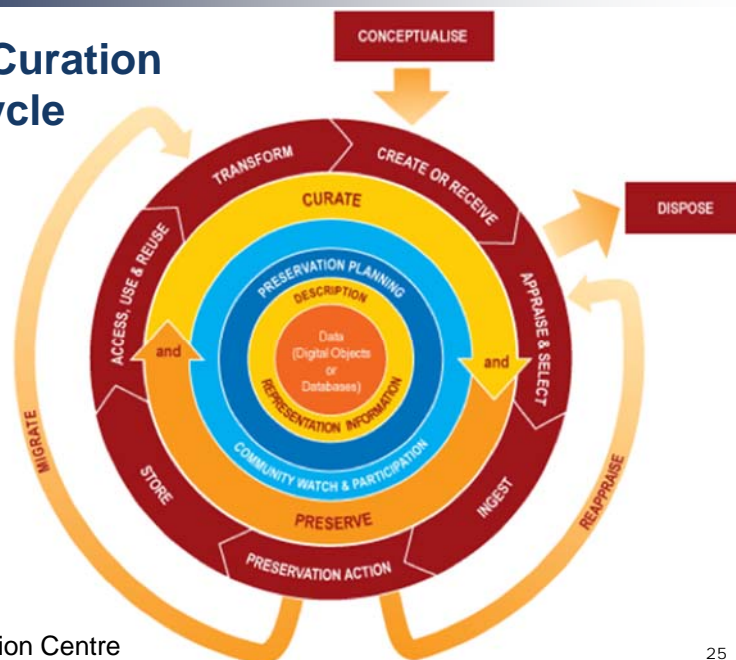
Digital Curation Centre

25

"…An Open Source Dynamic Digital Repository…"

Digital Repositories
Types

MacKenzie Smith et. Al., DSpace

26

Types of Digital Repositories

- Digital Library

- Digital Archive

- Institutional Repository

- Electronic Theses and Dissertation Repository

- Learning Object Repository

- Open Access Repository

27

Types of Digital Repositories

- Digital Asset Management Repository
- Preservation Repository
- Dark Archive
- Consortial Digital Repository
- Subject- or Discipline-Oriented Repository
- Metadata Repository

- Other Types?

28

Up Close: Institutional Repository

- A "digital collection capturing and preserving the intellectual output of a single or multi-university community"
- A non-exclusive institution or community-wide service
- Be actively taking submissions
- Have some mechanism for creator to submit work
- Departments, programs, units, formats, subjects may have IR-like, but scoped repositories

Case for IRs, SPARC; Cat McDowell, Evaluation IR Deployment   29

Up Close: SmarTech

Georgia Tech

Up Close: ETD Repository

■ A system for managing the life-cycle of electronic theses and dissertations (ETDs) from initial submission to final publication.

Texas Digital Library

Up Close: EmoryETDs

Emory University

32

Up Close: Digital Archives

- A digital collection for which an institution has agreed to accept long-term responsibility for preserving the resources on the collection and for providing continual access to those resources in keeping with an archive's user access policies.

Archive-It

33

Up Close:
West Texas Digital Archives

West Texas Digital Archives

Up Close:
Learning Object Repository

- Has a well-researched user interface and architecture that facilitates use by educators and learners
- Permits various levels of inter-activity:
  - ☐ Search
  - ☐ Submissions
  - ☐ Comments/reviews
  - ☐ Creating personal collections
- May contain LOs, or connect users to LOs stored elsewhere

GMU – Teaching with Technology Guides – Learning Resources

Up Close: MERLOT

Multimedia Educational Resource for Online Learning and Teaching

36

DR Functionality: A "Short" List

■ Authentication/Authorization
■ Content Deposit
■ Metadata Capture
■ File Management
■ Usage Controls
■ Security
■ Discovery and Access Tools
■ Disaster Recovery
■ Migration and Preservation Planning

37

"At the very basic level, the definition of a trusted digital repository must start with "a mission to provide reliable, long-term access to managed digital resources to its designated community, now and into the future"

Digital Repositories
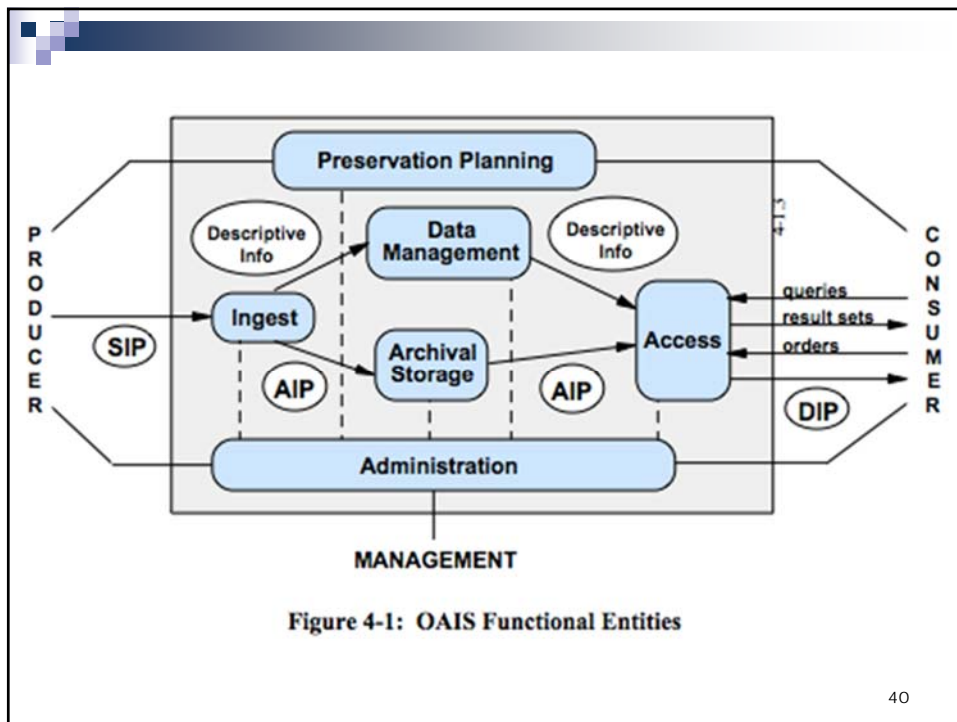# Standards and Best Practices

TDR, 2002

38

Identifying Components

- OAIS
- TRAC -> ISO/DIS 16363
- NISO's "Framework"
- CRL's "Ten Principles"

All often used as foundations, guides, and measures for digital repository initiatives

Center for Research Libraries
GLOBAL RESOURCES NETWORK

About CRL | Membership | Collections | Services | Foru...

Home » Archiving & Preservation » Digital Archives » Metrics

ARCHIVING & PRESERVATION

Metrics for Repository Asses...

Digital Archives
+ Digital Archive Reports
+ Certification and Assessment of Digital Repositories
Metrics
  TRAC/TDR
  Ten Principles
  Other Assessment Tools
  ISO Standard 16363 for Trusted Digital Repositories
+ Past Projects
Print Archives

Many scholars and scientists rely heavily upon d... research. Maintaining such digital resources ove... longitudinal or time lapse studies, and to preserv... support, and others to evaluate, their findings. T... collections of digital content and data.

Digital repositories are emerging to meet these n... specialize in data for particular fields or disciplin... content on a variety of subjects. A number of as... here to help libraries, scholars, publishers, and c... repositories and digital preservation services the... support. .

39

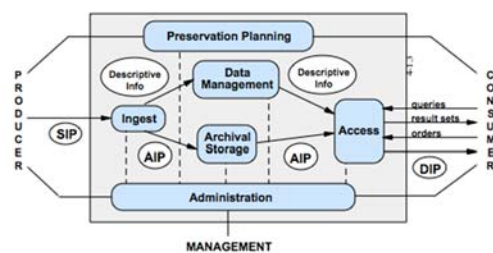Figure 4-1: OAIS Functional Entities

Open Archival Information System (OAIS)

- Conceptualizes Information Packages
  - ☐ Submission (SIP)
  - ☐ Archival (AIP)
  - ☐ Dissemination (DIP)
- Includes functions
  - ☐ Ingest
  - ☐ Archival Storage
  - ☐ Data Management
  - ☐ Administration
  - ☐ Preservation Planning
  - ☐ Access

Figure 4-1: OAIS Functional Entities

Evaluating Trustworthiness

Trustworthy
Repositories Audit &
Certification: Criteria
and Checklist

☐ Builds on Open
Archival Information
System (OAIS)
reference model

Trustworthy Digital Repositories Audit & Certification (TRAC), CRL.   **42**

Evaluating Trustworthiness

Covers three broad
aspects:

A/3. Organizational
infrastructure

B/4. Digital object
management
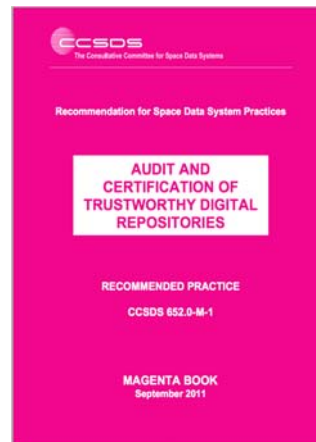
C/5. Infrastructure and
Security Risk Management

Audit and Certification of Trustworthy Digital Repositories, CCSDS.   **43**

Evaluating Trustworthiness

ISO/DIS 16363

☐ Formalizes TRAC into a Standard

Audit and Certification of Trustworthy Digital Repositories, CCSDS.   **44**

What the Framework Offers…

■ NISO Best Practice intended for two audiences:

1. Organizations planning and implementing digital collections initiatives

2. Funding organizations that want to encourage the development of good digital collections

A Framework of Guidance for Building Good Digital Collections, NISO.   **45**

What the Framework Offers…

- Includes sets of Principles as guidance
  - Collections
  - Objects
  - Metadata
  - Initiatives
- Identifies existing resources that support the development of sound local practices
- Encourages community participation

A Framework of Guidance for Building Good Digital Collections, NISO.   **46**

CRL's Ten Principles

The Repository…

1. Commits to continuing maintenance of digital objects for identified community/communities
2. Demonstrates organizational fitness (including financial, staffing, and processes) to fulfill its commitment
3. Acquires and maintains requisite contractual and legal rights and fulfills responsibilities

Ten Principles, CRL.   **47**

CRL's Ten Principles

The Repository…

4. Has an effective and efficient policy framework
5. Acquires and ingests digital objects based upon stated criteria that correspond to its commitments and capabilities.
6. Maintains/ensures the integrity, authenticity and usability of digital objects it holds over time.
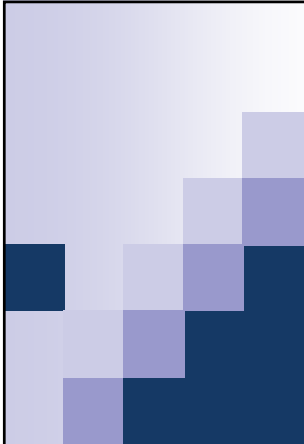
Ten Principles, CRL.

48

CRL's Ten Principles

The Repository…

7. Creates and maintains requisite metadata about actions taken on digital objects during preservation; … about the relevant production, access support, and usage process contexts before preservation

8. Fulfills requisite dissemination requirements

9. Has a strategic program for preservation planning and action

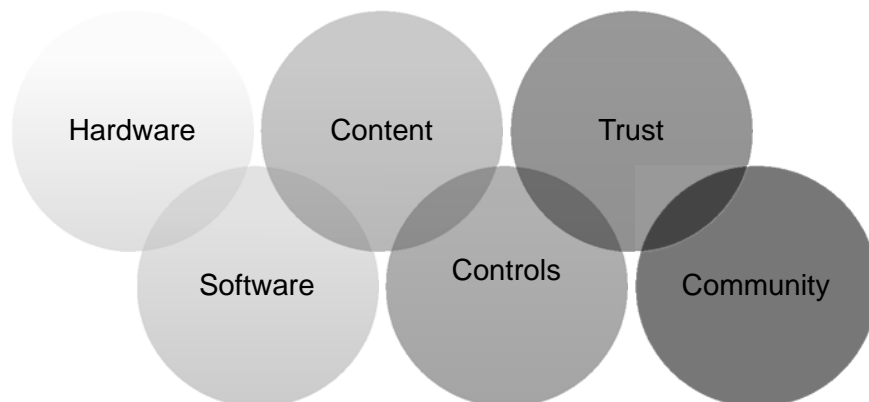10. Has technical infrastructure adequate to continuing maintenance and security of its digital objects.

Ten Principles, CRL.

49

Module 3:
Building a Repository Service:
Components

50

Hardware

Content

Trust

Software

Controls

Community

51

The Parts Nobody* Sees

Basic Hardware
- Servers
- Storage
- Back-up Systems

Basic Infrastructure
- Networking / Connectivity
- Power
- HVAC
- Security

ALMOST Nobody…

52

Software:
The Parts People See

May be used to support back-end, front-end, and "middleware" applications

- Commercial
- Open Source
- Hybrid
- Home-grown

53

Content:
The Things People Want

- Active Scholarship
- Records of Enduring Value
- Unique Resources
 - Digital Assets, Digital Objects, Items, Resources, Files, "Stuff"

54

"Born Analog" Content…

Records, ledgers, transcripts, photographs, film, diaries, maps, correspondence, paintings, artifacts, artwork, texts, audio, video, datasets, learning objects, institutional records, official documents, bulletins, directories and publications…
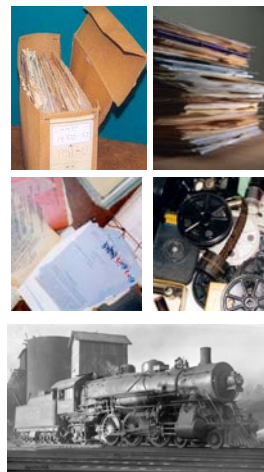
55

…Becomes Physical Items in Collections…
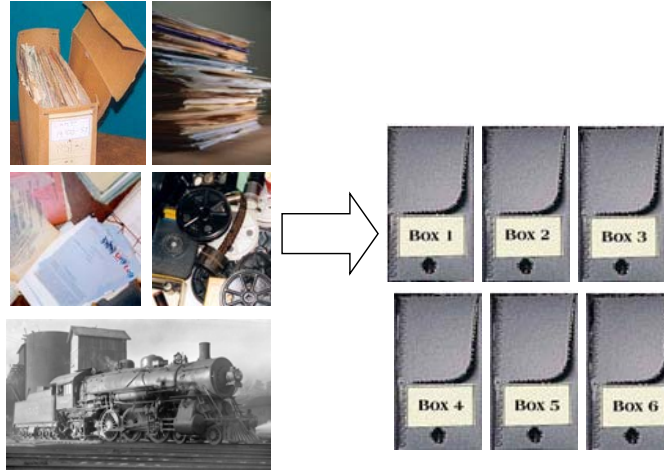
56

…And can be Transformed into Quality Digital Objects

- Images
  - ☐ Format, bit depth and resolution
- Audio
  - ☐ Format, sample rate and quantization (compression)
- Text
  - ☐ Format, levels of accuracy, levels of encoding,
- 3-D objects
  - ☐ Levels of detail, capture to support static vs. dynamic representation

57

**Handout 1:
Digital Capture
Quality Basics**

A brief review of quality choices for common
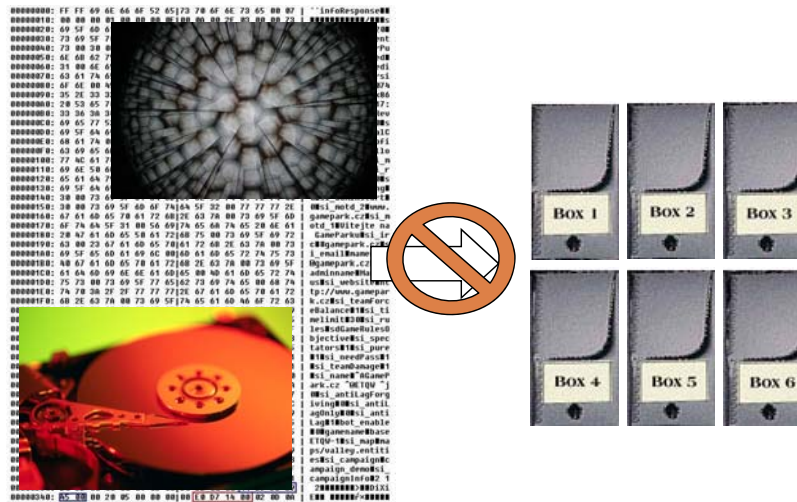digital file formats and types

58

"Born Digital" Content…

Electronic records and
documents, digital audio,
digital video, digital still
imagery, datasets, multi-
media works, online
publications, web sites,
interactives, blogs, wikis…

59

…Unlike Physical Collections…

…Follows a Different Path

- "at much greater risk of either being lost and no longer available as historical resources, or of being altered, preventing future researchers from studying them in their original form…"

- "Millions of digital materials, such as Web sites mounted in the early days of the Internet, are already lost—either completely or in their original versions."

Content Brings Its Own Metadata Needs

Metadata is often specific to a community, a content type, a format, or an audience.

Seeing Standards: A Visualization of the Metadata Universe, Jenn Riley. [62]
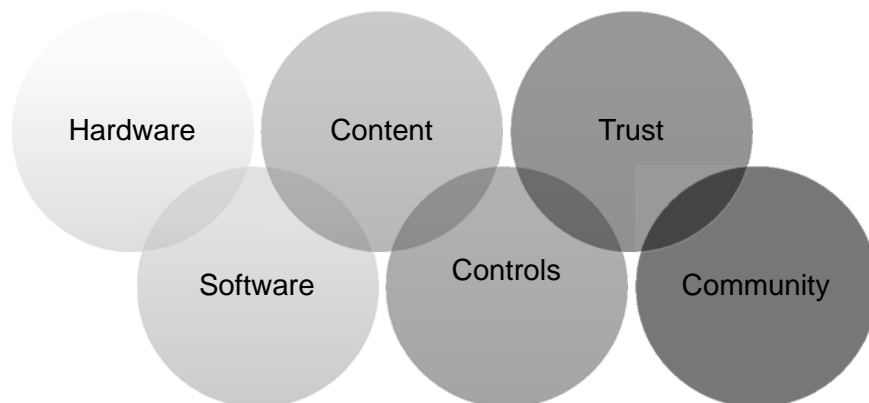

…Or No Metadata at All!

[63]

ARL Code of Best Practices and Fair Use for Academic and Research Libraries

Understanding Rights and Responsibilities

64

Key Components

Hardware

Content

Trust

Software

Controls

Community

65

Controls: Keep Things in Order

- Deposit
- Review
- Access
- Reproduction
- Security
- Risk and Disaster Mitigation

66

Trust: What Keeps the DR Going

- Reliable, long-term access now and in the future
- Responsibility for long-term maintenance of digital resources
- Commonly accepted conventions and standards
- Policies, practices, and performance that can be audited and measured (e.g.TRAC)

67

Community: What Makes it Happen

■ An ever-changing web of personal, organizational, cooperative, conceptual, and contractual understandings, agreements, and consensuses

68

Worksheet 2: Components

When inventorying your existing environment, what do you have, what will you need, and do you have a sense now of where and how you'd like to get it?

69

Worksheet 2: YOUR Component Inventory

This exercise expands YOUR digital repository profile, by identifying the following:
- What pieces and parts are already in place
- Who manages them
- What access do you have (or could have) to these resources

Component	Comments
Hardware: Servers, storage, backup, network, connectivity	
Software: Proprietary, open source, cloud services, homegrown	
Content: Born digital, born analog, digitization ready analog, metadata	
Community: Organizational support, user communities, funders and advocates	
Controls: Policies, administrative metadata	
Trust: Organizational reputation	

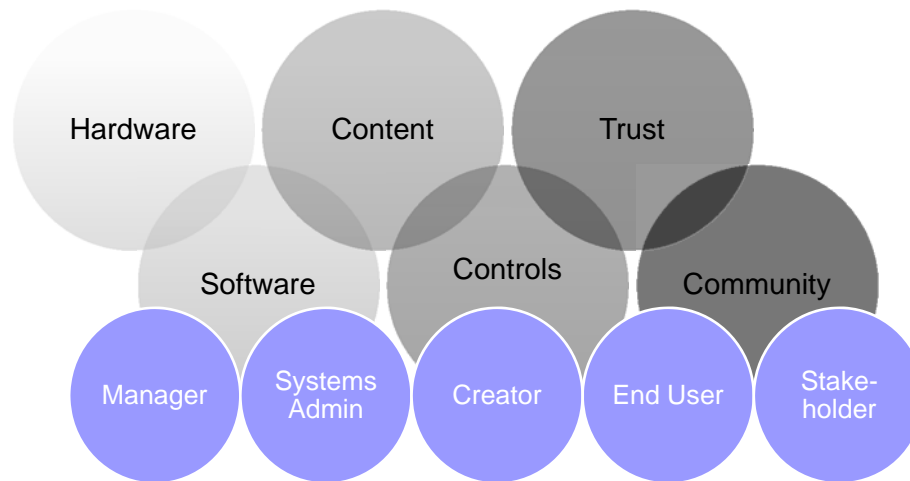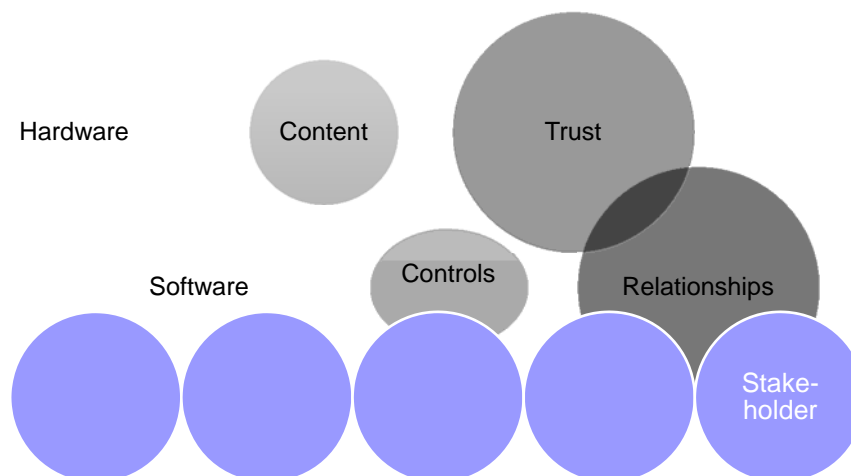**Module 4:**
Building a Repository Service:
Roles and Views

70

Key Components

Hardware

Content

Trust

Software

Controls

Community
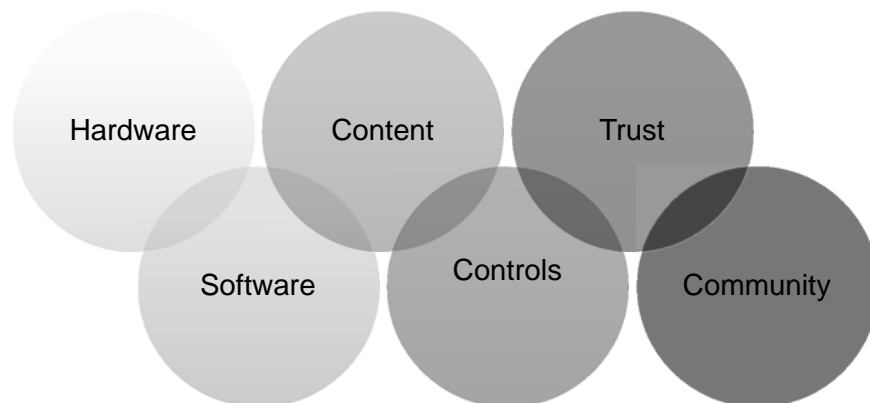
71

Key Components, Many Views

Hardware

Content

Trust

Software

Controls

Community

Manager

Systems Admin

Creator

End User

Stake-holder

72

Stakeholders

Hardware

Content

Trust

Software

Controls

Relationships

Stake-holder

73

Getting to Know Stakeholders

■ The mission-vision-money folks, focused on…
  - ☐ Benefits of digital stewardship
  - ☐ Resources commitment
  - ☐ Opportunity costs
  - ☐ Risk mitigation
  - ☐ Partnerships and collaborations
  - ☐ Quality of service

74

End User
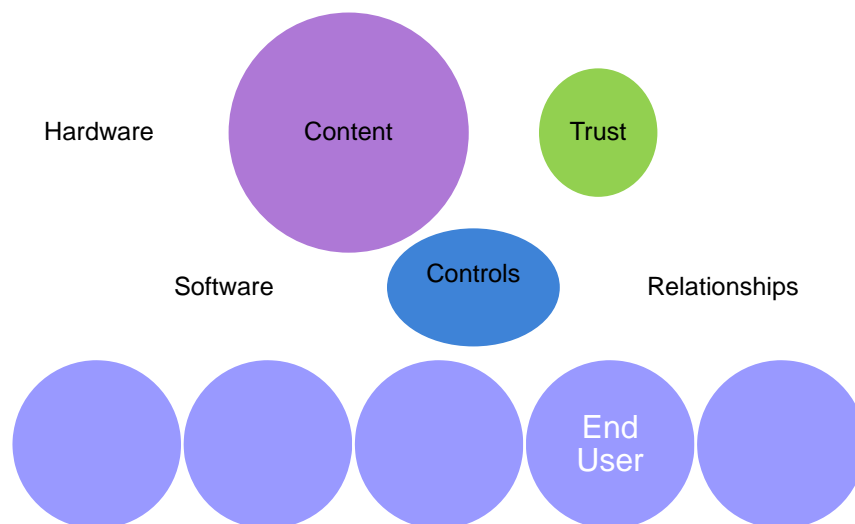
75

End Users

Want information quickly, with a minimum of fuss

Willing to use free information and trust it just as much as for-fee information

Are unlikely to know of digital repository initiatives, especially beyond their "communities"

- YOU are an End User
- Potentially, everyone you know is an End User

76

What People Want…

They want to…

- Work with co-authors
- Keep track of different versions of the same document
- Work from different computers and locations: Mac and PC
- Make their own work available to others
- Have easy access to other people's work
- Keep up in their fields
- Organize their materials according to their own scheme
- Control ownership, security, and access
- Ensure that documents are persistently viewable or usable
- Have someone else take responsibility for digital tools
- Be sure not to violate copyright issues
- Keep everything related to computers easy and flawless
- Not be any busier

Foster and Gibbons, Understanding Faculty
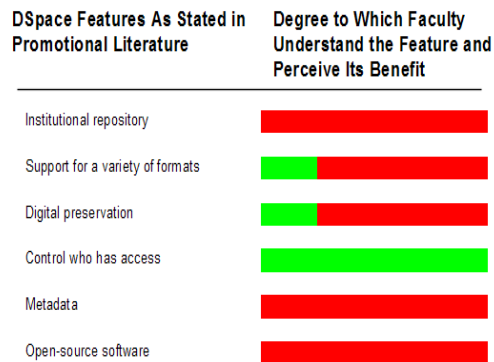
77

What People "Get"

DSpace Features As Stated in Promotional Literature	Degree to Which Faculty Understand the Feature and Perceive Its Benefit
Institutional repository	████████████
Support for a variety of formats	████████████
Digital preservation	████████████
Control who has access	████████████
Metadata	████████████
Open-source software	████████████

Figure 1. Perceived fit between DSpace features and needs from faculty/researcher perspective. Green indicates understanding while red indicates misunderstanding, lack of understanding, or disinterest.

Foster and Gibbons, Understanding Faculty

78

Software

How awareness of software functionality manifests in the End User:

- Can I find what I need?
- Can I use what I find?
- Can I keep a copy of what I find?
- Do I have to pay?
- Will it be there tomorrow?

79

Content

Simply, they want a select, yet comprehensive body of content, well described, and available

- Preferably for free?
- Preferably right now?
- Preferably at the same place forever?
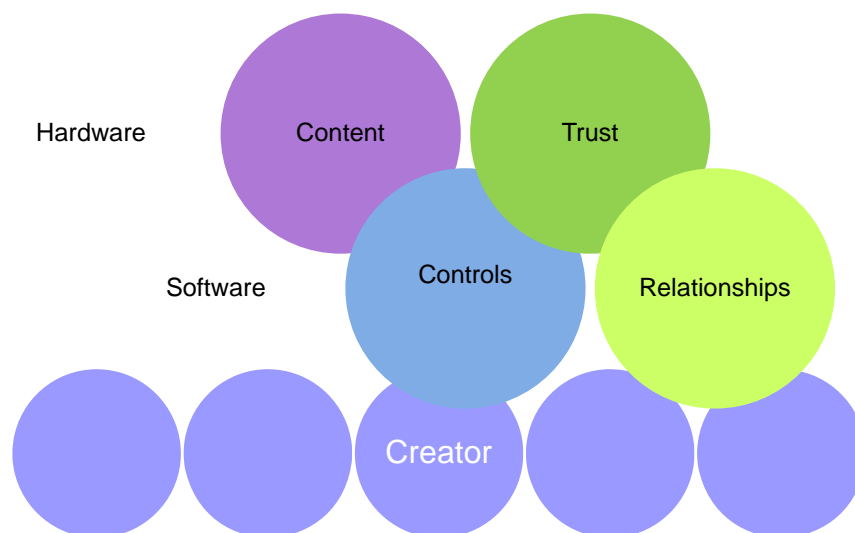
Creators

Repository Content Creators / Depositors

- Equivalent of "Donors"; may include Archivists
- Deposit content, or references to content
- Make decisions about use, rights, ownership
- Make decisions about "submission" or "native" formats
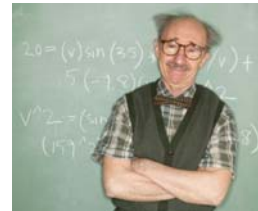
82

Content

Content Creators are "Content Big"

- They create/contribute to the repository:
  - ☐ Representations of analog materials
  - ☐ Born digital content
  - ☐ Initial descriptive metadata
- They may contribute:
  - ☐ Initial rights and use metadata
  - ☐ Initial access controls

83

Trust

Content Creators are "Trust Big"
- Why?

  "Preservation is the creation of digital products worth maintaining over time…"

Controls

- "Gray areas" of Content Deposit

- Ownership and right to publish and/or distribute can be confusing to Content Creators

- Resources like SHERPA can help

Systems Administrator

Hardware

Content

Trust

Software

Controls

Community
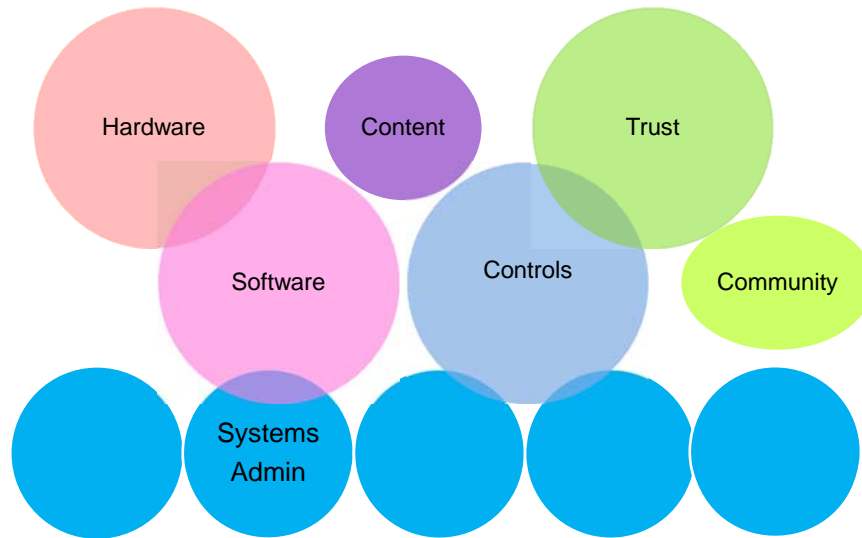
Systems
Admin

86

Repository Systems Administrators

- Manages hardware and infrastructure
- Responsible for security
- Coordinates software development and preservation activities at the system level
- Resource allocation – equipment, staffing, licensing, etc.

87

Hardware

Systems Administrators are "Hardware Big"
  Focus on
  - Capacity
  - Scalability
  - Redundancy
  - Stability
  - Disaster recovery
  - Security
  - Connectivity

Cost is never out of mind

88

Hardware and Infrastructure

- Servers
- Disk Array
- Connectivity
- Back-Up Appliances
- Power
- Cooling (not much heating)
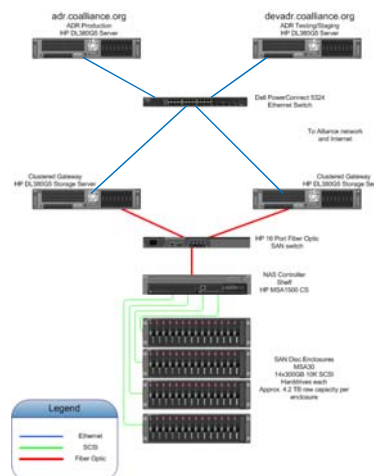
89

Hardware and Infrastructure Overview
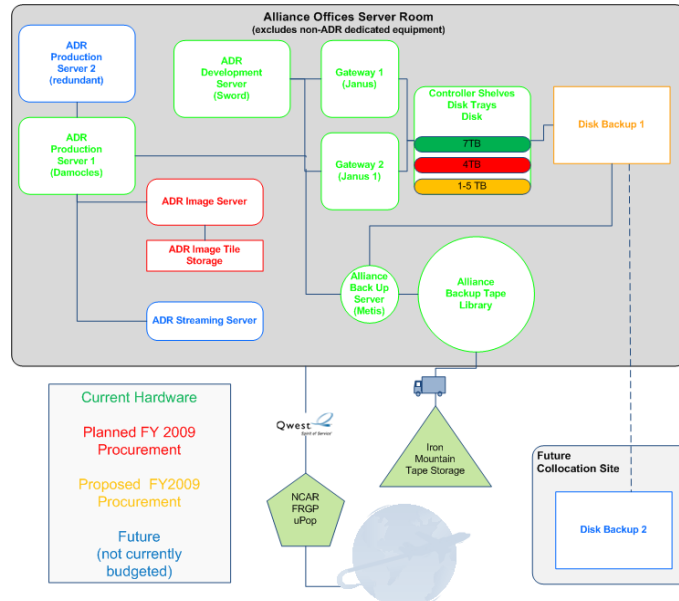
90

Infrastructure
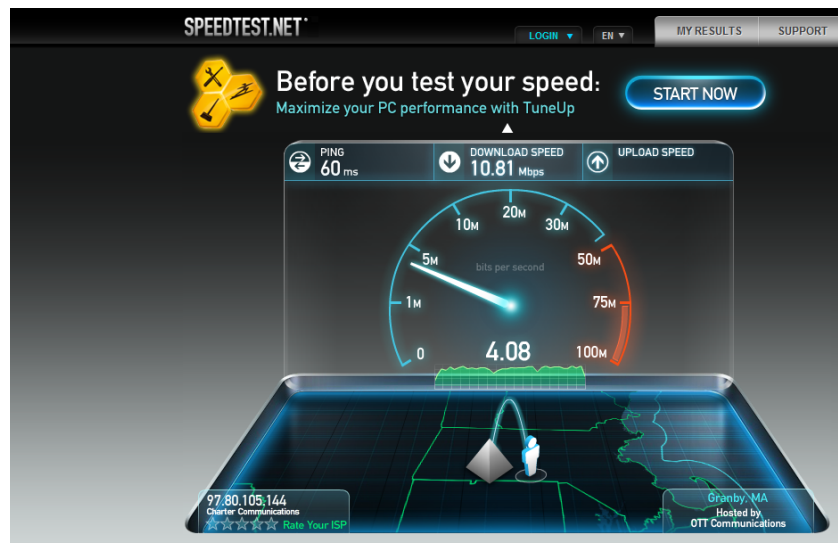
91

Infrastructure

Speedtest.net

92

Hardware and Infrastructure Strategies

- Redundancy within systems
- Mirrored sites
- Disaster recovery planning
  - Back-up systems
    - Tape
    - Disk
- Diversify funding sources
- Document activities
- Disseminate to community and review "peer paths"
- Maintain and grow skill sets and knowledge bases

93

And then there is this to consider…

94

And then there is this to consider…

95

Software

Systems Admins. are "Software Big"

- Translate functionality requests into repository services, by evaluating and using
  - ☐ Vendor-supplied repository software
    - Digital Commons, Digital Archive, Documentum, Cumulus, Artesia, DigiTool…
  - ☐ Open-source repository software
    - DSpace, Fedora, EPrints…

96

Software

With a variety of middleware and front-end tools to enhance functionality:

- XTF, Origami, IPaper, Omeka, Hydra, Islandora, Manakin, Blacklight, …

97

Software Myth

"If only we had *(fill in the blank)** everything would be fine"

> * Fedora, DSpace, CONTENTdm, Insight, Greenstone, Keystone, DLXS, Documentum, Cumulus, Islandora, etc…

98

Software Reality

No one software will
- Manage digital objects and metadata
- Provide public access
- Permit easy re-purposing
- Preserve digital assets
- Apply and Manage policies
- Make Quality choices
- Determine Access controls

CAUTION

NO SYSTEM WILL MANAGE ITSELF!

Software should be seen as a tool, not a "solution"

99

The "Systems View" of a Repository Service

Presentation

Access

Management

Storage
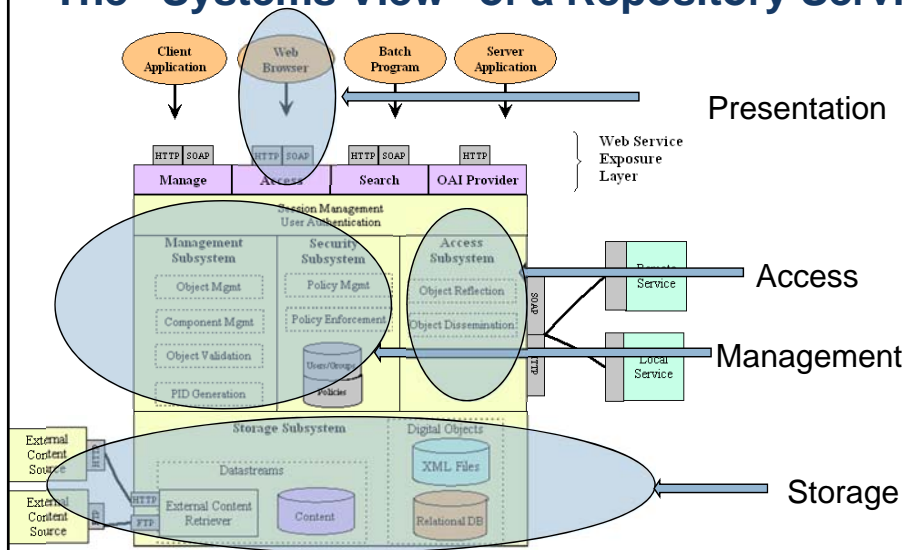
Staples, Thornton, Ross Wayland and Sandra Payette, "The Fedora Project: An Open- source Digital Object Repository System, "D-LIb Magazine", April 2003. http://www.dlib.org/dlib/april03/staples/04staples.html **100**

Controls

Systems Administrators are "Controls Big"

■ Enforce policies through use of technology and software

■ Work with Managers to establish a service level agreement (SLA) which can determine scope of

☐ Access controls

☐ Versioning

☐ Digital object identifiers

☐ Fixity checks

☐ Back-ups

101

Metadata Controls

Systems Administrators use creator, manager, and software generated metadata to programmatically apply and enforce controls

102

Server Controls

- Blocking (and unblocking) search engine crawlers
- Limiting access at the "port" or "domain"
- Limiting access through firewalls
- Maintaining security certificates
- Supporting authentication and verification processes

103

Trust

Systems Administrators are "Trust Big"

Their near-central role also requires a strong understanding or awareness of everything!

- NISO Framework Principles
- Collections
- Objects
- Metadata
- Initiatives
- TRAC- ISO 16363
- C. Technologies, technical infrastructure, and security.
- Some of A. Organizational infrastructure
- An awareness of B. Digital object management
- ISO 16919 -

104

Repository Manager
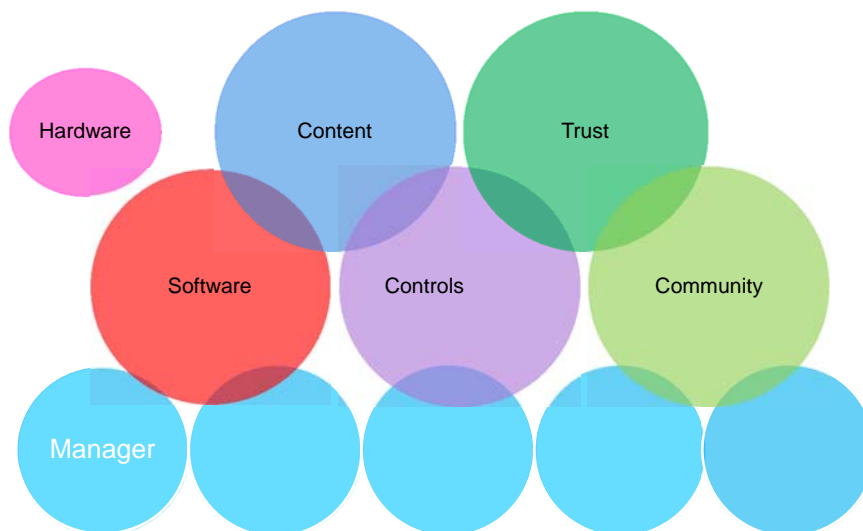
105

Repository Managers

- Stewards entrusted with content, collections and resource management
- Frequently (digital) archivists, (digital) librarians, (electronic) records managers, and information managers
- Maintain content and metadata in forms that are archival, reusable, and interoperable
- Enforce conditions of use, rights, ownership
- Guide all other repository community roles in some form or fashion
- Guide outreach, marketing collection development activities

Zuccala and Opennheim, Managing and Evaluating DRs   **106**

Repository Managers

The main role of the Manager…

"should be to recognise and define the *raison-d'être* of the repository so that depositors, users and members of the public will be familiar with its existence and purpose…"

Zuccala and Opennheim   **107**

Repository Managers

Not to be too obnoxious,
but these folks could be considered the
"Masters of the
[Repository] Universe!"

Hardware

- Managers should know enough about hardware to:
  - Make the case for hardware to stakeholders and funders
  - Negotiate agreements and contracts with vendors and service providers if necessary.

Software

- Managers are "Software Big"
- Functionality
  - Indexing
  - Discovery
  - Delivery
  - Management

Software

Answering questions like "How do I deposit my stuff?" leads to intersections with software architecture that accommodates Submission Information Packages (SIPs):

- Workflows
- Standards
- Best practices
- Scalability
- System-driven format transformations

Content

- Managers are "Content Big"
- or, more accurately "Actual Content May Vary…"
- You must be familiar with Standards and Best Practices
  - ☐ Metadata
  - ☐ Formats
  - ☐ Quality

112

Content Standards

- Follows NISO Framework's Collections and Objects Principles
- May create and structure objects and metadata according to accepted standards and best practices (i.e. METS, Digital Capture Recommendations from the Library of Congress)
- May identify file type candidates for migration to new formats at point of deposit (i.e. MS Word to PDF)

113

Metadata Standards

- Follows NISO Framework's Metadata and Objects Principles
- May opt to identify, require and support a normalizing schema (i.e. MODS, Dublin Core)
- Likely to maintain an OAI-Dublin Core compliant metadata record somewhere for each object
  - To support harvesting and aggregation of records in OTHER repositories

114

Digital Object Standards

- Be able to work with creators on SIP object profiles and standards
- Be able to work with Systems Administrators on Archival and Dissemination Information Packages object profiles (AIPs, DIPs)

115

What ARE you talking about?

- Managers straddle the content creation and the content management and preservation worlds
- They need to be as concerned with capture best practices as they are with digital repository and object design – as both ultimately ensure the long-term preservation and access of digital resources

116

OAIS Conceptual Model
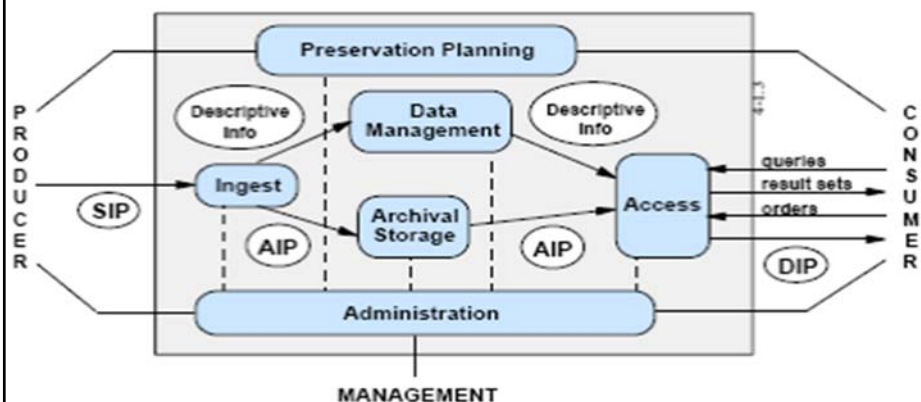
117

A Digital Object Model

Persistent ID PID	Digital Object Identifier
Object Properties	System Properties Manage and track the object
Datastream 1 **Datastream ...** **Datastream N**	Datastreams Aggregates content items
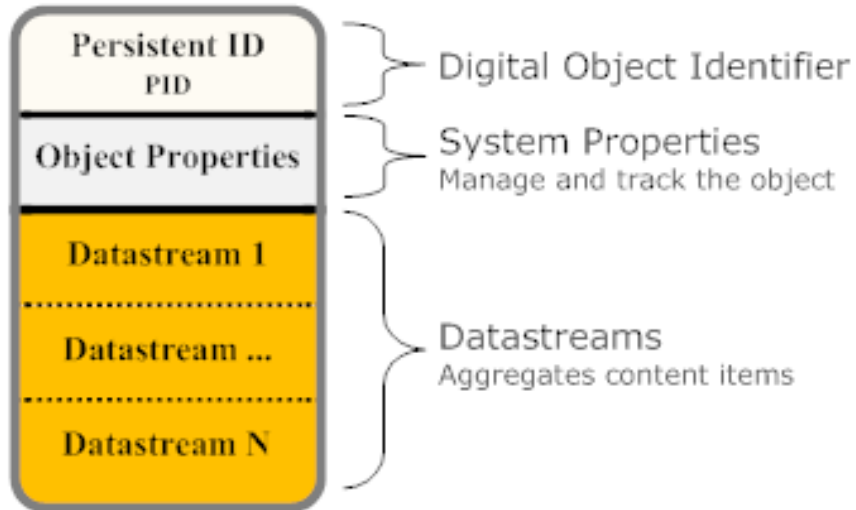
118

Controls

- General policies and responsibilities governed by a Policy Document
- Specific responsibilities defined in a service agreement or deed of gift
- Manage access, rights and usage through embedded information in files, metadata, and software, as instructed by policies and content creators

119

Community

Managers have relationships with a diverse community:

- ☐ Content Creators/ Depositors
- ☐ Stakeholders
- ☐ Systems Administrators
- ☐ End Users

and, of course, they may BE End Users!

120

Trust

Their central role requires a strong understanding or awareness of everything!

- ■ NISO Framework Principles
- ■ ISO 14721 - OAIS
- ■ Collections
- ■ Objects
- ■ Metadata
- ■ Initiatives
- ■ ISO 16363 - Audit/Cert. TDR
- ■ B. Digital object management
- ■ Some of A. Organizational infrastructure
- ■ An awareness of C. Technologies, technical infrastructure, and security
- ■ ISO 16919 – Auditor Reqs.

121

Trust within the Community

- Managers trust…
  - ☐ Content Creators/ Depositors
  - ☐ Stakeholders
  - ☐ Systems Administrators
  - ☐ End Users

122

Building Trust Outside the Community

- Managers often coordinate repository marketing and evaluation activities
- Both marketing and evaluation activities can be used to
  - ☐ Inform users and potential users of range of services and benefits
  - ☐ Demonstrate to stakeholders scope and application of services
- Press kits and embedded tools to support marketing and evaluation, and are meant to be locally adapted

123

Worksheet 3: Roles

Many repository services are run by just a handful of people focused on meeting the needs of the masses. At this point, what role or roles do you see yourself having in your local repository service?

124

Worksheet 3: Roles

What roles and activities will YOU be responsible for? If not you WHO will do it?

Role	I Do It	Others Do It	I Don't Know
Manager: Entrusted with content and collections			
Systems Admins: Hardware, software, and network infrastructure			
Creator/Depositor: Donors, producers, etc.			
End User: "Communities," both internal and external			
Stakeholder: Internal and external decision makers and funders			

Module 5:
Building a Repository Service:
Sustainability

125

Sustainability

"Ensuring that valuable digital assets will be available for future use is not simply a matter of finding sufficient funds.

It is about mobilizing resources—human, technical, and financial—across a spectrum of stakeholders diffuse over both space and time…"

Sustainable Economics for a Digital Planet. URL:
http://brtf.sdsc.edu/biblio/BRTF_Final_Report.pdf

126

Principles of __GOOD__ Digital Initiatives

A good digital initiative (a.k.a. project, program)…

☐ has a substantial design and planning component.

☐ has an appropriate level of staffing with necessary expertise to achieve its objectives.

☐ follows best practices for project management.

☐ has an evaluation component.

☐ markets itself and broadly disseminates information about the initiative's process and outcomes.

☐ considers the entire lifecycle of the digital collection and associated services.

A Framework of Guidance for Building Good Digital Collections, NISO.

127

Taking on a "Digital Mortgage?"

"Each organization must budget to transfer old files to new formats as software and hardware change and electronic media reach the end of their relatively short life expectancies…also a digital infrastructure including staff, contracts, equipment, and software"

Puglia, Lowry & Troll; Marcum in Vogt. Handbook for Digital Projects:
A Management Tool for Preservation and Access. URL:
http://www.nedcc.org/resources/digitalhandbook/dman.pdf

128

Insuring Digital Collections

- Three possible strategies
  - ☐ Actual value
  - ☐ Replacement value
  - ☐ Self-Insure
- Start by identifying risks
  - ☐ Organizational Risks
  - ☐ Content Risks
  - ☐ Infrastructure Risks

129

DRAMBORA

130

DRAMBORA

131

Defining Digital Preservation

Short Definition

■ Digital preservation combines policies, strategies and actions that ensure access to digital content over time.

(ALCTS, 2009)

132

Examples of Repository Policy Areas

- **Organizational**
  - ☐ Governance
  - ☐ Participation
  - ☐ Community Organization
  - ☐ Community Termination
  - ☐ Grant Support
- **Content and Metadata**
  - ☐ Deposit and Usage
  - ☐ File Format Support
  - ☐ Identification
  - ☐ Ownership
  - ☐ Metadata

- Access, Rights, and Usage
- Authentication
- Copyright
- Preservation and Infrastructure
- Preservable formats
- Repository Termination
- Technical Infrastructure

133

The Five Organizational Stages of Digital Preservation

The five stages of organizational response to digital preservation are:

1. **Acknowledge**
2. **Act**
3. **Consolidate**
4. **Institutionalize**
5. **Externalize**

The Five Organizational Stages of Digital Preservation. Kenney, A., McGovern, N.

134

Repository Systems and Preservation

■ Content <u>integrity</u> 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 requirements
  - ☐ Routine audits

135

Repository Systems and Preservation

Content <u>maintenance</u> includes:
  - ☐ A 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

136

Repository Service Threats

- Media Failure
- Hardware Failure
- Software Failure
- Power Failure
- Communication/Network Failure
- Organizational Failure

137

Content

- Concerned with growth and scalability
- Concerned with obsolescence
- Concerned with degradation and durability
- Focused on Preservation strategies

138

Principles of <u>GOOD</u> Digital Objects

A good object

- ☐ exists in a format that supports its intended uses
- ☐ is "preservable"
- ☐ is meaningful and useful outside of its local context
- ☐ will be named with a persistent, globally unique identifier that can be resolved to the current address of the object
- ☐ can be authenticated
- ☐ has associated metadata

A Framework of Guidance for Building Good Digital Collections, NISO. 139

Content and Data Preservation Strategies

- ■ Bit-level
  - ☐ Store
  - ☐ Refresh
  - ☐ Migrate
  - ☐ Preserve the Technology
- ■ Full
  - ☐ Emulation
  - ☐ Normalization
- ■ Lifecycle Management
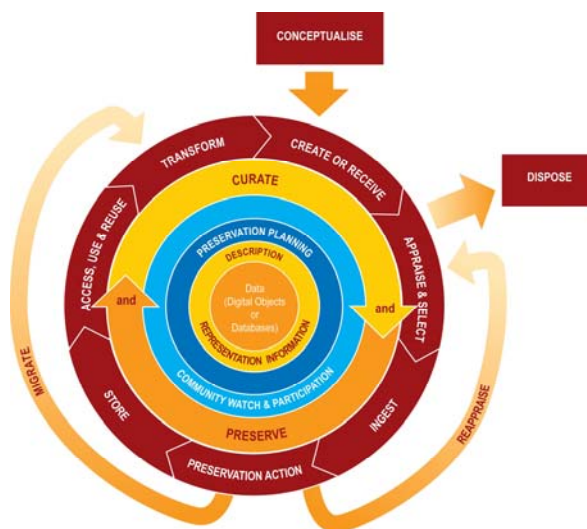
140

Thoughts…

"Stewardship is easy and inexpensive to claim; it is expensive and difficult to honor, and perhaps it will prove to be all too easy to later abdicate…"

Clifford Lynch

141

Worksheet 4: Sustaining YOUR Digital Repository

142

Worksheet 4: Sustaining

Five Organizational Stages of Digital Preservation

The five stages of organizational response to digital preservation are:

1. Acknowledge: Understanding that digital preservation is a local concern;
2. Act: Initiating digital preservation projects;
3. Consolidate: Seguing from projects to programs;
4. Institutionalize: Incorporating the larger environment; and
5. Externalize: Embracing inter-institutional collaboration and dependency.

(Anne R. Kenney & Nancy Y. McGovern. The Five Organizational Stages of Digital Preservation)

Questions to consider:

1. What stage do you believe your institution is currently in?

2. What components do you believe would need to change and how would they need to change to achieve the next stage?
   (ie. Planning, Content, Controls, Community, Resources, Infrastructure, and Applications)

3. What challenges to do you see to achieving the next stage locally?

4. What are five things you may be able to do today (or this year…) to move your institution forward to the next stage?
   (Remember, Rome wasn't built in a day…)

Module 6:
Building a Repository Service:
Discovery and Access

143

Repository Controls

■ Primarily Users are concerned with
  □ Open Access
  □ Copyright
  □ Conditions of Use
■ Users are interested in what they can
  □ Take
  □ Keep
  □ Use
  □ Alter
  □ Re-distribute

144

Controls and Social Media

- **Follow**
  - ☐ Syndication (RSS, Atom)

- **Aggregators**
  - ☐ local, regional, national

- **Harvesters**
  - ☐ OAIster, Internet Archive

- **Culture of participation**
  - ☐ Blogs & Wikis
  - ☐ Social networking
  - ☐ Social tagging

- **Openness**
  - ☐ Sharing
  - ☐ Syndication

145

Transformative Technology

- **The "Untethered Internet"**
  - ☐ Mobile devices and tablets

- **Mash-ups**

- **Augmented reality**

- **Cloud services**
  - ☐ Streaming delivery

146

Module 7:
Promoting and Measuring Use

147

Marketing Repository Services

- Develop a multi-platform "brand" using Web and print
- Find and engage YOUR End Users
- Recruit YOUR advocates
  - Enlist Stakeholders, Creators, and End Users in dissemination of marketing "messages"
- Identify potential partnerships
- Embrace "R&D"
  - "Rip-off and Duplicate"

148

Marketing Repository Services

- Web Strategies
  - Build "Community Spaces"
    - Mostly public, with "members only" areas
  - Develop mailing lists and Listservs,
  - Wikis, Blogs, and "Fan" Pages
  - Tweet, Like, Follow, Connect…
- "Print" Strategies
  - Flyers, brochures, bookmarks, stickers,
  - Annual reports, white papers, and presentations
  - Press Releases to a range of media outlets
    - Campus newspaper, Friends newsletter, Member magazine

149

IMLS C2C Online

150

Evaluation

"[Repositories] must be both useful and usable…"

- Make sure you know requirements of funders!
  - ☐ Include how you will USE findings!
- Adopt direct and indirect approaches
- Can be coordinated in-house or out-sourced

151

Evaluation

- Web Analytics
  - ☐ File downloads count
  - ☐ Anonymous user demographics
- Surveys
  - ☐ Online and Print
  - ☐ Incentives
- Focus Groups
  - ☐ "Early Adopters" are key
- Anecdotal evidence
  - ☐ Might not "help" officially, but it can't hurt!

Comparing: **All sites of similar size** ⑦ Open category list

3,000 ... 3,000

Friday, May 8, 2009
Visits (Benchmark) **131**

Friday, May 8, 2009
Visits **1,525**
Difference: **1,064.1%**

1,500 ... 1,500

April 13, 2009   April 20, 2009   April 27, 2009   May 4, 2009
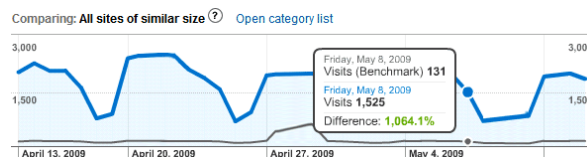
152

Evaluation

ISO 16363 – Audit and Certification of Trustworthy Digital Repositories

ISO 16919 – Requirements for Bodies Providing Audit and Certification of Candidate Trustworthy Digital Reposiories
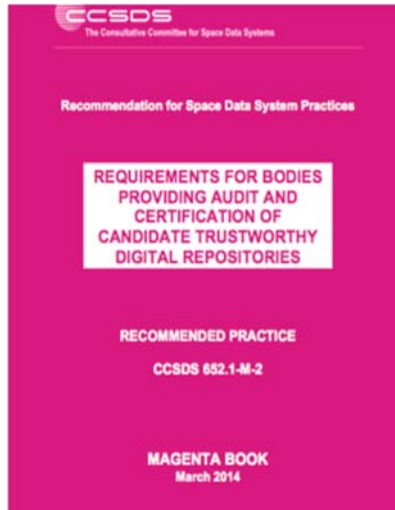
153

Workshop Wrap-Up

That's All Folks!

154

Parting Thoughts To Remember…

■ A digital repository is a service, as much as it is software, hardware, or content.

■ If you begin a digital repository service, you are committing - at some level - to providing the same type of service and support you provide in your physical institutions.

155

Parting Thoughts To Remember…

■ You are not alone!

YOU

+ Your collections

+ Your colleagues

+ Your users

+ Everyone here today

+ digitalpreservation.gov[10]

_____

Success! (or, at least a VERY GOOD start!)

156

Quality Primer

Images: Bit Depth and Resolution

1 bit image = two colors, black or white

2 bit image = four colors (1 byte)

8 bit image = 2x2x2x2x2x2x2x2=256 colors

24 bit image = 16.8 million colors

Bit Depth
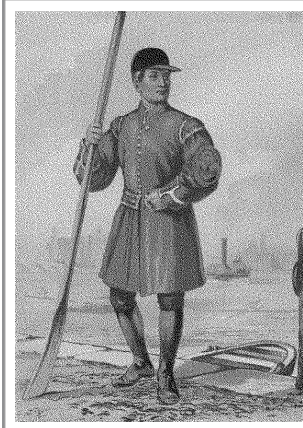
1 bit image   24 shades of gray   16 million colors

Looking Closer:

Resolution

600ppi   300ppi   75ppi
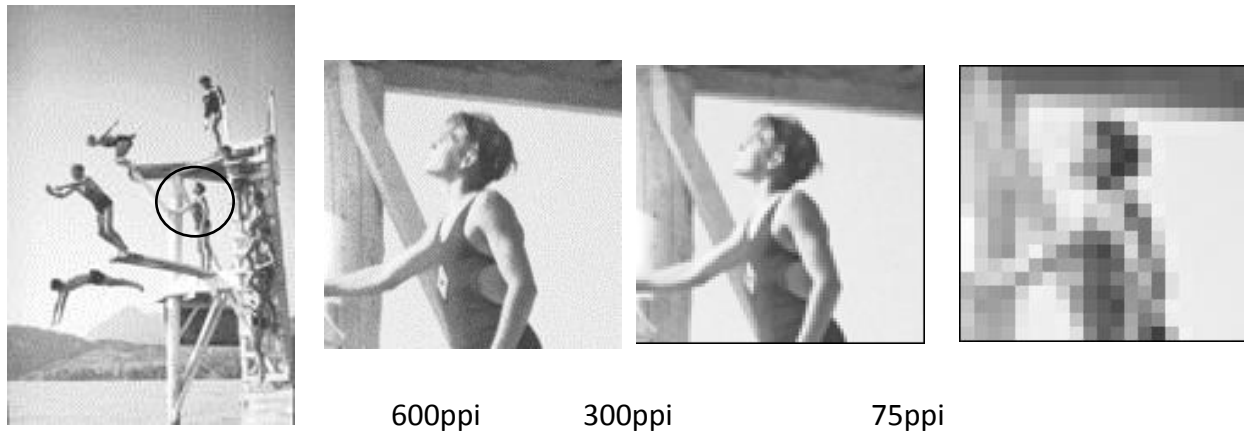
File Size and Storage: Doing the Math:

Image size formula: X pixels x Y pixels x bit depth ÷ 8

8 x 10 photo at 300 ppi and 8 bit grayscale= 7.2MB (((8x300) x (10 x 300))x 8) ÷ 8

The same 8 x 10 photo at 600 ppi = 28MB
8 x 10 at 300 ppi 24 bit color = 22 MB
8 x 10 at 600 ppi 24 bit color = 87 MB

Audio Quality: Sampling
From: Introduction to Computer Music: Volume One
http://www.indiana.edu/%7Eemusic/etext/digital_audio/chapter5_sample.shtml

Analog waveform represented as sound pressure (SPL) level over time

Same waveform quantized proportional to SPL
Each sample is a measurement of the instantaneous amplitude of source signal
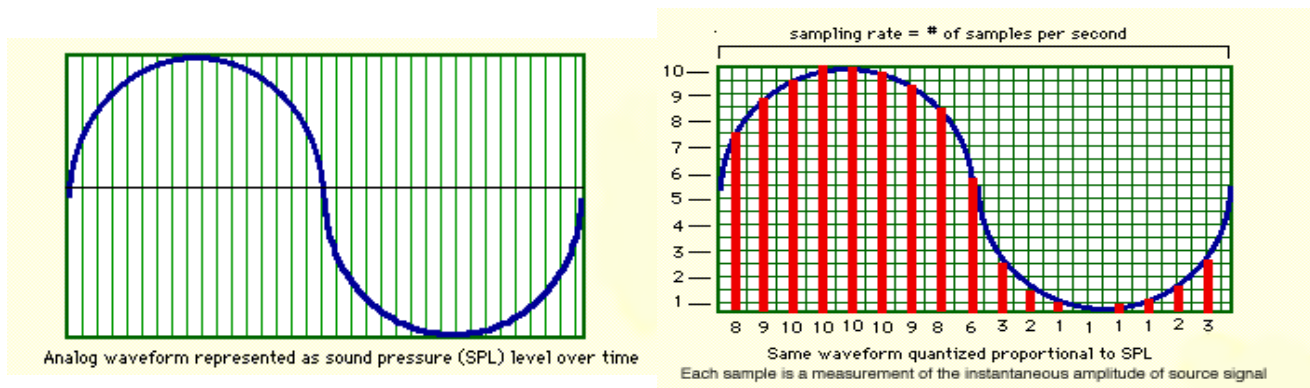
Audio Quality: Compression

Compression creates smaller file sizes with a resultant loss of quality as "insignificant" bits of
data are discarded and sound values are interpolated.

Handout 2: Roles and Components Chart

	Hardware	Software	Content	Controls	Trust	Relations
Managers	☞	👆	👆	👆	👆	👆
Sys. Admin.	👆	👆	☞	👆	👆	☞
Creators	👇	👇	👆	👆	👆	👆
End Users	👇	👇	👆	☞	👆	👇
Stakeholders	👇	👇	☞	☞	👆	👆

👆 Highly important requires either depth of knowledge or has a high value

☞ Something to know about, but not in any great detail

👇 Not consciously considered important or valued

Repositories, Copyright, and Fair Use

Association of Research Libraries (ARL)
Code of Best Practices in Fair Use for Academic and Research Libraries (2012)
http://www.arl.org/pp/ppcopyright/codefairuse/code/four-creating.shtml

FOUR: Creating Digital Collections of Archival and Special Collections Materials

Description
Many libraries hold special collections and archives of rare or unusual text and nontext materials (published and unpublished) that do not circulate on the same terms as the general collection. The copyright status of materials in these collections is often unclear. Despite the investments that have been made in acquiring and preserving such collections, they frequently are of limited general utility because they typically can be consulted only on-site, and in some cases using only limited analog research aids. The research value of these collections typically resides not only in the individual items they contain (although such items are often unique in themselves), but also in the unique assemblage or aggregation they represent. Special collections can have a shared provenance or be organized around a key topic, era, or theme. Libraries and their patrons would benefit significantly from digitization and off-site availability of these valuable collections. While institutions must abide by any donor restrictions applicable to their donated collections, and they will inevitably consider practical and political concerns such as maintaining good relations with donor communities, librarians will benefit significantly from knowing their rights under fair use.

Presenting these unique collections as a digital aggregate, especially with commentary, criticism, and other curation, can be highly transformative. Works held in these collections and archives will serve a host of transformative scholarly and educational purposes relative to their typically narrower original purposes. Materials in special collections typically include significant amounts of primary sources and artifacts (correspondence, institutional records, annotated volumes, ephemeral popular entertainment) whose value as historical objects for scholarly research is significantly different from their original purpose. The new value created by aggregating related documents in a single, well-curated collection is also significant. In addition to access for scholarly purposes, digitization facilitates novel transformative uses of the collection as a whole—see principle seven below regarding digitization for search and other nonconsumptive uses.

Principle
It is fair use to create digital versions of a library's special collections and archives and to make these versions electronically accessible in appropriate contexts.

Limitations
• Providing access to published works that are available in unused copies on the commercial market at reasonable prices should be undertaken only with careful consideration, if at all. To the extent that the copy of such a work in a particular collection is unique (e.g., contains

marginalia or other unique markings or characteristics), access to unique aspects of the copy will be supportable under fair use. The presence of non-unique copies in a special collection can be indicated by descriptive entries without implicating copyright.

• Where digitized special collections are posted online, reasonable steps should be taken to limit access to material likely to contain damaging or sensitive private information.

• Full attribution, in a form satisfactory to scholars in the field, should be provided for all special collection items made available online, to the extent it is reasonably possible to do so.

Enhancements
• The fair use case will be even stronger where items to be digitized consist largely of works, such as personal photographs, correspondence, or ephemera, whose owners are not exploiting the material commercially and likely could not be located to seek permission for new uses.

• Libraries should consider taking technological steps, reasonable in light of both the nature of the material and of institutional capabilities, to prevent downloading of digital files by users, or else to limit the quality of files to what is appropriate to the use.

• Libraries should also provide copyright owners with a simple tool for registering objections to online use, and respond to such objections promptly.

• Subject to the considerations outlined above, a special collection should be digitized in its entirety, and presented as a cohesive collection whenever possible.

• Adding criticism, commentary, rich metadata, and other additional value and context to the collection will strengthen the fair use case.

• The fair use case will be stronger when the availability of the material is appropriately publicized to scholars in the field and other persons likely to be especially interested