Dynamic Metadata & Data Management for Digital Collections

Sharing More than a Border
OLA/WLA Conference 2008

OCLC™ The world's libraries. Connected.

Data Curation

"Digital curation, broadly interpreted, is about maintaining and adding value to, a trusted body of digital information for current and future use." Lorcan Dempsey

"Data's future quality –richness, trustworthiness– is a function of investment in it." Philip Lord

Authenticity. Why is it Important?

The quality of being reliable or trustworthy

Libraries and Museums have trust – we don't want to blow it.

Program data

What are your sources of program data?

Which data sources are important?

Standards adopted & documented

Policies, regulations in effect

Best Practices adopted & documented

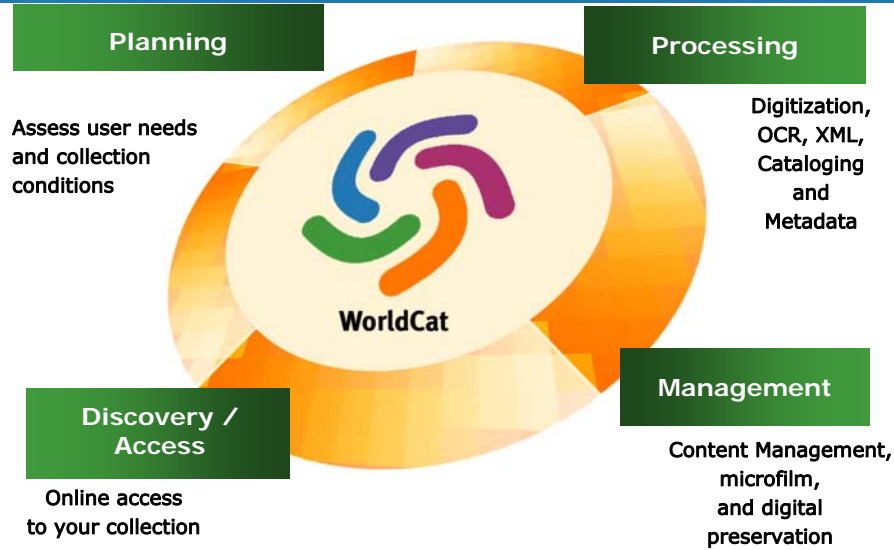Metadata of all types

Technical infrastructure documentation

How about some more?

Legal documentation

Intellectual property assets

Risk management policies

Evaluation & assessment metrics

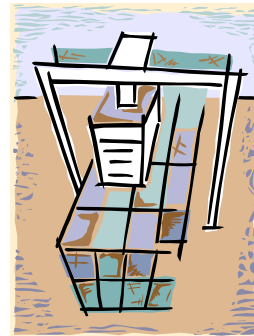Budget & financial data

User data

Digital collections lifecycle

OCLC

Planning

Assess user needs and collection conditions

Processing

Digitization, OCR, XML, Cataloging and Metadata

WorldCat

Discovery / Access

Online access to your collection

Management

Content Management, microfilm, and digital preservation

Data curation awareness

OCLC

Documentation

Start early, revise often

Metadata applications

Crossover to data driven, XML compliant web site---

Database framework in which data will be housed and disseminated.

Metadata

Title:	Cable car shuttle, Crystal Bay, Nevada
Photographer/Artist/Creator:	
Image Date:	Ca. 1970
Summary/Description:	Caption on image: This cable car is an exact replica of the world famous San Francisco cable cars and furnishes shuttle service between two great combined resort hotels Cal-Neva Lodge and Cal-Neva Biltmore, Crystal Bay, Lake Tahoe, Nevada. View of passengers on the cable car; postcard
Subject:	Vehicles-Tahoe, Lake, Region (Calif. and Nev.)
Location:	Crystal Bay (Nev.)
	Washoe County
	Nevada
Collection:	Images of Lake Tahoe
Image Number:	P92-03-70
Electronic Publisher:	Special Collections Department, University of Nevada, Reno Library
Access and Ordering Information:	Please see: http://www.library.unr.edu/specoll/photos.html
Date.Digital:	2003
Donor:	
Format.Creation:	These images exist as archived TIFF images and one or more JPEG versions for general use. They were scanned at 400 dpi, 24 bit from the original using an Epson 636 scanner, default color configuration.
File Name:	P92-03-70.jpg
Resource Type:	Image
Source:	
Holding.Institution:	University of Nevada, Reno

Descriptive Metadata

Administrative Metadata

Technical Metadata

Add to My Favorites   View My Favorites

Powered by
CONTENTdm

Technical metadata

Shortcuts... | 86% | Edit Pictures... | Auto Correct

Properties

distance%20learning.jpg

Picture properties

Type: JPG File
Dimensions: 825 x 654 pixels
Size: 317 KB
Modified: 10/2/2006 4:43:28 PM
Created: 4/7/2004 1:55:27 PM
Location: C:\Data\My Pictures\graphics\
Read-only: No
Title:
Description:
Keywords:
Horizontal Resolution: 72 dpi
Vertical Resolution: 72 dpi
Bit Depth: 24
Frame Count: 1

Camera properties

This file has no camera properties.

File Management Protocol

Example of file naming structure.

Meets standard of eight-digit file names

Three-digit extension

Compatible across computer platforms, hardware, software

File Path Naming Convention

Photos
└─1929
   └─21
      └─18
         └─08

ISO 9660 compliant file name options

03.tif
03ac.tif
03curtis.tif

Documentation for technology standards

There are standards relevant to:

- Data
- Networking
- Communication
- Imaging
- Preservation
- Metadata

Field properties in CONTENTdm

CONTENTdm Field Properties

Click on a field name to edit the properties for that field:

Field Name	Dublin Core Mapping	Data Type	Large field	Searchable	Hidden	Controlled Vocabulary
Title	Title	Text	No	Yes	No	No
Subject	Subject	Text	No	Yes	No	No
Description	Description	Text	Yes	Yes	No	No
Creator	Creator	Text	No	No	No	No
Publisher	Publisher	Text	No	No	No	No
Contributors	Contributors	Text	No	No	No	No
Date	Date	Text	No	No	No	No
Type	Type	Text	No	No	No	No
Format	Format	Text	No	No	No	No
Identifier	Identifier	Text	No	No	No	No
Source	Source	Text	No	No	No	No
Language	Language	Text	No	No	No	No
Relation	Relation	Text	No	No	No	No
Coverage	Coverage	Text	No	No	No	No
Rights	Rights	Text	No	No	No	No

Add new field
Back to Collection Administration page

Standards vs implementation

Implementation workflow—

Adoption of standards

Adoption of best practices models

Implementation within institutional policy framework

Metadata consistency through a local style guide

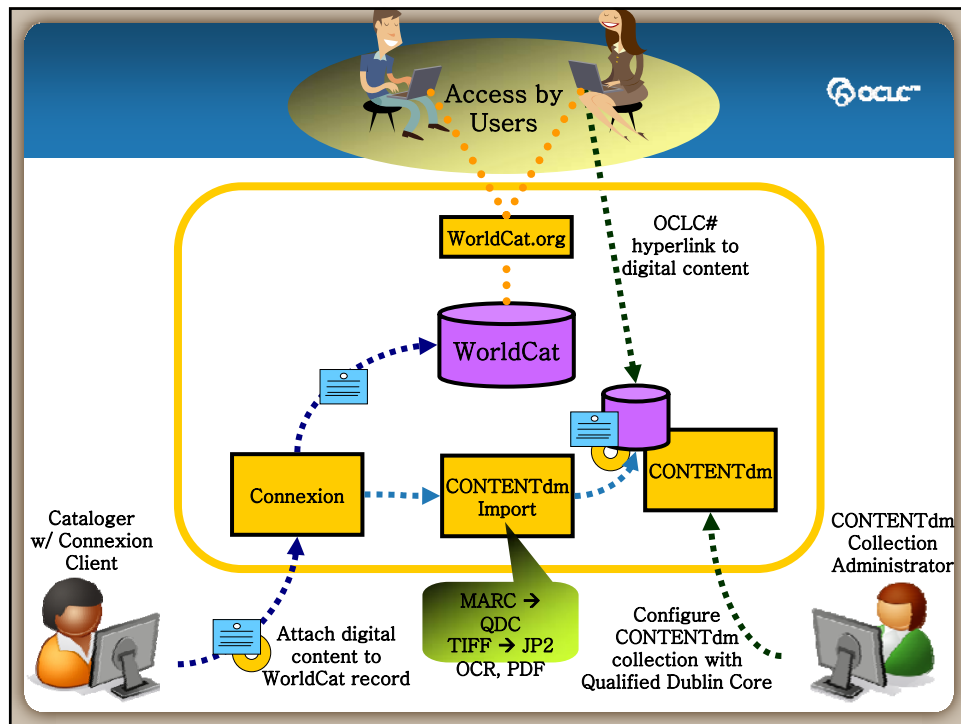Documentation of decisions and change management

Data interoperability

It can be achieved through:

- Schema
- Crosswalks/mapping
- Standardization
- Controlled Vocabulary

Next Generation Cataloging Services Pilot

Capture ONIX metadata from publishers and vendors upstream

Academic and public libraries

Publishers and vendors

On the Record

Report of
The Library of Congress Working Group
on the Future of Bibliographic Control

January 9, 2008

Data Preservation & Sustainability

OCLC

All digital preservation will depend on data curation

PREMIS: PREservation Metadata Implementation Strategies [OCLC - Projects] - Microsoft Internet Ex...

File   Edit   View   Favorites   Tools   Help   Links »

United States

☐ Community   ☐ Products and Services   ☐ Professional Development   ☐ Research

◄ Projects

PREMIS: PREservation Metadata Implementation Strategies

PREMIS (PREservation Metadata: Implementation Strategies) Working Group

PREMIS: PREservation ►
Metadata
Implementation
Strategies
· **Background**
· **Membership**
· **Preservation**
  Metadata
  Framework Working
  Group (archived)
· **Resources**

[NEW] PREMIS working group releases final products!

The final products of the PREMIS Working Group are now available for download:

▪ _Data Dictionary for Preservation Metadata: Final Report of the PREMIS Working Group_ (May 2005) (PDF:3.2MB/237pp.)

This publication includes the PREMIS Working Group Final Report, the Data Dictionary, and Examples. These sections are also available as separate documents:

○ _PREMIS Data Dictionary, version 1.0_ (PDF:2.1MB/101pp.)
A data dictionary for core preservation metadata needed to support the long-term preservation of digital materials.

○ _PREMIS Working Group Final Report_ (PDF:606K/53pp.)
Provides information on the objectives of the PREMIS working group, discussions of particular semantic units and special topics, and a glossary.

○ _Use of the Data Dictionary: PREMIS Examples_ (PDF:4.5MB/97pp.)
Illustrates use of the PREMIS data dictionary for a number of digital materials and preservation contexts.

▪ XML Schemas for the PREMIS Data Dictionary
A set of XML schema containing all the semantic units defined in the PREMIS Data Dictionary.

▪ _Implementing Preservation Repositories for Digital Materials: Current Practice and Emerging Trends in the Cultural Heritage Community._ (September 2004) (PDF:668K/66pp.)
Reports the findings of a survey focusing on key aspects of planned and existing digital preservation repositories, including mission, funding, preservation strategy, access policies, and current practice for managing preservation metadata in digital archiving systems.
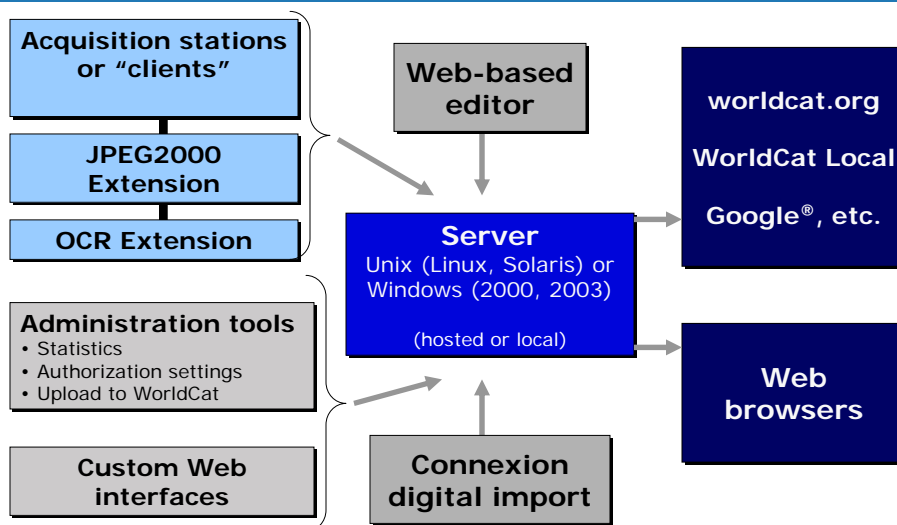
PDF:3.2MB/237pp.

PREMIS maintenance activity

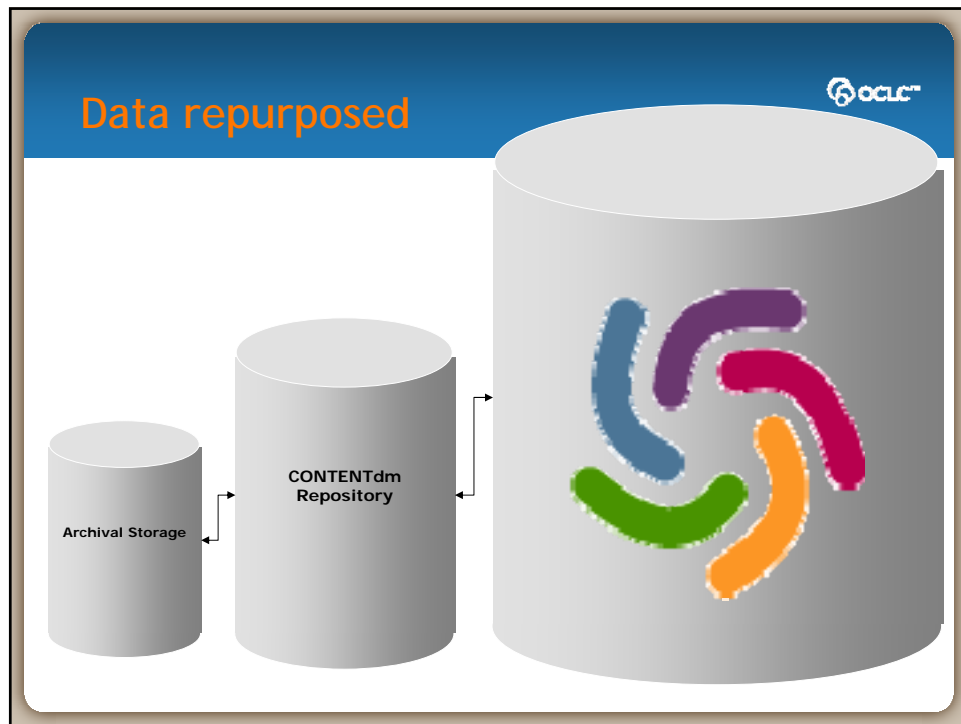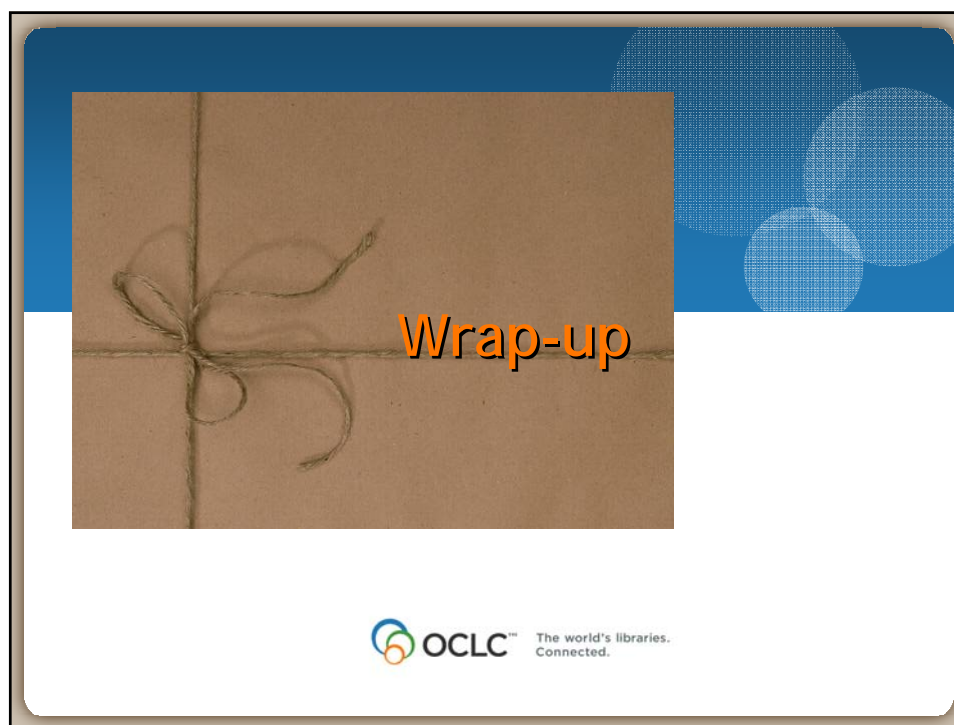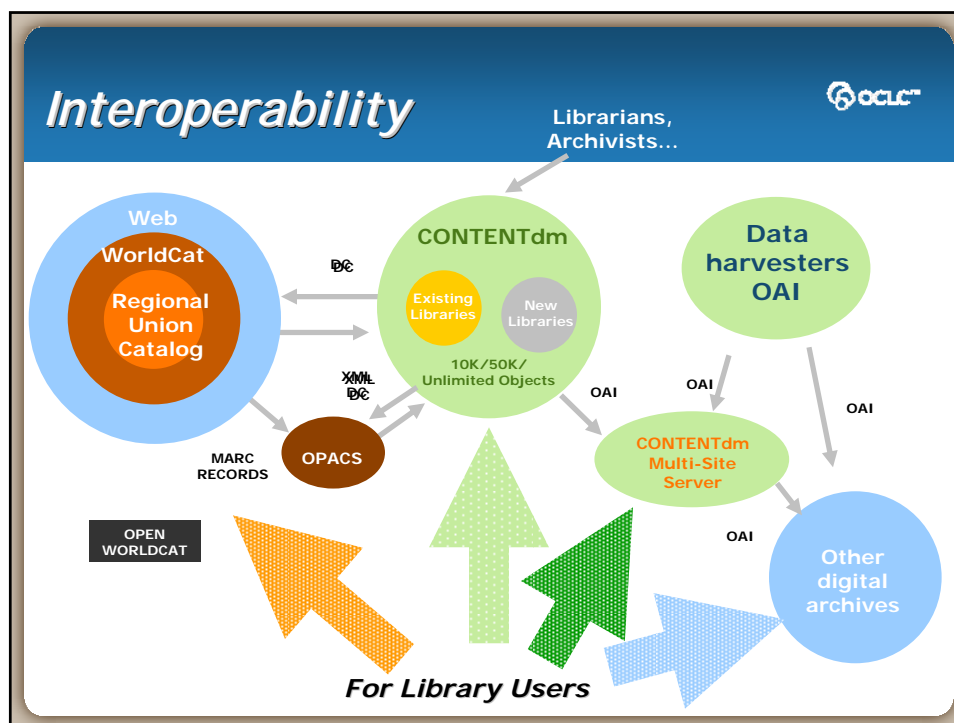## IT Data Sources

In information technology—

System architecture - CONTENTdm

Acquisition stations or "clients"

JPEG2000 Extension

OCR Extension

Administration tools
- Statistics
- Authorization settings
- Upload to WorldCat

Custom Web interfaces

Web-based editor

Server
Unix (Linux, Solaris) or Windows (2000, 2003)

(hosted or local)

Connexion digital import

worldcat.org

WorldCat Local

Google®, etc.

Web browsers

Data repurposed

Archival Storage

CONTENTdm Repository

Digital Repositories – multiple applications
University of Utah

On-going program management

Use and evaluation - continue to measure & collect data

Feedback to funding sources – a continuous loop

Marketing and publicity

Promotion of digital program goals

Organizations and people

Managing digital assets changes the organization

Supporting users requires everyone to be skilled

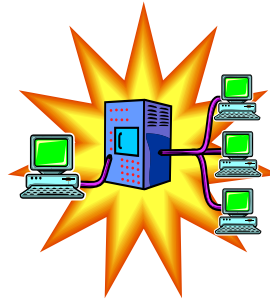New technologies means new skills but the principles remain the same

Training is the solution

Sustainability

- Data curation
- Backup plans
- Costs for upgrades
- Documentation
- Review and schedule upgrades

Documentation helps

Change management is vital to digital projects.

- Be flexible
- Be willing to change
- Document your decisions
- Create procedures, protocol manuals
- Technology is the most changeable factor

So, tomorrow...

Digital data repositories already sustained > 30 years
- How?
- Vision, leadership, commitment

Libraries, archives, museums sustained 100s of years
- How?
- Vision, leadership, commitment
- Aggregate value proposition
- Perception now under threat!

Thank You, ??s

Gayle Palmer
Digital & Preservation Program Manager
OCLC Western
Palmerg@oclc.org
800-854-5753, ext 4041