

**BEYOND DUBLIN CORE:
DEVELOPMENT OF THE WORKFLOW
MANAGEMENT SYSTEM AND
METADATA IMPLEMENTATION AT
RUTGERS,
THE STATE UNIVERSITY OF
NEW JERSEY**

**Presented by Mary Beth Weber and
Sharon Favaro,
Rutgers, The State University of
New Jersey
DigCCurr2007
April 19, 2007**

What is digital curation?

A few of the current definitions/interpretations
(continued):

- **Wikipedia:** "Digital curation, broadly interpreted, is about maintaining and adding value to a trusted body of digital information for current and future use. Digital curation encompasses all of the actions needed to maintain digitised and born-digital objects and data over their entire life-cycle and over time for current and future generations of users. Implicit in this definition are the processes of digital archiving and digital preservation but it also includes all the processes needed for good data creation and management, and the capacity to add value to data to generate new sources of information and knowledge."

A few of the current definitions/interpretations:

- **Digital Curation Centre:** "Digital curation is the management and preservation of digital data over the long-term. The increasing amount of data being created in digital formats...and the generation of new digital data in the sciences means that much of it is at risk of being unreadable in the future. Someone needs to confront this problem by firstly advising on what information to preserve and then on how best to preserve it, from the beginning of the information life-cycle to the end; hence digital 'curation'."
- **Digital Data Curation Task Force: Report of the Task Force Strategy Discussion Day (2002):** "Tony Hey took up the term...to distinguish the actions involved in caring for digital data beyond its original use, from digital preservation."

Development of a repository

- In 2003, Rutgers University Libraries (RUL) received a National Leadership Grant from the Institute of Museum and Library Services (IMLS) to develop the New Jersey Digital Highway (NJDH), a statewide repository and collaborative portal to provide a “one stop shop” for information about New Jersey’s history, culture, heritage, and immigration experience through photographs, books, documents, periodicals, three-dimensional objects, audio, and video
- Partners included the American Labor Museum/Botto House, the New Jersey State Library, the New Jersey Historical Society, the New Jersey State Archives, and smaller libraries, museums, archives, historical societies, public broadcasting organizations, and public schools

New Jersey Digital Highway (NJDH)

<http://www.njdigitalhighway.org/>

The screenshot shows the New Jersey Digital Highway website in a Mozilla Firefox browser window. The browser's address bar displays the URL <http://www.njdigitalhighway.org/>. The website's header features the logo "N.J. New Jersey Digital Highway DH" with the tagline "Where History, Culture, and Learning Merge". Below the logo, navigation links are provided for "Everyone", "Educators", "Students", and "Librarians & Curators". A vertical sidebar on the right contains buttons for "Home", "Site Search", "Sitemap", "Contact", and "About".

The main content area includes a paragraph stating: "The New Jersey Digital Highway is your 'one stop shop' for New Jersey history and culture, from the collections of NJ libraries, museums, archives and historical societies." Below this is a "Quick Search" section with a search box and a "Search NJDH" button. A "Working to Make NJDH Better!" banner is also present.

A "Thank you for evaluating the New Jersey Digital Highway" message is displayed, followed by a "NJDH Evaluation Update" section with a "read more >>" link. A "History Events Calendar" is shown with a calendar grid for the month of April, and a "Click here" link below it.

The "Interest Areas for Different Audiences:" section lists four categories:

- Everyone** can find information about New Jersey and about documenting your own family history and culture.
- Educators** will also find information, tools and activities for teaching about the immigrant experience in New Jersey.
- Students** will find information about New Jersey, resources to help them learn and prepare school projects, and activities to explore their own family history.
- Librarians & Curators** will find information about creating and managing their collections, as well as opportunities to participate in New Jersey Digital Highway.

At the bottom of the page, a "Special Feature" section is partially visible, showing a small portrait of a man.

Developing a Metadata Creation and Management Tool at RUL

So how did we get there...

- Repositories are an emerging technology, and a metadata creation utility was needed that could meet the changing needs of NJDH partners, RUL, RU faculty, students, and the public
- Commercial products were reviewed and the decision was made not to purchase a commercial product for content management

Developing a Metadata Creation and Management Tool at RUL Developing a Metadata (continued)

- The Workflow Management System (WMS), built in-house by RUL programmers, a metadata creation and management tool that feeds into **RUcore**, the institutional repository, uses a robust hybrid metadata schema to ensure scalability, interoperability, and extensibility

Top reasons for RUL creating in-house repository management software:

- Freedom to use a cross breed of metadata schema for interoperability with other systems and presentation platforms
- Ability to guide developments, enhancements, and adapt to changes
- Free access to source codes

WMS Features

- Enables creation and editing of metadata
- Creates a METS-XML wrapper for digital objects and associated metadata for export and ingest
- Employs an underlying metadata schema that draws on MODS, METS, and PREMIS to ensure scalability for various projects, and interoperability with other systems

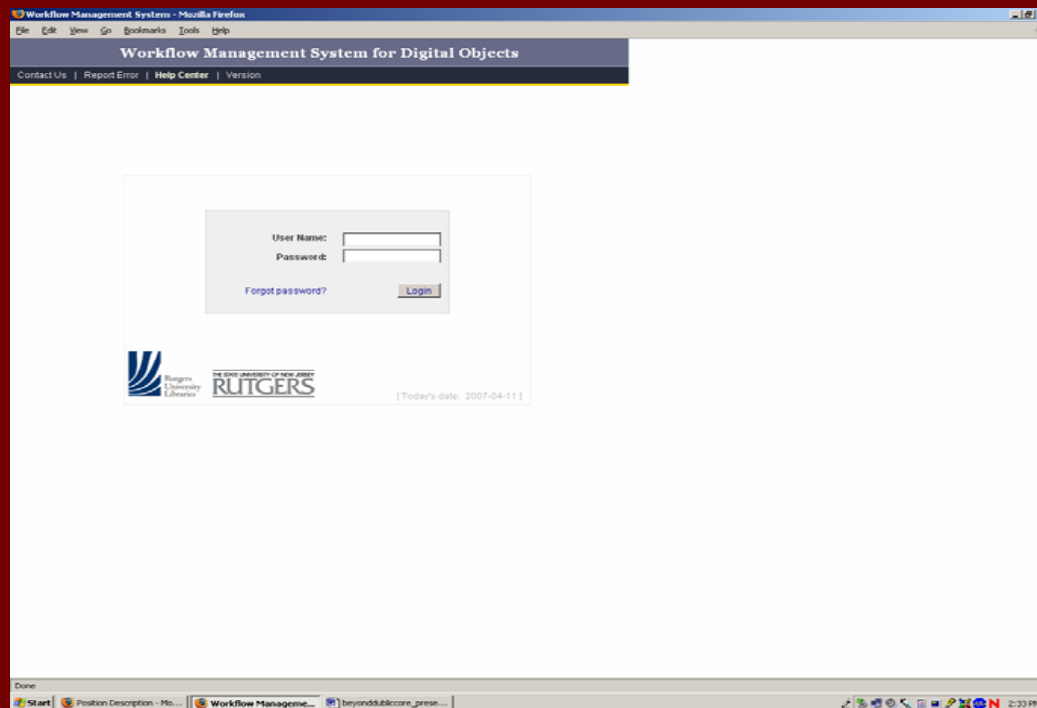
WMS Features

(continued)

- Metadata schema maps to Dublin Core, which is mandatory for participation in the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) for sharing information between repositories
- Supports multiple strategies for entering objects and creating metadata, including batch import of digital files, the ability to add an entire digital object collection through batch import, and the ability to migrate metadata from a collection owner's database into the WMS

Workflow Management System

Independent platform web accessible system to create metadata and is the front end to ingest digital objects into RUCore



Rutgers Institutional Repository

Software

- The goal was to build a trusted repository architecture with a strong preservation focus
- Fedora provides native support for METS, and RUL collects preservation, rights and provenance information as a critical part of a long-term preservation strategy

Rutgers Institutional Repository

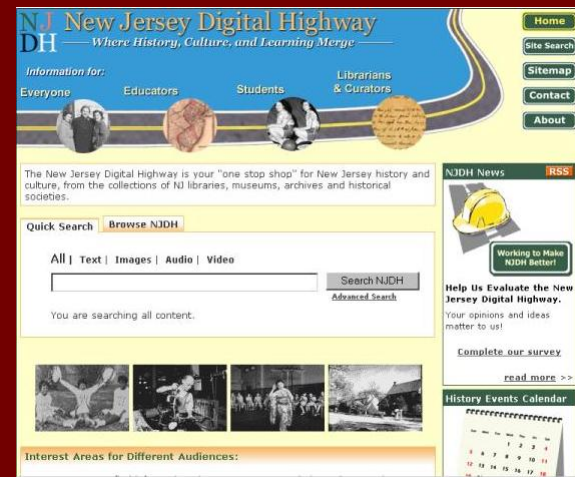
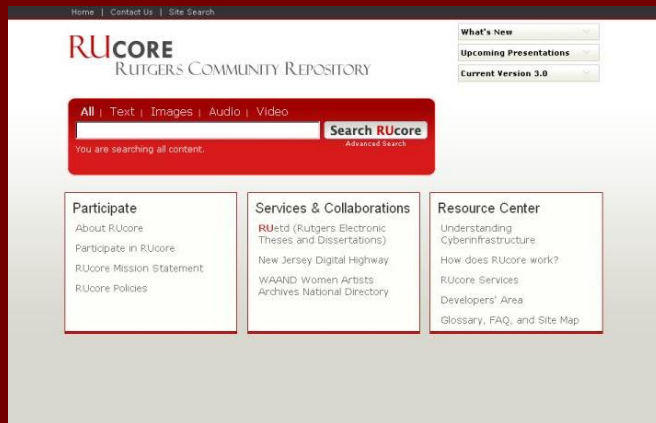
Software (continued)

- Fedora provides a flexible architecture, and was the best service oriented repository architecture.
- Fedora was open to collaborative development, and Rutgers became a Fedora development site. This enables us to benefit from shared collaboration and to extend our finite programming resources

Public View of the RUL Repository

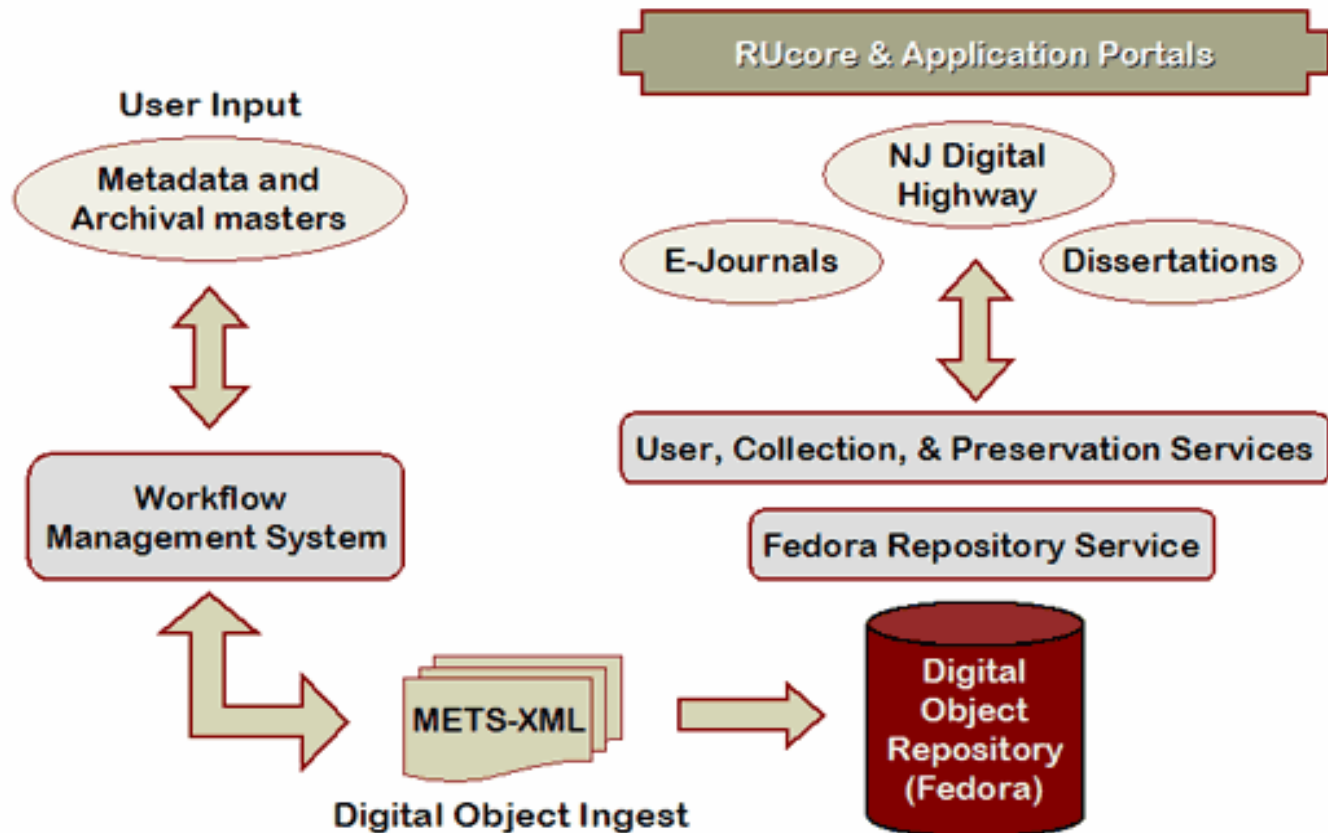
Workflow Management System

Fedora Repository Architecture

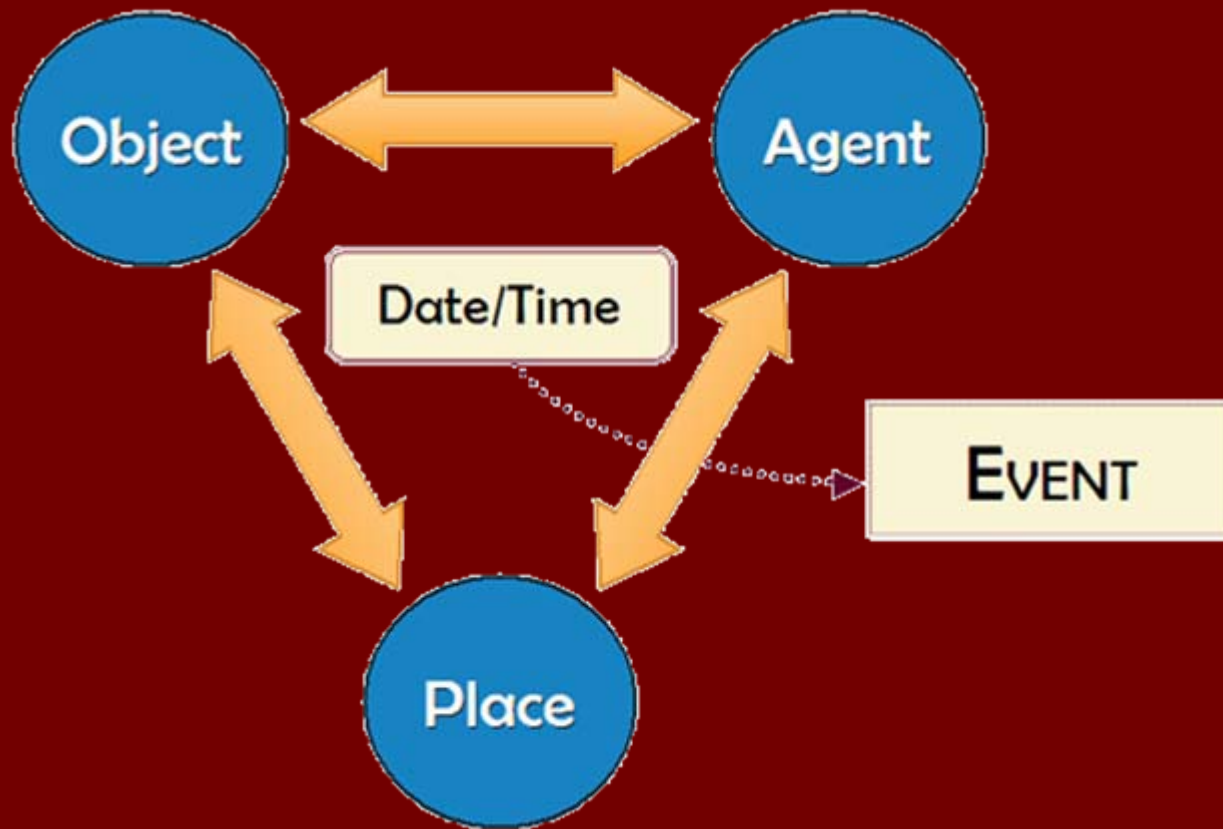


Technical view of RUL Repository

RUcore - How it Works



RUcore Data Model



RUcore Data Model

- Innovative “event-based” data model within a METS data architecture to capture information and ancillary documents for any facet of a resource’s lifecycle whether describing the intellectual content, the provenance, the digital lifecycle or the rights lifecycle:
 - Object** – representation of an information object (photograph, three-dimensional piece of artwork, or a digital image)
 - Agent** - person or organization responsible for creating, describing, managing, or using information about a particular object (photographer, artist, metadata manager, author)
 - Place** - context independent, and can be place of creation, display, location, etc.
 - When an object interacts with an **agent** or **place** at a specific point in **time**, an **event** occurs in the lifecycle of the object. Examples of events are provenance events (donation, acquisition), or preservation events (tape removal, digital restoration).

Digital Curation at Rutgers University

Digital curation process at RUL facilitates:

- Resource search and discovery through description and controlled vocabulary terms
- Use and re-use of valuable or fragile materials previously available to a limited privileged audience
- Access to objects that are not well served by traditional MARC cataloging
- Preservation of resources

Digital Workflow Team

So how do we get started . . .

- Digital projects at RUL are handled using a team approach: WMS Manager, Project Manager, Collection Manager/Curator, Object Creators, and Metadata Manager
- WMS Manager oversees all development of the WMS, including programming, specifications for new releases, coordination of testing for new releases, and metadata standards and applications
- Project Manager is the primary liaison to the owner/curator of a given collection and also coordinates the digitizing of the objects
- Owner/Curator of a collection is responsible for providing a description of the collection as well as descriptive information for each object in the collection (format, dates, restrictions on access, copyright information if applicable)
- Object Creator is a staff member who creates metadata for specific collections and projects

Digital Workflow Team: Metadata Manager

- Metadata Manager has primary responsibility for metadata design and creation, manages metadata timetables and deliverables, develops templates and controlled vocabularies, and trains the curator/owner and staff when metadata is created locally
- Metadata managers have full access to the all WMS administrative settings, establish collection and project records, create user accounts, create project specific templates, and design project workflow

Digital Workflow: Beginning a Project

- When new collections are acquired:
 - Project Manager and Metadata Manager meet with the Collection Owner to evaluate the collection
 - Workflow is established for scanning digital objects and metadata creation
 - Administrative information is gathered
 - Collection condition is evaluated
 - Rights restrictions are determined

Digital Workflow: Beginning a Project (continued)

- The Metadata Manager works with Collection/Curator Manager to:
 - Review or set up a finding aid, inventory or other resources containing an object description, date, and digital file number to assist with metadata creation for unprocessed collections

Metadata Manager's Role

- Metadata Manager creates an organization record that includes institutional information (MARC org ID, address, and a handle ID)
- Metadata Manager works with the Collection Curator to create a collection record including collection ID, collection title, and provides information such as name, affiliation, and email address for the Metadata Manager, Collection Owner/Curator, and Object Creators. Collection record also contains descriptive information and may be linked to other collections through the hierarchical structure

WMS Controlled Vocabulary Module

- Controlled vocabulary module restricted to Metadata Managers enables add vocabulary terms without intervention of a programmer
- Vocabulary currently featured in the WMS consists of:
 - Library of Congress Subject Heading (LCSH)
 - Art Architecture Thesaurus (AAT)
 - Rutgers University vocabulary (RULib)
- Previously, any changes to controlled vocabulary terms triggered a two-part process:
 - Updating a master spreadsheet
 - Working with a programmer to accommodate changes and additions

Templates for Digital Workflow

- Templates streamline workflow and automatically populate metadata records with required information specific to a given project
- Templates created at the project level free the object creator from the repetition of supplying required information
- Templates are particularly useful for individuals with little familiarity with the WMS since they limit the elements and information that must be provided
- Drawbacks:
 - Object Creator cannot change or modify a template or create another template set at project level
 - The current WMS release supports a single object architecture; expansion of this capability is under consideration for a future release

Batch Processing Functions

- The WMS provides the ability to add an entire digital object collection through batch import or to migrate digital files and metadata from a collection owner's database into the WMS
- Batch import may be used for any databases or spreadsheets that export records in text
- Challenges of batch import:
 - Reconciling subtle inconsistencies in vocabulary
 - Close attention to details such as capitalization
 - Resolving the issue of combined architectures on one spreadsheet since the current WMS implementation can handle one object architecture per spreadsheet for batch import

Batch Processing Functions

- Batch import used for two projects:
 - Integrated a public library's historic photograph collection into NJDH. The library had an existing Access database that required some tweaking such as mapping their metadata elements to the WMS proprietary metadata, and providing controlled vocabulary terms and authority control for names.
 - Imported an archive's collection from an Excel spreadsheet. The Metadata Manager created a basic Excel sheet for an archive project partner with required elements and suggested additional elements that the project archivist was likely to use.

WMS Metadata

- The WMS uses an underlying metadata schema that draws on MODS, METS, and PREMIS for scalability with various projects, and maps to Dublin Core for interoperability with other systems through OAI-PMH.
- The WMS is composed of five types of metadata that are drawn from various schemas:

WMS Metadata (continued)

■ Descriptive metadata

- Provides information for users to discover and obtain access to information resources
- MODS was chosen since it enables records to be mapped back to IRIS in MARC format
- MODS provides and retains standard bibliographic cataloging principles
- It enables RUL to provide information through multiple presentation standards in a schema easily understood by each user community

WMS Metadata (continued)

■ Source metadata

- Describes provenance, condition and conservation of analog source materials, such as photographs, books, maps, audio, and video
- Defines the nature of an analog source object and uses elements from the PREMIS (PREservation Metadata) schema.

WMS Metadata (continued)

■ Technical metadata

- Provides information about the digital master files that RUCore will maintain for long-term preservation and access
- Required for all digital objects, both born digital or digitized from an analog source object
- RUL uses the ANSI/NISO Standard Z39.87-2006 Data Dictionary – Technical Metadata for Digital Still Images for technical metadata. Metadata standards for sound and video are currently underway

WMS Metadata (continued)

■ Rights metadata

- Identifies the rights holder(s) for each information resource and identifies the permissions for use that the rights holder has granted, including any restrictions
- Refers to digitized objects that are available via the Internet
- RUL uses PREMIS for Rights metadata
- RUL has developed a simple rights schema based upon rights events that enables archivists or rights administrators to document every instance of research or permissions requests, rights transfer, copyright status, etc.

WMS Metadata (continued)

- The draft rights schema can be reviewed at:

http://rucore.libraries.rutgers.edu/collab/ref/doc_mwq_rulib_rights_md_draft.pdf

- **Digiprov metadata** or digital provenance metadata, provides a digital "audit trail" of any changes to the metadata

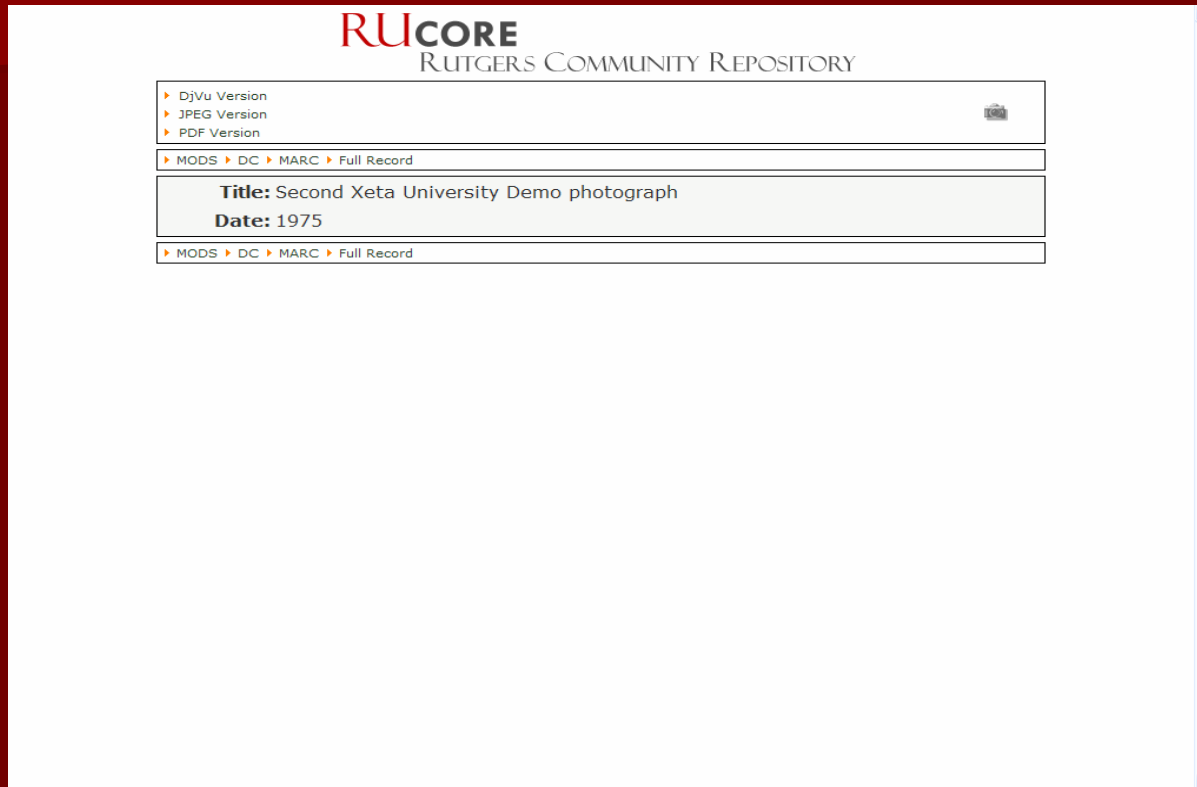
WMS Final Review: Completed Records

- When metadata creation is complete, objects may be viewed in FOXML, METS, or HTML. The HTML view breaks the metadata down by sections and displays descriptive metadata, source metadata, technical metadata, rights metadata, a structure map, information on files associated with the object, and Digiprov metadata.

WMS Output and Display

- Objects are available via different presentation formats after they are ingested into RUCore:
 - Still images are available as DjVu, PDF, or JPEG.
 - Sound objects are available as MP3 files.
 - Video resources are available as Quick Time or Flash Video
- File upload assigns a persistent identifier based on organization information. RUL uses the Corporation for National Research Initiatives' Handle System to assign, manage, and resolve persistent identifiers for digital objects. Clicking on the handle or persistent identifier, permits users to view records in MODS, Dublin Core, MARC, or a combined full record with all schemas

WMS Output and Display



The screenshot displays the metadata page for a digital object in the RUCORE system. At the top, the logo for RUCORE (Rutgers Community Repository) is shown. Below the logo, there are several navigation and metadata sections:

- A box containing links to different file formats: [DjVu Version](#), [JPEG Version](#), and [PDF Version](#). A small camera icon is visible on the right side of this box.
- A breadcrumb trail: [MODS](#) > [DC](#) > [MARC](#) > [Full Record](#).
- A central box with the following metadata:
 - Title:** Second Xeta University Demo photograph
 - Date:** 1975
- A second breadcrumb trail: [MODS](#) > [DC](#) > [MARC](#) > [Full Record](#).

WMS Releases and Testing

- Ongoing testing for new WMS releases occurs on a regular basis. Testing is intensive, consisting of daily three-hour sessions over a period of two or three weeks
- Programmers work on enhancements between releases and have target dates for new functionalities geared to specific future releases
- Testers use a free open source error reporting system to report problems to programmers and to share information with other testers
- Outstanding bugs are reviewed when testing has been completed to determine what can be resolved, and what functionality should be pushed to a future release

What's next?

RUcore version 6.0 (Summer 2007)

- Collection Owner's portal
 - use statistics for objects
 - administrative area to upload documents

Questions?

Thank you.

Email:

Mary Beth Weber

mbfecko@rci.rutgers.edu

Sharon Favaro

sbs49@rci.rutgers.edu