RUtgers COmmunity REpository (RUcore)

A FEDORA-based Institutional Repository To Support Multidisciplinary Collections

CNI Task Force Meeting
April 3 – 4, 2006
Arlington, VA

Ronald C. Jantz
Rutgers University Libraries
Welcome

Welcome to RUCore, the Rutgers Community Repository. RUCore preserves and makes available digital collections and the significant scholarly, research and administrative resources created by the Rutgers community and its collaborators.

RUCore is a cyberinfrastructure and trusted repository architecture that meets or exceeds the emerging digital information standards promoted by granting agencies and professional associations.

Learn More about RUCore

- Our mission statement
- Submitting or recommending collections to RUCore
- What collections does RUCore already include?
- What tools and services can RUCore provide to RU faculty and administrators?
- How does RUCore work?
RUcore Objectives

- To provide seamless, perpetual access to digital collections -- our resources and the resources of others.

- To develop a flexible framework of “core” capabilities providing the enabling infrastructure, interoperability, and sustainability.

- To create an information architecture which will support multidisciplinary collections.
Institutional Repositories

Institutional Repository

“... a set of services that a university offers to the members of its community for the management and dissemination of digital materials created by the institution and its community members.” – Clifford Lynch

Types of Materials Deposited in Repositories*

- Theses/Dissertations
- Pre-prints/post-prints
- Digital images
- Assets from Special Collections
- Technical reports/working papers
- Data sets

RUcore - How it Works

User Input
- Metadata and Archival masters
  - Workflow Management System

RUCORE & Application Portals
- E-Journals
- NJ Digital Highway
- Dissertations
  - User, Collection, & Preservation Services
  - Fedora Repository Service

Digital Object Ingest
- METS-XML

Digital Object Repository (Fedora)
Application Portals and Services

- **Rutgers University Community Repository**
  - [http://rucore.libraries.rutgers.edu](http://rucore.libraries.rutgers.edu)

- **Cultural Heritage Resources - The New Jersey Digital Highway**
  - Books, diaries, letters, photographs, maps, multimedia, etc.
  - [http://www.njdigitalhighway.org](http://www.njdigitalhighway.org)
  - Funded by an IMLS Grant

- **E-journals - an e-journal publishing service**
  - See Pragmatic Case Studies in Psychotherapy
  - [http://pcsp.libraries.rutgers.edu](http://pcsp.libraries.rutgers.edu)

- **Electronic Theses and Dissertations (ETDs)**
  - To be launched in May, 2006
The New Jersey Digital Highway (NJDH) is a new way to explore our state's history and culture. NJDH is a collaboration of the state's libraries, museums, archives and historical societies to provide a "one stop shop" for New Jersey's cultural treasures—photographs, books, documents, three-dimensional objects, video and audio.

*Search Our Collection*

Limit search results to:

- Text
- Images
- Audio
- Video

*Advanced Search*

*Clear Buttons*

*NJDH News*

**Fundraising for Preservation and Conservation Workshop**

Tuesday, October 11, 2005.

A one-day workshop cosponsored by Rutgers University Libraries held at Rutgers... read more >>

2005 NJDH Educators Seminar Program.
FEBRUARY 17, 2006, FROM THE EDITOR


The Edwards & Kannan case study is noteworthy, in part, for possessing interrelated levels of case study context. Specifically, this case involves Vumile, an individual in group therapy, with comparisons to other clients in the group. In addition, the therapy employs a manualized protocol that is a variant of one employed in randomized clinical trials. The case study, then, is able to explore the links among individual therapy, group therapy, manualized treatment, and randomized clinical trial research, which is associated with the empirically supported treatment (EST) movement.
In its most general sense, a digital collection is simply a grouping of objects according to some criteria.

Types of digital collections in Rucore

- Explicit – A digital collection whose object membership is specified explicitly within the descriptive metadata.
- Dynamic – A digital collection of objects which are grouped according to user specified criteria.
The Collection Architecture

Solid line – explicit membership
Dashed line – dynamic membership
The Digital Object

The digital object is the basic unit of management, encapsulating all essential information about the “document” to be disseminated and preserved, including:

- Descriptive metadata (based on MODS), administrative and event-based migration metadata
- Byte streams in both presentation and archival formats (non-proprietary, infrastructure independent forms)
- Persistent IDs and digital signatures
- Software for dynamic behavior
Digital Object - Examples

**Book Object**

- **Persistent ID**
- **Special Behaviors**
- **Metadata**
- **Data streams**
  - SMAP1 – StrMap (TOC)
  - DJVU1 - presentation
  - PDF1 - presentation
  - XML1 – OCR text
  - ARCH1 - Archival master (tiffs of each page)

**Audio Object**

- **Persistent ID**
- **Special Behaviors**
- **Metadata**
- **Data streams**
  - SMAP1 – StrMap (TOC)
  - MP3-1 - presentation
  - PDF1 – Transcript
  - XML1 – OCR text
  - ARCH1 - Archival master (wav files & transcript)

**Categories**

- Descriptive
- Technical
- Source
- Rights
- Digital Prov.
Life Cycle of the Digital Object

- **Document** (T₁)
- **Digital Original**
- **Migrated Derivative Objects** (T₂)
- **Output Device** (T₃)
A Trusted Repository

A trusted repository will safeguard the availability and authenticity of digital assets - pre-prints, datasets, e-journals, dissertations and more.

- Technical infrastructure
- Digital preservation services
- Procedures – workflow, roles
- Policies & Standards – scanning, validation, levels of preservation
- Administrative, management, financial

Certification as a trusted repository*

* RLG/NARA draft “An Audit Checklist for the Certification of Trusted Digital Repositories”
FEDORA Preservation Services
Working Group

Charter: Define the requirements and architecture for preservation services that can be integrated into Fedora.

Objectives
- Deliver preservation services as part of the Fedora framework.
- Enable a Fedora-based repository to become certified as a trusted repository.

Working Group Membership
Cornell, Harris Computing Services,
National Library of Australia, Northwestern, Rutgers, Tufts, Yale
FEDORA Preservation Services

- Audit trails and versioning*
- Signature/Checksum creation and validation
- Event logging and alerting for preservation services
- Object format validation
- Content model validation
- Enable Repository static/active states
- History service of major repository events

* - Available in Fedora Release 2.1
Certification Checklist

- Repository has an identifiable, written definition for each AIP or class of information preserved by the repository (the digital object architecture).

- Repository obtains sufficient physical control over the digital objects to preserve them – authentication and integrity checking (checksum creation/verification).

- Repository has and uses a naming convention that can be shown to generate visible unique identifiers for all AIPs (PIDs – CNRI Handles).

- Repository’s ingest process verifies each SIP for completeness and correctness (format and content model validation).

- Repository provides Producer/depositor with appropriate responses at predefined points during the ingest processes (event alerting services).
FEDORA Event Management

Systems & Applications (detect, gather)

- Crawler
- Fedora
- Apps

Logging Software

Fedora Core

Closed Event Pool

Active Event Log
- Event 1
- Event 2
- Event 3
- Event 4-6 (chain)
- Event n

Listening & Communication Services

- Preservation Services
- Collection Services
- User Services

take action, close event
## Collection Management Statistics Report - All Collections

### General Information
- There are currently 2 active objects in the FEDORA Repository.

<table>
<thead>
<tr>
<th><strong>Full Summary</strong></th>
<th>View a summary of all statistical data for a selected timeframe.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Search Results</strong></td>
<td>View a summary of all relevant search activity in the last ten days.</td>
</tr>
<tr>
<td><strong>[Last 10 Days]</strong></td>
<td></td>
</tr>
</tbody>
</table>

### Object Logs
- **Edit Logs**
  - Browse the log for edited objects in the selected collection.

- **Ingest Logs**
  - Browse the log for ingested objects in the selected collection.

- **Purge Logs**
  - Browse the log for purged objects in the selected collection.

- **Checksum Failure**
  - Browse the log for objects which failed a checksums test in the selected collection.

### Download Statistics
- **Top 25 Downloads**
  - View a list of the top twenty-five objects downloaded.

- **Full Summary**
  - View a list of all the objects downloaded.
### RUcore Open Source Software

<table>
<thead>
<tr>
<th>AREA</th>
<th>Software/Tool/Technology</th>
</tr>
</thead>
<tbody>
<tr>
<td>Portal &amp; user interface</td>
<td>PHP/MySql (by RUL)</td>
</tr>
<tr>
<td>Journal &amp; Dissertations</td>
<td>Open Journal System (PKP)</td>
</tr>
<tr>
<td>Search Engine (full text)</td>
<td>Amberfish (Etymon Systems)</td>
</tr>
<tr>
<td>Workflow Management</td>
<td>PHP/MySql (by RUL)</td>
</tr>
<tr>
<td>Management Services</td>
<td>PHP/MySql (by RUL)</td>
</tr>
<tr>
<td>Handles/persistent IDs</td>
<td>CNRI Handle server</td>
</tr>
<tr>
<td>Digital library framework</td>
<td>Fedora</td>
</tr>
<tr>
<td>OS/Webserver</td>
<td>Unix-Linux/Apache</td>
</tr>
</tbody>
</table>
Continuing Directions

- Institutional Repository
  - Develop a trusted repository
  - Provide faculty repository services
  - Establish new roles within the Library: digital preservation, creation-liaison services, digital publishing

- Digital Preservation
  - Develop preservation services as part of Fedora
  - Seek certification as a trusted repository

- Sustainability and modularity
  - Leverage Open Source software
  - Continue R&D in special areas
  - Share our work with others
URLs

Rutgers University Libraries
- Rutgers University Community Repository at http://rucore.libraries.rutgers.edu
- The New Jersey Digital Highway (funded by IMLS) at http://www.njdigitalhighway.org
- RUL published e-journals
  - “Pragmatic Case Studies in Psychotherapy” at http://pcsp.libraries.rutgers.edu
  - “Electronic Journal of Boundary Elements” at http://ejbe.libraries.rutgers.edu
  - “Journal of Rutgers University Libraries” at http://jrul.libraries.rutgers.edu

Fedora
- Fedora - developed by the University of Virginia and Cornell (grant funded by Mellon) at http://www.fedora.info
- The 2nd Annual Fedora Users’ Conference (June 19-20) at http://www.lib.virginia.edu/digital/fedoraconf/