CINET Registry

Mentor
Dr. Keith Bisset (NDSSL, VBI)

Aditya Agashe,
Harshal Ganpatrao Hayatnagarkar and
Sarang Joshi

6 May 2014
Presentation #8
CS 6604 - Digital Libraries
Virginia Tech, Blacksburg, VA, USA
Project Overview

• Hypothesis: To model CINET catalog in a Digital Object based repository.

• Migration Project: RDBMS Catalog to Digital Object Repository based catalog.
Objectives

- Modelling
- Migration
- Services and Visualizations
- Faceted Browsing
- CRUD Operations

Workflow
Technology stack

- Ruby on Rails
  - Your localized Web Applications
    - Models
  - Hydra-Head
    - Primary Gems
      - Hydra Model Logic
        - Hydra Access Controls
      - Supporting Gems
        - OM
        - ActiveFedora
        - Rubydora
    - Solrizor (Solrizor-Fedora)
  - Blacklight
    - Blacklight discovering & viewing objects (R)
      - Search/Facet Logic
  - Fedora
    - Fedora REST API
  - Solr
    - Solr REST API
Modelling

- Project Hydra transparently interfaces Fedora Commons repository.

- Metadata attributes + object relationships

- Domain-specific metadata
  - Use of taxonomy
  - Use of keywords (WIP)
Modelling: Digital Object Schema
Migration

• Ingesting data from RDBMS to Fedora Commons repository

• Export data from DB

• Data ingest is a two-pass process
  – Pass 1: Creating digital objects and populating attributes
  – Pass 2: Setting relationships between newly created digital objects by maintaining integrity constraints
SCRUD Interfaces

• **SCRUD**: Search, Create, Read, Update and Delete.

• Rails Scaffolding: For CRUD operations
  – Rails facility for generating RESTful web-services and skeleton HTML UI.

• Project Blacklight: For search and browse operations
  – Faceted browsing
  – Text search
  – Based on Solr indexing
Services and Visualization

**Incentivization**
- Highlight user contributions
- Networks and algorithms contributed. Analyses performed.
- Used coordinated visualization.

**Utilization**
- Network datasets
- Algorithms
- Computation resources

**Memoization**
- To avoid repetition of analyses
- Uniquely identify an analysis.
- Return results of an existing analysis when network, measure and parameters match.
- Saves time of execution.
Lessons learned

• Modelling
  – Domain-specific metadata handling through taxonomy and keywords.
  – Constrain handling: ‘has one’ relationship handling through ‘has many’ relationship.
  – Transparent handling of Solr indexes.
  – Federated queries on Fedora Commons are not flexible through Hydra.

• CRUD scaffolds
  – Object associations such as composition are not handled out-of-box.

• Hierarchical faceted browsing for complex models is not easy.
Future work

• Performance improvement for database
• Instance-specific metadata
• Pluggable metadata
• Disseminators for graph format conversion
• Authentication and authorization
• Integration with CINET
Thank you...