Fedora Relationships and Information Network Overlays

CS 431 - April 19, 2006
Carl Lagoze - Cornell University
Fedora Resource Index:
Using RDF and ontologies
Fedora 2.0 and RDF

- **Object-to-object and object-to-literal Relationships**
  - Ontology of common relationships (RDF schema)
  - Relationships stored in special datastream (RELS-EXT)

- **Resource Index (RI)**
  - RDF-based index of repository (Kowari triple-store)
  - Graph-based index includes:
    - Object properties and Dublin Core
    - Object Relationships
    - Object Disseminations

- **RI Search**
  - Powerful querying of graph of inter-related objects
  - REST-based query interface (using RDQL or ITQL)
  - Results in different formats (triples, tuples, sparql)
Uses of Object Relationships

• Define collections (e.g., collection objects)
• Assert critical relationships among object for management purposes
• Enable network overlay
  - Surrogate objects referring to external entities
  - Assert relationships among them
  - Assert other relationships (e.g., annotations)
• Enable navigation of repository (as tree or graph)
Fedora Relationship Ontology (RDFS)

- isPartOf / hasPart
- isMemberOf / hasMember
- isDescriptionOf / hasDescription
- hasEquivalent
- ... others
Demo:
Collection - Member Relationships

- **Collection Object** [smiley]
  - Datastream containing a query to Resource Index for all members of collection

- **Image Objects** [brush]
  - Use RELS-EXT datastream to assert relationship to collection object
Fedora Repository Service
Fedora Repository Service

- **Client App**
- **Web Browser**
- **Batch Program**
- **Server App**

**Web Service Exposure Layer**

**Management Subsystem**
- Object Mgmt
- Object Validation
- PID Generation

**Security Subsystem**
- Policy Mgmt
- Policy Enforcement

**Access Subsystem**
- Search
- Object Reflection
- Dissemination

**User Authentication**

**Storage Subsystem**

**Datastreams**
- External Content Retriever
- Content

**Digital Objects**
- RDBMS
- XML
- Triples

**External Content Source**

**Remote Service**

**Local Service**
Fedora Web Service APIs in a Nutshell

- **Management Service (API-M)**
  - Ingest Object
  - Export Object
  - Get Object XML
  - Purge Object
  - Modify Object
  - Get Next PID
  - Get Datastream(s)
  - Get DatastreamHistory
  - **Get DisseminatorHistory**
  - Get Disseminator(s)
  - Add/modify/purge Datastream
  - Add/modify/purge Disseminator
  - Set State
Fedora Web Service APIs in a Nutshell

- **Access Service (API-A and API-A-LITE)**
  - Describe Repository
  - Get Object Profile
  - Get Object History
  - Get Datastream
  - Get Dissemination

- Find Objects
- Resume Find Objects
Fedora Software Distribution

- **Open Source (Mozilla Public License)**
- **100% Java (Sun Java J2SDK1.4)**
- **Supporting Technologies**
  - Apache Tomcat and Apache Axis (SOAP)
  - Xerces for XML parsing and validation
  - Saxon for XSLT transformation
  - Schematron for validation
  - MySQL and Mckoi relational database
  - Oracle 9i support
  - Kowari for triple-store
- **Deployment Platforms**
  - Windows 2000, NT, XP
  - Solaris
  - Linux
  - Mac OSX
What is a digital library anymore, anyway
Information Network Overlays
Executive Summary

• Move beyond the one way information flow
  - catalog->index->search->access

• Enhance primary data with knowledge activity
  - “People as part of the information infrastructure”
    (Dolores Iorizzo)
Digital Libraries - Ingest Focus
Digital Libraries - Federation Phase

Z39.50
Dienst
SDLIP
OAI-PMH
SRW/SRU
We have been very successful!
So, are we done?

The primary goal of digital libraries has been often been misconstrued as providing accessibility to a massive volume of resources (e.g., Google, institutional repositories). The real opportunity is to:

• Make the library more inclusive of different types of information like data, products, computational services

• Establish the library as a knowledge environment where people organize around information, contribute new information, and collaboratively create new knowledge (i.e., the “wisdom of crowds”).
Information Flow in Traditional Library

In-Band

Out-of-Band

Knowledge
Information Flow in the Digital Library

In-Band

Out-of-Band

Knowledge

Dublin Core Used Here
Creating a Collaborative Knowledge Network
The World Is Flat: A Brief History of the Twenty-first Century (Hardcover)
by Thomas L. Friedman. This book offers a comprehensive look at the world as it has evolved since the Cold War.

List Price: $25.00
Price: $18.00 and eligible for FREE Super Saver Shipping on orders over $25. See details
You Save: $7.00 (28%)

Availability: Usually ships within 24 hours. Ships from and sold by Amazon.com.

Want it delivered on Tuesday, September 27 (order cut off at 11:59 PM and 25 minutes, and classic two-day shipping at checkout; see details)

Best Value
Buy The World Is Flat: A Brief History of the Twenty-first Century, and get History: Fiction or Science? at an additional 35% off Amazon.com's everyday low price.

Buy Together: $49.20
You Save: $13.25

Editors' Reviews
Amazon.com
Thomas L. Friedman is not so much a futurist, which he is sometimes called, as a presentist, a man who, in his new book, The World Is Flat, as in his earlier, influential works and the Olive Tree, is not to give you a speculative preview of the wonders that are sure to come in your lifetime, but rather to get you caught up on the wonders that are already here.

The world isn't going to be flat, it's flat, which gives Friedman's breathtaking narrative much of its urgency, and which also saves it from the Esquire-style polemic that often plagues such books. What Friedman means by "flat" is "connected": the lowering of trade and political barriers and the exponential technical advances of the digital revolution have made it possible to do business, or almost anything else, instantaneously with billions of other people across the planet. This in itself should not be news to anyone. But the news is that Friedman has to deliver is that just when we stopped paying attention to those developments--when the dot-com boom turned all eyes toward the Internet--when they actually began to accelerate. Globalization II, as he calls it, is driven not by major corporations or giant trade organizations like the World Bank, but by individuals: desktop entrepreneurs and innovative startups all over the world (but especially in India and China) who can compete--and win--not just for low-wage manufacturing and information jobs but, increasingly, for the highest-end research and design work as well. (She doesn't forget the "nifty supply chains" Bob Zoellick that take the small act of big in more destructive ways.) Friedman tells his eye-opening story with the catchy slogans and globe-hopping anecdotes that readers of his earlier books and his New York Times columns will know well, and also with a stern sort of optimism, like he wants to tell you how exciting this new world is, but he also wants you to know you're going to be tamed if you don't snap up with it. His book is an excellent place to begin. --Tom Vaneck

Customers who bought this item also bought:
The New World: A Blueprint for the Twenty-first Century
by Thomas L. Friedman

Explore Similar Books: Books, Music, and DVD

Add to Wish List
Don't have one? We'll set you up for free.

Add to Cart

Dealing with the “Item Problem”

Different kinds of Stuff

...that may take more than one form (polymorphic)

...and may be produced dynamically

... each of which may be related to lots of other stuff.
A bit of NSDL background

• Mission: “Improve Science, Math, Engineering education through digital libraries”
• Original NSDL solicitation in 1999
• Over 180 projects funded
• Core integration (Columbia, Cornell, UCAR) charged with providing organizational, technical infrastructure
• CI (Cornell) funding through 2006
• http://www.nsdl.org
The metadata repository is a resource for service providers.

It holds information about every collection and item known to the NSDL.
Characteristics of the Metadata Repository

- Oracle database
- Qualified Dublin Core
- Item records with collection association
- OAI-PMH ingest and exposure
- Current collection ~ 1,000,000
- Metadata quality issues
Broader Problems in this approach

• Access alone does not equate to educational value
  - Reeves *Impact of Media and Technology in Schools*
• Static metadata records don't capture changing and multiple contexts of use and applicability
  - Recker and Wiley *Designing Instruction with Learning Objects*
• Patterns of use, informal opinions, descriptions often more useful than taxonomic classification.
  - Collis and Strijker *Technology and Human Issues in Reusing Learning*
Information in Context

...how has it been aggregated

...what has it been used with

...how is it classified

...how is it described & rated

...how does it relate to standards

...how was it used

...who used it

STEM Resource
Characteristics of the Network Overlay

• Integrate local and distributed entities
• Entities are polymorphic (operational semantics)
• Ontology-based relationships
• Web service integration
Digital Libraries:
Beyond Search and Access

• Build on foundation of near universal access
  - “Google *”
• Provide context for:
  - **Content aggregation**: combining information entities in novel ways
  - **Knowledge integration**: capturing semantic relationships between information entities
  - **Information reuse**: allowing secondary, tertiary products
  - **Information transformation**: combining information entities with computational services
  - **collaboration and contribution**: blurring the line between authors, publishers, users, experts...
Translate to Technical Requirements

• **Rich information objects**
  - Integration of local and remote mixed-genre content

• **Dynamic information objects**
  - Integration with local and distributed services

• **Graph-based information model**
  - Nodes are information objects
  - Edges are relationships among those objects

• **Access and management API**
  - Exposing full functionality for programmatic access

• **Fine granularity access management**
NSDL Data Repository (NDR)

• Fedora-based implementation of information network overlay
• Content model to represent NSDL information entities and relationships
• Extensive use of resource index and new oai service
Fedora NDR

- **Objects:** agents, metadata items, resources, services (metadata providers), aggregations
- **Relationships:** metadataFor, providedBy, memberOf, representedBy + ontology-specific
- **Disseminations:** metadata transformations
- **OAI harvesting:** both static and generated metadata formats
- **Authentication/Authorization:** Collections and services manage their own repository content, contribution of annotations, new content
Types of Objects
- Agents
- Aggregators
- Metadata Providers
- Resources

Types of Relationships
metadataProviderFor (mdp4)
aggregatorFor (agg4)
providedBy (pBy)
metadataFor (m4)
memberOf (mOf)

- 1st. A recommended resource
- 2nd. Makes it a "blessed" NSDL Collection

NSDL FEDORA-BASED REPOSITORY

NSDL BigBang

NSDL Agent 1000

NSDL Collections 1002

Aggr 2002

Aggr 2005

MDP 3000

MDP Recommended 1005

MDP 3004

NSDL RS Agent 1004

Example 10006

Example Collection

Example Agent 1001C

Aggr 10012

M 10005

M 10007
Metadata in the NDR

• **Multiple formats**
  - static (ingested from provider)
  - generated/crosswalked

• **Multi-sourced**
  - de-dupped
  - Retain branding of metadata

• **OAI-PMH harvesting**
Resources, Metadata, Metadata Providers
Metadata Content Model

Harvest/Ingest

Representation

Dissemination

Exposure

native_dc

native_marc

RELSEXT

DC

format_native_dc

format_native_marc

format_nsdl_dc

format_native_dc_inf

format_native_marc_inf

nsdl metadata bdef

mf values:
native_dc
native_marc
nsdl_dc, oai_dc
nsdl_all

getMetadata(mf)

getMetadataAbout(mf)

getRecord(mf)
listRecords(mf)
listIdentifiers(mf)

proai
Collections and Aggregations

- Set basis
- Semantic basis
- Agent associated
Aggregation Model
Annotation/Reviews

- Unstructured metadata about a resource
- Exists as resource and annotation
- Separate agent provenance from annotated resource
Annotation Model