Fedora: A network overlay approach to federated searching

Leah Solla
Cornell University

ACS 228th National Meeting
Washington, DC
August 17, 2009
FEDORA: Flexible Extensible Digital Object Repository Architecture

• Originally built for storing, managing, accessing digital content

• Data model
  – Defines set of abstractions for expressing digital objects
  – Asserting relationships among digital objects
  – Linking “behaviors” to digital objects (actions)

• Data Objects as Aggregations
  – Multiple datastreams
  – Multiple descriptions
  – Multiple sources

  “Data is like onions, lotsa layers”
Fedora Object Model

• **Representational perspective**
  – Digital objects as rooted sub-graph
    • Multiple representations as nodes (i.e., multiple formats)
  – Relationships – links between objects

• **Functional perspective**
  – Digital objects (unique information entities)
  – Datastreams (information components of objects)
  – Disseminators (operations that establish links)
  – Relationship data (relationships b/w objects as RDF fragments)
Fedora Object Model

Figure 1 - Fedora representational view
NSDL: A network overlay approach to federated searching

Overlay Requirements:

• Resource-centric
• Representation of other information (not one-to-one mapping)
  – People, organizations, services using & contributing
  – Education standards
  – Applications of resources
• Local & distributed information
• Diversity of relationships among above
Fedora Repository as Network Overlay

- **Multiple Object Types:**
  - Resources (with local or remote content)
  - Metadata
  - Aggregations (collections)
  - Metadata Providers (branding)
  - Agents
- **Relationships with arbitrary graph queries:**
  - Structural (part of)
  - Annotation (relates to)
- **Maintains Fedora’s independence of datastreams - allows other metadata and content to coexist with repository**
Fedora Network Overlay Content Model

Figure 4 - Semantic aggregations

http://arxiv.org/abs/cs/0501080v2
The NSDL Data Repository (NDR), implemented as a set of digital objects and relationships in a Fedora repository.
Key aspects of this overlay (NCore)

- Vision is to represent contextual knowledge around web resources
- Serves as a forum for independent parties to contribute, discover, use, and re-use this context at will
- Allows libraries to construct a cohesive and vetted view of the contents therein
- Allows all these independent parties to go about their business and not step on each other's toes
- NOTE: The subset of resources on the web can be an existing repository!
NCore: The Technical Ecosystem

Protocol:
- OAI-PMH
- HTTP
- REST
- NDR API
How do we ensure Cornell’s information resources are discoverable and accessible?

- Greenstone
- LibGuides
- DLXS
- Luna Insight
- Fedora
- DSpace
- Filemaker
- Web sites
- ARTstor
- Voyager

- arXiv
- DPubS
- VIVO
- Google Books
- 500 licensed databases of journals, images, books, reference works, datasets…
Discovery/Access Overlay View

- Cornell User View
- API/UI
- Discovery layer with Relations & Annotations
- Resources from Multiple Providers

- Library Catalog
- Licensed DBs
- HathiTrust
More information on Fedora & NCore:


Contact & Thanks

Dean B. Krafft
Cornell University Library
201 Olin Library
Ithaca, NY 14853
USA
dean.krafft@cornell.edu