



THE NATIONAL SCIENCE DIGITAL LIBRARY

NSDL CI Tech Overview for BEN Collaborative



Dean Krafft, Cornell University
dean@cs.cornell.edu



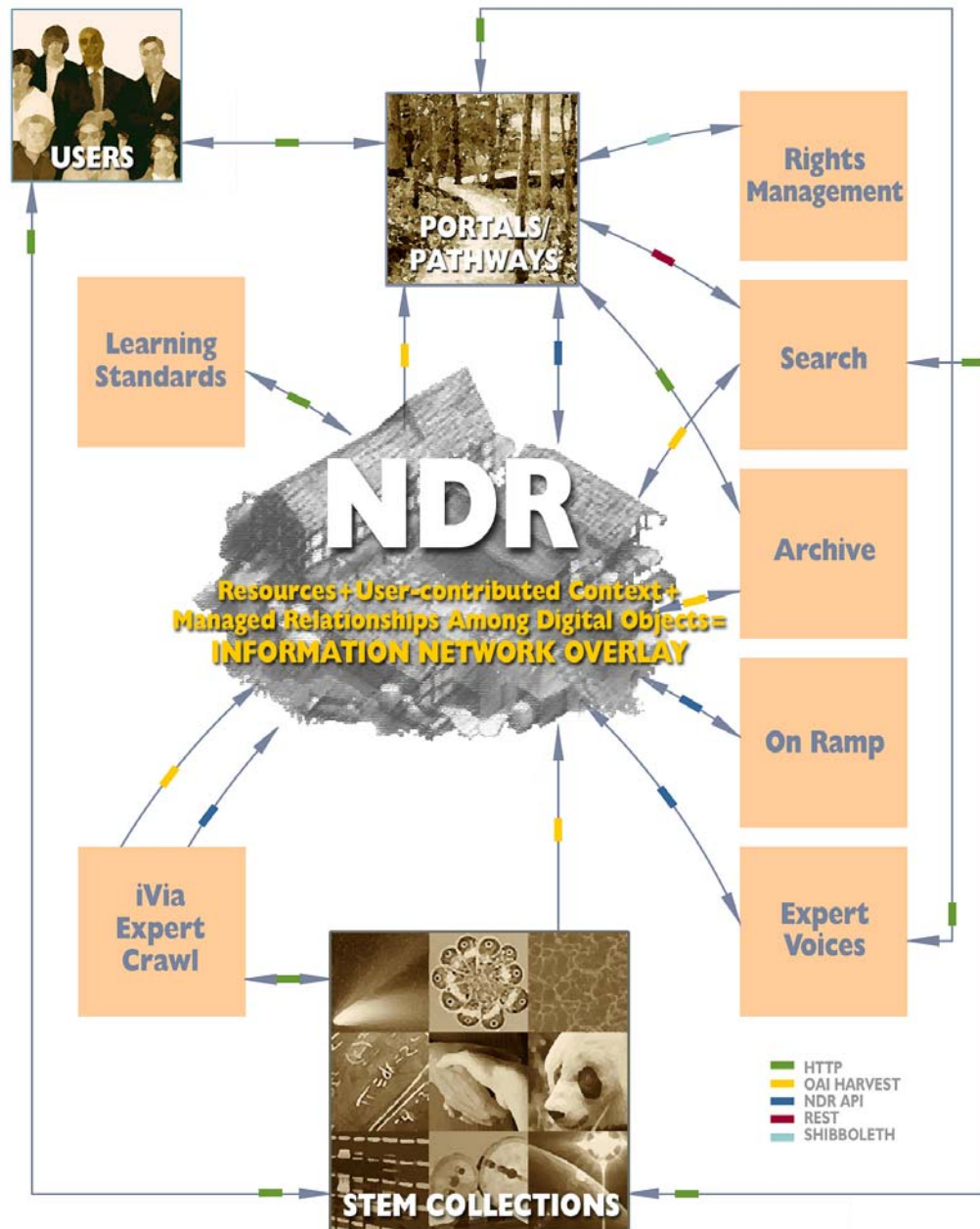
The Technical Vision

- Why not just use a search engine?
- Guide selection and use, not just resource discovery
- Support creating “context” for resources
- Present resources in context: in a lesson plan; with ratings; correlated with standards
- Tools to enable the community: structuring, evaluation, tagging, contribution, collaboration
- Create a dynamic, living library

Metadata Repository to Data Repository

- Currently NSDL is based on Metadata Repository created with OAI harvests
- Limited model
 - Metadata-centric orientation
 - No content – only metadata
 - Limited relationships – collection/item
 - Limits on context, structure, and access
- Solution: the NSDL Data Repository (NDR)

NSDL Data Repository Architecture (NDR) Nov. 2005

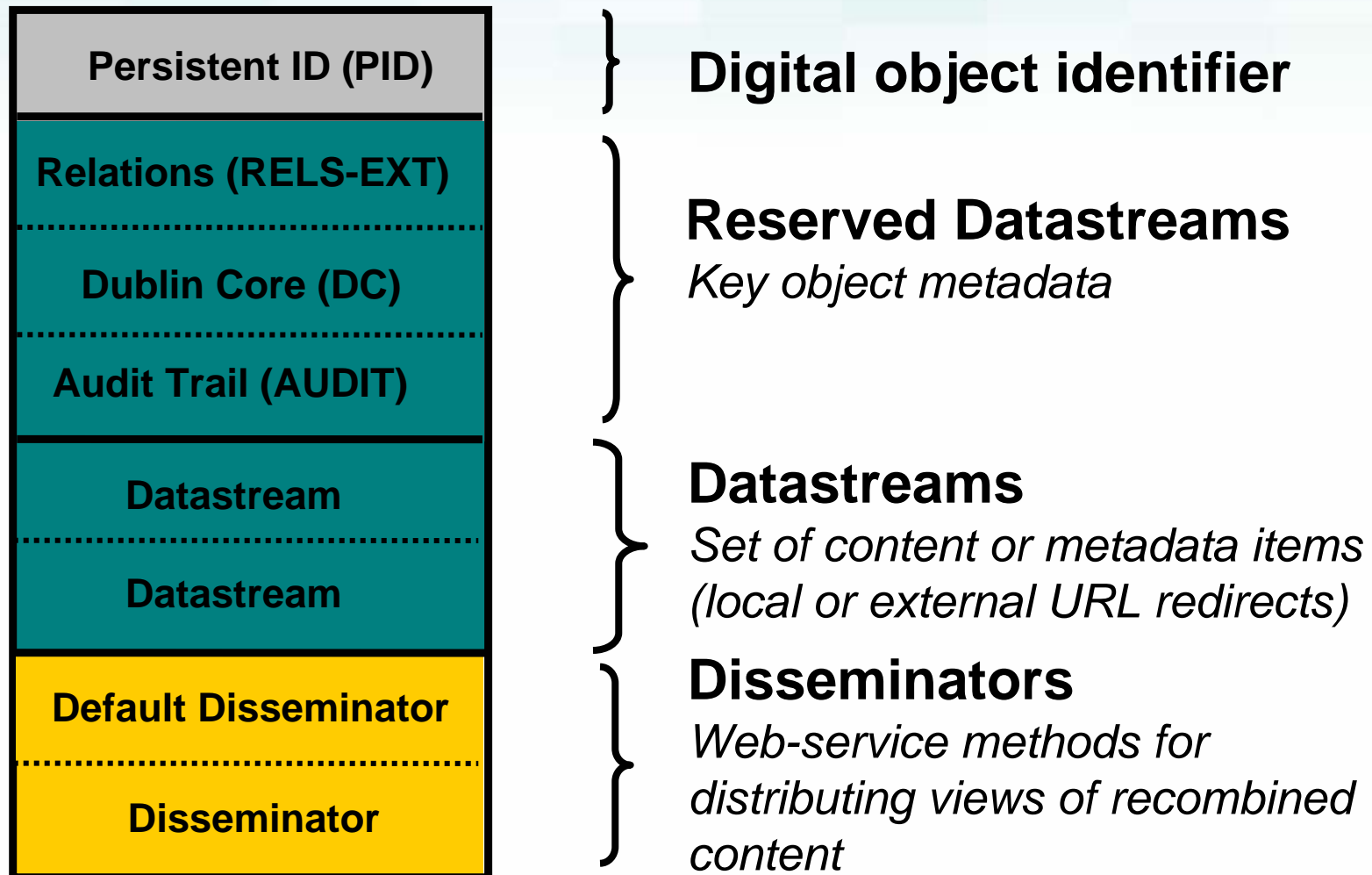


Implementing the NDR: Fedora

- A **F**lexible, **E**xtensible **D**igital **O**bject **R**epository **A**rchitecture
- An architecture and toolkit (like IIS or SQL Server), middleware, not a vertical application
- DSpace in contrast: a vertical application with a fixed workflow targeted at users
- Arbitrary internal and external digital objects, disseminations (transformations and combinations), relationships among objects

Fedora Digital Object Model

Component View



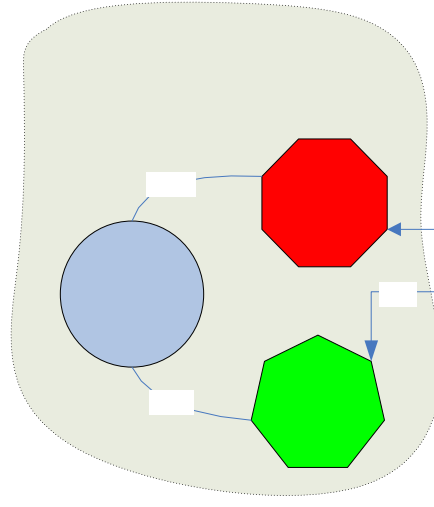
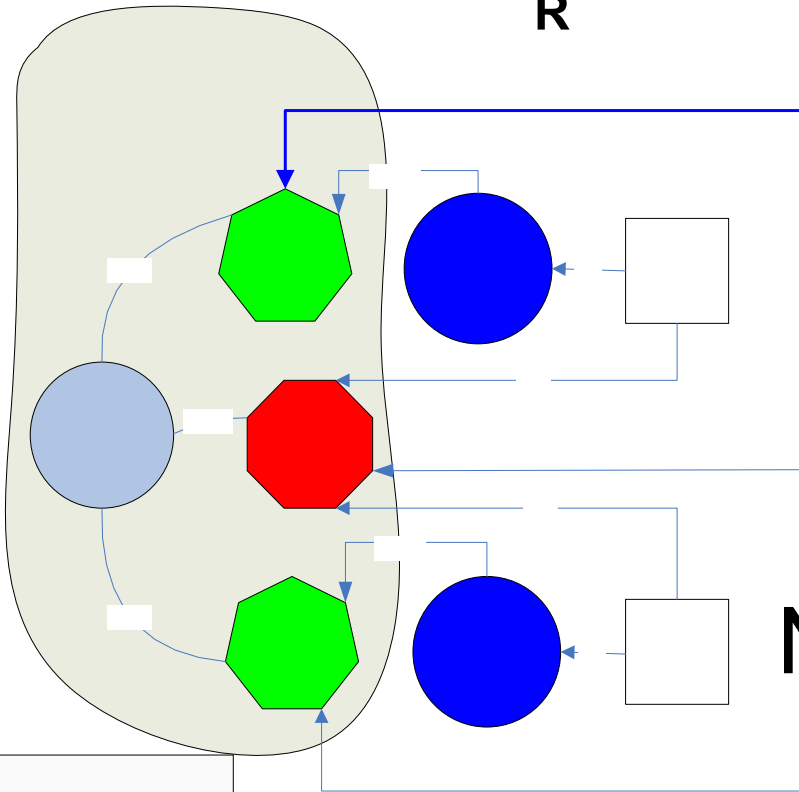
Fedora Repository Architecture

- Set of SOAP/REST services: Manage, Access, Search, Query
- Fundamental store is XML, with RDBMS cache (Oracle, MySQL), and RDF triple store for relationship queries
- Modular architecture: Manage, Access, Storage, Dissemination, Authentication, Authorization, RDF Resource Index

Achieving the Vision with Fedora

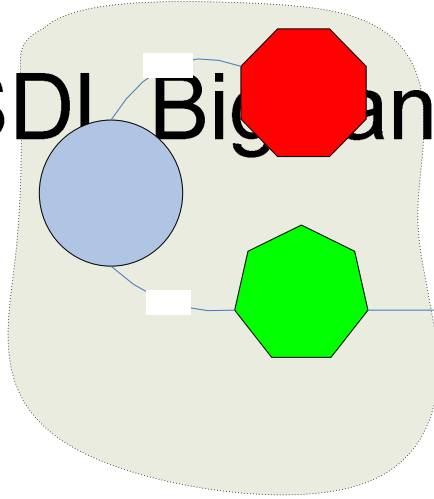
- Network overlay architecture: A lens for viewing science content on the net, whether content is local, remote, or archived – it all has a repository-based URL
- Objects: Aggregators (collections), Metadata Providers (branding), Agents, Resources (with local or remote content), Metadata
- Relationships: Structural (part of), Equivalence, Membership, with arbitrary graph queries
- Web services: disseminations are arbitrary recombinations of content
- Authentication/Authorization: Collections and services manage their own repository content

NSDL F
R

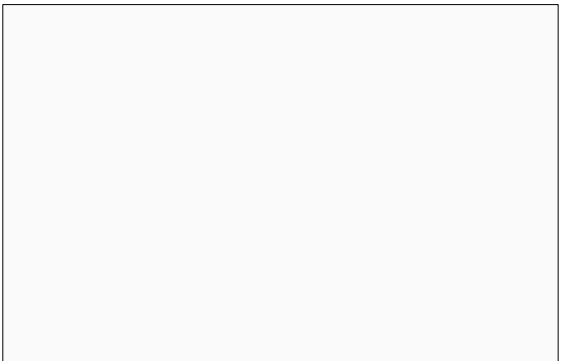
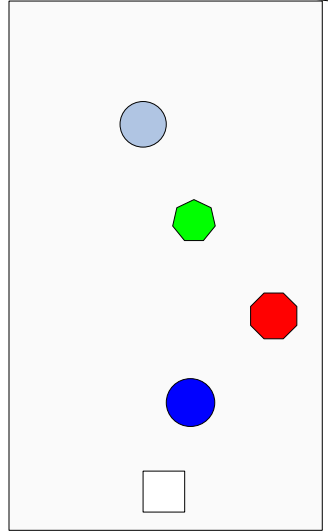


BAS

NSDL Bigang



repBy



g4

Aggr
2002

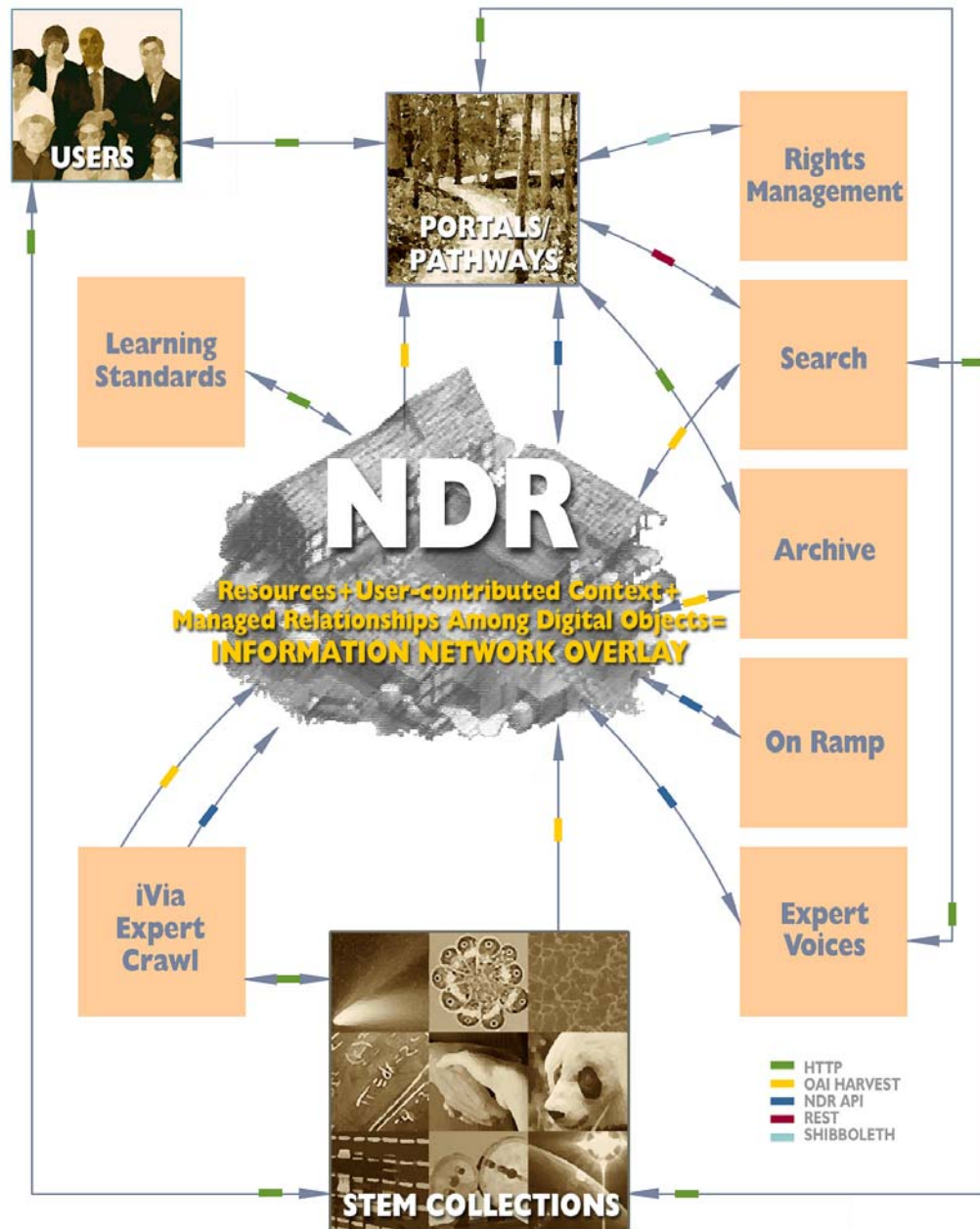
NSDL Data Repository (NDR)

- Repository in production load with over 875,000 metadata records and over 2 million digital objects
- Over 163 million RDF triples causing
- Scaling challenges: moved to 64-bit architecture with 32GB memory; carefully structuring RDF queries
- Estimating fully operational beta version of new NDR in February

Upcoming Challenges

- Continued scalability: can probably go about a factor of five over current size with current technology. Fedora team working on Oracle-based triplestore
- We can create and store context and relationships for resources. How can we take advantage of this for the end user?

NSDL Data Repository Architecture (NDR) Nov. 2005



NDR Application: ExpertVoices

- Blogging system fully integrated with NDR resources, supporting:
 - Topic-based discussion (e.g. tsunamis) with pointers to related resources
 - WordPress plug-in implementing NDR API
 - Integration with NSDL single sign-on
 - Resource recommendations, ratings, reviews, annotations, and simple metadata
 - Discussion forum /QA with StoryStarters
- Alpha version now, full beta 1Q06

Expert Voices Blogging

The screenshot shows the NSDL Expert Voices Blogging website. At the top left is the logo for 'EV NSDL' with the tagline 'Expert Voices: Speaking of Something Interesting'. To the right is a search bar and navigation links for 'NSDL', 'Blogs', and 'more options'. Below the header, the breadcrumb 'NSDL.org > Expert Voices' and a 'Larger Text' link are visible. The main content area is titled 'Who Says' and is divided into four categories: 'K-12 TEACHERS', 'UNIVERSITY FACULTY', 'LIBRARIANS', and 'NSDL LIBRARY BUILDERS'. Each category contains a post titled 'NSDL Expert Voices Intro' with a date of 'Nov. 2nd, 2005 9:35 pm' and author 'admin'. The posts include a welcome message and a 'Read more >' link. To the right of the main content is a sidebar with a red link 'Return to the Annual Meeting Blog', a notice 'Coming to a browser near you in Jan'06', a 'Categories' list (Current News, Education, Health, Mathematics, Science, Social Studies, Technology), and a 'Top Resources' section with a table header 'Resource Title' and 'Resource Description'. The footer contains 'THE NATIONAL SCIENCE DIGITAL LIBRARY' and the NSDL logo.

EV NSDL
Expert Voices: Speaking of Something Interesting

SEARCH

NSDL Blogs more options

NSDL.org > Expert Voices Larger Text

Who Says

K-12 TEACHERS

NSDL Expert Voices Intro
Nov. 2nd, 2005 9:35 pm
Written by: **admin**
Welcome to the NSDL Expert Voices blogging project!
[Read more >](#)

UNIVERSITY FACULTY

NSDL Expert Voices Intro
Nov. 2nd, 2005 9:35 pm
Written by: **admin**
Welcome to the NSDL Expert Voices blogging project!
[Read more >](#)

LIBRARIANS

NSDL Expert Voices Intro
Nov. 2nd, 2005 9:35 pm
Written by: **admin**
Welcome to the NSDL Expert Voices blogging project!
[Read more >](#)

NSDL LIBRARY BUILDERS

NSDL Expert Voices Intro
Nov. 2nd, 2005 9:35 pm
Written by: **admin**
Welcome to the NSDL Expert Voices blogging project!
[Read more >](#)

Return to the Annual Meeting Blog

Coming to a browser near you in Jan'06

Categories

- Current News
- Education
- Health
- Mathematics
- Science
- Social Studies
- Technology

Top Resources

Resource Title	Resource Description
----------------	----------------------

THE NATIONAL SCIENCE DIGITAL LIBRARY **NSDL**

NDR Application: OnRamp

- NDR-integrated multi-user, multi-project content management system
- Supports NSDL single sign-on and group management
- Decentralized workflow for the creation and distribution of both simple and complex content
- Disseminates content in multiple publication and online forms
- Delivery estimated early 2Q06

OnRamp: Content Management

ONR NSDL
On Ramp: Creating and Sharing Great Stories

SEARCH OPTIONS LOGIN

WORK IN PROGRESS **RELEASED WORK** ADMIN

Targets

- nsdl-homepage-highlights-rss
- nsdl-2005-annual-meeting-blog
- white-board-report-rss
- white-board-extended-report-rss
- outreach-fliers
- outreach-materials

Pending Packages

- How "Content Alignment" Tools CAT from the Center for Natural Language Processing (CNLP) of Syracuse University School of Information Studies Help Teachers Utilize Digital Library Resources in Classrooms** (created: November 9, 2005 3:37:59 PM EST) (modified: November 9, 2005 3:37:59 PM EST)
[Properties](#) [View](#) [Activate](#) [Comments](#) [History](#)
- How the Macaulay Library Collection—largest collection of animal sounds in the world with more than 160,000 recordings, 67 percent of the world's birds, with rapidly increasing holdings for insects, fish, frogs, and mammals—Enables Outreach and Collaboration with K12 Students and Teachers** (created: November 9, 2005 3:37:59 PM EST) (modified: November 9, 2005 3:37:59 PM EST)
[Properties](#) [View](#) [Activate](#) [Comments](#) [History](#)
- How User Information Such as the Pilot ARL DigiQUAL NSDL.org User Survey Enable Outreach and Collaboration with Audience Groups** (created: November 9, 2005 3:37:59 PM EST) (modified: November 9, 2005 3:37:59 PM EST)
[Properties](#) [View](#) [Activate](#) [Comments](#) [History](#)
- Find Over 26,000 Species of Birds, Mammals, Reptiles and Amphibians With WWF "Wildfinder"** (created: November 9, 2005 3:37:59 PM EST) (modified: November 9, 2005 3:37:59 PM EST)
[Properties](#) [View](#) [Activate](#) [Comments](#) [History](#)

Active Packages

- Content Assignment Tool (CAT) to be Released at NSDL's Annual Meeting** (created: November 9, 2005 3:37:59 PM EST) (modified: November 9, 2005 3:37:59 PM EST)

Other NDR-integrated applications

- Resource-centric search
 - alpha REST service now
 - Integrated end-user beta mid-February
- iVia-based Expert-Guided crawl
 - Already in use for some production collections
 - Beta available for Pathways experts use/testing now
- Educational Standards integration – initial version projected 2Q06

Summary

- OAI-PMH harvesting now (production)
- NSDL Single Sign-On now (production)
- iVia crawls possible now (beta)
- ExpertVoices integrated blogging 1Q06
- NDR API interactions available 2Q06
- OnRamp workflow/CMS available 2Q06

Questions for Pathways on NDR

- What will you consume?
- What will you produce?
- How should you communicate?
- What should we add to the NDR?
 - New relationships?
 - New objects?
 - Special indexes?

Initial Internal NDR API (samples)

- addAgent: creates an agent object
- addResource: requires info:URI
- addAggregator: requires resourcePID, agentPID, and service description
- addMetadata: requires resourcePID and metadata XML, returns metadataPID
- findResource: takes info:URI and returns resourcePID
- getHandle: takes resourcePID, returns handle
- listMembers: takes, e.g., aggregatorPID and returns list of resource PIDS with aggregatedBy relationship

NDR Initial API Characteristics

- Uses REST calls for all interactions
- Specializes Fedora for NDR objects/relationships
- Objects and relationships are directly manipulated
- No authentication or authorization: API “owns” every object in repository
- API does not prevent creating semantically inconsistent NDR; applications must guarantee consistency

Fedora Key Features

- **Content aggregation**
 - Digital object model to combine information entities in novel ways
- **Knowledge integration**
 - Ontology-based relationships among objects
- **Information reuse**
 - Create secondary, tertiary objects
- **Information transformation**
 - Combine objects with computational services
- **Collaboration and contribution**
 - Enable annotation, info sharing, workflow, contextualization
- **Information management and preservation**
 - XML-based object storage
 - Service-oriented architecture; web services
 - Store relationships and service linkages with objects

Fedora 2.1 Capabilities

- **Object-to-object 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 and Object Disseminations**
 - Powerful querying of graph of inter-related objects
 - REST-based query interface (using RDQL or ITQL)
 - Results in different formats (triples, tuples, sparql)
- **Authentication**
 - Plug-in Authentication modules
 - Fine-grained Authorization using XACML XML-based policies