Planning for Flexible Integration via Service-Oriented Architecture (SOA)

APSR Forum – The Well-Integrated Repository
Sydney, Australia
February 2006

Sandy Payette
Co-Director, Fedora Project
Researcher, Cornell Information Science
Flexible integration and soa

- What are the motivating contexts?
  - Signs of change: e-research, e-science, advanced scholarly pub
  - Anticipation of next moves (e.g., semantic web; web 2.0)
- How do we position for the future?
  - Goals
    - Functional requirements placed upon repositories
- What do we mean by flexible integration?
  - SOA and Service frameworks
  - Understanding why repositories play a central role in service frameworks
- Example SOA work at Cornell
First... motivating contexts
Disconnected networks:
- formal publication network
- social network (actors)

Hybrid network
- documents (formal and informal)
- data
- services
- actors

The Future:
Rich Scholarly Information Networks
• E-research, e-scholarship
• Scholarly workbench
• Advanced scholarly communication/publication
Already signs of change...

- **Grid computing in sciences**
  - Share computing resources
  - Share services and distributed virtual file systems
  - Examples
    - Enabling Grids for E-Science (http://public.eu-egee.org/)
    - National Virtual Observatory (http://www.us-vo.org/)

- **Humanities computing**
  - Hyperlinked historical documentary editions
  - New Forms of Digital Scholarship
    - Rossetti archive (http://www.rosettiarchive.org/)
    - Perseus (www.perseus.tufts)
    - Pompeii Forum (http://pompeii.virginia.edu)
  - Tibetan and Himalayan Digital Library (thdl.org)
What’s out ahead?

- Semantic Web
- Web 2.0
- Expectations of new generation of scholars
Web 2.0 concepts

Implications of web 2.0

• Key themes
  - Services (not packaged apps)
  - Architecture of participation
  - Remix/transform data sources
  - Harness collective intelligence

• Emergent Behavior
  - Upcoming generations of scholars will have a completely different paradigm and expectations regarding technology
  - Collaborative classification (e.g., flickr)
  - Power of collective intelligence (amazon)
  - Alternative trust models (reputation - ebay; open-source)

• Plan now for technologies to be service-oriented, flexible, decomposed
Positioning repositories for the future...
Well-integrated repository should...

1. Support the creation and publication of new forms of "information units"

2. Integrate with the processes (e.g., workflows) of research, collaboration, and scholarly communication

3. Enable knowledge integration: capture semantic and factual relationships among information entities

4. Promote information re-use and contextualization

5. Facilitate collaborative activity and capture information that is created as a byproduct of it
Support the new "information unit"

- Documents
- Text
- Data
- Simulations
- Images
- Video
- Computations
- Automated Analyses

Aggregations
Enabling technologies:
Selected SOAs (Cornell University)

- Fedora Service Framework
- National Science Digital Library
- NSF Pathways (Cornell/LANL)
Fedora Digital Objects

• **Flexible object model can support**
  - Documents, articles, journals
  - Electronic Scholarly Texts
  - Digital Images
  - Complex multimedia publications
  - Datasets
  - Metadata
  - Learning objects
  - More...

• **Create “networks” of objects**
  - Define object relationships and other properties via RDF
  - Collection/member; part/whole; etc.
Network of Digital Objects in a Fedora Repository
Pathways Project
Challenges - Phase 1

• **Current situation**
  - Heterogeneous repository systems
  - Heterogeneous object models (or no object model)
  - Multiple protocols and service APIs
  - Services lacking formal interface definitions

• **Can these resources ever play nicely together?**

• **Need common abstractions...**
  - Information model
  - Service model

Cornell University and Los Alamos National Lab
(http://www.infosci.cornell.edu/pathways)
Most things can be represented as a graph of nodes and arcs.
Building Block: Repository Integration
(Access-perspective)

Repositories

Fedora Services
- Preservation Monitoring
- Federation PID Resolution
- Event Notification
- Fedora Repository Service
- Preservation Integrity
- Directory Ingest
- Federation PID Resolution

Apps
- Fedora Search
- Fedora Workflow
- Fire Client
- POLICY
- FIRE Client
- PolicyBuilder
- InterDisseminator Service
- OpenURL
- OpenURL

Pathways
- InterDisseminator Service
- OpenURL
- OpenURL
- OpenURL

Interoperable Access
(ontology-based)

Semantic Service Matching
(OWL-S)

Client

Other Services
- JHOVE
- GDFR
- External Workflow

Cornell CIS and LANL Library
Information Model:
“Article” Example Core-1 Ontology
NSDL: Creating a Collaborative Knowledge Network

Repositories

Annotations
Quality Assessments
Relationships
Reviews

Knowledge Network
Conclusions:
Planning for integration
What basic changes can occur now?

- Think in terms of service frameworks
- Expose repositories as services
- Enable easy integration of repository with other services
- Support compound digital objects
  - Local and remote content
  - Any media type
  - Provide a way to associate services with objects (dynamic views)
- Provide ability to assert relationships among objects
- Look toward ontology-based metadata
Ongoing Challenges

- Low barrier to entry
  - Simple protocols (e.g., like OAI)
  - Light-weight (REST vs. SOAP?)
  - Simple tools to create overlays
  - Note complexity in setting up Grid-based services

- Service matching (object-to-service)
  - Ontologies to expose objects with formats and semantics
  - OWL-S for semantic service description
  - Matching-making algorithms

- Security and Trust
  - Authentication and trust among repositories and services
  - Interoperability of authorization policy

- Preservation
  - Distributed and dynamic digital objects a reality