Fedora 4 as a Shared Linked Data Repository for the AIC Collections

Stefano Cossu, Director of Application Services – The Art Institute of Chicago – scossu@artic.edu
Open Repositories Conference 2014, Helsinki, Finland
The AIC Collections

- 250,000 art objects in Collections
- Several millions of image and text assets
- Many different departments access these data
  - 11 Curatorial depts.
  - Registrar
  - Conservation
  - Imaging
  - Publications
  - Help Desk
  - Etc.
Goals

- Build a scalable asset repository
- Create a pool of shared, linked information
- Move toward a de-centralized, asynchronous architecture
- Maintain current CMS as the main management platform
- Standardize data formats
LAKE – Linked Asset and Knowledge Ecosystem

LAKE
(Fedora Cluster)

Objects

Events

Places

Actors

Assets
LAKE Gateways to Shared Data
AIC Departments with Access Roles

Gateways: CItI - Phenix - Drupal - APIs

Web Collections

Curatorial

OSCI + LaunchPad

Registrar

Mobile Apps + Special Projects

Conservation

AV

Imaging

LAKE
(Fedora Cluster)

Objects

Events

Places

Actors

Assets
Non-Shared Data

Gateways: CITI - Phoenix - Drupal - APIs

LAKE (Fedora Cluster)

- Objects
- Events
- Actors
- Places
- Assets

- Website Resources
- Web Collections
- OSCI + LaunchPad
- OSCI + LP Resources
- Mobile + SP Resources
- Mobile Apps + Special Projects
- AV Resources
- AV
- Imaging Resources
- Imaging
- Registrar Resources
- Registrar
- Curatorial Resources
- Curatorial
- Conservation Resources
- Conservation
Why Fedora?

- Scalable and reliable
- Interface agnostic
- Content agnostic
- Modular, distributed
- Community driven
Fedora 4: a hard decision

• We built a F3 proof of concept, then moved to F4
• F4 features are very helpful to our mission
• F3 had the guarantee of stability
• Some features won't be available soon in F4
• Starting with F3 and migrating to F4 later would have been very time-consuming
Fedora 4 Key Features for AIC

- Federation of external sources
- Asynchronous content processing (Sequencers)
- Powerful REST API
- Clustering
- Completely RDF-based
- HTML management interface
Use case proposals

Functionality we plan to build around F4 features:
• Large file ingestion
• Metadata extraction
• Content Modeling and Access Policies
Use Case 1: Ingesting Large Files

- Staff (John Doe)
  - Upload high res and master files to remote folder (NFS/SMB/WebDAV)
  - Upload to remote folder

- Staff (Joan Dow)
  - Upload to remote folder

- Staff (Jane Dough)
  - Upload to remote folder

Imaging server with drop box:
- ~jdoe/batch1234
- ~jdoe/batch2345
- ~jdoe/batch3456
- ~jdoe/batch4567

Images (permanent location)
Use Case 1: Ingesting Large Files

1) Fill in metadata form

2) Generate and send manifest file

3) Ingest master files

4) Move high res files to permanent location and delete temp batch folder

5) High res files are available in DAMS as federated content
Use Case 1: Ingesting Large Files

1) Send processing request
2) Pull master files
3) Create Derivative(s)
4) Send Metadata extraction request
5) Extract metadata
6) Store metadata

1) Send processing request
2) Pull high-res files

Fedora Server

image processor

Metadata extractor

images

imagingserver
Use Case 2: Content Modeling
Use Case 2: Content Modeling
Fedora 3 features that we would like to see redesigned and improved in Fedora 4:

- Enhanced Content Modeling
- Disseminators