Open Technologies for Durable Digital Content

Sandy Payette
Chief Executive Officer

Thornton Staples
Director of Community Strategy and Alliances
Introducing DuraSpace™

an organization creating open technologies for durable digital content.

Fedora Commons and the DSpace Foundation, two of the largest providers of open source software for managing and providing access to digital content, are joining their organizations to pursue a common mission. Jointly, they will provide leadership and innovation in open source technologies for global communities who manage, preserve, and provide access to digital content.

The joined organization, named “DuraSpace,” will sustain and grow its flagship repository platforms - Fedora and DSpace. DuraSpace will also expand its portfolio by offering new technologies and services that respond to the dynamic environment of the Web and to new requirements from existing and future users. DuraSpace will focus on supporting existing communities and will also engage a larger and more diverse group of stakeholders in support of its not-for-profit mission.
DuraSpace Governance

Board of Directors
Technology Portfolio
DuraSpace Solution Communities

- preservation and archiving
- small archives
- data curation
- scholar’s repository
- open access publishing
Fedora
New and Notable Fedora Integrations

- Scholarly and E-Research Platforms
  - Hydra
  - Islandora (Drupal+Fedora) and VRE
- Preservation and Archiving
  - RODA (Portuguese National Archives)
  - Sun Archive Appliance (Fedora/OpenStorage)
- Digital Collections and Media Archives
  - WGBH OpenVault
  - Sloan Kettering
Sun Archive Appliance “Fedora Inside”

Net@Work - NEW Active Archive “Appliance”

Distributed as Solution Packs:
- Institutional Repositories
- Document Management
- Image/Audio/Video Repositories
- Digital Books/Magazines/News
- Virtual Research Environments (VREs) - Bioscience

Note: Islandora = Drupal/Fedora module + additional components
Fedora 3.x Performance

- 14 million objects
- 10 ingests per second
- 750 million triples

Scalability

- 150 million objects
- Ingests – flat response time
- Disseminations – nearly flat
- 5 times faster than prior tests
Fedora Roadmap – Next

✓ Theme: Improved Modularity
  • Modern builds using Maven2
  • Modern plug-in architecture
    • OSGi
  • Promote shared modules
  • Attract new developers to contribute

RoadMap URL:
fedora-commons.org/confluence/display/FCREPO/Fedora+Roadmap
Fedora Roadmap – Next

✓ **Theme: Storage Pluggability**
  - DuraCloud Plug-in
  - Akubra storage abstraction with multiplexing
  - Replication and Migration

✓ **Theme: Security**
  - Fedora Enhanced Security Layer (FESL)

RoadMap URL:
[fedora-commons.org/confluence/display/FCREPO/Fedora+Roadmap](fedora-commons.org/confluence/display/FCREPO/Fedora+Roadmap)
Key Features of DSpace

Turn key application

Library centric model

Focus on end user features

Easy-entry solution for Institutional and Open Access Repositories
DSpace Community of Institutional and Open Access Repositories

DSpace registry just reached 700 installations worldwide
# DSpace - Notable New Features

<table>
<thead>
<tr>
<th>New DSpace Features</th>
<th>1.5 (current)</th>
<th>1.6 (Fall09)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SWORD module version 1.3.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>New authentication methods:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>hierarchical LDAP and Shibboleth</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>UI language translations:</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>German, Italian, Ukrainian, Greek, Thai</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Graceful resolver for URN</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Repository Statistics</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Embargo feature on content</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Bulk metadata editing</td>
<td></td>
<td>✓</td>
</tr>
</tbody>
</table>

**DSpace - Notable New Features**

- SWORD module version 1.3.1
- New authentication methods: hierarchical LDAP and Shibboleth
- UI language translations: German, Italian, Ukrainian, Greek, Thai
- Graceful resolver for URN
- Repository Statistics
- Embargo feature on content
- Bulk metadata editing
DSpace Futures

✓ DSpace 2.0 Service Manager demonstrated path to modularity
  ✓ StorageService
  ✓ MetadataSchemaService
  ✓ AuthN/AuthZ
  ✓ Session support

✓ Now strategizing on synergies with Fedora
  ✓ Shared modules
  ✓ DSpace on Fedora
DuraCloud
DuraCloud Basics

✓ New service fronting a network of cloud providers
✓ Replicate content to multiple storage providers; multiple geographic areas
✓ Monitor and audit digital content
✓ Run services in the cloud next to content

✓ Hosted by DuraSpace not-for-profit org
✓ Partnerships with cloud providers
✓ “Pay for use” for services and storage
DuraCloud

Trusted management of and access to durable digital assets in the cloud

DuraCloud Mediating Service

Durable Store Service Layer (Foundation Service)
Durable Store Service Layer (Commercially Operated)
Durable Store Service Layer (Locally Operated)

Foundation provided Software and Configuration (Open Source)

Amazon
Rackspace
EMC
Private Cloud

Sun
Microsoft
Offline Backup

Compute and Storage [Cloud/Grid] Providers
NYPL pilot

Digital Gallery Collection

- back up copy 700k images (50 TB data)
- transformation from Tiff to JPEG 2000
- run image server in cloud
- run video streaming in the cloud?
BHL pilot

BioDiversity Heritage Library

- back up copy entire corpus (40 TB data)
- have multiple copies including Europe
- do compute intensive data mining over corpus
Timeline

- Begin pilots – September 2009
- DuraCloud Alpha release to Pilot Partners - Oct 2009
- Pilot data loading and testing – Fall 2009
- Survey complete - October 2009
- Beta for repository community - Q1 2010
- Strategic partnerships in place - Q1 2010
- Pricing Model determined - Q1 2010
- Pilot testing with software services Q1 2010
- Report pilot results – Q1 2010
- Launch production service Q2 2010
More info:
http://duraspace.org

Stay tuned this year for new DuraSpace developments!!