Fedora Update:
Building Communities, Building Software

CNI 2013
December 9, 2013

Robert Cartolano, Columbia University
Tom Cramer, Stanford University
Jonathan Markow, DuraSpace
Robin Ruggaber, University of Virginia
What is Fedora?

Flexible Extensible Digital Object Repository Architecture

- open source digital repository software.

- community-built, installed in 300+ locations.

https://wiki.duraspace.org/display/FF/Fedora+Repository+Home
CNI 2012 - Fedora Futures Objectives

- Preserve strengths of the architecture and community
- Address needs for robust and full-featured repository services
- Provide a platform in the repository ecosystem for the next 5-10 years
CNI 2013 - One Year Later...

**Fedora Futures = Fedora**

- Expanding community
- Active development
- Improving long-term sustainability
Fedora - One Year Later

- Dec. 2012 - CNI 2012, Fedora 4 development begins
- June 2013 - Andrew Woods hired as Tech Lead
- July 2013 - Fedora 4 Alpha Release, OR 2013 Update
- August 2013 - Fedora 4 Beta Development Begins
- October 2013 - Fedora 3.7.1 Released
- Dec. 2013 - CNI 2013 update
- January 2014 - Fedora 4.0 “pre-Beta” Release
- July 2014 - Scheduled Fedora 4.0 Release at OR2014
Fedora - One Year Later

● Building Community
  ○ Fedora Steering, Fedora Advisors
  ○ Developer contributions
  ○ Engage Fedora Users

● Building Software
  ○ Fedora 4 strategy
  ○ Maintain Fedora 3
  ○ Improve development platform

● Building Sustainability
  ○ Increase funding
  ○ Develop staffing model
  ○ Refine DuraSpace services
Building Community

- 300+ Fedora Users
- 39 Fedora Sponsors
- 19 Active Developers
  - fedora-tech@googlegroups.com
- 17 Members of Fedora Advisory Group
  - fedora-advisors@googlegroups.com
- 10 Members of Fedora Steering Group
  - fedora-steering@googlegroups.com
Fedora Steering

Strategic planning, developer contributions, financial support, outreach
Fedora In-Kind Developer Contributors

- Columbia University
- discovery garden inc.
- FIZ Karlsruhe
- Max Planck Digital Library
- Media Shelf
- Stanford University
- University of California, San Diego
- University of New South Wales
- University of North Carolina, Chapel Hill
- University of Prince Edward Island
- University of Virginia
- University of Wisconsin
- Yale University
Engage Fedora Community

- Multiple in-person meetings
- Scheduled phone calls
- Mailing lists
- Quarterly update
- Web site redesign
- Gather use cases

Coming Soon:
http://fedorarepository.org
Building Software

- 3 year development effort
- Hired full-time Tech Lead
  - Andrew Woods - June 2013
- Community-sourced development
- Use-case driven
- Single product roadmap, backlog of features
- Frequent releases & acceptance testing
Code Base: Clean, Modern, Growing

http://www.ohloh.net/p/fcrepo4

- Fedora 4 Alpha 1 (July ‘13) had 80% of the Fedora 3 API functionality
  - in 7% the lines of code
  - with 72% test coverage (vs. 10% for F3)
Development Process
“Changing Everything, Two Weeks at a Time”

Distributed Team → Two-Week Sprints → Regular Releases

- Alpha 1: July ‘13
- Alpha 2: Oct ‘13
- “Pre-Beta” 1: Jan ‘14
- Beta 1: Mar ‘14
- Fedora 4.0: Jun ‘14

Acceptance Testing
- UAT by sites, for specific features
- Ongoing builds of Hydra & Islandora...

Jul 2013 - Dec 2013 Sprint Schedule

<table>
<thead>
<tr>
<th>Sprint</th>
<th>Dates</th>
<th>Andrew W (DuraSpace)</th>
<th>Greg J (UNC)</th>
<th>Osman D (Yale)</th>
<th>Yu J (DK)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>7/29 - 8/9</td>
<td>#</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>2</td>
<td>8/12 - 8/23</td>
<td>#</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>8/26 - 9/6</td>
<td>#</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>9/9 - 9/20</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>9/23 - 10/4</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>10/7 - 10/18</td>
<td>#</td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>7</td>
<td>10/21 - 11/1</td>
<td>#</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>11/4 - 11/15</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>11/18 - 11/29</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Sprint Summaries

1. Sprint B1
   - Authentication and Authorization
   - Policy-driven storage
2. Sprint B2
   - Triplestore
   - Large files
   - Policy-driven storage
   - Islandora
3. Sprint B3
   - Authentication and Authorization
   - Fedora 3 to 4 Upgrade
   - Backup and Restore
4. Sprint B4
   - Authentication and Authorization
   - Fedora 3 to 4 Upgrade
   - Triplestore
   - Large Files
   - Customizable Search Index
5. Sprint B5
   - Fedora 3 to 4 Upgrade
   - Islandora
   - Clustering
5. Sprint B6
   - Authorization
   - Clustering
   - Large Files

https://wiki.duraspace.org/display/FF/Beta+Development
Fedora 4 Component Stack

REST Framework
Fedora Services
ModeShape
Infinispan
Storage (Objects and Datastreams)

Access & Preservation Services
Repository Services
Caching, Clustering & Storage Services
Infinispan Options

- Read-biased cache
- Storage options
  - Filesystem
  - Several Dbs
  - Cloud
- Clustering Options
  - Local, non-clustered
  - Replication
  - Invalidation
  - Distribution
ModeShape Options & Features

- JCR implementation
- Storage options
  - Infinispan
  - File system
  - Several Dbs
- Features
  - Versioning
  - Transactions
  - Content hierarchy
  - Authorization
  - Workspaces
  - Events
  - Backup/Restore
  - Import/Export
  - Federation/Projection
  - Basic search
  - Namespaces
Fedora Integration Features
Key Features of Fedora 4

- Auditing & fixity services
- Simplicity
- Clustering
- Content modeling
- Batch operations
- Native linked data (RDF) support
Key Features of Fedora 4 (cont’d.)

- Event-driven architecture
- Transactions
- Advanced storage capabilities
- External data source “projection”
  - aka “virtual ingest”
- Admin and external search
- Admin and external Sparql endpoints
Ease of Deployment

This includes both the ability to deploy a repository instance without having pre-established system resources beyond the language platform (currently Java 7) and a JavaEE web-application container (not a full application server), as well as enabling repository administrators to (re)define application configuration and services at either deploy-time or run-time.

- Provide a double-clickable repository bundle
- Provide a WAR file that can be deployed to an existing servlet container
- Easily-Customizable Repository Configuration
  - Self-Hosted Repository Configuration
  - Configuration auto-discovery
  - Package default configuration as a repository
- Provide mechanisms for wiring components at packaging-time, at deployment-time, or at run-time.
Design For Scalability

Challenge:
● Exponential growth of data
● High availability, usage spikes
● Increased file sizes

Approach:
● Minimize local application state
● Minimize functions that are memory-bound
● Leverage distributed processing models
Building Sustainability

- **Finance**
  - Year-over-year growth - $173,000 to $411,333
  - 97 total DuraSpace Sponsors

- **Staffing**
  - Full-Time Tech Lead - Andrew Woods
  - Full-Time Product Manager - interviews underway

- **DuraSpace**
  - Strategic planning, budgeting support
  - Membership drives
  - Outreach, marketing, web site redesign
  - Administrative Support
Fedora Community - 300 Strong!

Of 97 DuraSpace sponsors shown above, 39 are Fedora-specific.
Upcoming Events

Fedora In-Person:

● Wed. Dec. 11 - Fedora Advisors
● March 10, 2014 - D.C. Fedora Users
● Upcoming Conferences - DCC, iPres, Code4Lib, DuraSpace Sponsors, RDA
Get Involved

- Review sponsorship level
- Contribute use cases
  - these will map to planned features
  - provide feedback
- Test Fedora 4
- Deploy a Fedora 4 pilot
- Provide in-kind developers
- Participate in upcoming meetings, events
Questions