Fedora Commons Update

Sandy Payette, Executive Director

Sun PASIG
Baltimore, Maryland
November 2008
Fedora 3.0 and 3.1 Released

3.0 (July 2008)
- Content Model Architecture
- JMS Messaging Support (repository publishes events)
- Mulgara semantic triplestore support
- New REST API (experimental version)
- Easy Relationships API (add/get/purgeRelationship)
- Ingest/Export using Atom syndication format
- Architectural grooming, streamlining and bug fixes

3.1 (October 2008)
- SPARQL support via Mulgara
- Bug fixes
Now stand back and strategize...
Waves of Repository-Enabled Applications

- **Institutional Repo and Digital Library Apps**
  - IR: Scholars to deposit articles, etc.
  - DL: Digital library search and access; collections

- **Collaborative “Web 2.0”**
  - collaborative filtering
  - annotate; discuss scholarly materials

- **E-Science, E-Research, Data Curation**
  - Publications linked to data
  - Data aggregation from distributed sources
  - Use of “Live” archives
We can build amazing private islands

But should we?
Emergence of Infrastructure

Systems
- Integrate components
- Central control
- Dedicated/specialized gateways
- More closed
- More preconceived

Networks
- Integrate systems
- Distributed control
- Generic gateways
- More open
- More reconfigurable

Repositories and infrastructure; let us contemplate....

- Single vs. multi-institutional perspective?
- Web fit?
- Impact of “cloud” and virtualization technologies?
- Best path to real interoperability?
- Complexity vs. simplicity?
- Positioning in emerging value networks?
Strategic Direction ...

New Ways to Expose Content
Fedora 2008-09: new Web and RDF Exposures

New Web APIs: Atom Publishing, SWORD, Other TBD


New: SPARQL/Mulgara, Linked Data?
Networks of Digital Objects ...
( aggregations related to aggregations)
Strategic Direction ...

Storage Abstraction
Akubra Project ... backend storage abstraction

Fedora Repository Service

New Web APIs | Fedora APIs | Registry Search | RDF Query

Store

Akubra Storage Abstraction

Plug-in 1 | Plug-in 2 | Plug-in ...

Cloud Storage:
- Amazon S3
- Others to come

Akubra 1.0 - Dec 2008

Cloud Storage: - Amazon S3 - Others to come

File System

registry db

More ...
“Content Service Ladder”

motivation: interoperability

Level 0: Storage Delegation

Storage Abstraction

Plug-in 1  Plug-in 2  Plug-in ...

Content Service Ladder

Bit streams
Blobs

Commercial Cloud
Enterprise SAN/Fabric
Local Storage

Fedora Commons

semantic aggregations blobs
Strategic Direction ...

Performance/Scalability Measurement
Welcome to the Fedora Performance and Scalability Wiki

Overall goal

It is of paramount importance for large-scale applications that Fedora Commons can handle huge amounts of data efficiently. While Fedora is generally known to be stable and reliable, there appears to be a lack of data and experience regarding large-scale installations and the performance implications thereof.

There are some known performance issues, namely regarding the ingest of new objects. It remains to be investigated if there can be some room for improvement. A faster ingest would be beneficial for many large projects using Fedora Commons.

The aim of this Wiki therefore is to gather data, document limits and constraints, devise recommendations and help improve Fedora Commons evolve further by contributing the findings to the community.

Latest Findings

We have successfully loaded 14 million objects (see Test Data) with roughly 750 million triples into Fedora. The ingest rate of about 10 objects/second was steady across the whole ingest process, which took roughly 21 days. Please find more details here.

About this Wiki

The Wiki consists of four distinct sections:

- Understanding Fedora explains certain aspects which are not currently covered by the Fedora documentation, namely functionality deep inside the system which is normally not important for the end user to know but bears significance performance-wise.
- Measurement documents some aspects about the basis of the measurement approach itself. In order to understand the test runs and the findings better, a context has to be provided.
- Test Runs documents the executed tests and their outcome. This is the core of the wiki.
Ingest Full Repository Part 2

 milliseconds

# ingest

7,000,001 - 14,000,000
Developer Forum: new collaboration space!
Solution Community Startups (talk to Thorny Staples)

Preservation and Archiving Solution Community

Steward: Ron Jantz, Rutgers University Libraries
Knowledgebase Gardener: Nancy McGovern, Interuniversity Consortium for Political and Social Research (ICPSR)
Evangelist: Chris Erickson, Brigham Young University

The mission of the Fedora Preservation and Archiving Solution Community is to establish the community framework to develop and support the preservation needs of the users of Fedora-based software including scientists, scholars, archivists, librarians, educators, and managers. The scope of the community's effort will be the complete life cycle of the digital object, independent of format, structure, or content. Through a collaborative process, we will leverage existing capabilities from the Fedora community and from the larger community that is involved in digital preservation. Our ultimate objective will be to develop the preservation architecture and comprehensive preservation capabilities including frameworks and supporting technologies that can be used selectively.

- [Goals]
- [Assumptions]
- [Community Documents and Guidance]
- Software Tools and Frameworks
- [Standards]
- [Storage Strategies]
New strategic partnership
DuraSpace
by DSpace Foundation and Fedora Commons

New Mellon grant for 6 month planning phase
More Info:

http://fedora-commons.org/
http://fedora-commons.org/confluence

Please join us in our new work!