Building the Future of Fedora

Edwin Shin (eddie@curationexperts.com) & Andrew Woods (awoods@duraspace.org)

11 July, 2013 • Open Repositories, Charlottetown

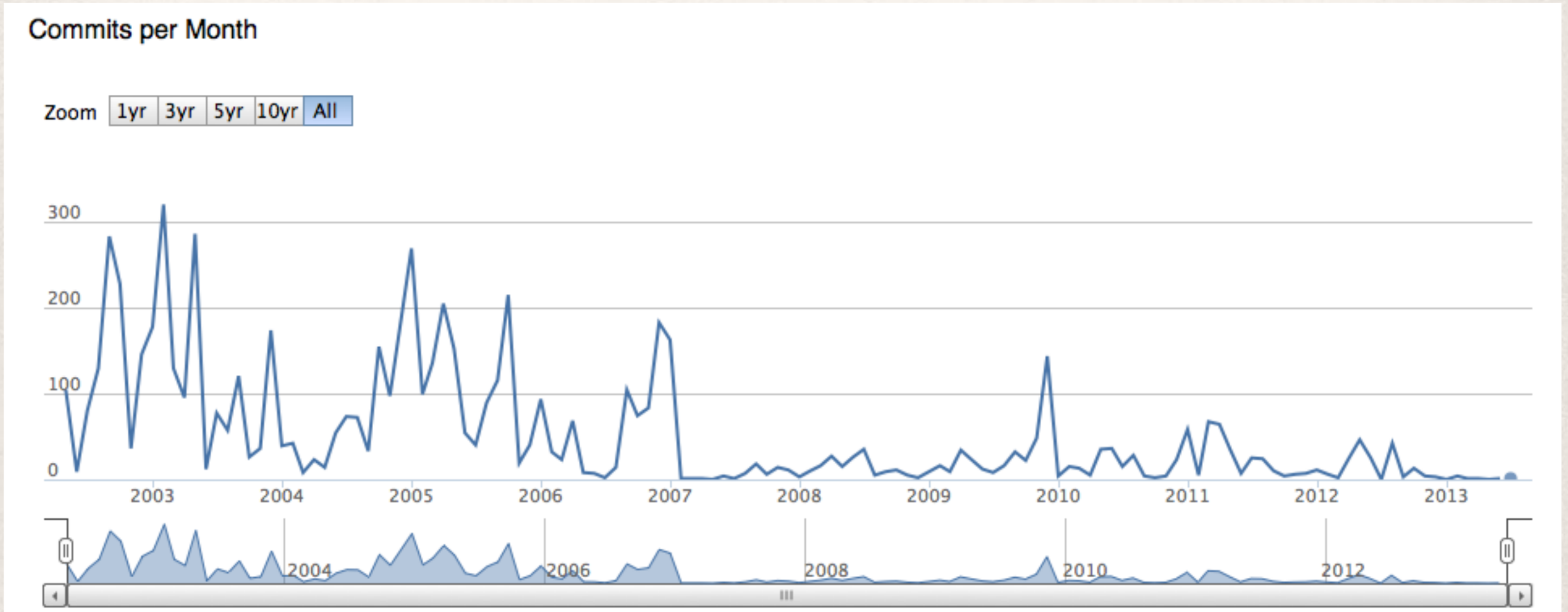
The Problem

- A large, aging codebase +
- Declining year-on-year number of developers +
- Declining year-on-year number of commits +

= slow to develop new features, hard to attract new developers

A strong and engaged developer community is an essential part of a preservation repository's success and sustainability

Fedora 3 Commits Over Time



Building Lean

Build - Measure - Learn

- Regular, short deliverables, validated with customers
  - A feature is delivered when it's made user-visible
  - A change in the development culture: customer-driven, data-driven
- Continuous integration, code quality, metrics gathering
- Profiling, benchmarking test suite

Fedora 4: Use Cases

Identified over 30 initial use cases

Large overlap, four major topics

1. manage research data
2. improve administrability
3. handle heterogeneous data more efficiently
4. interact with linked open data/semantic web

See: https://wiki.duraspace.org/display/FF/Use+Cases

Building Lean, cont'd

- Q: Reuse or Rewrite?
- A: Reuse *and* Rewrite

- Just 1 week to implement the minimum feature set to support running Hydra and Islandora on top of Fedora 4

Validation	Feature
• Hydra (rubydora, sufia fork) • Islandora (tuque)	• REST APIs

• SCAPE
  ◦ billions of Google Books scans
  ◦ > 90TB
	• Clustering for performance
	• Projection over HDFS
	• Deployment

Fedora 4: Features

Durability	• Self-healing • Transactions • Clustering for high availability • Metrics and reporting
Performance	• Batch operations • Clustering for scalability • Projection, aka "instant ingest"
Flexibility	• HATEOAS support • Eventing, messaging, & web hooks • Policy-driven storage • More storage options • Easy install & deployment • CMIS* • WebDAV* • OAuth 2*

* experimental

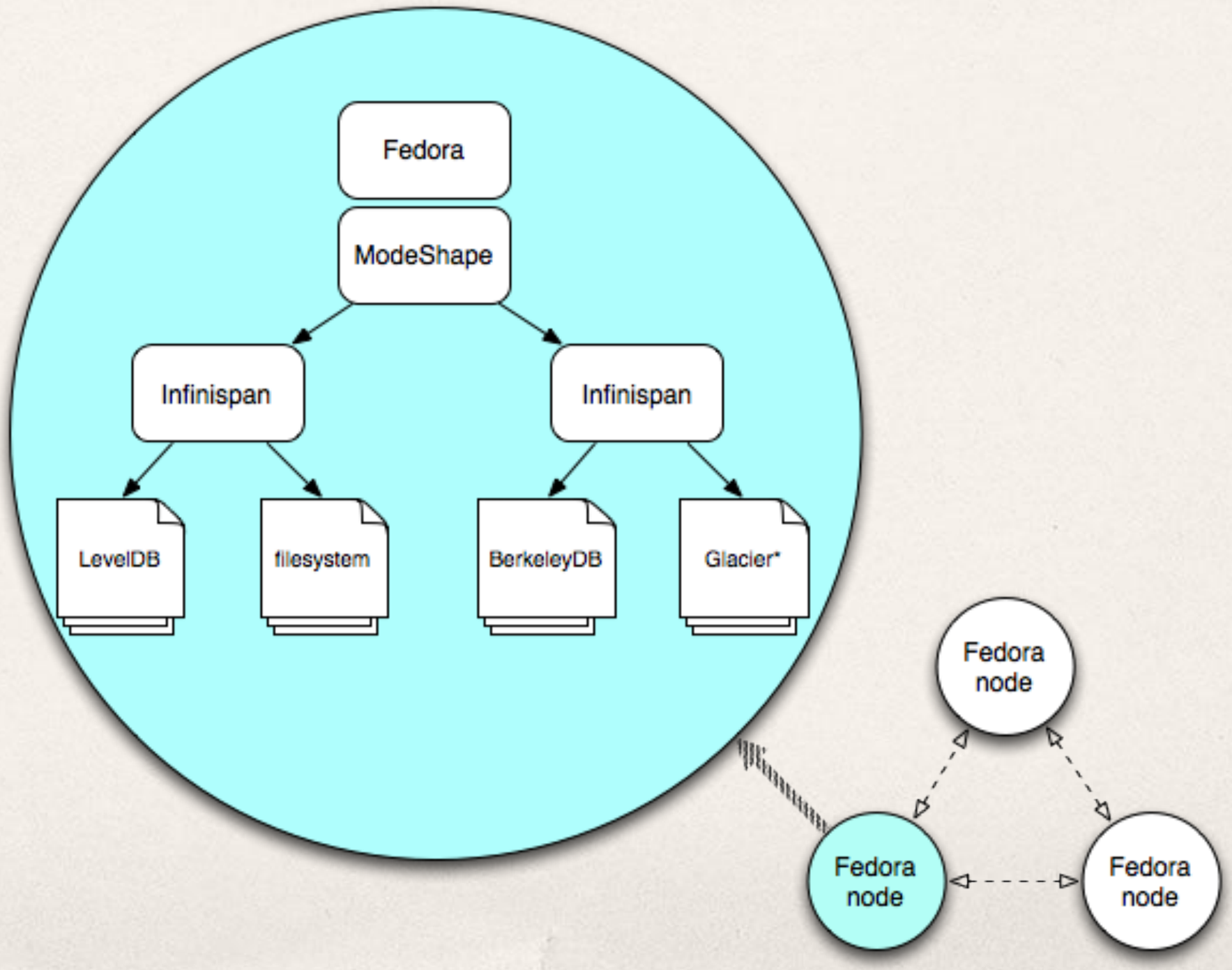
Fedora by the Numbers

	Fedora 3.6.2	Fedora 4 (alpha)
Lines of code	128,381	8,641
Test coverage	10.2%	71.8%
Public, documented API	44.4%	99.8%
Commits (12 months)	73	970
Contributors (12 months)	6	14

Sources:

- http://sonar.fcrepo.org/
- https://www.ohloh.net/p/fcrepo/
- https://www.ohloh.net/p/fcrepo4/

Architecture



Who Should be Using Alpha 1?

Early adopters

- Institutions with specific pain points with Fedora 3, e.g.
  - performance, scalability, storage flexibility, storage cost, high availability
- Institutions new to Fedora
- Institutions building out new (greenfield) Fedora applications, e.g.
  - research data managment
  - multimedia/video

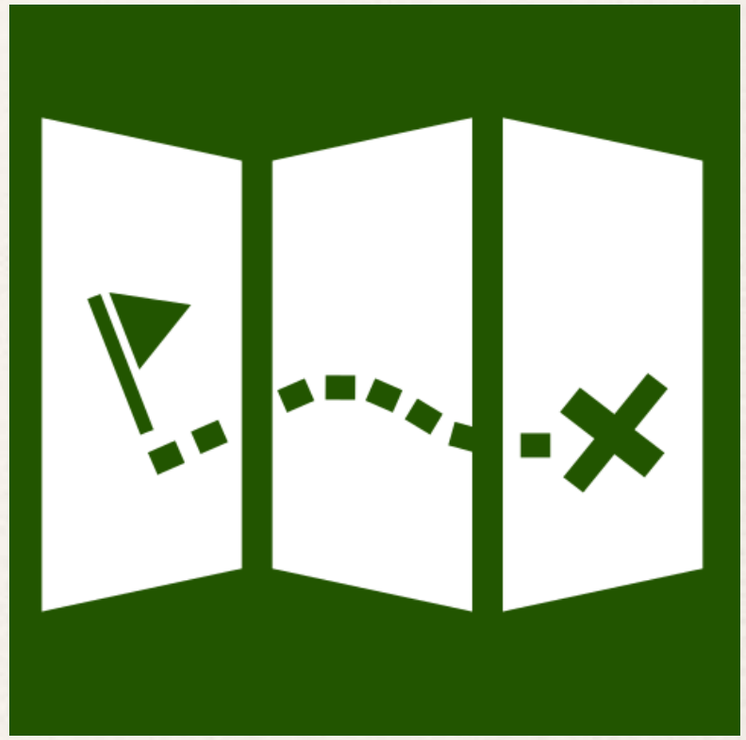
Solid Foundation In-Place

- Software infrastructure has been established
  - Code base
  - Agile process
  - Continuous integration environment

- Governance infrastructure has been established
  - Steering committee
  - Advisory working groups (technical and other)
  - Development team

Process Map

1. Minimize base feature set
   - Core features (examples)
     - Stable API
     - Versioning
     - Authentication / Authorization
     - Hardening Alpha capabilities
     - ...
   - External features (examples)
     - Fedora 3 --> Fedora 4 migration
     - Search
     - Triplestore
     - ...



2. Stakeholder validation of feature sprints
3. Aggressive release schedule

Be a Part of the Solution

- Provide sponsorship funding
- Provide skilled developers
- Provide use cases
- Spread the word

Thanks to our great devs!

- Chris Beer, Stanford University
- Ben Armintor, Columbia University
- Adam Soroka, University of Virginia
- Frank Asseg, FIZ Karlsruhe
- Paul Pound, University of Prince Edward Island
- Nigel Banks, Discovery Garden
- Esmé Cowles, University of California, San Diego
- Anusha Ranganathan, Oxford University
- Vincent Nguyen, Centers for Disease Control
- Greg Jansen, UNC Chapel Hill