Transition to a Broader Participation: Experience from the DSpace Project

MacKenzie Smith
MIT Libraries
DSpace is two things…

- **Service model** for Open Access and long-term archiving of digital scholarship

- **Open source software technology platform**

  Used in various settings to implement different service models
Where Are We?

- Mar 2004: First user group meeting 120 attendees, 7 countries
- Mar 2005: DSpace workshop Bangalore, India
- Feb 06: DSpace UG meeting, Sydney

- Nov 2000 - Nov 2002: HP-MIT development of DSpace 1.0
- Nov 2002: DSpace 1.0 Released
- Nov 2002 - Mar 2004: HP-MIT development; community support
- Apr 2004: Committer group formed
- Mar 2004 - Open Source community research/development
- Jul 2005: 2nd user group meeting 140 attendees, 22 countries
- Mar 2006: Governance advisory board meeting
Where Are We?

**Phase one (2003)**
- Lots of places experiment, “pilot” IR
- HP and MIT support and enhance software

**Phase two (2004)**
- With experience, other people answer questions and help out
- But they also complain… “why can’t DSpace do x?” and “when are you going to add y?” (vendor syndrome)

**Phase three (2005)**
- Many organizations go “live” with their IR
- Serious development outside MIT or HP, important contributions
- Community grows, wants structure, stability
DSpace Community

- Mostly higher education, related businesses
- Large library constituency for “IR” service model
- Many DSpace sites lack large IT departments
  - Need more technical support than typical OSS communities
  - Lack spare programmers to help with non-essential work (e.g. core infrastructure)
- Some commercial support options available
  - HP Global Solutions Group
  - Biomed Central Open Repository
  - CILEA in Italy, other national service bureaus
  - Private consultants
DSpace Community

Atypical open source software project

- Users are *organizations*, not individuals
- DSpace is an *entire application*, not a tool
- DSpace is an *end-user application*, not middleware or productivity tool
- Features and functions decided by *domain experts*, not programmers
Open Source Software and the Total Cost of Ownership
Total Cost of Ownership

- The *cost model* for DSpace depends on the specific *service model*

- Cost Categories for typical IR
  - Software
  - Product (or Service) Management
  - User Support
  - Ancilliary or “added-value” services
  - Collection Management
Total Cost of Ownership

Software costs for DSpace

- **Acquisition**
  - OSS free, commercial software for fee

- **Ongoing system administration**
  - Big expense in all cases, whether hosted locally or remotely

- **Local customizations and improvements**
  - Potentially big expense, but avoidable

- **Ongoing software maintenance/improvements**
  - Applies to OSS, can apply to commercial software (e.g. paying vendor for new features)
  - Can run into “free rider” problem here
Total Cost of Ownership

Hardware costs (open source and commercial)

- Servers
  - Usually not most expensive, since archives are high-value, low-access

- Storage
  - Can be very expensive depending on content policies

- Network connectivity

- Backup and disaster recovery
Total Cost of Ownership

Product management (open source and commercial)

- Defining local service model
  - What problem is more urgent for your institution?
  - What are the library’s priorities?
- Defining local policies and procedures
  - Just faculty research, or other types of material too?
- Developing business plan
  - What support can be provided for free and what needs to be cost-recovered?
- Performing outreach and marketing
  - What will faculty and administrators find compelling?
Total Cost of Ownership

User support (open source and commercial)

- Establishing communities
  - Setting up new users
  - Capturing legacy content and metadata

- User training and documentation
  - For metadata provision, format support, etc.

- Ongoing end-user support
  - Who do people call when things go wrong?
Total Cost of Ownership

“Added-value” services (open source and commercial)

- Scanning
- Cataloging
- Uploading
- Copyright checking

❯ Not required, but many service models choose to offer
Total Cost of Ownership

Collection Management (open source and commercial)

Least well-understood, but includes

- Setting, maintaining policies (>50 categories!)
- Description practices
- Digital preservation
- Reporting
- Legal responsibilities
Total Cost of Ownership

- Open source software free to acquire, *not to use*

- Commercial software costs to acquire, *and to use*

- Open source software
  - can outsource work (e.g. hire consultant, HP, BMC Open Repository, etc.)
  - or do yourself with local IT staff
Bringing it back together...

- DSpace and other open source software allow *unlimited customization, enhancement*

- Cost to maintain core system (infrastructure) is *shared by the community*
  - Currently borne by small number of institutions
  - Sustainability requires commitment to support the core

- New challenge for higher education, libraries, archives, and research organizations
DSpace Community

- Need to more clearly distinguish
  - Goals
    - Open Access agenda, digital preservation
  - Service models
    - IRs, LORs, DLs, etc.
  - Applications
    - DSpace, Eprints, FEZ
  - Infrastructure
    - Fedora, Cocoon, SRB
DSpace Community

Successful evolution

- Continuing good will
- Lots of ideas for improved service models
- Solid community tech support
- Increasing contributions of new features

But...
DSpace Community

- Need well-defined community management, governance processes
- Need more commitment from repository managers to maintaining, improving core system over time
Will This Work?

“When several villages are united in a single complete community, large enough to be nearly or quite self-sufficing, the state comes into existence, originating in the bare needs of life, and continuing in existence for the sake of a good life.”

Aristotle