DSpace development from MIT's Digital Library Research Program

MacKenzie Smith
Associate Director for Technology
MIT Libraries
Digital Library Research

- Many, many hard problems
- Can’t wait for perfect solutions
- Released DSpace and began using it in 2004
- Build research projects around it to
  - Learn
  - Improve DSpace
  - Provide related tools
Digital Library Research

- Not just MIT
  - HP Labs
  - Universidade do Minho
  - Indian Institute of Science
  - UCSD Libraries
  - Katholieke Universiteit Leuven
  - Several others

- Other types of contribution from working archives, tend to be local requirements-driven
DSpace 1.x Architecture
Digital Discovery

How will scholars find relevant content on the Web?

SIMILE

- Applied research on applications of Semantic Web standards and technologies to digital libraries
- Collaboration with MIT Libraries, CSAIL, W3C
  - and HP Labs initially
- Create *tools* for discovering and navigating digital resources on the Web
  - Data model *scalable* to the entire Web
  - *Interoperable* across systems, institutions, domains of practice
Digital Discovery

How will scholars find relevant content on the Web?

SIMILE tools

- Longwell – Web-based faceted browser for RDF metadata
- Piggy Bank – Firefox extension for desktop metadata management (similar to EndNote)
- Solvent – scripting language for HTML Web page scraping into RDF
- Semantic Bank – tool for “publishing” Piggy Bank collections to a group or to the world
- Welkin – RDF viewer, data exploration tool
- Gadget – command line XML inspector (RDF conversion aid)
- What else is there?
Digital Discovery

How will scholars find relevant content on the Web?

SIMILE in DSpace 2.0 achieves

- transparent multi-schema metadata support
- cross-schema edit, search and browse
- semantic interoperability between different DSpace sites
Digital Collections

How will libraries preserve digital data over archival time frames?

DSpace@Cambridge

- Cambridge-MIT Institute funding; Cambridge University, MIT collaboration
- Technical focus on digital preservation
  - Checksum checker, jhove, other tools for data integrity
  - Preservation strategies for HTML, XLS, others
Digital Collections

How will libraries preserve digital data over archival time frames?

FAÇADE (Future-proofing Computer-Aided Design)

- Focus on architecture and urban planning
- 3D CAD models (parametric, kinematic, solid)
- 2D CAD drawings
- Related material (e.g. materials lists, communications)
- Support annotations
- Work with Frank Gehry, Gehry Technologies, Dassault Systemes, other projects
Digital Collections

How will digital libraries scale up?

**DSRB (DSpace + SRB)**

- NARA funding; SDSC, UCSD Libraries collaboration
- SRB middleware defines abstract “storage utility” for DSpace assetstore (extremely large-scale)
- SRB content can be widely distributed, replicated
- Integrated with DSpace 1.3, test storage available from SDSC
- Possible model for other abstract storage utilities (e.g. HP, EMC)
Digital Collections

How can we federate our collections and services?

PLEDGE (PoLicy Enforcement in Data Grid Environments)

- NARA funding
- Collaboration with SDSC, UCSD Libraries
- Define policies across the entire information life cycle and from the enterprise to the file level
- Specify policy encoding standards
- Develop policy engine for DSpace, SRB (and potentially other external systems)
Digital Education

How should libraries manage digital teaching materials?

CWSpace

- OpenCourseWare – 2,000 MIT courses freely available on the Web
- Courses evolve, need archiving
- “Learning Objects” hard to find, reuse
- Data standards immature, evolving (IEEE LOM, IMS CP)
Digital Education

*How should libraries manage digital teaching materials?*

**DSpace@Cambridge**

- Also collaborating with University of Cambridge CARET (academic computing)
- Interoperability with Virtual Learning Environments (e.g. SAKAI roadmap)
Partnerships

Digital Library Community

- Technology partnerships
  - Digital Library Abstract Service Framework
  - Global Digital Format Registry

- Digital Scholarly Communication partnerships
  - SPARC, Creative Commons, Science Commons
  - DPubS publishing system (Cornell, Penn State)