DSpace@MIT
A Case Study in Academic Institutional Repositories

Sean Thomas, Program Manager for Scholarly Repository Services; Product Manager, DSpace@MIT

Repositories in Science & Technology: Preserving Access to the Record of Science.

11/30/11

MITLibraries

What Am I Doing Here?

What is DSpace?

❑ Developed in 2002 by MIT and HP Labs, DSpace is an open source solution for long-term archiving and open access to a variety of digital resources.

❑ DSpace provides facilities for capturing, storing, indexing, preserving, and distributing digital content.

❑ DSpace is used by around 1000 institutions globally.

Why was DSpace Developed?

❑ Support for long-term archiving of MIT's scholarly output

❑ Stewardship of born-digital content (e.g., format registration, migration, preservation needs)

❑ Support for heterogeneous formats

❑ Dissemination platform for technical reports, working papers, and other MIT-produced research

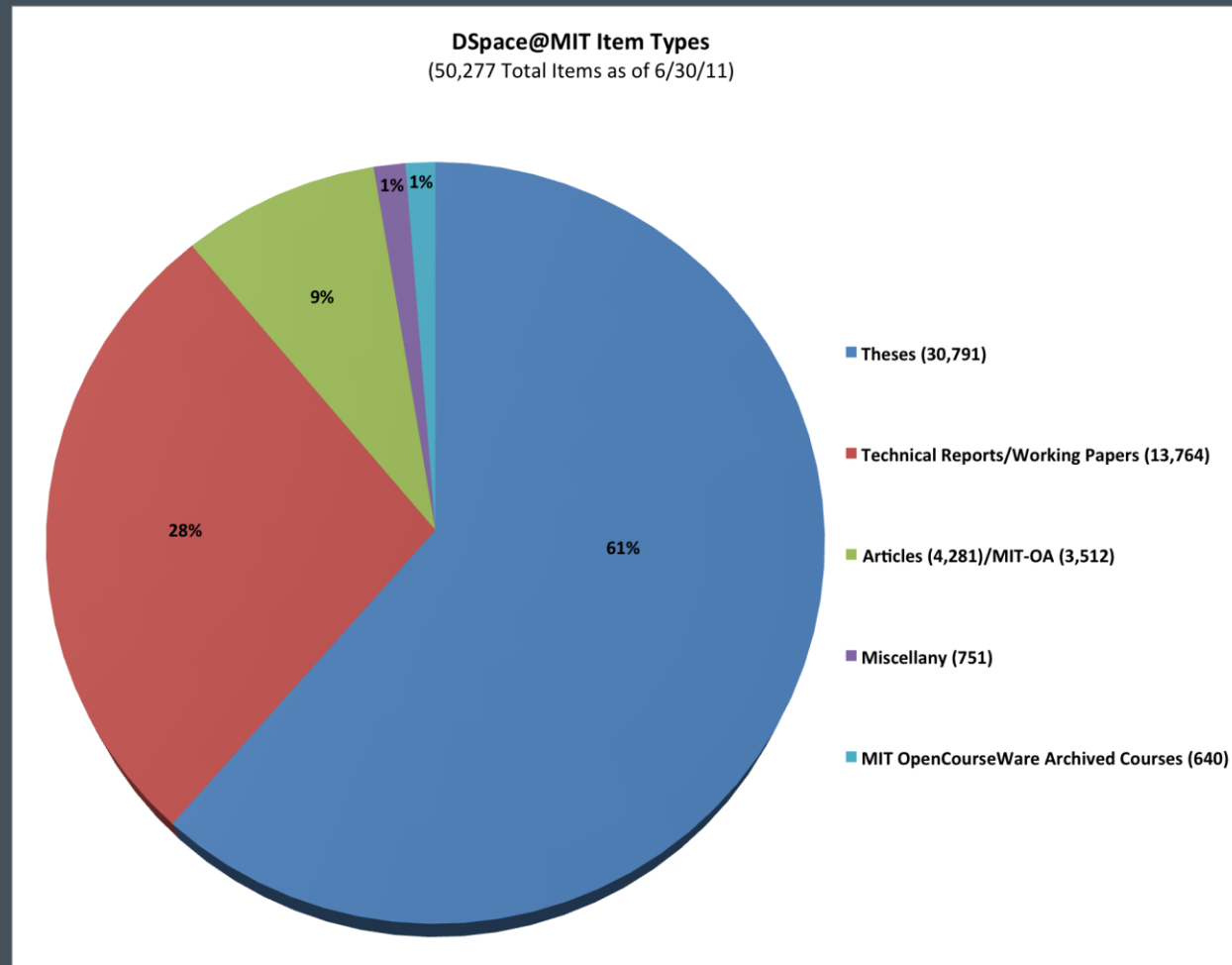❑ Open access content discovery

DSpace@MIT (Statistics)

- 50,227 deposited items
- 90+ (MIT DLCPs)
- 600+ collections

- 11.7 Million downloads FY11
- 2.1 Million additional views
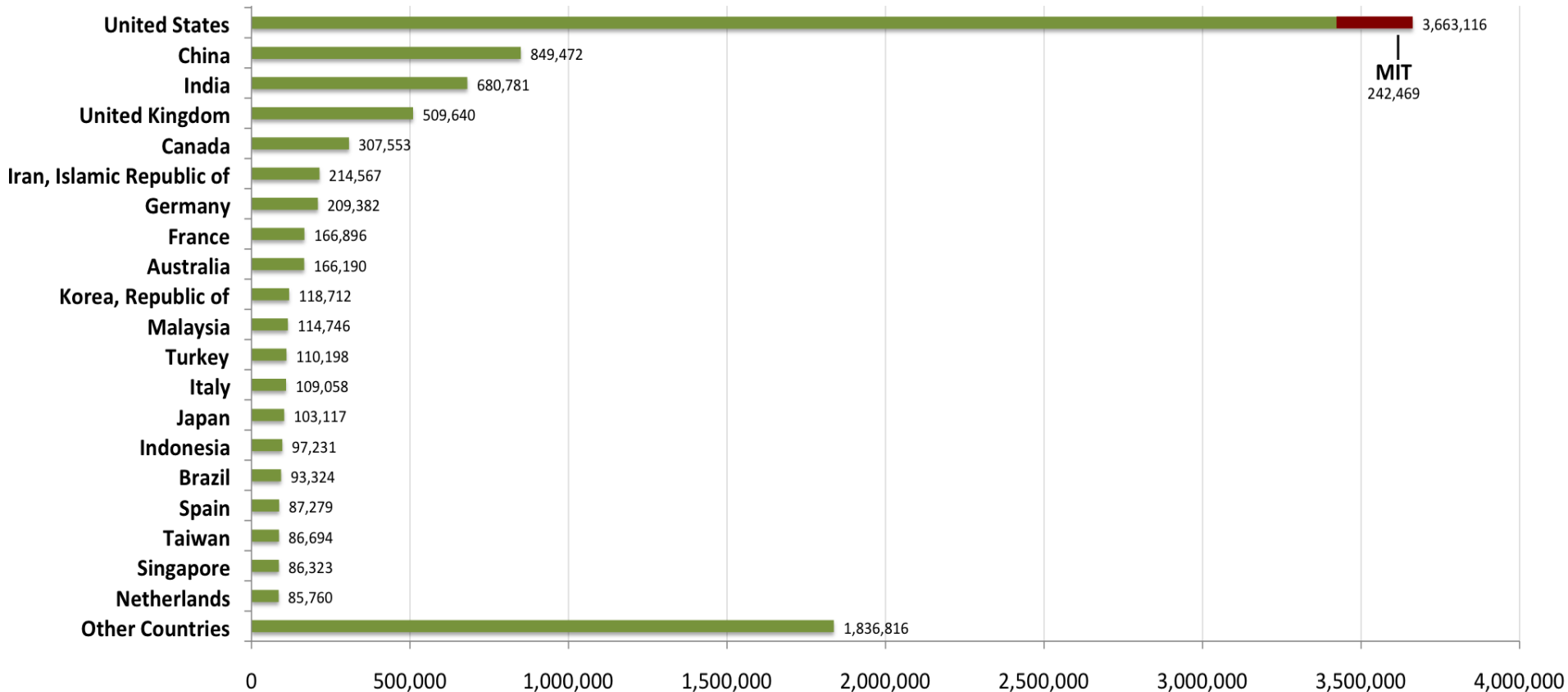- ~1M downloads/month; 32K downloads/day

DSpace@MIT Item Growth
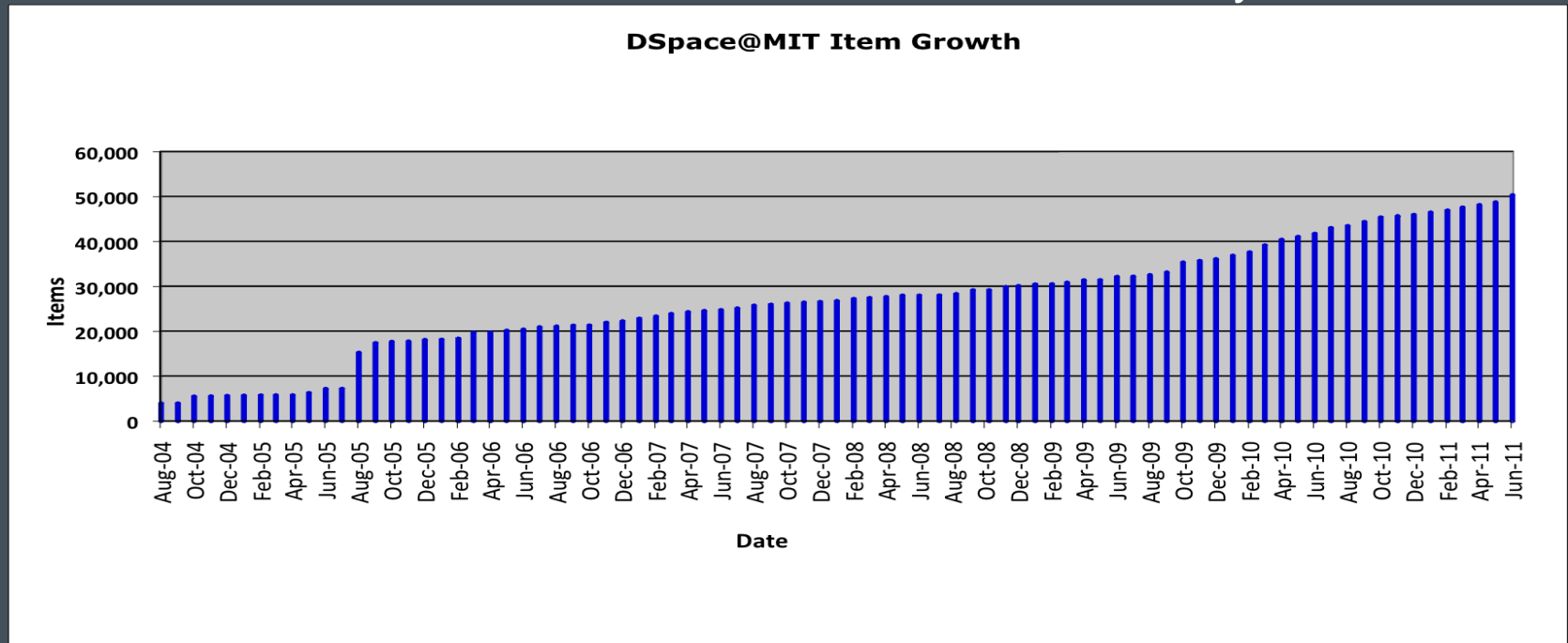
DSpace@MIT (Content)

- ❑ E-theses
- ❑ Technical Reports
- ❑ Working Papers
- ❑ Peer-reviewed Literature
- ❑ Research Datasets
- ❑ Archived OCW courses
- ❑ Images
- ❑ Audio/Video
- ❑ Websites

DSpace@MIT Item Types
(50,277 Total Items as of 6/30/11)

61%

28%

9%

1% 1%

- ■ Theses (30,791)
- ■ Technical Reports/Working Papers (13,764)
- ■ Articles (4,281)/MIT-OA (3,512)
- ■ Miscellany (751)
- ■ MIT OpenCourseWare Archived Courses (640)

DSpace@MIT (Usage by Country)

DSpace@MIT Usage (FY2011 File Downloads by Country)

Country	Downloads
United States	3,663,116 (MIT 242,469)
China	849,472
India	680,781
United Kingdom	509,640
Canada	307,553
Iran, Islamic Republic of	214,567
Germany	209,382
France	166,896
Australia	166,190
Korea, Republic of	118,712
Malaysia	114,746
Turkey	110,198
Italy	109,058
Japan	103,117
Indonesia	97,231
Brazil	93,324
Spain	87,279
Taiwan	86,694
Singapore	86,323
Netherlands	85,760
Other Countries	1,836,816

MIT Faculty Open Access Policy

"The Faculty of the Massachusetts Institute of Technology is committed to disseminating the fruits of its research and scholarship as widely as possible. In keeping with that commitment, the Faculty adopts the following policy: Each Faculty member grants to the Massachusetts Institute of Technology nonexclusive permission to make available his or her scholarly articles and to exercise the copyright in those articles for the purpose of open dissemination."
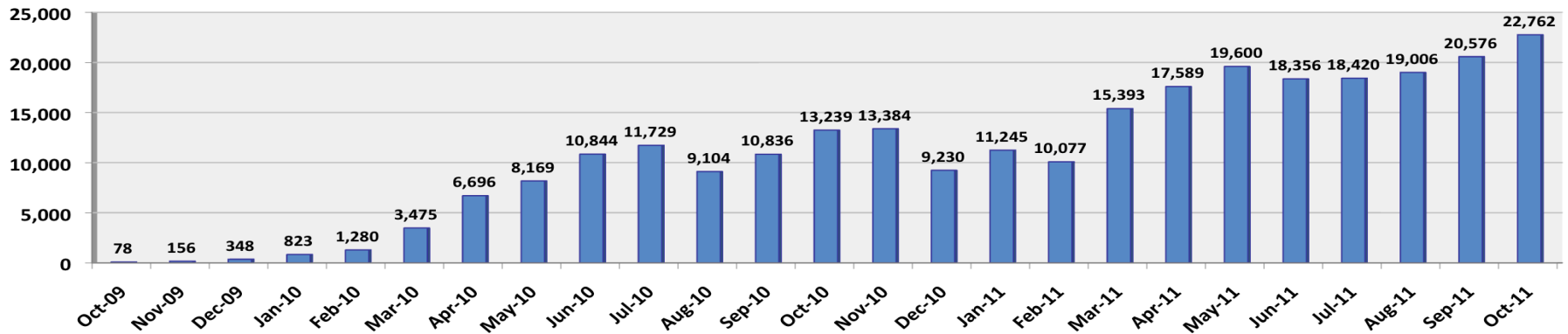
**Adopted by unanimous vote of the faculty on 3/18/2009**

MIT Open Access Articles

- ❑ 4400+ articles live; hundreds more queued
- ❑ 270,000+ downloads in two years

DSpace@MIT - (Browser-based Article Downloads)
MIT Open Access Articles
(Monthly - Oct. 2009 - Sept. 2011)

Month	Downloads
Oct-09	78
Nov-09	156
Dec-09	348
Jan-10	823
Feb-10	1,280
Mar-10	3,475
Apr-10	6,696
May-10	8,169
Jun-10	10,844
Jul-10	11,729
Aug-10	9,104
Sep-10	10,836
Oct-10	13,239
Nov-10	13,384
Dec-10	9,230
Jan-11	11,245
Feb-11	10,077
Mar-11	15,393
Apr-11	17,589
May-11	19,600
Jun-11	18,356
Jul-11	18,420
Aug-11	19,006
Sep-11	20,576
Oct-11	22,762

MIT Open Access Articles

MIT Mission: "The Institute is committed to generating, disseminating, and preserving knowledge, and to working with others to bring this knowledge to bear on the world's great challenges."
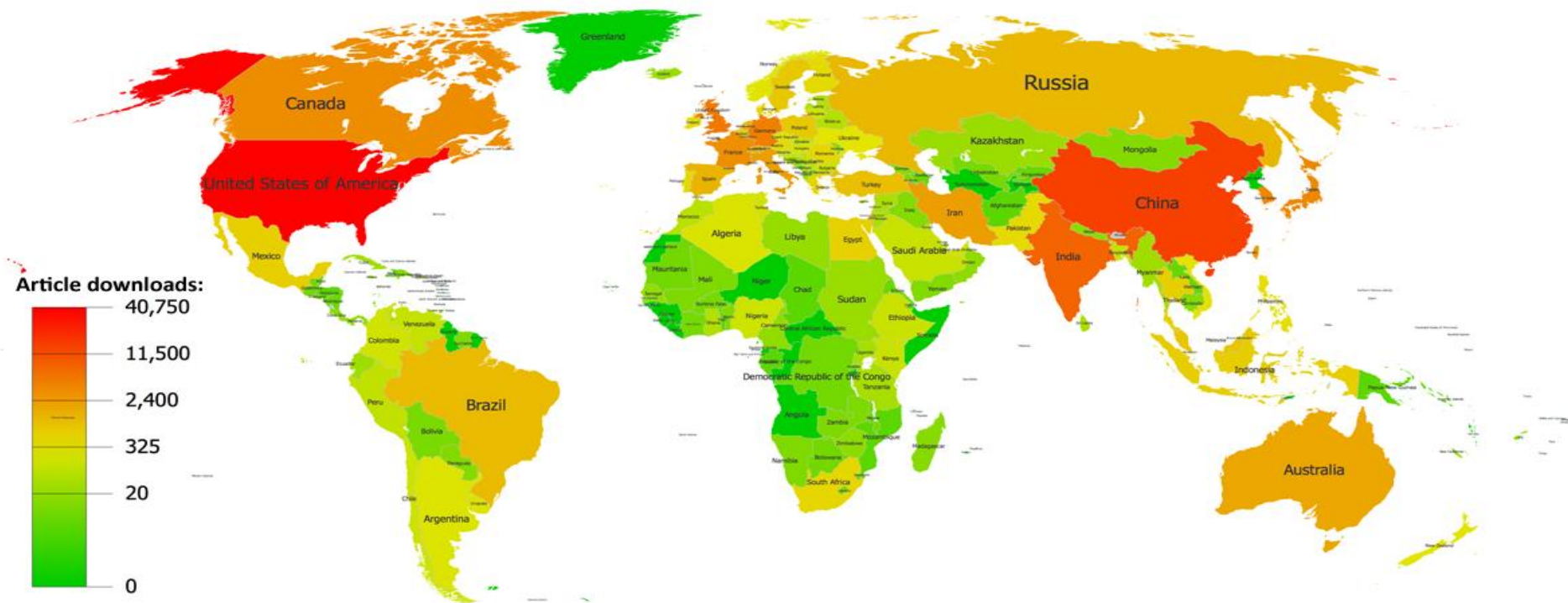
Inside OA

- ❑ Input streams
  - ○ From publisher
    - ⊙ Harvest from publisher (as policies allow)
    - ⊙ Direct submission by publishers via SWORD protocol(Contains METS manifest, SWAP metadata → DC)
  - ○ 'Harvesting' from MIT web domain
  - ○ Recruitment by Libraries liaison staff
  - ○ Direct submission by author
- ❑ Building sustainable workflows
  - ○ Minimize faculty effort
  - ○ Leverage existing systems to aid in author disambiguation
  - ○ Leverage existing systems for metadata (e.g., CrossRef)

Challenges & Lessons Learned

1. Competing Development Priorities

- ❑ Streamlining content acquisition processes
- ❑ Search engine optimization
- ❑ Discovery user interface
  - ○ User experience (UX) enhancements
  - ○ Additional interface functionality
- ❑ Preservation-related initiatives
- ❑ Local development vs. DSpace open source community commitments

2. Sustained Community Engagement

- ❑ Churning staff in research community
  - ○ Depositing administrators, database owners
- ❑ Dual/multiple deposit workflows
  - ○ IR, subject repositories, DLC websites
- ❑ Difference between exposure and persistence
  - ○ Personal and DLC websites are *fragile*
- ❑ Perspective – libraries not traditionally thought of in the research exposure context

Currently piloting mediated deposit for additional content and expanding engagement through our DLC liaisons.

3. Metadata for Discovery

❑ Personal and organizational name variants

- ○ For DISCOVERY
  - ◉ Can't build a true Author page
  - ◉ Can't hyperlink to other works from an item page
  - ◉ Can't show our 'dirty laundry' in a facet
  - ◉ Can't share out usage statistics to an author
- ○ For REPURPOSING
  - ◉ Can't give web admins an easy way harvest content
  - ◉ Can't harvest and index in other Libraries' systems
  - ◉ Can't build useful semantic indexes
- ○ For EFFICIENT CATALOGING (authors, orgs, sponsors)

Beyond DSpace@MIT
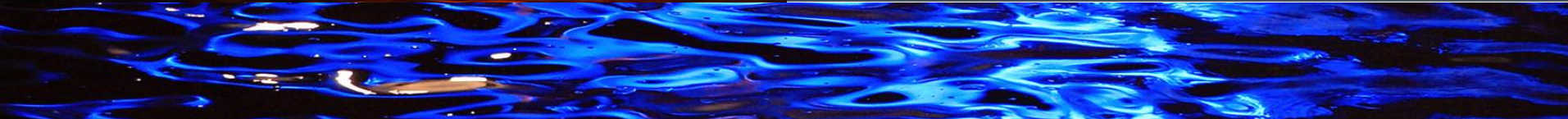
Future Opportunities?

❑ Google Refine – for data normalization
- ○ Clustering algorithms with mass editing functionality
- ○ Reconciliation to external systems of authority via APIs
- ○ Save editing history as JSON to apply to future data

❑ ORCID
- ○ Help alleviate the name disambiguation problem

❑ VIVO
- ○ Improved serendipitous discovery of content of interest
- ○ Exposed semantically rich data aids reuse in external systems
- ○ Could become a local resource for *authorities*

❑ ???

Additional Information

- ❑ DSpace Homepage:
  - ○ http://dspace.mit.edu
- ❑ More Information on DSpace@MIT:
  - ○ http://libraries.mit.edu/dspace-mit/
- ❑ MIT Faculty Open Access Policy:
  - ○ http://info-libraries.mit.edu/scholarly/faculty-and-researchers/mit-faculty-open-access-policy/
- ❑ MIT Open Access Articles Collection
  - ○ http://dspace.mit.edu/handle/1721.1/49433
- ❑ Sean Thomas (DSpace@MIT): sthomas@mit.edu

Photo Credits

- http://www.flickr.com/photos/cobalt/6343737548/
- http://www.flickr.com/photos/eandjsfilmcrew/5100556161/
- http://www.flickr.com/photos/bsaren/3694059797/
- http://www.flickr.com/photos/cdevers/5701852547/
- http://www.flickr.com/photos/chanc/374386314/