

Patricia Renfro  
Columbia University  
January 28, 2010

**SUBJECT,  
INSTITUTIONAL AND  
OTHER REPOSITORIES:  
A US PERSPECTIVE**

- ◎ Five major subject repositories
  - Characteristics of success
  - Impact of Public Access policies
  - Issues of sustainability
- ◎ Institutional repositories
  - Characteristics
  - OA mandates
  - Success?
- ◎ Data
- ◎ A new social networking model

# Five transformative subject repositories

- **arXiv** – high energy physics, maths, statistics, financial economics, computer science
- **CiteSeerx** – computer science
- **RePEc** – economics
- **SSRN** – social sciences
- **PubMed Central** – biomedical and life sciences



Cornell University  
Library

arXiv.org

- 575,000 e-prints, 60,000 new articles per year
- 2008: 30 million downloads
- Founded 1991 by Paul Ginsparg
- 2001 – at Cornell University
- Open deposit – light quality control – advisory committee & moderators
- Physicists choose arXiv link over publishers link 82% of time
- Interoperates with SPIRES, PMC, RePEc, SSRN ...

# CiteSeer<sup>x</sup> beta

Scientific Literature Digital Library and Search Engine

- 1,424,256 items
- Based at Penn State University
- Harvests and indexes research articles from publicly available Websites
- Autonomous citation indexing
- Currently has NSF funding
- Previous sponsors include Microsoft

# *RePEc*

- ◉ 315,000 working papers
- ◉ Links to 490,000 journal articles
- ◉ Collaboration of hundreds of volunteers from 70 countries
- ◉ Institutional & personal submissions



- 219,000 full text documents
- 268,100 abstracts
- Contents are Open Access
- For profit company – makes money from services offered
- Strong focus on use and citation data



- 1, 608,559 peer-reviewed journal articles
- Developed and administered by the US NIH & NLM, integrated with GenBank
- Publisher participation voluntary - can opt for delay in release – 6, 12 months
- NIH Public Access Policy Spring 08 – major impact
- UK PubMed Central & PubMed Central Canada



# Characteristics of success?

- ⦿ Tightly linked to disciplinary practice and culture
- ⦿ High energy physics, economics – prior practice of sharing print preprints
- ⦿ Researchers identify with discipline before institution
- ⦿ Researchers see direct value of deposit – date stamping new work, being read by colleagues, knowing about use.

# Impact of Public Access Policies

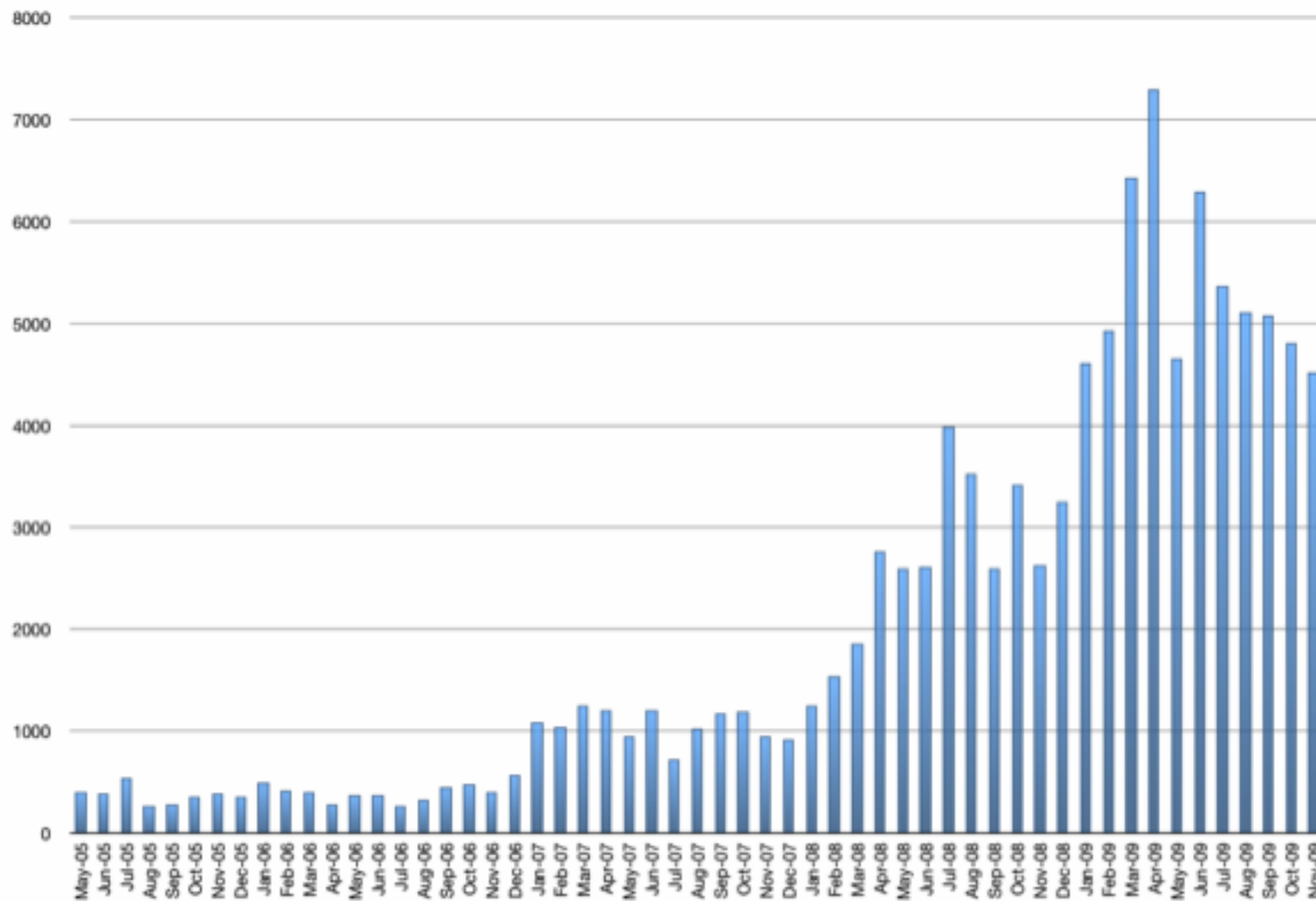
- ◉ NIH Public Access Policy passed Spring 2008
  - Requires deposit in PMC of journal articles resulting from NIH-funded research - embargo can be no longer than 12 months
  - Before Policy – 19% of 80,000 papers per year collected. First 5 months of policy 56% collected

# NIHMS Statistics

## Monthly Aggregate Submission Statistics

This chart represents the number of manuscripts received by NIHMS and approved for processing by the author/PI for a given month.

[Download](#) the data used to create this chart in CSV format.



# Funders requiring OA deposit

- NIH
- National Center for Atmospheric Research (NCAR)
- MacArthur Foundation
- Howard Hughes Medical Institute

# Federal Research Public Access Act FRPAA

- ⦿ Would require that publications resulting from federally funded research (11 largest funders, NSF, DOE ...) be made publicly available within 6 months of publication
- ⦿ Legislation currently in the Senate
- ⦿ Office of Science & Technology Public Comment blog – 500 comments – policy recommendations to follow
- ⦿ How/where to deposit under discussion
- ⦿ Johns Hopkins grant to explore development of repository for NSF funded research output

# Sustainability

- ⦿ Government funding - PubMed Central –cost \$4.5 million per year
  - Other US gov funded repositories – e.g. NASA Astrophysics Data System
  - Will FRPAA result in agency-funded solution
- ⦿ Institutional funding - arXiv cost \$400,000 p.a.
- ⦿ Community-based alternative – recent arXiv announcement
- ⦿ Commercial model – SSRN
- ⦿ Sponsorship - CiteSeerX

# US institutional repositories

- ◉ ROAR listings by size – few US IRs compare in size with Cambridge, Ghent, Amsterdam, Queensland.
- ◉ University's of Michigan's Deep Blue repository follows Tilburg, Adelaide and Southampton in size.
- ◉ Commitment, resource allocation varies widely
- ◉ No consistency in coverage – e.g. economics department representation spotty.

# Characteristics

- ◉ Content generally full text - no equivalent to the UK's RAE & incorporation of current research listings
- ◉ But – some interest In faculty profiles – e.g. bepress's “Selected Works” feature; & Cornell's Vivo
- ◉ Systems in use: DSpace, Fedora, bepress Digital Commons; eprints, & locally developed.



# Progress of OA mandates

- ◉ Feb 08 - Harvard OA mandate from Arts and Sciences faculty, followed by Law School and Kennedy School
- ◉ MIT, Boston University, University of Kansas, Trinity University, Oberlin College, Stanford School of Education followed
- ◉ Discussions underway in many institutions
- ◉ OA for dissertations – becoming more common

# Success?

- ⦿ Major concern – lack of faculty buy-in – strong contrast with subject repositories
- ⦿ Result – refocusing of IRs
  - University of Rochester – in depth study of users – newly developed system – new features – Web space for collaboration, support for authoring, co-authoring, self-publication
  - California Digital Library asked – what's success? eScholarship = 30,000 items but UC system estimated to publish 26,000 research papers a year. Refocus on publishing – disseminating scholarship – IR deposit a by-product

# Data deposit

- ⦿ AER and other economics journals – data publishing requirement
- ⦿ New data archiving policy for ecology and evolution journals – Journal of evolutionary biology, Molecular Biology etc
- ⦿ NSF grant for DRYAD – digital repository for ecology ++
- ⦿ NIH – requires a data sharing plan – but no enforcement  
Other federal funding agencies – no requirement
- ⦿ Recent NSF grant to Johns Hopkins to develop framework to understand data practices

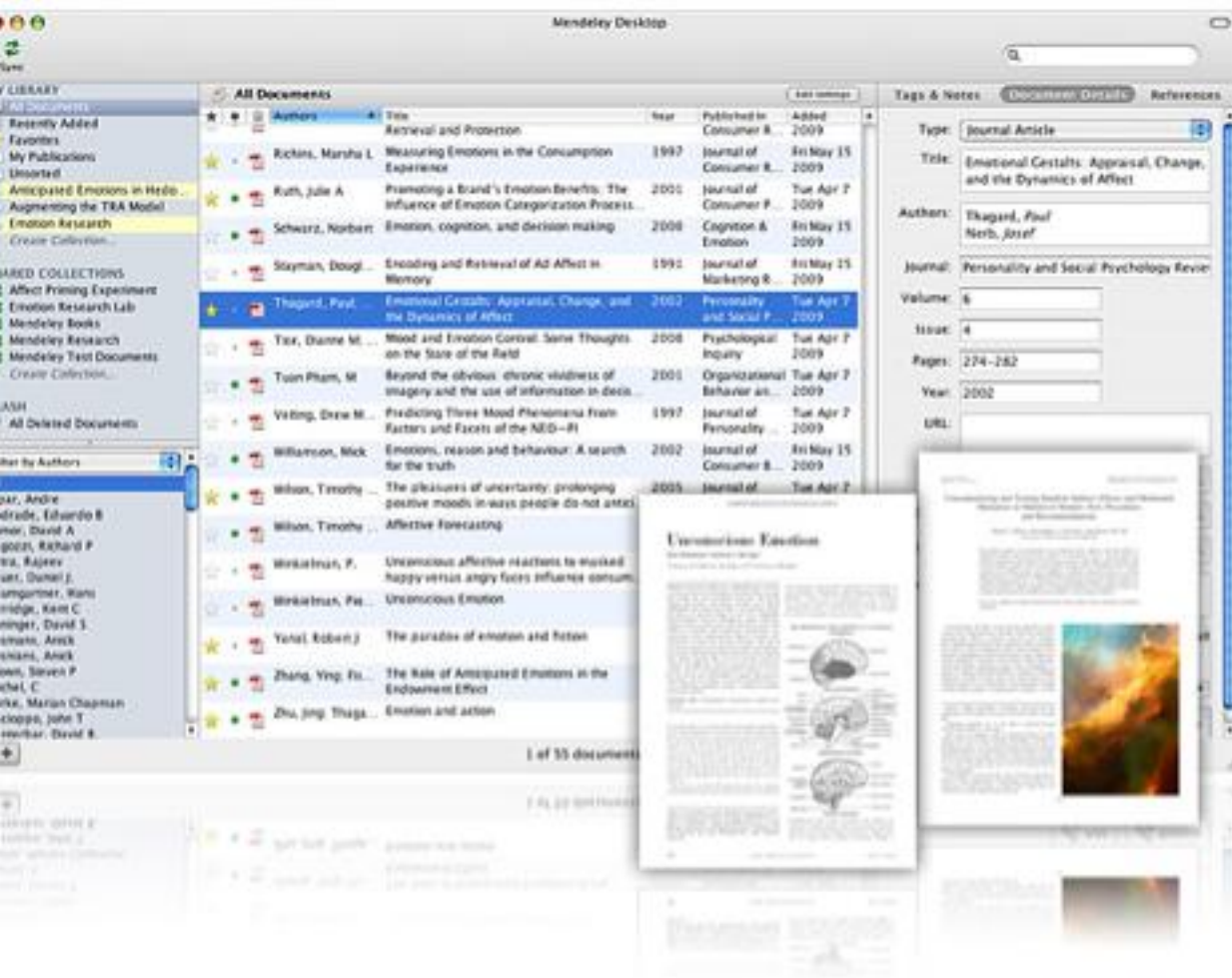
Get Mendeley

How it works

About us

Support

Blog



# Organize, share and discover **research papers**

Like iTunes™ for research papers.  
Mendeley is a free research  
management tool for desktop & web.

Share papers  cite & create bibliographies

in Word & OpenOffice  collaborate

search & annotate PDFs  sync across PCs

**Get Mendeley**  
FREE DOWNLOAD

Available for **Windows**, **Mac OS X** and **Linux**

**Discover**

new research and trends

**Share**

with other researchers

**Organize**

your own research library

# Mendeley – a very different model

- ⦿ Mendeley – “iTunes for research papers” to organize, share and discover research papers
- ⦿ Social network becomes repository
- ⦿ Simple tools, lots of value, instantaneous results (drag & drop papers onto site; recommender service/annotation feature, use analysis, creation of shared communities)
- ⦿ 100,000 users & over 8 million research articles – doubles in size every 10-12 weeks
- ⦿ London-based, team drawn from UK, German and US institutions – significant investment capital
- ⦿ <http://www.mendeley.com/>

Questions?