Research and Education Data Management

ITPB 2007-08 Priority Area

Agenda for Discussion at November 1, 2007 ITPB Meeting

Christine L. Borgman
Issue framing

☐ #2. *(44 VOTES)* The campus research and educational data management - i.e. data management as infrastructure

☐ help for researchers to create data management plans

☐ develop an institutional response to the directions of grant agencies regarding data

☐ policy and standards for setting up databases

☐ who owns the problem
e-Research infrastructure: Layered Model

- **Information & knowledge layer**
  - Content
    - Digital Libraries
  - Scientific DBs

- **Middleware**
  - services layer

- **ITC Infrastructure**
  - Processors, memory, network

Applications Space

User Interfaces & Tools

Slide courtesy of Stephen Griffin, NSF, and Norman Wiseman, JISC
Scholarly Information Infrastructure

- Goal: enable new forms of scholarship that are
  - information-intensive
  - data-intensive
  - distributed
  - collaborative
  - multi-disciplinary
- Means: use information technology to
  - improve access to scholarly information
  - manage the “data deluge”
  - leverage data as a form of scholarly capital
  - conduct computationally intensive research
Driving Forces

- **Technology push**
  - Distributed access to content and computing resources
  - Tools and services for data collection, mining
  - High performance computing, grid services

- **Collaboration pull**
  - Virtual organizations
  - Share distributed resources

Content layer

- **Documents**
  - Publications: books, journals, conference papers, ...
  - Semi-formal: technical reports, working papers, proposals…
  - Unpublished: websites, blogs, wikis…

- **Data**
  - Observational
  - Computational
  - Experimental
  - Records

- **Composite objects**

http://www.medscape.com/content/2004/00/46/81/468129/art-mgm468129.fig1.jpg
Value chain of information

**Links**
- Cited/citing documents
- Publications to data sources
- Data to publications in which reported

**Across boundaries**
- Repositories
- Publisher databases
- Disciplines
- Countries

Open Access Movement

- Access to scholarly publications
  - Bundles to libraries, at growing expense
  - Limited access outside top tier universities
  - Internet lowers barriers to entry, encourages new publishing models

- Open science, open scholarship
  - “Author self-archiving”
  - Institutional repositories
  - Open access journals

- Open access to documents and data
Role of data in the value chain

- Scholarly capital
  - Human capital
  - Instrumentation
  - Data

- Leverage research investments
  - Replicate, verify research findings
  - Ask new questions with extant data
    - Computational biology, chemistry
    - Longitudinal and comparative social research
    - Mining large bodies of literary texts

- Collaborative research
  - Data creation
  - Data sharing, reuse
Policy environment

- Publicly funded research
  - Access to scholarly products
    - Publications
    - Data
  - Legislation and policy
    - U.S. Congress - bills pending
    - European Union policies and initiatives for open access
  - Funding agency policies (NIH, NSF, JISC…)
    - data management plans
    - Data release, data deposit
    - Clinical trials databases
Action plan: next steps

- Campus research and educational data management - i.e. data management as infrastructure
  - help for researchers to create data management plans
  - develop an institutional response to the directions of grant agencies regarding data
  - policy and standards for setting up databases
  - who owns the problem