Data Sharing Policies Social and Organizational Issues and Challenges

Paul Wouters, Anne Beaulieu, Geof Bowker, Kathleen Casey, Colin Reddy

Nerdi, NIWI-KNAW; Science Studies, UCSD
E-study on data sharing policy

- Web scan on US policies
- Websites 11 US org.
- Email survey among non-US organizations
- Funded by Min OCW and support ESF

What is the state of the art in data sharing policies?

What are the views of research organizations?
Web scan results

- Basic policy principle US:
  - Public availability and
  - accessibility of research data
- Trend towards enforcing data sharing (NIH)
- Federal laws main context:
  - Privacy Act
  - Bayh-Dole Act
  - Freedom of Information Act
  - Fair use exemption in Copyright regulation
- Political tradition of openness
Email survey results

- Data sharing policies in their infancy
- Under development in 12 out of 21 countries
- Not an issue in 9 countries

Legislation: France, Poland
Part of policy documents: Australia, Canada, Hungary, Iceland, Netherlands, Norway
In discussion: Estonia, Germany, Italy and Slovenia
## Email survey results 2

<table>
<thead>
<tr>
<th></th>
<th>Data policy developed</th>
<th>No data policy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Funding agencies / Research councils</td>
<td>Australia, Canada, Iceland, Netherlands, Norway</td>
<td>Belgium, Denmark, Germany, Italy, Spain, Slovenia, Sweden, Turkey, UK</td>
</tr>
<tr>
<td>Academies / societies</td>
<td>Hungary, Norway, Slovenia</td>
<td>Austria, Estonia, Ireland, Slovenia (Med.), Czech Republic</td>
</tr>
<tr>
<td>Research institutions</td>
<td>France, Italy</td>
<td>France</td>
</tr>
</tbody>
</table>
## Email survey results 3

<table>
<thead>
<tr>
<th>Type of problem</th>
<th>Number of responses</th>
</tr>
</thead>
<tbody>
<tr>
<td>legal problems (among others privacy)</td>
<td>9</td>
</tr>
<tr>
<td>technical problems</td>
<td>9</td>
</tr>
<tr>
<td>standards</td>
<td>8</td>
</tr>
<tr>
<td>institutional barriers</td>
<td>3</td>
</tr>
<tr>
<td>prohibitive cost</td>
<td>3</td>
</tr>
</tbody>
</table>
## Relation policy - practice?

<table>
<thead>
<tr>
<th></th>
<th>Nat. pol.: yes</th>
<th>Nat. pol.: no</th>
</tr>
</thead>
<tbody>
<tr>
<td>Org. pol.: yes</td>
<td>10</td>
<td>0</td>
</tr>
<tr>
<td>Org. pol.: no</td>
<td>3</td>
<td>14</td>
</tr>
</tbody>
</table>
Web scan results 2

- Scientific databases as “non-linear accelerators of research” (Cerf 1999)
- Paradox of digital technologies:
  - Increased possibility of sharing
  - Increased possibility of control
- Economic roles of data and research
- Shaping institutional context research as policy objective
Motivations data sharing

- data collected with public funds belong in the public domain
- obligation to scientific openness and accountability
- the obligation to abide by the law
- improve US competitiveness
Motivations for data sharing 2

- the advancement of science
- further discovery
- verification
- replication and secondary analyses
- reduce unnecessary duplication
- reduction of the need for new data collection economies of scale
- improve the productivity and cost-effectiveness
- answer research questions that cannot otherwise be addressed
- multidisciplinary research teams
- prevent neglect of fruitful avenues
- access to data for new but talented researchers
- improve training for graduate and undergraduate students
Dilemmas and limitations

- Funding schedules
- Data sharing condition for funds?
- Types of data
- Researcher main responsible
- Information skills: new support staff?
- Protection of the rights of persons and research subjects (including privacy protection);
- Protection of intellectual property rights;
- Concerns over the integrity of the research process;
- Considerations of national economic and security interests
Relation policy-practice

Viable policies depend on:

Five interdependent contexts:
- Institutional and managerial
- Budgetary
- Technology
- Policy/Legal
- Cultural

Trust and standardisation as key issues

Four case studies:
- CERN
- EBI
- fMRI
- GBIF
Policy cycle for data access

- Agenda setting: getting attention
- Formulation solution to the problem
- Adoption: support for solution
- Implementation: infrastructure/software
- Evaluation: feedback from scientists, articulate new agenda; incentives
Key issues from case studies

- CERN: production – not sharing – main issue
- GBIF and EBI: funding schemes
- EBI: Public – private domains
- GBIF: Information management
- Archiving criteria/standards
- fMRI: Data “mobility”:
  - Metadata
  - Re-use
- Incentives
Key issues from case studies

- Good Thing <-> Extra Work
- Mom and apple pie -> controversy
- Peer review -> data sharing
- Quality control of data
- Sharing and power relationships
- Tools for sharing
- Privacy (human subject data)
- Making public: WHAT does it mean?
Peer to Peer Data Sharing

- Discrete research groups
- Data location not self-evident
- Researcher is keeper/steward of the data
- Data tied to specific research project
- Trust at face to face level
Data Archives

- Centralized repository
- Data annotated and formatted (metadata)
- Focused on one field or sub-field
- Uncertain budgets due to system of research funding
- Trust at institutional level
Centralized data production

- “Big science” institutions or networks
- Close coordination of data production
- Trust part of production process
- Data sharing *not* a separate issue: data availability limited to groups involved in production
- Highly processed (interpreted) data available for public (education)
Data Sharing Configurations

Different actors:
- Peer to peer
- Data archives and repositories
- Centralized data production

Different mechanisms:
- Face to face
- Mediated by ICTs

Different data types