The OECD Follow up Group on Issues of Access to Publicly Funded Research Data: A Summary of the Interim Report

Peter Arzberger
University of California San Diego

Global Research Village IV
Importance of ICT for Research and Science: Science Policies for Economies in Transition
Warsaw 11 October 2002
Building Upon Previous Work

Scientific Access to Data and Information
Scope of Working Group Report

- **Included**: Data produced from 100% public funding, e.g., University research and data collected by governments (nonclassified, relevant to research)

- **Excluded**: Data produced from private-sector (industry); Data produced from mixed sources for funding, restricted government data

- **Minimal treatment** in report:
  - Data produced from mixed sources for funding, restricted government data
  - Issues of integrating data from developing countries
  - Impact of national security issues not been fully explored

- **Note**: Phase 1
NIH Draft Statement on Sharing Research Data

- The NIH will expect investigators supported by NIH funding to **make their research data available to the scientific community for subsequent analyses**.

- The NIH will require that **data sharing be addressed in grant applications** (e.g., in sections related to significance, budget, and the end of the research plan) and in the review of applications.

- Applicants whose research will produce data that are not amenable to sharing **should include in the application reasons for not making the data available**.

  Sharing as the default
National Science Foundation
Social Behavioral and Economic Sciences

- NSF is committed to the principle that the various forms of data collected with public funds belong in the public domain.
- Purpose of policy is to advance science by encouraging data sharing among researchers.
- Grantees … will develop and submit specific plans to share materials collected with NSF support.
  - Include how and where these materials will be stored at reasonable cost, and how access will be provided to other researcher, generally at their cost.
- Policy explicitly recognizes that many complexities arises across the range of data collection supported by SBE programs, and that unusual circumstances may require modifications or even full exemptions.
Premises

- **Publicly funded data are a public good**
  - Should be as open as possible, with as few restrictions as possible, on a non-discriminatory basis, and available freely or at the lowest possible cost
  - *Good stewardship of public knowledge*

- **Data are central to the scientific research process**
  - Data is basis of value chain of science and technology, optimum return on public investment
  - *Strong value chains of innovation*

- **Data sharing issues are international in scope**
  - ICT makes multidisciplinary and international collaborations possible
  - Key scientific and social problems are global: health, environment
  - *The creation of value from international cooperation*
Core Principle

Publicly funded research data should be openly available, subject only to compelling superseding considerations and policies

- Limited period of exclusive use by principal investigators,
- Protection of confidentiality and privacy,
- National security, and
- Respect for intellectual property rights.
Framework to Fulfill Vision of Open Access

Synthesize from Case Studies; Interdependent

• Technical:
  – Hardware, software, equipment, people. Grid, cyberinfrastructure, e-science vision. Data integration, interoperability. Quality (security, authenticity)

• Institutional and Management:
  – Funding agencies, government departments, governing boards of large activities; universities and research institutes

• Budgetary:
  – Infrastructure upgrade, sustainability of data resources

• Legal and Policy:
  – Privacy, National Security, Database protection, IPR

• Cultural and Behavioral:
  – Reward structures, incentives, career paths; community value of sharing; disciplinary traditions
Key Conclusions

- **Technical**: Broad access to data and optimum exploitation of data sharing opportunities require an appropriately designed infrastructure.

- **Institutional and Management**: Heterogeneous data require tailored data management approaches.

- **Budgetary**: Scientific data infrastructure requires continued and separate budgetary support.
  - Social Informatics research suggests mismatch in research review applied to infrastructure.
Key Conclusions (2)

- **Legal and Policy**: International and national laws and policies directly affect data access and data sharing practices

- **Cultural and Behavioral**: Reward structures and mechanisms are a necessary component for promoting data access and data sharing practices
Some Recommendations: Governments and International Organizations

- OECD promote core principles and operating principles
- OECD should promote discussion of data access at next Ministerial
- Consider data resources and collections developed from public funding as a public good investment; preferable use-policy
- Data producing agencies and government departments should avoid full cost recovery pricing policies and restrict charges for data to marginal cost recovery
Some Recommendations:
Research Funding Organizations

- Continue level of investments in technical research
- Require technical details of data sharing be part of science project
- Promote work on international technical standards
- Data should be considered core infrastructure (funding, review)
- Funding for data infrastructure should be long-term
Some Recommendations: Research Institutions and Professional and Scholarly Associations

- Data underlying articles should be made available
- Data sharing should be discussed by professional societies
  - Value of contributing to the data repositories
- Open data sharing should become default
Next Steps

- Finalize Report
  - March 2003
  - http://dataaccess.ucsd.edu
    - Input welcomed

- Other Issues for Future Reports
  - Mix Models (Public / Private)
  - Interactions with developing countries
  - National Security

- Study
  - Economics of open access in information age
Global Biodiversity Information Facility (GBIF)

• The purpose of GBIF is to make the world biodiversity data freely and universally available.
  Biodiversity (biological diversity) refers, collectively, to all the world’s species of plants, animals and microorganisms.

• GBIF aims to “create an interoperable” whole of all the different facilities that house biodiversity data.
  • GBIF will leverage existing national and regional nodes into one data access facility through efforts “to design, implement, co-ordinate, and promote the compilation, linking, standardization, digitisation, and global dissemination of the world’s biodiversity data . . .”
  • GBIF does not fund the nodes.
GBIF: Contexts

- **Technical**
  - Protocols for interoperability; quality of data

- **Institutional and Managerial**
  - Coordination [responsibility] within countries (nodes) for national institutions for funding and activities, e.g. funding to digitize data and develop interoperable databases. Done by country [flexibility].

- **Budgetary**
  - Funding levels and cycles need to be substantial enough to develop software for heterogeneous users and uses (e.g. the public)[efficiency, sustainability]]

- **Legal and Policy**
  - Issues of ownership and IPR are barriers to moving data across boundaries. Non technical.

- **Cultural and Behavioral**
  - Access to biodiversity data - benefit to the community but the value of data access needs to be reflected in reward and incentive structures; attribution, trust.
Synchrotrons

Molecules

Macromolecular Complexes, Organelles, Cells

Microscopes

Organs, Organ Systems, Organisms

Magnetic Resonance Imagers

Source: Mark Ellisman
Integrating Knowledge

Vision of e-science

Interoperability  Technical  Quality

Institutional and Management
Transparency
Responsibility
Accountability

Budgetary
Efficiency

Legal and Policy
Property
Legality

Cultural and Behavioral
Professionalism
Appropriateness
Quality

Publicly funded research data should be openly available, subject only to compelling superseding considerations and policies

Our Common Journey: A Transition Toward Sustainability
National Research Council
Acknowledgements

Peter Schroeder, Paul Uhlir, Teri Simas, Kathleen Casey, Paul Wouters, David Moorman Leif Laaksonen, Anne Beaulieu, Zara Mirmalek, Colin Reddy

- CODATA
- Netherlands' Ministry of Education, Culture and Science and the Netherlands Institute of Scientific Information Services (NIWI-KNAW)
- National Science Foundation

- http://dataaccess.ucsd.edu