# 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



18<sup>th</sup> CODATA Conference: Frontiers of Scientific and Technical Data Montreal 30 September 2002

## Access to Publicly Finance Research Global Research Village III Amsterdam 2000



- Conference Conclusions, Hans Franken
  - Highlights Issues
  - Recommends
    - OECD/CSTP explore formulating commonly agreed principles for consideration at GRV IV
    - Steering group start, expand

Starting point for this working group

### Task of Working Group:

- Produce a Report
  - Current practices and future trends of Access to and Sharing of Research Data from Public Funding
  - Effects of current practices on quality of research and progress of science
  - Principles of making data access policies on data produced from publicly funded research
- Present for consideration at GRV IV

# Swimming in Data:

Exploding ability to capture and manipulate data

• Across Scales, Disciplines, National Boundaries







• From Acquisition, Refinement, Reduction and Deposition



Pointing to opportunity of analysis and linking across these interfaces



I ne internet provides a global postal system, global shopping center, common global library, common global university... Thomas Friedman, *The Lexus and the Olive Tree* 



## **Global Grid Activities**



# 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 explore
- Note: Phase 1

## Working Group Participants

#### • Members

- Peter Arzberger, USA, Chair
- **Peter Schroeder**, *Netherlands, Vice-Chair;*
- Geoffrey Bowker, USA
- Sigrun Eckelmann, Germany
- Tim Hubbard, England
- Koji Kamitani, Japan
- Leif Laaksonen, Finland
- Gudrun Maass, OECD/CSTP
- Doug McEachern, Australia
- David Moorman, Canada
- Masamitsu Negishi, Japan
- Adminstrative Coordinator – Teri Simas, USA

- Members (cont.)
  - Paul Uhlir, USA
  - Mitsutoshi Wada, Japan
  - Andrzej Wierzbiki, Poland
  - Jan Windmuller, Denmark
- Experts
  - Anne Beaulieu, Netherlands
  - Kathleen Casey, USA
  - Colin Reddy, Netherlands
  - Paul Wouters, Netherlands
- **Observers** 
  - Tony Mayer, ESF
  - Jackie Bax, Netherlands
  - Hugo von Linstow (now at GBIF), *Denmark (obs)*
  - David Schindel, NSF (obs)

# Activities

- Web and e-mail Surveys:
  - Data Sharing Policies
- Case Studies:
  - Specifics
    - European Organization for Nuclear Research (CERN)
    - European Bioinformatics Institute (EBI)
    - fMRI Data Center (Dartmouth, New Hampshire)
    - Global Biodiversity Information Facility (GBIF) and biodiversity science
  - Committee's contacts
- Dissemination and Input:
  - Publications: D-Lib Magazine
  - Web Site: http://dataaccess.ucsd.edu
  - Presentations (CODATA, GRV IV, Society for Social Studies of Science Conference)

### Importance of ICT for Research and Science: Science Policies for Economies in Transition



- Session 4 Access to and sharing of research data from public funding (http://www.kbn.gov.pl/GRV/)
  - Background and Principles: The WhatWhyWhereWhenWhom,
     P.Schroeder
  - Data Sharing Policies: Social and Organizational Issues and Challenges. Anne Beaulieu, Geoffrey Bowker, Kathleen Casey, Colin Reddy, Paul Wouters
  - Preliminary Report on the Recommendations of the CSTP Follow Up Group on Sharing of Publicly Funded Research Data, David Moorman and Peter Arzberger

# Building Upon Previous Work





# Scientific Access to Data and Information

# **CODATA** Principles

- In November 2000, CODATA formulated six "principles for science in the internet era" to support "full and open access to data needed for research and education". These principles are:
  - science is an investment in the public interest
  - scientific advances rely on full and open access to data
- a market model for access to data is unsuitable for research and education
- publication of data is essential to scientific research and the dissemination of knowledge
- the interests of database owners must be balanced with society's need for open exchange of ideas
- legislators should take into account the impact intellectual property laws may have on research and education.

# Data Sharing Policies

#### Paul Wouters

Networked Research and Digital Information NIWI-KNAW Email paul.wouters@niwi.knaw.nl

- Reviewed US Policy; Surveyed ESF and other agencies about data sharing policies
- Findings
  - Significant proportion of OECD countries has no policies on sharing of data from publicly funded research, however most national research organizations expect that data sharing will become a policy issue in the next five years
  - US ahead of other countries in policies regarding data access and sharing
  - International and national laws and policies directly affect data access and sharing practices



# 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 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.

# **Core Principle**

Publicly funded research data should remain publicly available, subject only to compelling superseding considerations and policies

- Fair prior use by principal investigators,
- Protection of confidentiality and privacy,
- National security, and
- Respect for intellectual property rights.

## Premises

- Publicly funded data are a public good
  - Should be as open as possible and available free of charge or at the lowest possible cost (subject to restrictions)
  - 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

### Framework for Discussion

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:
  - International and national laws and policies directly affect data access and data sharing practices
- Cultural and Behavioral:
  - Reward structures, incentives, career paths; community value of sharing

# Global Biodiversity Information Facility (GBIF)

•The purpose of GBIF is to make the wor 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.

# Key Findings

- **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 Findings (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

### Desired Outcomes More details in Report

- Governments of OECD member countries and International Organizations
  - Continue discussions at Ministerial
  - Endorse core principle
- Research Funding Organizations
  - Adopt core principle, design flexibly
  - Revise review and funding mechanisms for infrastructure
- Research Institutes, Professional and Scholarly Associations
  - Value data sharing and management
  - Data sharing is the default



Publicly funded research data should remain publicly available, subject only to compelling superseding considerations and policies

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

## Acknowledgements



• CODATA

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

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



http://dataaccess.ucsd.edu



**OC**enW

