SCIENCE AND GOVERNMENT

An International Framework to Promote Access to Data

Peter Arzberger,[1]* Peter Schroeder,[2] Anne Beaulieu,[3] Geof Bowker,[1] Kathleen Casey,[1] Leif Laaksonen,[4] David Moorman,[5] Paul Uhlir,[6] Paul Wouters[3]

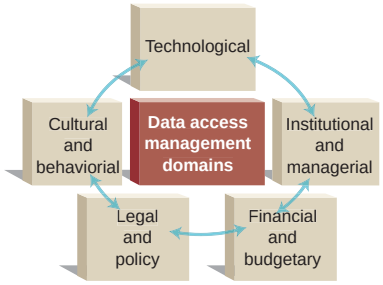
Recent national and multinational investments (*1*) in networking and continued gains in information technological capability (*2*) have given rise to a complex cyberinfrastructure that is rapidly increasing our ability to produce, manage, and use data (*3*). As research becomes increasingly global (*4*), data-intensive, and multifaceted (*5*, *6*), it is imperative to address national and international data access and sharing issues systematically in a policy arena that transcends national jurisdictions. Open access to publicly funded data provides greater returns from the public investment in research, generates wealth through downstream commercialization of outputs, and provides decision-makers with facts needed to address complex, often transnational, problems. This article summarizes key findings of an international group that studied these issues on behalf of the Organisation for Economic Cooperation and Development (OECD) (*7*), which resulted in a ministerial-level declaration (*8*).

OPERATING PRINCIPLES FOR DATA ACCESS REGIMES

- Openness
- Transparency and active data dissemination
- Assignment and assumption of formal responsibilities
- Technical and semantic interoperability of databases
- Quality control, data validation, authentication, and authorization
- Operational efficiency and flexibility
- Respect for intellectual property and other ethical and legal requirements
- Management accountability, including funding approaches

Legitimate restrictions on open access, and strong disincentives to sharing exist, based on concerns of protecting national security, privacy and confidentiality, intellectual property, and time-limited exclusive use by the scientific investigator. The lack of clear funding-agency policies in the face of strong competing interests, often far removed from academic research, poses problems for scientists in developing and developed countries and inhibit the advance of science for the public good. For example, research on cholera outbreaks and their relation to environmental factors (*9*) or on understanding global climate change (*10*) requires access to data drawn from many disciplines and sources. This issue has been a topic of recent debate and its resolution is a high priority in many scientific and policy-making communities (*11*–*17*).

Analysis of these, and other examples (*18*), suggests that successful data access and sharing arrangements exhibit a number of key attributes and operating principles (see table, this page). Administrative and organizational management "domains" (see figure, this page) provide a framework for locating and analyzing where improvements can be made. Diversity in science suggests that a variety of institutional models and tailored data management approaches will be needed.



Domains of a data access regime.

Establishing and maintaining this infrastructure requires continued and dedicated budgetary planning, with appropriate financial support. The use of research data cannot be maximized if access, management, and preservation costs (including cost of documentation and metadata creation) are an afterthought or are insufficiently or inconsistently funded in research projects (*19*). D. Atkins *et al*. (*3*) recommend that roughly one-third of the provisioning and operations of cyberinfrastructure and applications be used to support data repositories and digital libraries.

Appropriate professional and career reward structures are necessary (*20*–*22*). The way scientists are being evaluated and how their careers are shaped are at stake. For example, researchers who have spent years on building new databases, such as the Sloan Digital Sky Survey in astronomy, have effectively put their scientific careers on hold even though these databases are critical for the future development of the field. These considerations apply equally to those who produce, manage, and reuse research data.

At this point there is considerable heterogeneity in policies. In the United States, federal government databases are not copyright protected, whereas in the European Union government databases are eligible for protection under several database protection laws. Even within countries, different funding agencies have different stated policies; for example, in Canada, with three major science funding agencies, one follows the principles in the OECD declaration, one states access should not be a barrier, and a third has no policy (*23*). National laws and international agreements can directly affect data access and sharing practices.

At the last meeting of the OECD Committee for Scientific and Technological Policy (CSTP) at the ministerial level, ministers endorsed a declaration (*8*) based on the principle that research data from public funding should be openly available. Furthermore, they invited OECD to develop a set of guidelines based on commonly agreed principles (similar to those in the table) to facilitate optimal cost-effective access to digital research data from public funding. It can be expected that these future guidelines will influence national and international regulation of research data, much as the OECD Guidelines on the Protection of Privacy (*24*), which have been a model for legislation all around the Western world.

Although the involvement of researchers in resolving these issues is critical, many scientists remain ignorant about existing policies at their institutions or nations, let alone those of other countries. To better inform future policies, several topics require further examination:

[1]University of California, San Diego, La Jolla, CA 92093, USA. [2]Ministry of Education, Culture and Science, Zoetermeer, Netherlands. [3]Networked Research and Digital Information, Royal Netherlands Academy of Arts and Sciences, Amsterdam, Netherlands. [4]CSC-Scientific Computing Ltd., Espoo, Finland. [5]Social Sciences and Humanities Research Council, Ottawa, Canada. [6]National Research Council, Washington, DC 20418, USA.

*To whom correspondence should be addressed. E-mail: parzberg@ucsd.edu

1) Sharing of research data is far from a universal norm (*25*). What reward structures might lead to better access and sharing practices (*26*, *27*)?

2) Scientific disciplines differ in their needs for data reuse. Increasingly, research questions demand access to data from different disciplines, and it can be difficult to use those data sets without sufficiently descriptive and understandable metadata. To what extent should data be made amenable to use in other areas of research or for other types of applications (*19*)?

3) Several studies (*28*–*30*) and recent reports (*31*, *32*) have documented the beneficial effects of open access to public data on social and economic progress, at both the national and international levels. More empirical analyses would be very useful to help research-funding managers and policy-makers understand these issues better.

4) Various contractually based models of open community data sharing are emerging (*33*), as are international federated data management networks (*13*, *34*). More study and experience with such approaches are needed, however, before they can be broadly implemented across national boundaries and disciplines.

5) Data access is particularly important in the international context of global issues such as health, environmental change, and food production, with particular challenges for data and researchers in developing countries. A systematic examination of barriers and best practices would document the current situation and offer guidelines for further action (*35*).

6) Public-private research collaborations in which partners have different motivations for producing data pose unique challenges for data policy and practice, for example, requiring access to data to validate published research. What are key principles to guide public policy in this mixed-mode funding of research?

Making data sharing and the principle of open access the default rule rather than the exception within the scientific community will be a significant step forward. Only with a sustained national and international effort, one that includes scientists, funding agencies, and other national and international bodies in the research enterprise, will the vision of using, analyzing, and synthesizing the vast amounts of data being produced by new technologies be realized.

References and Notes

1. There are many such investments, e.g., the Global Ring Network for Advanced Application Development, GLORIAD, is a multinational effort between China, Russia, and the United States that promises greater collaboration in e-science; see www.nsf.gov/od/lpa/news/03/pr03151.htm.
2. G. Stix, *Sci. Am.* **284**, 80 (January 2001).
3. See, e.g., D. E. Atkins *et al.*, *Revolutionizing Science and Engineering Through Cyberinfrastructure: Report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure* (National Science Foundation, Arlington, VA, January 2003); available at www.cise.nsf.gov/sci/reports/toc.cfm.
4. In Canada, for example, roughly half of the large-scale research projects funded in the social sciences involve international collaborations. Without an international agreement on data sharing or access, each project must formulate its own regime and make compromises on jurisdictional differences. It is only a matter of time before serious conflict arises. Internal analysis of Social Sciences and Humanities Research Council, (SSHRC)-funded Major Collaborative Research Initiatives (Policy and Planning Division, SSHRC, 2003). See also (*19*).
5. H. Newman, M. Ellsiman, and J. Orcutt [*Commun. ACM* **46**, 68 (November 2003)] document the increased funding by agencies for large-scale projects that produce large data volumes.
6. S. H. Koslow [*Nature Rev. Neurosci.* **3**, 211 (2002)] outlines complexities of sharing data in Neuroscience.
7. Authors of this article were participants in the study, presented in Paris, 19 March 2003, DSTI/STP(2003)20, "Promoting access to public research data for scientific, economic and social development," which has not been published yet.
8. *Science, Technology, and Innovation for the 21st Century*, meeting of the OECD Committee for Science and Technological Policy at Ministerial Level, Paris, 29 to 30 January 2004, Final communiqué, Annex 1. Declaration on access to research data from public funding; available at www.oecd.org.
9. R. Colwell, "A global thirst for safe water: The case of cholera," paper presented at the National Academy of Sciences, 25 January 2002; available at www7.nationalacademies.org/wstb/2002_Wolman_Lecture.pdf. This study drew upon data from epidemiology, U.S. National Aeronautics and Space Administration remote sensing, marine biology, microbiology, genomics, and the social sciences.
10. See, e.g., Intergovernmental Panel on Climate Change; http://www.ipcc.ch/.
11. The UK Medical Research Council's new policy on Data Sharing and Preservation outlines the benefits of data sharing and drew on preliminary work of this group; www.mrc.ac.uk/index/strategy/strategy-science_strategy/strategy-strategic_implementation/strategy-data_sharing/strategy-data_sharing_policy-link.
12. See also NIH's data-sharing policy http://grants2.nih.gov/grants/policy/data_sharing/.
13. The Global Biodiversity Information Facility (www.gbif.org), an initiative of the Global Science Forum of the OECD, depends on open access to data across all countries to meet its mission of making biodiversity data available.
14. Data sharing issues in the human genome community also have been publicly and frequently discussed at www.genome.gov/page.cfm?pageID=10506376.
15. E. Marshall, *Science* **299**, 990 (2003).
16. S. Salzberg, E. Birney, S. Eddy, O. White, *Nature* **422**, 801 (2003).
17. J. Lubchenco, S. Iwata, *Science* **301**, 1443 (2003).
18. In *Promise and Practice in Data Sharing*, P. Wouters and P. Schroeder (the series editors) present four case studies, conducted by the group, which review practices and policies of the European Organization for Nuclear Research, the European Bioinformatics Institute, the Functional Magnetic Resonance Imaging Data Centre, and the Global Biodiversity Information Facility. [*The Public Domain of Digital Research Data* series, Netherlands Institute for Scientific Information Services–De Koninklijke Nederlandse Akademie van Wetenschappen (NIWI-KNAW), Amsterdam, 2003]; available at www.niwi.knaw.nl/en/nerdi2/pub/pa/pw/datasharing/toonplaatje.
19. S. M. Maurer, R. G. Firestone, C. R. Scriver, *Nature* **405**, 117 (2000).
20. A. Finkbeiner, *Science* **292**, 1472 (2001).
21. In Finland, universities are implementing reward structures for people creating digital material. For job applicants, the digital material production will be weighed with the number of publications.
22. Committee for a Pilot Study on Database Interfaces, *Finding the Forest in the Trees: The Challenge of Combining Diverse Environmental Data* (National Academy Press, Washington, DC, 1995), provides examples of incentives (pp. 97–98).
23. See the Grants Guides and program policies of the Social Sciences and Humanities Research Council (http://www.sshrc.ca), the Natural Science and Engineering Research Council, (http://www.nserc.ca), and the Canadian Institutes for Health Research (http://www.cihr.ca).
24. *OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data*; available at www.oecd.org.
25. K. McCain, *Sci. Technol. Hum. Values* **16**, 491 (1991).
26. J. Sieber, *Law Hum. Behav.* **12**, 199 (1998).
27. "Incentives for data sharing need to be offered that offset the investigators' loss of control over their databases. … Ultimately, there has to be a procedural framework that makes sharing sensible, efficient, and value-added. If all those pieces are in place, fewer external or coercive forces are needed to convince researchers to share." Minutes of the 196th National Advisory Mental Health Council Meeting, Bethesda, MD, 14 September 2000; available at www.nimh.nih.gov/council/min900.cfm.
28. P. Weiss, "Borders in cyberspace: Conflicting public sector information policies and their economic impact" (U.S. National Weather Service, February 2002); available at www.weather.gov/sp/Borders_report.pdf.
29. "Commercial exploitation of Europe's public sector information: Final report for the European Commission Directorate General for the Information Society, prepared by Pira International Ltd., University of East Anglia, and KnowledgeView Ltd., Surbiton, Surrey, UK, 20 September 2000.
30. Commission of the European Communities, "Public sector information: A key resource for Europe" [COM(98)585 European Commission, Brussels, 1999], final adopted on 20 January 1999; available at http://europa.eu.int/ISPO/docs/policy/docs/COM(98)585/.
31. J. M. Esanu, P. F. Uhlir, Eds., *The Role of Scientific and Technical Data in the Public Domain: Proceedings of a Symposium* (National Academies Press, Washington, DC, 2003), p. 226.
32. J. M. Esanu, P. F. Uhlir, Eds., *Open Access and Public Domain in Digital Data and Information for Science: Proceedings of an International Symposium* (National Academies Press, Washington, DC, in press).
33. J. H. Reichman, P. F. Uhlir, *Law Contemp. Probl.* **66**, 315 (winter-spring 2003).
34. D. Butler, *Nature* **427**, 275 (2004). The Ministerial Meeting (*8*) also endorsed the establishment of an International Neuroinformatics Coordinating Facility, the beginnings of a new neuroscience network (Annex 4).
35. Key barriers for involvement of data providers in developing countries include the basic physical infrastructure and control by governments of public information. Positive data access examples in developing countries are discussed in (*32*).

36. Participants from the United States were supported by the NSF through grant ACI-9619020; cofunded by the Office of International Science and Engineering, the Divisions of Advanced Computational Infrastructure and Research, Biological Infrastructure, and Social and Economic Sciences and the Directorate for Math and Physical Sciences. The Netherlands' Ministry of Education, Culture and Science sponsored the two studies on data access published by NIWI-KNAW. Other support was provided by the Secretariat of the international CODATA (Committee on Data for Science and Technology), Polish State Committee for Scientific Research, and the European Science Foundation. The views expressed by the authors are their own and not necessarily those of their employers or supporting agencies.