Notes: Northwest Environmental Data-network Steering Committee and PNAMP Data Management Work Group Meeting

Time: 4/5/2006 from 9:00 to 4:00

Location: Columbia River Inter-Tribal Fisheries Commission, 729 NE Oregon St, Ste 200, Portland Oregon.

Participants: Present: Bruce Schmidt (StreamNet), Tom Pansky (BPA), Peter Paquet (NPCC), Tom O'Neil (NWHI), Stewart Toshach (NED), John Piccinini (BPA), Jill Leary (Lower Columbia River Estuary Partnership), Bobby Riggers (OWEB), Jim Turner (NOAAF). For part of the afternoon discussion we were also joined by Amy Langston (CBFWF), Eric Lowrance (BPA), and Dana Collins, (BPA).

On phone: Nancy Tubbs (USGS), Cedric Cooney (StreamNet-ODFW), Jennifer Bayer (USGS-PNAMP)

Apologies: Cy Smith (OGIC), Joy Paulus (WaIAC)

A. NED AGENDA ITEMS

1) Introductions (see above) for those present.

2) Short Updates from NED Workgroup Leads.

SubBasin Planning, (Peter Paquet and Phil Roger)

Work is nearing completion on providing access to EDT data sets. Consideration is being to how to include data management planning in future updates to Sub-basin plans. Phil is completing a write-up of the challenges and problems met in completing this work. It is understood that regional decision makers are now aware of the need to attach data reporting standards to ongoing data collection efforts. A particular issue is how to develop technical capabilities to post data. Bruce Schmidt said that StreamNet could help with developing common Access databases that could be installed on Subasin planners computers for collecting information.

Upland Habitat, (Tom O'Neil)

Work is ending on wrapping up the EDT and CRITFC data and other data received from Keith Wolf. A last push is being made to collect information on the Subbasin GIS data sets and the cross-walk for Habitat Classifications.

Water Quality (John Piccininni)

John reported that the group had completed a further conference call to discuss changes to the data base inventory spreadsheet and was working to complete this product.

Spatial Temporal

The *Best Practices for Reporting Locational and Time Related Data* product is now complete and has been posted on the NED web site. Further editorial work completed by Joy Paulus and staff at the Wa IAC was noted and appreciated. The next step is for NED SC members to take the *Best Practices* to their executive with a recommendation for use within their agency or entity.

Salmonid Data Management Group (Stewart Toshach)

Stewart, working with BPA staff has completed a cross-walk between the PCSRF and PISCES data element definitions and units of measure. There are many inconsistencies in the data elements and units of measure. The next step is for agency executive and managers to decide what actions they want to take to correct these inconsistencies.

Jon Bonkoski at the State of the Salmon has been invited to use NED as a forum to help coordinate the North Pacific Hatchery Database project.

A detailed description of needed work for the Salmon Data Tracking Project has been developed and circulated within NOAAF. Once it is finalized it will be circulated through the NED SC to StreamNet partners, and others, for comments they may have on the planned work.

Stewart and Bruce plan on attending an April 10th PNAMP Effectiveness Monitoring Project Database Coordination Meeting to discuss needs for managing regional project data.

3) Planning for Regional Data Management Workshop for May 23 and 24, 2006. The most recent agenda was discussed. Most discussion concerned the emphasis of the workshop, the desired audience, and how to best get needed participants to attend. In particular, how would the workshop help to meet the needs of fish and wildlife data collectors who are critical to data collection? What would these participants get from participation? Could we be more specific about including these users' needs into the workshop agenda?

It was agreed that the workshop planning group would discuss and redraft the workshop agenda. The NED Steering Committee members were asked to provide more contact information about appropriate workshop presenters. Jennifer Bayer will be able to circulate the agenda to the participants from the CSMEP/PNAMP workshop, or alternatively, make the mailing list available.

4) Status of organizational arrangements for EPA Challenge Grant Application. Peter Paquet suggested that we now need to approach the States directly to see if they are interested in making an application on behalf of the region. The current situation is that the EPA does not accept applications from Consortia of States. The NPCC has had direct discussion with the EPA about the potential eligibility of the Council to apply and while the EPA is considering this question a decision, even if it was positive, is not likely to be timely. In 2005, in discussions with OWEB

(Ken Bierly and Doug Terra), OWEB expressed interest in helping to coordinate arrangements for a State of Oregon application for EPA Challenge Grant Funding. Bobbi Riggers to follow up with Ken to see if the State of Oregon still has an interest and if so what the next steps would be.

5) Development of NED proposal for regional information management. Peter and Stewart reported that no action had been taken on this item during the last month. Work would move ahead this month – recognizing that the upcoming May workshop could also be used to gather some input.

6) Award of EPA Challenge Grant to National Congress of American Indians (NCAI) and relationship to NED and other NW regional efforts. Phil Roger provided some information on this grant but was not yet certain, how it would apply to the region. He will provide further information if it becomes available. CRITFC has an existing effort, supported with EPA Challenge Grant funding, to establish node sharing of toxicology and other water quality data and would like to apply node sharing to the EDT River Reach Archiving Project. A second proposal, from CRITFC, Umatilla and Nez Perce Tribes has been submitted and is intended to improve water quality data sharing in general.

Other topics:

July SC Meeting

Currently the July NED meeting is scheduled for July 5th, one day after July 4th. We briefly discussed whether we wanted to stick with this date or schedule an alternative – further from the Independence Day holiday. We need to make this decision at our next meeting in May 3rd.

NED Web-site

Currently NED meeting agenda, meeting notes and discussion documents and products are being routinely posted to http://www.nwcouncil.org/ned/. Most of the material is organized in relation to NED meetings. For pending meetings you will be able to view the agenda and any materials relating to the meeting. For past meetings you will also be able to view the meeting notes, once they are posted. We have been working with Eric Schrepel at the NPCC who has been providing web support for NED. If you have ideas about how you would like the information to be presented please let us know.

Promoting Access to Public Research Paper

Stewart recommended that the NED SC members look at the following Paper: Promoting Access to Public Research Data for Scientific, Economic, and Social Development by *P Arzberger,1* P Schroeder,2 A Beaulieu,3 G Bowker,4 K Casey,1 L Laaksonen,5 D Moorman,6 P Uhlir,7 P Wouters3* (Copy attached). The paper provides an overview of many of the data management issues and policies that the NED SC is challenged with.

B) PNAMP DATA MANAGEMENT AGENDA ITEMS

1) Report on PNAMP/CSMEP Workshop. There was consistent emphasis from key speakers at the workshop to improve data management to help support regional monitoring efforts: "We need better methods for communicating and sharing data"; "We need more effective focus on meeting data needs"; "We need data sharing"; "Improve Transparency of information – "Information is power"; "We need to share better – we don't share well"; and, "We need a well thought out data management information system".

2) PNAMP Inventory Task**:** Update and issues.

Bruce reported that Mike Banach had been working with Jennifer Bayer to narrow down the list of data that would be collected under the project. A further meeting with Russell Scranton and others was planned and once the data collection needs are finalized the data collection would begin. It appears that the data will not be collected in time to meet BPA's interest in using the data to help to inform decisions for 2007-09 mainstem funding.

C) AFTERNOON SESSION 1:00 through 4:00

1) Discussion of Data Partnership Agreement Material. Stewart briefed the SC on different types of data sharing, data exchange, portal, sharing directives and other types of agreements related to data management. There was agreement that NED would first work on agreement materials for the NED Portal and that shorter and simpler versions should be used amongst competing alternatives.

2) Discussion of Regional Data Stewardship roles. Discussion focused on the Geospatial One-Stop Channel Steward Handbook at pages 1-8 and definitions in Appendix C. The purpose of was to introduce the task of developing a description of the needed Data Steward Role for NED. It was agreed that the Channel Steward Handbook contained suitable material for development of data stewardship roles for the NED portal and represented a good starting point.

3) Discussion of CBFWA "Status of the Resource" data needs with Amy Langston, and potential collaboration with NED Portal. Amy provided a one page description of the CBFWA Status of the Resource Project and some mocked-up content. The goals are to

- "Develop an interface to display species status information as it related to identified biological objectives (subbasin plan, management plans, recovery plans) at the subbasin level", and;
- Add limiting factors and all-agency activity information at the subbasin level.

Much of the discussion concerned the likely source of data and where needed data would be sourced from in order to complete the report.

One concern, since it could involve a duplication of other efforts was that one of the proposed steps involved the development and population of a database. There was discussion about how much of the data needed for the report be extracted directly from existing databases. Amy said she had hoped to be able to source a lot of the information through the NED Portal. Tom Pansky said that the goal of the NED Portal was to make it easier for data users to locate data, (by competing queries on available metadata) including the types of data of interest to CBFWA, however the Portal had not yet been populated with all the needed metadata records and this would take some time. Amy expected that would be a need for further discussion. Amy said that the CBFWA effort also planned to identify existing data sources that are not already a part of existing databases, and that are not currently available on line. There was interest, when those data resources are identified, in helping to support access to them and referencing their existence through the NED portal.

Promoting Access to Public Research Data for Scientific, Economic, and Social Development

P Arzberger,[1]* P Schroeder,[2] A Beaulieu,[3] G Bowker,[4] K Casey,[1] L Laaksonen,[5] D Moorman,[6] P Uhlir,[7] P Wouters[3]

*[*1]University of California San Diego, 9500 Gilman Dr., San Diego, California, USA, Email: parzberg@ucsd.edu, kcasey@ucsd.edu*
[2]Ministry of Education, Culture and Science, The Netherlands, Email: p.schroeder@minocw.nl
[3]Networked Research and Digital Information (Nerdi), The Netherlands, Email: Anne.Beaulieu@niwi.knaw.nl, Paul.Wouters@niwi.knaw.nl
[4]Santa Clara University, Santa Clara, California, USA, Email: gbowker@scu.edu
[5]CSC - Scientific Computing Ltd, Finland, Email: Leif.Laaksonen@csc.fi
[6]Social Sciences and Humanities Research Council, Canada, Email: DAVID.MOORMAN@SSHRC.CA
[7]U.S. National Academy of Sciences/National Research Council, USA, Email: puhlir@nas.edu

ABSTRACT

Access to and sharing of data are essential for the conduct and advancement of science. This article argues that publicly funded research data should be openly available to the maximum extent possible. To seize upon advancements of cyberinfrastructure and the explosion of data in a range of scientific disciplines, this access to and sharing of publicly funded data must be advanced within an international framework, beyond technological solutions. The authors, members of an OECD Follow-up Group, present their research findings, based closely ontheir report to OECD, on key issues in data access, as well as operating principles and management aspects necessary to successful data access regimes.

Keywords: Data access, Science policy, Data sharing, Data management, Database, Archives, Scientific infrastructure, Global e-science, OECD, Public domain

1 INTRODUCTION AND SUMMARY

It is now commonplace to say that information and communications technologies are rapidly transforming the world of research. We are only beginning to recognize, however, that management of the scientific enterprise must adapt if we, as a society, are to take full advantage of the knowledge and understanding generated by researchers. One of the most important areas of information and communication technology (ICT)-driven change is the emergence of e-science, briefly defined as increased access, via desktop or other interface via the Internet, to distributed resources, global collaboration, and the intellectual, historical, analytical, and investigative output of a range of scientific communities (Atkins, Droegemeier, Feldman, Garcia-Molina, Klein, Messerschmitt, et al., 2003; Research Councils UK, n.d.).

In recent years, the debate on e-science has tended to focus on the "open access" to the digital *output* of scientific research, namely, the results of research published by researchers as the articles in the scientific journals (Access all Areas, 2004; for recent discussions on open access see Cook (2004), Suber (2004) and House of Commons Science and Technology Committee (2004)). This focus on publications often overshadows the issues of access to the *input* of research - the research data, the raw material at the heart of the scientific process and the object of significant annual public investments. In terms of access, availability of research data generally poses more serious problems than access to publications.

Ensuring research data are easily accessible, so that they can be used as often and as widely as possible, is a matter of sound stewardship of public resources. Moreover, as research becomes increasingly global, there is a growing need to systematically address data access and sharing issues beyond national jurisdictions. The goals of this article and its recommendations are to ensure that both researchers and the public receive optimum returns on the public investments in research, and to build on the value chain of investments in research and its data resource.

To some extent, research data are shared today, often quite extensively within established networks, using both the latest technology and innovative management techniques. The Follow Up Group, which is identified in the Acknowledgments section of this paper, drew on the experiences of several of these networks to examine the roles and responsibilities of governments as they relate to data produced from publicly funded research. The objective was to seek good practices that can be used by national governments, international bodies, and scientists in other areas of research. In doing so, the Group developed an analytical framework for determining where further improvements can be made in the national and international organization, management, and regulation of research data (Arzberger, Schroeder, Beaulieu, Bowker, Casey, Laaksonen, et al., 2004).

The findings and recommendations presented here are based on the central principle that ***publicly funded research data should be openly available to the maximum extent possible.*** Availability should be restricted only by legitimate considerations of national security restrictions; protection of confidentiality and privacy; intellectual property rights; and time-limited exclusive use by principal investigators. Publicly funded research data are a public good, produced in the public interest. As such they should remain in the public realm. This does not preclude the subsequent commercialization of research results in patents and copyrights, or of the data themselves in databases, but it does mean that a copy of the data must be maintained and made openly accessible. Implicitly or explicitly, this principle is recognized by many of the world's leading scientific institutions, organizations, and agencies. Expanding the adoption of this principle to national and international stages will enable researchers, empower citizens and convey tremendous scientific, economic, and social benefits.

Evidence from the case studies and from other investigations undertaken for this report suggest that successful research data access and sharing arrangements, or regimes, share a number of key attributes and operating principles. These bring effective organization and management to the distribution and exchange of data. The key attributes include: openness; transparency of access and active dissemination; the assignment and assumption of formal responsibilities; interoperability; quality control; operational efficiency and flexibility; respect for private intellectual property and other ethical and legal matters; accountability; and professionalism. Whether they are discipline-specific or issue oriented, national or international, the regimes that adhere to these operating principles reap the greatest returns from the use of research data.

There are five broad groups of issues that stand out in any examination of research data access and sharing regimes. The Follow Up Group used these as an analytical framework for examining the case studies that informed this report, and in doing so, came to several broad conclusions:

- Technological issues: Broad access to research data, and their optimum exploitation, requires appropriately designed technological infrastructure, broad international agreement on interoperability, and effective data quality controls;
- Institutional and managerial issues: While the core open access principle applies to all science communities, the diversity of the scientific enterprise suggests that a variety of institutional models and tailored data management approaches are most effective in meeting the needs of researchers;
- Financial and budgetary issues: Scientific data infrastructure requires continued, and dedicated, budgetary planning and appropriate financial support. The use of research data cannot be maximized if access, management, and preservation costs are an add-on or after-thought in research projects;
- Legal and policy issues: National laws and international agreements directly affect data access and sharing practices, despite the fact that they are often adopted without due consideration of the impact on the sharing of publicly funded research data;
- Cultural and behavioural issues: Appropriate reward structures are a necessary component for promoting data access and sharing practices. These apply to those who produce and those who manage research data.

The case studies and other research conducted for this report suggest that concrete, beneficial actions can be taken by the different actors involved in making possible access to, and sharing of, publicly funded research data. This includes the Organization for Economic Cooperation and Development (OECD) as an international organization with credibility and stature in the science policy area. At the March 2003 meeting of the OEDC Committee of Science and Technology Policy, the Follow Up Group recommended that the OECD consider the following:

- Put the issues of data access and sharing on the agenda of the next Ministerial meeting (see Section 7, and Declaration on access to public research data from public funding (2004));
- In conjunction with relevant member country research organizations,

- Conduct or coordinate a study to survey national laws and policies that affect data access and sharing practices;
- Conduct or coordinate a study to compile model licensing agreements and templates for access to and sharing of publicly funded data;

With the rapid advances in scientific communications made possible by recent developments in ICTs, there are many aspects of research data access and sharing that have not been addressed sufficiently by this report, would benefit from further study, and will need further clarification. Accordingly, further possible actions that could be considered include:

- Governments from OECD expand their policy frameworks of research data access and sharing to include data produced from a mixture of public and private funds;
- OECD consider examinations of research data access and sharing to include issues of interacting with developing countries; and
- OECD promote further research, including a comprehensive economic analysis of existing data access regimes, at both the national and research project or program levels.

National governments have a crucial role to play in promoting and supporting data accessibility since they provide the necessary resources, establish overall polices for data management, regulate matters such as the protection of confidentiality and privacy, and determine restrictions based on national security. Most importantly, national governments are responsible for major research support and funding organizations, and it is here that many of the managerial aspects of data sharing need to be addressed. Drawing on good practices worldwide, the Follow Up Group suggests that national governments should consider the following:

- Adopt and effectively implement the principle that data produced from publicly funded research should be openly available to the maximum extent possible;
- Encourage their research funding agencies and major data producing departments to work together to find ways to enhance access to statistical data, such as census materials and surveys;
- Adopt free access or marginal cost pricing policies for the dissemination of research-useful data produced by government departments and agencies;
- Analyze, assess, and monitor policies, programs, and management practices related to data access and sharing policies within their national research and research funding organizations.

The widespread national, international, and cross-disciplinary sharing of research data is no longer a technological impossibility. Technology itself, however, will not fulfill the promise of e-science. Information and communication technologies provide the physical infrastructure. It is up to national governments, international agencies, research institutions, and scientists themselves to ensure that the institutional, financial and economic, legal, and cultural and behavioural aspects of data sharing are taken into account.

2 BACKGROUND TO THE STUDY

At its March 2001 meeting, the OECD Committee on Scientific and Technology Policy (CSTP) accepted a proposal from The Netherlands to establish a working group on issues of access to research information. The plans of the working group were presented at the October 2001 CSTP meeting. Subsequently, the Committee narrowed the scope of activities to access to and sharing of research data produced from public funding. Participation in the group was broadened to include Australia, Canada, Denmark, Finland, Germany, Japan, Poland, the Netherlands, the United Kingdom, and the United States. The CSTP asked the working group to:

- Report on current practices concerning access to and sharing of research data and their underlying principles on the basis of case studies;
- Report on the effects of selected current data sharing practices on the quality of research and the progress of science;
- Suggest principles for making policy on data sharing within the relevant national and international policies and regulatory frameworks.

The report's core principle is that ***publicly funded research data should be openly available to the maximum extent possible***. Adoption of this principle will promote good stewardship of public knowledge, develop strong value chains of innovation, and maximize benefits from international cooperation (see Box 1). The report's findings and recommendations are addressed to: CSTP members as representatives from the governments of OECD member countries; and professional and scholarly associations. The objective is to contribute to a better understanding of the importance of research data access and sharing, and to offer suggestions on how the new digital challenges should be met.

Building on a number of case studies and a great deal of other research, the report focuses on issues related to the access and sharing of publicly funded research data, in digital form, across all disciplines in the natural, health, and social sciences (Wouters & Schröder, 2003). Attention is paid to the international aspects of access and sharing relevant to scientific cooperation among OECD member states. Three significant topical areas fell outside the charge of this working group, however, and will require separate follow-up: issues particular to developing countries; issues related to data produced by a mixture of public and private funding; and the issue of national security restrictions in light of recent global events since 11 September 2001 (on issues of national security and data access, see Mervis & Stokstad (2002)).

Box 1: This core principle guides many public scientific institutions and scientists. However, it remains unevenly implemented. Most recently, it was adopted by the United Kingdom's Medical Research Council. After a workshop hosted by the European Science Foundation, the MRC drafted the following statement: MRC promotes the creation of a diverse range of datasets, many of which are rich in informational content, unique and cannot be readily replicated. Sharing allows scientists to extend the value of these datasets through new, high quality, ethical research and exploitation. It also reduces unnecessary duplication of data collection. Building preservation systematically into routine data management is part of good research practice: it strengthens quality, enables replication and audit, and provides a sound basis for data sharing Medical Research Council (n.d.).

In this report, we define "access to data" as the act of making the data available for use by others; by "sharing" we mean a researcher allowing one or more other individuals to use data, typically with the implicit, if not explicit assumption that it is on a reciprocal basis. The sharing of data involves providing specific access, whereas the act of providing access by itself does not necessarily involve any sharing arrangement. In this article, data sharing focuses on data exchanges between individuals or groups of researchers rather than institutions, while access may be provided at any level. Sharing in our view also reflects the cooperative norms of public science as practiced within many disciplines by many researchers in OECD countries. The U.S. National Institutes of Health on the other hand, use the term "data sharing" throughout all of their formal, institutional regulations on the use of research data generated by NIH funding (NIH Office of Extra Mural Research, n.d.). We define data as in the U.S. National Institutes of Health definition of final research data: "the recorded factual material commonly accepted in the scientific community as necessary to validate research findings". In the OECD documents on the subject, "data are defined as the factual records (numerical scores, textual records, images and sounds) used as sources and base material for scientific research." For other definitions of terms involved with data, see Westbrook & Grattidge (1991), National Research Council (1997) and Esanu & Uhlir (2004).

3 KEY ISSUES IN DATA ACCESS AND DATA SHARING

3.1 The changing information technology context for scientific research and innovation

Information and communication technologies (ICTs) are enabling the rapid transformation of an increasing number of research areas as well as the broader society: witness the growth in the number of Internet hosts per person, in the percentage of computers per household, and in the continued rate of growth of chip, storage, and network technology capacity (Stix, 2001). Concurrently, there has been an explosion in the amount of data produced across all types of scientific endeavour. Examples of an this explosive increase in data production range from genetic sequence and protein structure data in bioinformatics, to various types of brain imagery in neuroscience, to sky surveys and virtual observatories in astronomy, and geospatial data such as Global Spatial Data Infrastructure.

Continuing ICT advances, such as the development of grid computing, large-capacity optical transmission networks, wireless networks of sensors and devices, and complex imaging systems, promise to push these transformations farther and faster. ICT-dependent research, such as geographic information systems, data visualisation systems, and realistic modelling, are adding tremendously to our ability to study and understand the world in which we live. These developments provide researchers in OECD countries, and increasingly in developing countries, with the opportunity not only to be more efficient, more effective and better connected, but also to dramatically expand the scope and nature of their investigations. This expansion of scope of scientific investigation results from activities such as combining data from multiple data sources to gain a greater statistical power to resolve hypotheses (for example, see Biomedical Informatics Research Network (n.d.)) and obtaining real-time global measurement on environmental observations. Together they create the possibility of an "e-science infrastructure." The growing activities in data collection, storage, processing, distribution, and preservation are, however, only loosely connected. They require systematic planning to realize the full potential of the emerging e-science infrastructure.

3.2 The benefits of data access and sharing in public research

Within this new technological context, more widespread and efficient access to and sharing of research data can be expected to have substantial benefits for public scientific research (see Box 2). Open access to, and sharing of, data reinforces open scientific inquiry, encourages diversity of analysis and opinion, promotes new research, makes possible the testing of new or alternative hypotheses and methods of analysis, supports studies on data collection methods and measurement, facilitates the education of new researchers, enables the exploration of topics not envisioned by the initial investigators, and permits the creation of new data sets when data from multiple sources are combined. Sharing and open access to publicly funded research data not only helps to maximize the research potential of new digital technologies and networks, but provides greater returns from the public investment in research (Fienberg, Martin, & Staf, 1985; National Research Council, 1999).

BOX 2: Access to international data has helped produce a better understanding of public health issues and worldwide disease prevention and control. For instance, research on cholera outbreaks and their relationship to numerous environmental factors relied upon data drawn from epidemiology, NASA remote sensing, marine biology, microbiology, genomic data, and social science data. This research—an example of 'biocomplexity' studies supported by the U.S. National Science Foundation—would have been impossible without access to numerous databases. The effect of this interdisciplinary and international research project is an increased scientific and sociological understanding of cholera outbreaks and their prevention (Colwell, 2002).

Moreover, improving and expanding the open availability of public research data will generate wealth through the downstream commercialisation of outputs, provide decision-makers with the necessary facts to address complex, often trans-national problems, and offer individuals the opportunity to better understand the social and physical world in which we all live. For example, a recent analysis demonstrated the economic benefits of providing open access to government meteorological data without any restrictions on re-use (Weiss, 2003; Weiss, 2002; European Union Green Paper, 1998; PIRA International, 2000). The "value adding" meteorological information industry in the United States has revenues in excess of $500M annually. The public meteorological data also support a rapidly growing weather risk management industry that underwrites financial risk management instruments valued at approximately $8B. In contrast, the private-sector value adding industry for meteorological information in the European Union is very small, largely attributable to the highly restrictive data policies of most national governmental meteorological services. What are harder to measure, but certainly occur, are the lost opportunities for researchers, students, and various other potential public users who find the high costs of the E.U.'s public data to be too great to access and use.

As a key link in the value chain of investments in research, open access to factual data plays an increasingly important role in all these areas.

3.3 Roles and responsibilities of governments

If researchers throughout the world are to take full advantage of ICTs to improve and expand access to, and sharing of, research data, existing technological, institutional and managerial, financial and budgetary, legal and policy, and

BOX 3: Poor stewardship and lost opportunity costs for data access is exemplified by the case of Statistics Canada, which attempted to recover costs for its data management by charging data users. The effect of this form of management of these public data was a dramatic decrease in their use. In a study of the case, it was found that "Cost recovery was supposed to introduce a market type discipline on the demand for and supply of goods and services provided by the government. Since in economic terms Statistics Canada's outputs are public goods, the type of discipline envisioned by this policy is impossible to attain. Instead we have users who complain, refuse to pay and generally attempt to find alternative sources for their information needs. This policy fails the improved management of resources test (McMahon, 1996)"

cultural and behavioural aspects must be addressed comprehensively and in an integrated way. To date, these aspects have often been treated on an ad hoc, project-specific basis. Given that OECD countries spend tens of billions of dollars each year collecting data that can be used for research and for other social and economic benefits, ensuring that these data are easily accessible so that they can be used as often and as widely as possible, is a matter of sound stewardship of public resources (see Box 3).

Scientists, research institutions, and research funding agencies around the world are increasingly engaging in large-scale, data-intensive projects. Such projects require data-management infrastructure, data-exchange protocols and policy frameworks, and a broad professional understanding that more extensive availability and use of the data is both necessary and desirable. Over the past decade, numerous studies, disciplines, research programs, and agencies have begun to address the complexities and benefits of open data access and sharing arrangements (National Research Council, 1997; Medical Research Council, n.d.). As scientists become better connected with each other, particularly through the Internet, and as research focuses on issues of global importance, such as climate change, human health and biodiversity, there is growing need to systematically address data access and sharing issues beyond national jurisdictions and thereby create greater value from international co-operation. The goal should be to ensure that both researchers and the broader public receive the optimum return on public investments, and to build on the value chain of investments in research and research data (Stiglitz, Orszag & Orszag, 2000).

4 CORE PRINCIPLE AND PREMISES

The findings and recommendations that follow are based on the central principle that ***publicly funded research data should be openly available to the maximum extent possible***.

As a general principle, publicly funded research data should be as open as possible and available at the lowest possible access cost, subject only to legitimate restriction and considerations. Restrictions may be necessary for reasons of national security, for the protection of privacy of citizens, or the confidentiality of trade secrets. Access to and use of research data may be limited by the respect for private intellectual property rights. Finally, there may be reasons for granting periods of temporary exclusive use to those who collected the data. But the guiding principle should be openness.

In order to derive the maximum benefit from public investments in research data, access, use, management and preservation must be an integral part of the research process. Conversely, data should not be considered an expendable by-product of research. In many cases, data have value beyond the project and anticipated use for which they were originally collected. The re-use of publicly funded data for research and other types of applications should be promoted and not restricted.

The accessing and sharing of data is not merely a technical matter, but also a complex social process in which researchers have to balance different pressures and interests. Purely regulatory approaches to data sharing are not likely to be successful without consideration of these factors. Various approaches to data access and sharing are therefore necessary, including the establishment of regulations and incentives, and the dissemination of best practices.

The following three premises complement and support the core principle of this report:

4.1 Data from publicly funded research are a public good produced in the public interest

Both the data from publicly funded research and research itself have strong public good characteristics, as elaborated by Kaul, Grunberg & Stern (1999), that support their open availability to the public, and especially to other researchers.

4.2 Factual data are central to the scientific research process

The production, open dissemination, and unfettered use of factual data are essential attributes of, and inputs to, modern systems of scientific research and technological innovation. Recognizing the role of digital data as fundamental to the value chain of science, technology, and innovation will enable an optimum return on public investments.

4.3 Data access and sharing issues are international in scope

To more fully exploit the possibilities of global digital networks, and to capture their benefits for the global community, policy issues concerning access to and sharing of publicly funded scientific research data must be addressed, not only at the institutional and national levels, but also at the international level.

5 DATA ACCESS OPERATING PRINCIPLES AND ATTRIBUTES

Data access and sharing requires effective organization and management. The necessary components that make up this organization and management may be characterized as "data access regimes." In their ideal form, these regimes enable all participants in the scientific research process to freely and efficiently access and share data. Adequate data access regimes may require both distributed and centralised responsibilities across different management domains that include the technological, institutional and managerial, financial and budgetary, legal and policy, and cultural and behavioural.

Although no single approach to developing an effective data access regime is possible, a list of operating principles for and attributes of effective data access regimes and resources can be offered. This list of attributes and operating principles is based on a broad set of experiences, and supported by the case studies conducted for this article. The operating principles evolved out of recommendations developed by Franken (2001). Key attributes are listed below, and illustrated with an example from the case studies.

5.1 More explicit access regimes

There is a universal requirement for the formalisation of institutional rules and data management policies. The need for this formalisation follows from the growing complexity and scale of scientific research and the increasing expenditure on research data. At the moment, it is often not clear who is authorised to distribute data across the globe. To reach the necessary transparency in the tasks and responsibilities of those involved, terms of access to and use of data that rest on tacit agreements should be made explicit and formalised. A systematic and institutionalized approach is needed to help address operating characteristics of data access and to take advantage of the opportunities arising from publicly funded research.

5.2 Operating Principles

5.2.1 Openness

Open availability of publicly funded research data to the maximum extent possible is the core principle..

5.2.2 Transparency of access and active dissemination

Open data access requires actively disseminating where the data can be found, what the context and structure of the data collection is (metadata), how long the resource will be accessible, and what protocols and standards are employed. In short, this principle refers to the systematic visibility and traceability of data resources.

5.2.3 Assignment and assumption of formal responsibility

Formal responsibility for tasks associated with data access must be assumed by the appropriate participants in the global science system. The various individuals and institutions involved in the chain of data-related activities all have specific manifest and latent duties and obligations. These are founded in formal legal and professional normative standards and in the regulations of various agencies. Responsibility must also be assumed for various rights in the data supply, such as authorship, producer credits, ownership, financial arrangements, licensing terms, and, where appropriate, restrictions on use.

5.2.4 Professionalism

Codes of conduct, and related normative standards, of professional scientists and their communities can help to promote good practice and simplify the regulatory aspect of access regimes.

5.2.5 Interoperability

Technical and software standards and protocols are required to ensure the access and usability of data. These should be clear to the user and adopted by as many data management organizations as possible.

5.2.6 Quality

Quality refers to the proper description of uncertainties surrounding the production of the data (e.g., the techniques employed in their collection and archiving, and the measuring instruments and their calibration), the ability to ensure that the cited source and value are ***authentic***, that the data retain ***integrity*** (complete and absent from introduced errors), and that they are ***secure*** against loss, destruction, modification, and unauthorized access.

5.2.7 Operational Efficiency

Open access to data increases the efficiency of research by avoiding unnecessary duplication of data collection and permitting the creation of new data sets by combining data from multiple sources. Coupled with open access, comprehensive documentation of data sets and how to access them provides a more efficient use of resources.

5.2.8 Flexibility

In general, scientific communities will approach data management requirements more consistently within their discipline internationally, than they will across other disciplines on a national level. Data access regimes need to be sufficiently flexible to take account of this variation.

5.2.9 Property

Institutional intellectual property rights as well as the individual rights of researchers are considerations of property interests. Unlike the private sector, public research operates on a principle of collective property interests, which are promoted by the open access and sharing of data resources.

5.2.10 Legality

Legal restrictions may limit access to and use of data. Examples of legal restrictions involve national security, privacy, and trade secrets. Restrictions will apply primarily to 'secondary' data sets compiled for purposes other than scientific research. In some cases, the sensitive parts of data sets can be left out without rendering them useless. Specific types of legal restrictions include: national security, privacy and the protection of trade secrets.

5.2.11 Accountability

Accountability involves measuring the cost, benefit, and performance of data access and sharing regimes and taking appropriate actions in response to the results.

5.3 Building a Data Access Regime: the Global Biodiversity Information Facility (GBIF)

The Global Biodiversity Information Facility (GBIF), which began under the auspices of the OECD Megascience Forum, has sought to implement these principles as a means to achieve the larger goal of providing worldwide access to biodiversity data. GBIF's goal is to make "the world's scientific biodiversity data freely available to all [**openness**]"(Global Biodiversity Information Facility, n.d.). The fundamental motivation for GBIF is to enable access to a vast amount of biodiversity data housed in databases distributed in numerous countries and institutions. By bringing all these data into one interoperable network, and producing a registry of biodiversity information resources, GBIF will produce systematic visibility and traceability of data resources [**transparency**].

Formal responsibilities of different participants involved in the task of building GBIF's organisation and legal relationships have been established in GBIF's Memorandum of Understanding. GBIF's Secretariat is responsible for carrying out work programmes that are approved by the Governing Board, which consists of representatives of GBIF's Participants. This structure enables GBIF to have a legal identity as an international body, and to manage financial contributions and work programmes, while drawing upon the additional separate efforts and resources of Participants. The establishment of GBIF's activities occurred through contact with existing scientific and political bodies to maintain and establish professional codes of conduct, gain consensus about scientific outcomes, and negotiate with government representatives about GBIF's larger social and economic roles [**professionalism**].

Participants will provide stable gateways, or "nodes," to databases that contain primary or meta-level biodiversity data. These nodes must provide documentation and metadata about the data in the databases, vouch for data **quality**, ensure data authenticity and security. GBIF will help develop standards for database **interoperability** through one of its 4 work programmes, Data Access and Database Interoperability (DADI). GBIF aims to develop an interoperable network of distributed databases by coordinating and leveraging existing national and international programs and projects, which allows for **operational efficiency** and more cost-effective basis for making biodiversity data freely and easily available to a heterogeneous user community.

The databases and the data accessed through GBIF are in most cases owned and developed by other organisations and thus will not entail any assertion of IPRs by GBIF itself [**property**]. GBIF intends to provide best practices on how to deal with IPRs, particularly since it will be drawing from databases hosted by different institutions and countries with different legal frameworks, with a view to promoting open access and sharing to the maximum extent possible. GBIF also asserts in its MOU that biodiversity data will be properly used and acknowledged by its participants **[legality].** Further, its efforts are consistent with the Global Taxonomic Initiative of the Convention on Biological Diversity concerning the proper and equitable use of biodiversity data and the resources to which they refer.

During the establishment of GBIF, the OECD provided the forum to assess the level of support for this new scientific collaboration, to bring together related proposals and to develop detailed plans that could then be taken up by interested countries. According to paragraph 11.2 in GBIF's MOU, in the third year of its initial five-year period of existence, "an independent review of its operations, financial mechanisms, legal basis, governance structure, and links to other organizations will be conducted to determine if any changes are needed. The lessons learned will be used to evaluate the effectiveness of the governance structure and to recommend any necessary changes" [**accountability**]. That review is currently being conducted.

6 DATA ACCESS MANAGEMENT: FIVE DOMAINS

Efficient data access can only take place with the proper administration and organization of different management domains within data access regimes. These domains include technological, institutional and managerial, financial

and budgetary, legal and policy, and cultural and behavioural considerations (see Figure 1). The domains provide a framework for locating and analyzing where improvements to data access and sharing can be made.

The five domains differ in character across the traditions and practices of specific scientific disciplines, e.g., astrophysics, biology. Thus, data access regimes may vary in significant ways. There is no single model for how data access should take place. The implementation of the core principle of open availability, however, requires a systematic approach that recognizes the necessity of implementing improvements across the interdependent management domains. This approach also requires the involvement of actors from various levels: governments, funding agencies, research institutions and professional societies, as well as individual scientists themselves.

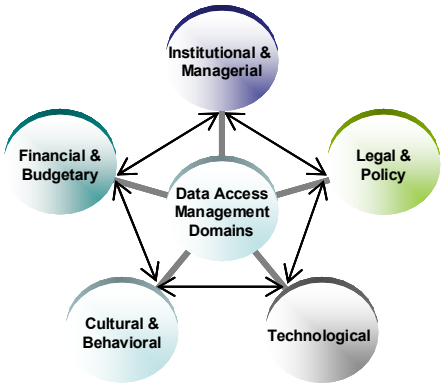

Figure 1. Components of a Data Access Regime

6.1 Technological domain

Broad access to research data, and their optimum exploitation, requires appropriately designed technological infrastructure, broad international agreement on interoperability, and effective data quality controls.
A technical infrastructure that supports user needs is necessary to derive maximum benefits from data access and sharing. This infrastructure must be robust enough for long-term use and, when appropriate, for diverse uses. It also must be flexible enough to respond to the continuous and rapid changes in scientific research and technology. While there are many technical issues to be resolved to take full advantage of past, current, and future investments in ICT infrastructure, the main barriers to effective data access and sharing are no longer technical, but are institutional and managerial, financial and budgetary, legal and policy, and cultural and behavioural.

6.1.1 Data Preparation and Metadata: ICPSR

In 1995, the Inter-University Consortium for Political and Social Research (ICPSR) initiated the development of the Data Documentation Initiative (DDI), an international criterion and methodology for the content, presentation, transport, and preservation of metadata about datasets in the social and behavioural sciences. DDI, which is in XML format, helps enhance users' ability to acquire and use data while it assists producers in packaging and disseminating them. After a period of beta-testing with participating international organisations, DDI is now in use by a number of organisations, including Networked Social Science Tools and Resources (NESSTAR), Health Canada, and ICPSR. ICPSR continues to assist data producers in preparing their data through its "Guide to Social Science Data Preparation and Archiving," a guide with broad appeal for individuals and organisations searching for easy and effective ways to technically manage and prepare data so that they can be easily and effectively placed into network environments (For more information on ICPSR and DDI, see Data Documentation Initiative (n.d.) and ICPSR (n.d.). For information on the importance and development of DDI, see Norwegian Social Science Data Services (1999)).

Technical operating principles for data access regimes include interoperability (of protocol and software to ensure the access and usability and multiple use of the data); and quality (including technical components of authenticity, integrity, and security) of data.

6.2 Institutional and managerial domain

While the core open access principle applies to all science communities, the diversity of the scientific enterprise suggests that a variety of institutional models and tailored data management approaches are most effective in meeting the needs of researchers.

Because scientific data have many different characteristics and uses, there is no monolithic institutional and management approach that can be applied universally. Key characteristics of data production and use include whether the data are (1) government-generated or generated at a research institution using public funds; (2) useful

only within the discipline or across many disciplines; (3) useful over the very long term or only within short-term horizons; (4) have public-policy implications; or, (5) have significant broader economic and social value, among other factors.

Institutional and managerial operating principles for data access regimes include transparency (systematic visibility of the data source); responsibility (explicit formal institutional rules on data management); and accountability (rendering public account for the performance of data access regimes).

6.2.1 Negotiated collaborations: CERN

The European Organisation for Nuclear Research, CERN, is one of the world's largest scientific laboratories, presently financed by twenty European countries. CERN overtly subscribes to the core principle and premises outlined in this report. However, the raw experimental data set does in itself not make much sense outside of the context of the specific experiment. The sheer size also necessitates heavy processing. Experiments at CERN are typically run by large-scale collaborations. Within each collaboration access to the data is unproblematic. In this stage the data are protected, however, partly because of technical issues (size and interpretation) and partly because of the competition between researchers. Availability of this type of data to other researchers depends on negotations. At a higher level of interpretation CERN puts its data in the public domain. The cleaned up and interpreted data are made available to the international physics community in the form of a Data Summary Table. The type of data produced and the method of processing used will therefore play a large part in deciding upon the most effective management model to adopt. This flexibility of management approach is a key factor in the data production and sharing environment at CERN.

6.3 Financial and budgetary domain

*Scientific data infrastructure requires continued, and dedicated, budgetary planning and appropriate financial support. The use of research data cannot be maximized if access, management, and preservation costs are an add-on or after-thought in research project*s.

In many areas of public research, there are indications of discrepancies between the funding of the specific research itself and the related data-management requirements (which do not necessarily benefit the individual scientist, but which are necessary for data re-use). Generally, research organizations fund the former well, but pay scant attention to the latter. In the digital environment, scientific data sets must be viewed as a key element of the broader research infrastructure and as an investment in the future capacity to innovate and solve pressing problems. Adequate support is essential for data-management functions, such as the development of sufficient explanatory documentation (i.e., metadata) for each data set, conversion of old formats onto new media, adaptation to new standards, and long-term preservation, archiving, and maintenance.

Budgetary operating principles for data access regimes include operational efficiency (maximizing the return on investment by promoting re-use of data, and providing proper documentation, specialists, and effective data management facilities).

6.3.1 Funding schemes “on a rolling basis:” the European Bioinformatics Institute (EBI)

The official mission of the EBI is to ensure that the growing body of information from molecular biology and genomic research is placed in the public domain and is freely accessible to the scientific community in ways that promote scientific progress. Like other scientific bodies, the EBI has a major problem in the funding for its building, maintaining and making available databases and information services even though they represent only a small fraction of the total research costs. The key issue is that funding for data sharing infrastructures needs to be constructed “on a rolling” or on-going basis to maintain effective data management. These funding requirements are very different from the funding schedules of research, which are usually project oriented. These differences in budgeting constitute the main threat to the EBI’s commitment to maintaining the public availability of its data.

6.4 Legal and policy domain

National laws and international agreements directly affect data access and sharing practices, despite the fact that they are often adopted without due consideration of the impact on the sharing of publicly funded research data.
Intellectual property laws, information policies, institutional guidelines, and contracts at the national and international levels often impose terms and conditions on data access and sharing practices. Laws and policies governing data access and sharing practices may vary among different countries, resulting in barriers to scientific cooperation and progress. Based on a recent Web survey (Wouters, 2002), most of the national research organization managers who responded expected that data sharing will become a major policy issue in the next five years. This situation requires greater attention by the science policy community at all levels. In particular, restrictions on re-use of public data by the research community must be eliminated or minimised as much as possible. Research grant provisions and licensing templates for promoting open access and unrestricted re-use of public research data already exist, but have not yet been broadly adopted.

Legal and policy operating principles for data access regimes include property (balance intellectual property rights of investigator and institution versus public good); and legality (lawful data management, respecting national security, privacy and trade secrets).

6.4.1 Policy interconnections: functional MRI and the Institutional Review Boards

The functional Magnetic Resonance Imaging Data Center's (fMRIDC) principal endeavour is to promote data sharing in brain mapping. The Western tradition of informed consent in bio-medicine operates according to the principle that the 'most specific consent is the best consent.' When data are to be gathered for submission to databases, the specificity of consent may run counter to the goals of meta-analysis or re-analysis by third parties, to investigate issues different from those for which the data was originally gathered. The creation of infrastructures for data sharing, therefore, has to conform to the rules of regulatory bodies, such as institutional review boards (IRBs), whose approval must be obtained to share data. As such, these bodies function as gatekeepers to the circulation of data. International coordination may also be necessary. Researchers submitting or requesting data across national boundaries may find it especially difficult to act in accordance with the various ethical guidelines that exist in different countries. The fMRIDC has been hesitant to accept data from non-US settings because of concerns regarding IRB compliance.

6.5 Cultural and behavioural domain

Appropriate reward structures are a necessary component for promoting data access and sharing practices. These apply to both those who produce and those who manage research data.

Although formal policy frameworks and regulations are necessary to make research data publicly available, they need to be supplemented by appropriate community-based norms and incentives for researchers to share and provide access to their data and for appropriate recognition of their data-related work. In many cases, there is a general lack of reward structures and mechanisms to promote open access to, and sharing of, data from public research.

Cultural and behavioural operating principles for data access regimes include quality (trust that data are what they purport to be); professionalism (build on codes of conduct and ethics of the scientific community); flexibility (there is no single model on how data access must be provided.)

6.5.1 Incentives: the Protein Data Bank

To publish in scientific journals, U.S. scientists involved in the field of crystallography must deposit their data in the Protein Data Bank (PDB) and acquire an accession number. As PDB Director Helen Berman explains, "By requiring everyone to submit data, the community is assured of having the most up to date information possible. Now, increasingly, under our regime, a lot of [data] depositors have come to realize that the practice that we use has some advantages for them in that we check things and we find errors and inconsistencies. That actually improves the quality of the product they produce."

7 POSSIBLE FOLLOW-UP STEPS FOR INTER-GOVERNMENTAL AND GOVERNMENTAL BODIES

Our findings from the case studies and from other research indicate a number of action areas by the different actors involved in making possible open access to, and sharing of, publicly funded research data. In this section we recommend possible action areas for the OECD and national governments.

7.1 OECD

As an international organization with credibility and stature in the science policy arena, the OECD, through the CSTP, can play a crucial role in promoting access to, and sharing of, data from publicly funded research. Central to this role is the gathering and sharing of information on successful practices in data related activities and policies. At the international level, only a few organizations have undertaken to do this, usually in the context of a specific discipline or research program. The recent, and vast, expansion of research data assets and the trend towards issue-based, interdisciplinary research, however, suggests that all countries and all fields of science stand to benefit from greater attention and an organized and coordinated approach to effective policy actions.

In its report to OECD, the Follow-up Group concluded:

The OECD should put the issues of data access and sharing on the agenda of the next Ministerial meeting. ICT advances have created the ability to transform science. New tools allow researchers to find data in seconds that would have taken months just a few years ago. Effective data access and sharing requires a comprehensive policy approach for implementation by public research institutions. Monitoring progress and devoting attention to the public research data issues and activities would assist decision-makers and research support agencies in developing appropriate policies and allocating resources.

This recommendation was made in March 2003. At the meeting of the OECD Committee for Scientific and Technological Policy (CSTP) at Ministerial Level on 30 January 2004 ministers responsible for national science and technology policies of OECD countries endorsed the *Declaration on Access to Research Data from Public Funding* (2004). In the Declaration, CSTP was invited to formulate OECD guidelines for *Access to Research Data from Public Funding*. A CSTP Working Group has been installed to draft these international guidelines.

7.2 National Governments

Although the OECD, UNESCO, ICSU, CODATA, and other international bodies can play a role in improving the current situation regarding research data access and sharing, it is at the national level that many important decisions and actions must be taken. National governments provide the resources for making data accessible, establish the overall policies for data management, regulate matters such as confidentiality and privacy, and determine restrictions based on national security. Most important, it is national governments that are responsible for the major research support and funding organizations, and it is here that many of the managerial aspects of data sharing need to be addressed.

The national governments of OECD countries should consider:

1. **Adopting, and effectively implementing, the principle that data produced from publicly funded research should be openly available to the maximum extent possible.** The public investments made in research data collection can only be maximized if the data are preserved, managed, and made accessible. This requires coordinated attention by governments at all levels, and adequate policy and financial support. The starting point for these actions, however, is the affirmation that data collected using public funds should be openly accessible to all.
2. **Encouraging their research funding agencies and major data producing departments to work together to find ways to enhance access to statistical data, such as census materials and surveys.** Many countries have taken steps to facilitate access to census and survey materials by developing catalogues, user-friendly repositories, off-site research facilities, training programs, and regulatory frameworks for providing appropriately guarded access to confidential information. Such steps have proven

enormously effective in maximizing the use of national surveys and producing insights into the functions of economies and societies.

3. **Adopting free access, or marginal cost pricing, policies for the dissemination of research-useful data produced by government departments and agencies.** The use of information collected through public funding should be freely accessible for research purposes. This maximizes the use of such information for public policy and public knowledge development.
4. **Analyzing, assessing and monitoring policies, programs, and management practices related to data access and sharing policies within their national research and research funding organizations.** This information would be useful to national governments so that they may assess the implementation of the previous three considerations. The resources, support programs, policies, and regulations related to research data sharing are, in large part, developed and implemented by research funding organizations. The operations of these organizations play a crucial role in determining the degree to which data are made accessible and shared between researchers. Many organizations, such as NSF and NIH in the United States, the Social Sciences and Humanities Research Council in Canada, and the European Science Foundation are now developing, or have developed, policies, regulations and support programs that promote data sharing. Issues such as establishing protocols for the collection and release of confidential information, developing technical infrastructure, agreeing on metadata standards, requiring data preservation strategies within individual research projects, and including data management costs as eligible expenditures in grant applications have been dealt with by one or more of these agencies. It would benefit the global scientific community if decision-makers within national governments had a clear understanding of where their respective agencies stood in relation to those in other countries.

7.3 Areas for Further Examination

The OECD/CSTP Working Group currently engaged in drafting international guidelines will consider the other recommendations to OECD from the report of the Follow-up Group (OECD Follow-up Group on Issues of Access to Publicly Funded research Data, n.d.). The recommendations concern the following activities:

1. **Consider conducting or coordinating a study to survey national laws and policies that affect data access and sharing practices.** This relatively simple undertaking could determine what policies exist, how accessible they are, and result in listing of the web sites where these policies are posted. This study would be of considerable benefit to science policy-makers, research administrators, and information resource managers in all countries, both within OECD and beyond. The study could look at the feasibility of developing a central and easily accessible repository of national laws and policies that affect data access and sharing practices. Such a compilation does not currently exist, and could be useful to facilitate international research collaborations (for a preliminary survey, see Wouters, 2002).
2. **Consider conducting or coordinating a study to compile model licensing agreements and templates for access to and sharing of publicly funded data.** Depending on the context, numerous factors need to be considered in data access and sharing arrangements. Nevertheless, many contractual models already exist that have been developed by research funding organisations, research program managers, university administrators, librarians, and others. The OECD, as a global organization, is ideally suited to span national domains where examples do exist, and thereby bring an international perspective. The study could compile and review existing agreements and models to find exemplary approaches. Having readily available models on hand would be of considerable benefit to researchers, universities, and research institutions, as well as data centers and archives, and could facilitate international research collaboration.
3. **Governments from OECD countries should consider expanding their policy framework of research data access and sharing to include data produced from a mixture of public and private funds.** Collaborative public/private research projects, and the resulting data, have their own unique set of characteristics and issues. As more national governments promote public-private partnerships in research, these issues will be of increasing importance to both public researchers and the companies that are involved. A further examination of the state of data sharing and access in these types of research arrangements needs to be made to develop sound science policy guidance.
4. **Consider examinations of research data access and sharing to include issues of interacting with developing countries.** The increase of participation in the research enterprise benefits the global science system and innovation. Providing developing countries with access to data from publicly funded research increases their participation in science. Further, as United Nations Education, Scientific and Cultural

Organization (UNESCO), the International Council of Scientific Unions (ICSU), private foundations, and other organizations have emphasized, access to scientific knowledge by developing countries is vital to the progress of the entire world. This access is particularly important in the context of global issues such as population health, environmental change, and food production. Of course, open access to data from publicly funded research in developed countries can provide a valuable resource for economic development, education, and scientific capacity building. Many efforts are already underway to improve access for researchers in developing countries (e.g., providing free or below-costs access to data and scientific information), as well as establishing optimal data regimes for developing countries to share their data (e.g., addressing issues of data repatriation). A systematic examination of barriers and best practices would provide both a picture of the current situation and a set of guidelines for further action.

5. **Consider promoting further research, including a comprehensive economic analysis of existing data access regimes, at both the national and research project or program levels.** To date, no one has yet undertaken a comprehensive, economic analysis of different data access regimes. Several key issues have not been closely examined, including the relative costs of providing data openly or not, the impact of cost recovery on the use of those data, and the positive externalities and network effects from providing open access to publicly funded research data. The OECD should consider conducting this type of analysis or encouraging member country research organizations to fund such studies.

8 CONCLUSION

Improving access to and sharing of publicly funded research data is an issue that touches on all aspects of the research enterprise and the development of knowledge, and involves all participants in the conduct of research. For the individual researcher, the sharing of data, particularly prior to publication, can be burdensome, time consuming, and unrewarding if the necessary measures are not taken to provide funding, facilities, and a social context that emphasises its value to the research community and to society (Sacrifice for the greater good?, 2003).

Advances in ICTs, the internationalisation of science, and the trend toward issue-based research hold great potential for the advancement of knowledge and for the benefit of all people. This potential will not be fully realized unless all of the major elements of data access regimes identified in this report are properly developed. To do so will take considerable discussion, understanding, and commitment on the part of all those involved in research, particularly at the policy and funding levels.

Agreement among OECD governments on a set of general principles to shape specific data access regimes, as well as adoption of the recommendations set forth above, would be enabling for scientists, empowering for citizens, and provide an important contribution to fulfill the promises of e-science.

9 ACKNOWLEDGMENTS

This article is a slightly updated version of the report of the same name (DSTI/STP(2003)20 dated 19 March 2003 and submitted to the Committee for Science, Technology and Industry (CSTP) of the Organization for Economic Cooperation and Development (OECD) at its March 2003 meeting by the *OECD Follow-up Group on Issues of Access to Research Data from Public Funding to the OECD*.

The Follow-up Group was established shortly after the March 2001 OECD Committee on Science and Technological Policy (CSTP) meeting, and delivered its report to the CSTP in March 2003. The March 2003 OECD Report Appendix of case studies, not included here, has been expanded and published elsewhere (Wouters & Schröder, 2003). The term "Follow-up" refers to recommendations that came from the Global Research Village III Conference on Access to Publicly Financed Research (December 2000, Amsterdam), to be followed-up by the fourth Global Research Village Meeting held in 2002 in Warsaw.

The recommendations of the report led to the establishment of a CSTP Drafting Group that prepared the Declaration on Access to Research Data from Public Funding (2004) for the ministerial CSTP meeting and subsequently (in 2004) to the establishment of the CSTP Working Group that is currently preparing the OECD guidelines on the subject. The issues raised by the recommendations of the report were addressed by the Drafting Group and will be further explored by the current Working group.

On the basis of this March 2003 Report (DSTI/STP(2003)20), a CSTP Drafting Group prepared the documents that led to the *Declaration on Access to Research Data from Public Funding* endorsed by ministers at the meeting of the OECD Committee for Science and Technological Policy at Ministerial Level, Paris, 29 to 30 January 2004, Final Communiqué of *Science, Technology, and Innovation for the 21st Century,* Annex 1, on 30 January 2004 . In the Declaration, OECD Ministers invited CSTP to formulate OECD guidelines for *Access to Research Data from Public Funding.* A CSTP Working Group has been installed to draft these international guidelines.

The authors would like to acknowledge participants in the members of the OECD Follow-up Group Sigrun Eckelmann, Deutsche Forschungsgemeinschaft (DFG) – Germany; Tim Hubbard, Wellcome Trust Sanger Institute – UK; Koji Kamitani, MEXT (Ministry of Education, Culture, Sports, Science & Technology) – Japan; Doug McEachern, Australian Research Council – Australia; Masamitsu Negishi, National Institute of Informatics – Japan; Andrzej P. Wierzbicki, National Institute of Telecommunications – Poland; Jan Windmueller, Ministry of Science, Technology and Innovation – Denmark. The following people contributed their expertise to the project: Jacky Bax, Ministry of Education, Culture and Science - The Netherlands; Kathleen Cass, CODATA - Paris, France; Gudrun Maas, OECD, Directorate for Science, Technology and Industry Science and Technology Policy Division – France; Tony Mayer, European Science Foundation – France; David Schindel, National Science Foundation – France; Hugo von Linstow, Global Biodiversity Information Facility – Denmark; Colin Reddy, Networked Research and Digital Information (Nerdi) - The Netherlands; and, Teri Simas, University of California, San Diego – USA.

Participants from the United States were supported by the NSF through grant ACI-9619020; co-funded by the Office of International Science and Engineering, the Divisions of Shared Cyberinfrastructure, 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.

10 REFERENCES

Access all areas (2004) *The Economist.* 372 (8387), 73.

Arzberger, P., Schroeder, P., Beaulieu, A., Bowker, G., Casey, K., Laaksonen, L., Moorman, D., Uhlir, P., & Wouters, P. (2004) An international framework to promote access to data. *Science* 303, 1777-78.

Atkins, D. E., Droegemeier, K.K., Feldman, S.I., Garcia-Molina, H., Klein, M.L.,Messerschmitt, D.G., Messina, P., Ostriker, J.P., & Wright, M.H. (2003) *Revolutionizing Science and Engineering through Cyberinfrastructure: Report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure.* Arlington, VA: National Science Foundation.

Biomedical Informatics Research Network (n.d.) Homepage of Biomedical Informatics Research Network. Available from: http://www.nbirn.net

Colwell, R. (2002) A Global Thirst for Safe Water: The Case of Cholera. *Abel Wolman Lecture at the National Academy of Sciences*, Washington, D.C. Retrieved February 2, 2003 from the World Wide Web: http://www7.nationalacademies.org/wstb/2002_Wolman_Lecture.pdf

Cook, A. (2004) Open access to the US government urged. *The Scientist* (July 21). Retrieved July 21, 2004 from the World Wide Web: http://www.biomedcentral.com/news/20040721/01/

Data Documentation Initiative (n.d.) About the Data Documentation Initiative Alliance. Retrieved from the *Data Documentation Initiative* website: http://www.icpsr.umich.edu/DDI/ORG/index.html

Declaration on access to public research data from public funding (2004) Annex I to Final Communique of OECD Committee for Science and Technological Policy at Ministerial Level: *Science, Technology, and Innovation for the 21st Century,* Paris, 29 to 30 January 2004, Final Communiqué, (www.oecd.org).

Esanu, J. M., Uhlir, P. F. (Eds.) (2004) *Open Access and Public Domain in Digital Data and Information for Science: Proceedings of an International Symposium,* Washington, D.C.: National Academies Press.

European Union Green Paper (1998) Public Sector Information: A Key Resource for Europe, COM (98) 585. Retrieved March 3, 2002 from the World Wide Web: europa.eu.int/ISPO/docs/policy/docs/COM(98)585/

Fienberg, S. E., Martin, M. E., & Straf, M. L. (Eds) (1985) *Sharing Research Data*, Washington, D.C.: National Academy Press.

Franken, H. (2001) Conference Conclusions. In Schröder, P. (Ed.), *Access to Publicly Financed Research,* Amsterdam, NIWI, 75-82. Retrieved August 10, 2004 from the Web: http://www.esf.org/sciencepolicy/172/report.pdf

Global Biodiversity Information Facility (n.d.) Homepage of the Global Biodiversity Information Facility. Available from: http://www.gbif.net

House of Commons Science and Technology Committee (2004) Scientific Publications: Free for all? 10th report of session 2003-2004. Retrieved January 26, 2004 from the World Wide Web: http://www.publications.parliament.uk/pa/cm/cmsctech.htm

ICPSR (n.d.) Guide to social science data preparation and archiving. Retrieved from the *ICPSR* website: www.icpsr.umich.edu/ACCESS/dpm.html

Kaul, I., Grunberg, I., & Stern, M. (Eds.) (1999) Defining Global Public Goods. *Global Public Goods: International Cooperation in the 21st Century*, New York: Oxford University Press.

Medical Research Council (n.d.) Draft MRC statement on data sharing and preservation policy. Retrieved July 19, 2004 from the *Medical Research Council* website: http://www.mrc.ac.uk/index/strategy/strategy-science_strategy/strategy-strategic_implementation/strategy-data_sharing/strategy-data_sharing_policy-link>

McMahon, R.C. (1996) Cost Recovery and Statistics Canada. *Government Information in Canada* 2(4). Retrieved February 10, 2003 from the World Wide Web: http://www.usask.ca/library/gic/v2n4/mcmahon/mcmahon.html

Mervis, J. & Stokstad, E. (2002) NASA censors report on agriculture threats. *Science* 297, 1973-75.

NIH Office of Extra Mural Research (n.d.) NIH Draft Statement on Sharing Research Data. Retrieved form the *NIH* website: http://grants2.nih.gov/grants/policy/data_sharing/

National Research Council (1997) *Bits of Power: Issues in Global Access to Scientific Data*, Washington, D.C.: National Academy Press.

National Research Council (1999) *A Question of Balance: Private Rights and the Public Interest in Scientific and Technical Databases*, Washington, D.C.: National Academy Press.

Norwegian Social Science Data Services (1999) Providing Global Access to Distributed Data through Metadata Standardisation -- The Parallel Stories of NESSATAR and DDI, submitted by the Norwegian Social Science Data Services to the Conference of European Statisticians, UN/ECE Work Session on Statistical Metadata, Geneva, Switzerland, 22-24 September 1999. Retrieved from the World Wide Web: http://www.nesstar.org/papers/GlobalAccess.html.

OECD Committee for Science and Technological Policy (2004) *Science, Technology, and Innovation for the 21st Century. Meeting of the OECD Committee for Science and Technological Policy at Ministerial Level, 29 - 30 January 2004 - Final Communiqué, Annex 1*. Retrieved July 19, 2004 from the OECD website: http://www.oecd.org/document/15/0,2340,en_2649_34487_25998799_1_1_1_1,00.html.

OECD Follow-up Group on Issues of Access to Publicly Funded research Data (n.d.) The public domain of digital research data. Retrieved July 19, 2004 from the World Wide Web: http://dataaccess.ucsd.edu

Pira International (2000) Commercial exploitation of Europe's public sector information: Final report for the European Commission Directorate General for the Information Society, UK, 30 October 2000. Retrieved from the web January 20, 2004 from the World Wide Web: http://europa.eu.int/information_society/topics/multi/psi/docs/pdfs/commercial_exploitation/commercial_final_report.pdf

Research Councils UK (n.d.) About the UK e-Science programme. Retrieved July 19, 2004 from the World Wide Web: www.research-councils.ac.uk/escience/

Sacrifice for the greater good? (2003) *Nature* 421, 875.

Stiglitz, J., Orszag, P. & Orszag, J. (2000) *The Role of Government in a Digital Age*. Retrieved March 2, 2002 from the World Wide Web: http://www.sbgo.com/Papers/CCIA%20FINAL.pdf.

Stix, G. (2001) Triumph of Light. *Scientific American* January, 80-85.

Suber, P. (2004) UK House Committee releases its report on open access. *Open Access News*. Retrieved July 19, 2004 from: http://www.earlham.edu/%7Epeters/fos/2004_07_18_fosblogarchive.html#a109028453862279542

Weiss, P. (2002) *Borders in Cyberspace: Conflicting Public Sector Information Policies and their Economic Impact*, Washington, D.C.: U.S. National Weather Service. Retrieved February, 5 2003 from the World Wide Web: www.weather.gov/sp/Borders_report.pdf.

Weiss, P. (2003) Conflicting International Public Sector Information Policies and their Effects on the Public. In Esanu, J. M. & Uhlir, P. F. (Eds.), *The Role of Scientific and Technical Data in the Public Domain: Proceedings of a Symposium,* Washington, D.C.: National Academies Press, 29-132.

Westbrook, J. H. & Grattidge, W. (1991) A Glossary of Terms Related to Data, Data Capture, Data Manipulation, and Databases. *CODATA* 23 (1, 2).

Wouters, P. (2002) *Policies on Digital Research Data: an International Survey*. Amsterdam, Netherlands Institute for Scientific Information Services, NIWI-KNAW.

Wouters, P. & Schröder, P. (Eds.) (2003) *Promise and practice in data sharing*. Amsterdam, Netherlands Institute for Scientific Information Services, NIWI-KNAW.