ISSUES IN THE TRANSBORDER FLOW OF SCIENTIFIC DATA

Paul F. Uhlir (PUHLIR @ NAS.EDU) and Shelton S. Alexander (SHEL @ GEOSC.PSU.EDU)

The US National Committee for CODATA is conducting an interdisciplinary study of international access to scientific data. The primary focus is on data in electronic forms, a topic of increasing complexity and importance in scientific research and international collaboration. The study is characterizing the technical, legal, economic and policy issues that have an impact -- whether favorable or negative -- on access to data by the scientific community. Special attention is being given to the specific conditions inherent in the transborder transfers of electronic scientific data among the academic, governmental and private sectors. The study is also identifying and describing those barriers that have the most adverse impact in each of the discipline areas within CODATA's purview -- the physical, astronomical, biological and geological sciences -- and across those disciplines, using representative examples. Finally, it is attempting to identify trends that are likely to have significant discipline-specific and interdisciplinary impacts on the use of scientific data, particularly in electronic forms, and will suggest approaches that could help overcome both generic and specific barriers to access in the international context.

1 Introduction

The Committee on Data for Science and Technology (CODATA) is an interdisciplinary committee organized under the International Council of Scientific Unions, a nongovernmental organization created in 1931 to promote international scientific activity in the different branches of science and their applications to humanity. According to CODATA's charter, the committee is concerned with all types of quantitative data resulting from experimental measurements or observations in the physical, biological, geological, and astronomical sciences. CODATA's general objectives include the improvement of the quality and accessibility of data, as well as the methods by which data are acquired, managed, and analyzed; the facilitation of international cooperation among those collecting, organizing, and using data; and the promotion of an increased awareness in the scientific and technical community of the importance of these activities.

The U.S. National Committee for CODATA (USNC/CODATA) is a standing committee organized under the National Research Council (NRC). The Council is the principal advisory body to the federal government on scientific and technical matters. The USNC/CODATA serves as a bridge between the scientific and technical community in the United States and the international CODATA regarding data issues in the natural sciences. Consistent with the objectives of CODATA and the role of the National Research Council, the USNC/CODATA has established a subcommittee, chaired by R. Stephen Berry of the University of Chicago, to undertake an interdisciplinary study of international access to scientific data. This paper reviews the background, scope, and methodology of this study, and solicits input from all interested parties.[1]

2 Background to the Study

Scientists commonly encounter difficulties in gaining access to data relevant to their research because of both technical and non-technical barriers. The issues related to adequate access to the mounting volumes of data in all scientific disciplines--particularly data in electronic forms--have been a topic of considerable concern in recent years. The integration of multidisciplinary data on an international basis to address problems such as global environmental degradation or disease epidemics raises new and even more challenging problems in this regard.

One National Research Council study, *Sharing Research Data*, published by the Committee on National Statistics in 1985, provided a comprehensive analysis of the issues related to broadening the access to social science research data. Many of the conclusions and recommendations set forth in that report are equally relevant in the context of natural science research, although this was not expressly addressed. The focus of the 1985 study, however, was primarily on the sharing of data within the U.S., rather than on an international basis.

A seminal work on Scientific and Technical (S&T) data access and dissemination in the international context, *Study on the Problems of Accessibility and Dissemination of Data for Science and Technology*, was published by a CODATA task group in 1975 (CODATA Bulletin 16). This report provided an excellent overview of the problems associated with the transborder flows of S&T data at that time. However, the focus of that study was

[1] Note: portions of this paper are also being published in the forthcoming "Proceedings of the Fourteenth International CODATA Conference," J. Dubois and N. Gershon, eds., CODATA, Paris.

mainly on analog data in hard copy formats rather than on the large and growing volumes of digitized data in electronic form that now support scientific research activities. Therefore, the 1975 study, while still relevant in some respects, is substantially outdated because it did not expressly address the problems inherent in the transborder flow of electronic S&T data.

A more recent attempt to examine the barriers to data access at the international level was done by CODATA in conjunction with the International Council for Scientific and Technical Information (ICSTI) through an informal survey in 1990. The survey solicited comments from approximately 70 producers, distributors, and users of scientific databases on barriers to data access in the following categories:

1. Restrictions on transmission of scientific data/information across national boundaries.

2. Impediments resulting from pricing policies and differing national practices regarding subsidies for database development.

3. Special problems of academic scientists who face high prices for data that have high commercial importance as well as basic research interest.

4. Barriers that might result from efforts of database owners to protect their intellectual property from unauthorized redistribution or other illegal practices.

5. The particular problems of developing countries.

The survey identified several barriers, but did not provide details on the nature and extent of these problems. However, from this study and other experiences within the scientific community it is evident that access to scientific data continues to be restricted by technical, legal, economic, and policy constraints on both a national and international basis. The constantly increasing use of computers and telecommunications networks in the creation, maintenance, and dissemination of scientific data significantly changes the context in which these constraints on effective access apply.

These problems also are exacerbated by the emergence of a global economy where competitiveness among technologically advanced nations leads to increased emphasis on the real and perceived value of scientific data and intellectual property to national interests. Scientists in developing countries commonly have additional or different concerns than their colleagues in the wealthier nations, leading to asymmetries in research relationships and related transfers of data. Apparent solutions to problems in one context may not be appropriate or applicable in another.

During the five years that have elapsed since the CODATA/ICSTI survey, the USNC/CODATA has held several seminars in conjunction with its regular meetings to gather background information on these issues from government policy experts and knowledgeable individuals in the international scientific community. A session on the "Social, Political and Legal Aspects of Databases" was also held at the October 1992 International CODATA Conference in Beijing. These preliminary investigations have made it clear to the USNC/CODATA that a broad study on the transborder flow of scientific data, with a primary focus on data in electronic forms, could help bring the problems and challenges in this area into sharper focus. In addition, the committee has been encouraged to do this study by the agencies that provide its core support.

3 Scope and Methodology of the Study

The USNC/CODATA study now in progress has undertaken a study of technical, legal, economic, and policy issues that impact the transborder flow of scientific data, with a special emphasis on data stored and transmitted in electronic formats (e.g., on the Internet). Because of the study's interdisciplinary focus, the standing committee has augmented its existing expertise in the natural sciences with experts in the legal, economic, and policy aspects of international transfers of scientific data.

The study will be performed by two panels with sub-panels organized according to a matrix of disciplines (physical, astronomical, biological, and geological sciences) and issues (technical, legal, economic, and policy). Each discipline area will be represented by 2-3 experts (including the panel and sub-panel chairs). The panels of 11-12 individuals will meet separately in sub-panels as well as together. The same individuals will constitute both panels but in different groupings on sub-panels. This matrix approach will enable the discipline panel to take a cross-cutting issue perspective, and the issue panel to take a multidisciplinary perspective.

The study will be performed under the following Terms of Reference:

* Outline the needs for data in the major research areas of current scientific interest that fall within the scope of CODATA--the physical, astronomical, geological, and biological sciences.

* Characterize the legal, economic, policy, and technical factors and trends that have an influence--

whether favorable or negative--on access to data by the scientific community.

* Identify and analyze the barriers to international access to scientific data that may be expected to have the most adverse impact in discipline areas within CODATA's purview, with emphasis on factors common to all the disciplines.

* Recommend to the sponsors of the study approaches that could help overcome barriers to access in the international context.

The panels will hold three meetings. Following the panel's initial meeting, input will be solicited from: other units at the National Research Council, including the Committee on Geophysical and Environmental Data and the various U.S. National Committees to ICSU, with significant involvement in these issues; other CODATA National Committees and Task Groups; other ICSU Scientific Unions, Associates, and Interdisciplinary Bodies; national, foreign, and international professional societies in the relevant disciplines; government agencies and intergovernmental bodies involved in scientific research; and other individuals, groups, and institutions. This major fact-finding activity will be supplemented by extensive research, as well as by follow-up interviews. The study members also will be encouraged to take advantage of opportunities to obtain additional inputs at international scientific conferences. The following questionnaire will be used to elicit specific information from these various sources:

1. Barriers to Data Access. Some restrictions on access to scientific data frequently are considered necessary to protect various interests as well as the integrity of the data. In your experience, have restrictions on data been a problem? Can you identify any specific impacts or trends? Please explain.

2. Pricing of Data. If you use data for scientific research, please tell us:

(a) What data sets you have recently used for which you or your institution paid nothing, and in what form did you get these data (e.g., paper, film, tape, diskette, CD-ROM, on-line, etc.)?

(b) What data have you recently used for which you paid any amount (including the cost of reproduction or communication connectivity); in what form did you get these data, how were you charged (e.g., flat rate, charge per use, etc.), and how much?

(c) What data would you like to use for your research, but consider them too expensive/costly? What is the cost of such data and what is their value (apart from cost)?

(d) For the data listed under (c) above, what arrangements could help make these data available to you? In what form would you like to be able to get these data?

If you supply data for scientific research (and perhaps for other uses), please tell us:

(e) Are you a profit-making enterprise; if not, what is the form and intent of your organization?

(f) What kind of data do you supply that are used by scientific researchers?

(g) Besides scientific researchers, what kind of other users of your data are there, if any?

(h) Do you provide special pricing for research/academic users? If so, what is your pricing policy?

(i) What are the media you use to distribute your data (e.g., paper, film, tapes, diskettes, CD-ROMs, on-line, etc.)?

(j) If you sell or otherwise market your data, what is your perception of the price elasticity and demand for the data you distribute. What changes would you make to your data products and services if demand were to increase?

3. Protection of Intellectual Property. (a) What are the principal legal and technical mechanisms for protecting unauthorized uses of data in your country/institution/discipline area?

(b) Can you provide any information about how such legal or technical mechanisms are implemented or enforced? What are the positive and negative impacts?

4. Less Developed Countries. (a) In your experience, what have been the principal problems associated with transferring data into or out of "less developed countries," including those nations from the former Soviet Union?

(b) What can be done to help alleviate these problems, especially by the international scientific community?

5. Electronic Networks. (a) Has the development and growth of the Internet and other electronic networking services affected the way you access or distribute data internationally? Please give specific examples if you can.

(b) How do you think the situation with electronic networks will change in the next ten years or so, and what are the likely impacts to your activities?

6. Other Technical Issues. (a) Besides those associated with electronic networks, what are the most important technical benefits or problems you

have experienced in either disseminating or accessing data internationally?

(b) What changes do you anticipate over the next ten years, and what are the likely impacts to your activities?

7. Scientific Data for Global Problems. (a) In your view, what is the role of international scientific data for addressing global problems, now and in the future? Please elaborate.

(b) What can be done to enhance the availability or exchange of scientific data to better address these concerns?

8. Other Issues. Do you have any specific concerns or examples of successes that you believe should be considered in this study? In addition, we would welcome your suggestions for other institutions or individuals to contact with regard to these questions, as well as any references to key documents.

Responses to these questions should be sent to Paul F. Uhlir, Director, U.S. National Committee for CODATA, National Research Council, 2101 Constitution Avenue, N.W., Washington, D.C. 20418, U.S.A. Telephone: (202) 334-3061; Internet: (BITS@NAS.EDU).

The results of the fact finding and research will be reviewed at the second meeting, in the fall of 1995, and a significant portion of the drafting of the report will be done at that time. The members will convene one more time to complete the report, which will be reviewed by the USNC/CODATA at its spring 1996 meeting. The report will be published electronically in the fall of 1996.

Biographical Sketches of Authors

Paul F. Uhlir is Associate Executive Director of the Commission on Physical Sciences, Mathematics, and Applications, and Director of the U.S. National Committee for CODATA at the National Research Council, 2101 Constitution Avenue, N.W., Washington, D.C. 20418.

Shelton S. Alexander is the recent past Chair of the U.S. National Committee for CODATA, and Professor of Geophysics at The Pennsylvania State University, Geosciences Department, 537 Deike Building, University Park, PA 16802.