Draft 10/8/04

ICSU Priority Area Assessment on Scientific Data and Information

Preface

Data and information are essential building blocks of science. The technologies for scientific data and information production, management, and dissemination are rapidly changing as are the scientific needs. In parallel to this there are evolving policy and management issues that need to be addressed in an international context. Hence, ICSU has identified scientific data and information as a priority area in developing its strategic plan for the coming years. In this regard, an international panel of independent experts was appointed by the Committee on Scientific Planning and Review (CSPR) to perform a Priority Area Assessment (PAA) on Scientific Data and Information. This panel was charged with both assessing the strategic issues in this arena and reviewing ICSU's current activities.

The approach of the PAA panel was to first consider the needs and priorities for scientific data and information – production, management, access and dissemination - and then to analyze the existing ICSU activities and structures in this area. This has resulted in a large number of recommendations – too many to be effectively implemented by any single organization. As is often the case, it is relatively easy to identify the problems but much less straightforward to come up with simple solutions. Hence, the various recommendations in this report are addressed to multiple audiences: the ICSU Executive (Secretariat, Board and CSPR); specific ICSU interdisciplinary bodies and committees; the ICSU membership; science funding bodies; governmental organizations; commercial publishers and data producers; and individual scientists[1]. It is hoped that all of these stakeholders with a shared interest in strengthening science for the benefit of society will find something to stimulate them in this report. What is very clear is that all these actors need to work together if the identified priorities and recommendations are to be taken forward.

One of the major challenges for the assessment panel was to maintain a balance between specific focus and in-depth analysis on one hand and broad generalities on the other. It was decided at the outset to concentrate mainly on generic issues that apply across scientific disciplines rather than specific disciplinary topics. The panel also agreed that because of the breadth of data and information issues and the way that they are interrelated, it was better to attempt at least a cursory consideration of the major issues rather than selecting a few isolated topics for the in-depth analysis that each issue ultimately demands. This should not be viewed as an expert report on astronomical data, bioinformatics, new scientific publishing paradigms or intellectual property rights. Such reports exist elsewhere and require specific disciplinary or professional expertise. This assessment cuts across all these subjects and should be

[1] In order to help different readers identify which recommendations may be most relevant to themselves or any particular constituency, a table of recommendations (by no.) versus audiences has been developed by the PAA panel and is attached at <u>annex</u> 5 of this report.

considered as a complement to (and on some topics, a stimulus for) more specific in depth studies. Because of ICSU's major involvement in environmental and global change research and the particular expertise of several panel members, this is perhaps the area most often cited in the assessment report to illustrate specific issues. However the issues themselves are applicable across many areas and it is hoped that, whatever his or her background, the reader can easily make this extrapolation.

This report, as published here, represents the views of the independent assessment panel. ICSU has commissioned similar assessments for the Environment in Relation to Sustainable Development (published in December 2003) and Capacity Building (to be published in 2005). The conclusions and recommendations of each of these reports will be considered, in the context of ICSU's overall strategic plan 2006-2012, by the ICSU General Assembly in Suzhou, China, in October 2005.

Priority Area Assessment on Scientific Data and Information

Executive Summary

The nature and use of scientific data and information[2], the conditions under which scientific data and information are produced, distributed, and managed, and the role of scientists and other actors in these processes have been changing rapidly in recent years. These changes are partly a result of the revolution in computational capacity and connectivity and advances in hardware and software that together have expanded the quality and quantity of research data and provided scientists with a greatly increased capacity for data gathering, analysis, and dissemination. They are also related to the emergence of new questions in scientific research that require different types of data, the integration or combination of existing data into new kinds of data sets, the digitization of scientific publications to permit on-line links to data and references, and widespread data sharing. Taken together, these changes are providing scientists throughout the world with more and enhanced access to research data and information. The benefits of this include the growing involvement of scientists in international research projects and increased scientific and policy interest in global scale and comparative research activities. Meanwhile, the legal concept of intellectual property as applied to scientific data and information is also in a state of flux and there are still major obstacles to data access in many parts of the world.

Because of the critical importance of data and information in the global scientific enterprise, the international research community must address a series of new challenges if it is to take full advantage of the data and information resources available for research today. Equally if not more important than its own data and information needs, today's research community must also assume responsibility for building a robust data and information infrastructure to provide for and protect the data and information resources required by future generations of scientists.

The ICSU Priority Area Assessment (PAA) Panel on Scientific Data and Information strongly recommends that ICSU assume an international leadership role in identifying and addressing critical policy and management issues related to scientific data and information [1][3] and that it create a new global framework for data and information policy and management. ICSU played a seminal role in the 1980's and 1990's in establishing an interdisciplinary and internationally coordinated research program on global environmental change. Our current scientific understanding of environmental change and our capacity to advance the scientific

[2] For the purposes of this report, the panel considered data and information as a continuum and both "data" and "data and information" are used interchangeably to refer to the whole of this continuum. Where appropriate the distinction is made between "data" and "scientific publications", which are a specific sub-set of scientific information that raise particular issues.
[3] The numbering of recommendations has been conserved between the Executive Summary and the main report, for ease of reference. Likewise, all the key recommendations have been brought together in the Executive Summary although the reader should refer to the full report for specific details.

agenda in this field could not have been accomplished without strong and capable ICSU leadership and judicious use of limited ICSU resources. Today there is a need for an equally strong ICSU role in establishing an international infrastructure and capacity for scientific data and information management and access that meets the needs of scientists in all countries and protects the interests of current and future generations of scientists.

A Long-term Strategic Framework and new Forum for Scientific Data and Information

The production, management, and dissemination of scientific data and information have become increasingly critical functions within the scientific research enterprise. Professional standards and practices must be employed in order to properly perform these functions. Data must be preserved over long periods of time so that the scientific records and observations obtained today will be available for use in research in the future. Different fields of science and other actors, in both the public and commercial sectors, need to come together to develop a more coordinated approach to data and information.

The Panel recommends that, on the basis of the many specific recommendations in this report, ICSU develop a long-term strategic framework for scientific data and information (policies, practices and infrastructure). An essential part of the development of this framework should be the closer coordination, and in some instances the transformation, of ICSU's current data and information activities [57]. The framework should build on existing data and information structures and services where it is advantageous to do so, but ICSU and its partners should be prepared to rethink, re-orient, and replace existing structures and bodies where it is necessary.

ICSU should foster greater communication, coordination, and collaboration within and across members of the ICSU family and other partners on issues, policies, practices, and structures for scientific data management. More specifically, there should be widespread participation in the development of the future framework for scientific data and information. **The panel recommends that, in parallel to the development of the long-term strategic framework, ICSU establish an international Scientific Data and Information Forum (SciDIF) involving all the key stakeholders: ICSU members, interdisciplinary bodies, science funding bodies and other data providers and users. Through SciDIF, ICSU should aim to ensure that the full benefits of new data and information technologies and capabilities are extended to scientists throughout the world [58]**.

Recognizing the importance and urgency of these issues, ICSU should **establish a Strategic Data and Information Committee to oversee the development of a long-term integrated framework for data and information and a Scientific Data and Information Forum (SciDIF). Membership of this committee should include representatives of relevant ICSU bodies and unions, and experts in information technology and professional data management. This should be an *ad hoc* committee, with members (10-12max.) appointed for a period of three years [2]**. The remit for the committee should be to oversee the implementation of the recommendations in this report; to develop a long-term integrated strategic framework; to establish a Scientific Data and Information Forum; and to work with the Forum to address the policy, financial and operational issues involved in implementing the strategic framework

The Public Sector Role in the Production of Scientific Data and Information

Public sector support for data collection has been a major factor in the advance of science over the past half century. Because of the high cost of data collection, particularly data collection over long periods of time, the scientific community is now almost totally dependent upon various types of public sector support to obtain data for its research. Without this support, the scientific research enterprise could no longer function. However decisions on data collection at both national and international levels are often taken in the absence of consultation with the scientific community. In extreme cases, scientifically valuable data series can be interrupted, discontinued, or altered for administrative, political, or budgetary reasons, regardless of their scientific importance.

3. ICSU should be a strong advocate to multilateral organizations for the data and information interests and needs of the international scientific community.

4. ICSU unions and national members should emphasize the benefits of scientific research and data management to national and multilateral governments and the need for regular consultation with scientists on policy and decision making involving data collection and management.

5. ICSU should work with its members and key partners to establish guidelines on good practice in public sector data management.

The Private Sector Role in Scientific Data and Information

The private sector plays an increasingly important role in scientific data and information production, publication, and dissemination. Examples of commercial data or information production exist in many fields, including Earth observation, genome sequencing, and scientific publishing. One consequence of this private sector role is that data collection and preservation priorities may be determined by market demand rather than scientific priorities. Another consequence is that there may be restrictions on access to research data because of the commercial need for confidentiality. When scientific data must be purchased or licensed, the cost of obtaining data becomes a larger proportion of the research budget.

6. The scientific community, through ICSU national and union members, should seek to persuade governments and private sector data providers that research data produced commercially or through public-private partnerships should be made available for free or for the cost of reproduction for purposes of research and education.

7. Government authorities who grant licenses to private sector firms for data production should ensure that data originating in the private sector have long-term research utility, are managed according to the same high professional standards as other types of scientific data, and are ultimately offered to a publicly supported archive for scientific and educational use.

8. Governments involved in public-private partnerships to produce data should ensure that the data are made available for research and education use in a timely manner.

Data and Information Rescue

Many types of data are at risk. They are not being used for scientific research because they are not available in digital formats or they are in danger of being lost because the media on which they are recorded may decay, become corrupted, or be superceded by new software. This is a particularly acute problem in developing and transitional countries.

9. Scientists should inventory major collections of extant data and information and should set priorities for the rescue and permanent preservation of the data and information that are most valuable and at greatest risk.

10. ICSU and its members should draw the attention of scientists, public policy makers, and research foundations to the issue of data at risk and ways to deal with the issue.

11. Scientists or countries that undertake significant efforts in the rescue of data at risk should consider the advantages of public/private partnerships in this effort.

Scientific Publications

Scientific publications are increasingly being produced and disseminated electronically and are playing new roles in research. At the same time, the economic foundations of scientific journal publishing are threatened both because of the demand for free access to on-line publications and because library budgets have not been able to keep pace with rising journal prices and the growth of the literature.

12. Many ICSU members are directly involved in scientific publishing. The panel recommends that ICSU continue to work with a broad based organization such as the International Council for Scientific and Technical Information (ICSTI) that includes both the publishing industry and scientists to promote new opportunities provided by information and communication technologies and address key issues related to the transition in the scientific publication process.

13. Journal publishers should encourage authors to make the source data for their articles available in electronic repositories that are stable, widely accessible and professionally managed.

14. ICSU should work with its members and relevant bodies in encouraging the coordinated development of digital libraries and their integration with journal publishing and data systems.

15. Because of the importance of extending the benefits of digital publications to all scientists worldwide, ICSU should encourage its member organizations to work with the International Network on the Availability of Scientific Publications (INASP) to build cost-efficient and sustainable publishing capacity and journal access in developing countries.

Professional Data and Information Management

16. The panel recommends that ICSU play a major role in promoting professional data management and that it foster greater attention to consistency, quality, permanent preservation of the scientific data record, and the use of common data management standards throughout the global scientific community.

Scientific data and information management can no longer be viewed as a task for untrained amateurs or as part of the routine "clean up" at the completion of a research project. The use of advanced information technology in scientific data management and dissemination makes it essential that data management be the responsibility of professionals. Scientific data centers and archives require stability in their financial resources so that they can make institutional commitments to data management and preservation over many decades.

17. Recognizing that scientific data management is undergoing rapid innovation and change, information technology specialists, librarians, research scientists, government data producers, donors, and others should be involved in a concerted effort to develop standards and curricula for professional training for scientific data managers.

18. Financial support for data and information management should become a routine component in all research budgets and the evaluation criteria for assessing research funding proposals should include evaluation of data management.

19. All scientists should receive training in data management as part of their graduate education. ICSU should encourage the development of guidelines for data management by working scientists and their institutions.

20. Scientists should be recognized and given credit for the scientific contribution of the data sets that they produce as well as for the analysis of those data.

21. ICSU, its members and associated bodies should raise awareness of the increasingly important role that institutional repositories play in relation to scientific information management and preservation and the need to ensure that such repositories are properly resourced, developed and maintained.

Metadata

The use of common metadata standards across fields of science facilitates the identification, re-use, and integration of scientific data and provides information that future scientists can use to evaluate the data. Metadata should be the principle vehicle for documenting known data quality issues.

22. ICSU should work with its members to promote the development and use of flexible, open, and easy to use community standards for metadata. These standards should be interoperable and independent of specific hardware and software platforms. Guidelines for their use should be widely circulated and incorporated into data management training courses.

23. Data repositories and publishers should ensure that standard metadata are available for all databases and records.

24. Metadata should be archived and made freely available electronically in multidisciplinary metadata catalogues.

25. To foster the efficient production of metadata, ICSU should encourage the development of software for writing metadata that can be made available to scientists throughout the world.

Archiving

Permanent archiving of scientific data and information is essential. In some fields, there are institutions that have a clear responsibility for data archiving, but this is not always the case and varies from one scientific field to another. There is a distinction between data centers, which are responsible for providing immediate access to scientific data and information, and archives, which provide for permanent preservation and management of DI.

26. ICSU and its members should raise awareness of the need for long-term institutional support for data archives both at the national and international level.

27. ISCU should foster discussion within the scientific community, including its members and interdisciplinary bodies, on criteria, institutional structures, and models for decision making related to the permanent preservation of scientific data and information.

28. Ways to reduce the costs of archiving, such as sampling from extant data bases or establishing multiple classifications to prioritize levels of archival support, should be examined by the scientific community.

Privacy and Security: Ensuring the Integrity of Data and Information

The ease of combining or integrating electronic research data obtained from multiple sources heightens the need for protection of individual, national, and corporate privacy and confidentiality in scientific databases.

29. ICSU should work with its member unions and other partners to conduct a substantive review of the new ethical and regulatory issues related to use of individual and population data, including the integration of data from multiple sources and variations in national legislation.

30. Data security and integrity must be addressed in the context of procedures for data management. A professional information technology staff is required for both archives and data centers if they are to maintain adequate system and database security and integrity.

31. Data that are disseminated by scientists as part of research projects should have metadata that describe the security and integrity measures employed in the collection and management of the data.

Interoperability

32. ICSU and its partners should "promote interoperability principles and metadata standards to facilitate cooperation and effective use of collected data and information," as recommended in the Science in the Information Society Agenda for Action.

Systems for Data Dissemination

The ease of disseminating data via the Internet encourages individual scientists to make their research data available to others through personal or project web pages or through distributed federations of data providers. Yet these "virtual repositories" provide the illusion rather than a guarantee of long term access to data. There is no means of ensuring that data managed by individuals or research groups in voluntary distributed systems on the Internet will continue to be available and well-managed over time.

33. ICSU should bring together representatives of voluntary data confederations and distributed data systems to discuss what has been learned over the past ten years about what contributes to the success of voluntary data confederations, what undermines them, and what must be done to preserve and enhance access to the data in the future.

The Digital Divide

The digital divide is most evident in low- and medium-income countries, but it has an impact on scientists throughout the world because it limits potential research and access to data in many regions.

34. Key issues relating to the digital divide in scientific data and information are addressed in the Science in the Information Society Agenda for Action. The panel strongly endorses the actions relating to the provision of high-speed internet access, human capacity building, electronic publishing and initiatives to make scientific information more accessible. ICSU should work with its members and partners to encourage the implementation of these actions.

35. Emphasis should be placed on the need for professional data and information management in those countries where the scientific data infrastructure is being constructed. ICSU bodies, such as the International Network for Access to Scientific Publications (INASP) and CODATA, have a key education and training role to play in these countries.

Equitable Access to Data and Information

Because of the scientific and budgetary benefits of reusing other scientists' data for research and the need to be able to test and retest hypotheses using the same body of data, science has long been best served by a system of minimal constraints on the availability of data and information.

36. ICSU should continue to stand firmly behind the principle of full and open access to scientific data.

37. With regard to scientific publishing, ICSU should ensure that the principle of universal and equitable access to scientific publications is upheld.

Who Pays?

Data production and management are costly. Collection of data, preparation of metadata, and provision of professional data management expertise and institutional support for data dissemination and permanent archiving will add to the overall expense of specific research projects and maintaining the larger research infrastructure. There are a number of economic models for supporting the necessary scientific data and information management, but none of them is completely satisfactory.

38. Ensuring the long-term accessibility of increasing quantities of scientific data and information will necessitate increased public (and private) investment in data management and long-term institutional support. ICSU and its members should explore various solutions to meet the financial challenge of providing full and open access to scientific data and universal and equitable access to publications.

Intellectual Property Rights

Scientific data and information are increasingly considered both as input to research, decisions, policy, and management, but also, like digital data in entertainment and commerce, property in and of themselves. Science has been well served by a system of minimal restraints (*e.g.,* those based on privacy considerations) on access to and use of data. Recent trends towards the appropriation of data, such as genetic information, and the protection of databases under *sui generis* regimes, as well as limitations to the fair use of digitized data (*e.g.*. anti-circumvention measures) pose serious obstacles to full and open access to data for scientific purposes.

39. Governments and other bodies concerned with international and national policy development, should ensure that IPR legislation recognizes the value of ensuring full and open access to data for scientific research and education purposes.

40. ICSU and its members should investigate appropriate mechanisms to ensure that science is fully represented in international treaty negotiations that might have an impact on access to data for scientific purposes.

Review of ICSU Bodies

ICSU has established a number of bodies that specialize in scientific data and information issues at an international level and whose activities were reviewed as part of this Priority Area Assessment (PAA). In considering the recommendations relating to specific bodies, it is important to note that these recommendations do not constitute a judgment on the past performance of those bodies but are, rather, an assessment of their future role and potential relative to the strategic priorities identified in the overall PAA. Implementing these recommendations on specific ICSU bodies will be a key step in the development of a long-term strategic framework and the proposed Scientific Data and Information Forum, in which the relevant bodies are expected to be central players.

The Committee on the Dissemination of Scientific Information (CDSI)

41. CDSI should be disbanded. ICSU should look to the International Council for Scientific and Technical Information (ICSTI) for the advice formerly expected from CDSI.

The International Network on the Availability of Scientific Publications (INASP)

44. INASP should be formally recognized as an ICSU interdisciplinary body.

Committee on Data for Science and Technology (CODATA)

In the context of the development of a long-term strategic framework and international Scientific Data and Information Forum:

48. CODATA should develop a clear long-term strategy that focuses on key international data management and policy issues and should place a strong emphasis on eliminating the digital divide.

Monitoring and Observation Bodies: The Global Ocean, Terrestrial and Climate Observing Systems (GOOS, GTOS, GCOS) and the Integrated Global Observing Strategy-Partnership (IGOS-P)

52. ICSU should play a stronger role in the Global Observing Systems by fostering the development and implementation of appropriate policies and data management procedures and representing scientific data user needs. There is a need for both operational archives for dissemination of data and a strategy for long-term data preservation.

53. ICSU should be actively involved in the Group on Earth Observations (GEO) as a representative of the international scientific community, advocating appropriate policies for scientific consultation, data collection, data access, and professional management of Earth observation data.

Federation of Astronomical and Geophysical Data Services (FAGS)

54. FAGS should be no longer be an ICSU interdisciplinary body. The responsibility for FAGS should be devolved to its user constituency, the three co-sponsoring Unions.

Panel on World Data Centers (WDC)

55. ICSU should re-examine its entire data center infrastructure in light of technological and scientific changes in data collection, use, and management. Planning for an updated system should build upon the successful accomplishments of the World Data Centers, but should go beyond current practice to take advantages of new technologies and capabilities. This effort should be integrated into the development of a long-term strategic framework and international Scientific Data and Information Forum (SciDIF). It must take account of the needs of existing ICSU programs and other new initiatives, such as the Group on Earth Observations (GEO), the forthcoming Polar Year in 2007, and the electronic Geophysical Year (eGY).

Scientific Committee on Frequency Allocations for Radio Astronomy and Space Science (IUCAF)

56. ICSU should encourage the three main partners of IUCAF--IAU, URSI and COSPAR-- to organize a joint meeting to define specific procedures and actions that will ensure that IUCAF becomes a truly interdisciplinary committee that has the necessary expertise to coordinate the efforts of *all* the passive radio sciences in frequency management matters.

Priority Area Assessment on Scientific Data and Information

I Introduction

Data and the interpretation of research results disseminated as information have always been essential components of scientific research. In recent years, the nature and use of scientific data and information, the conditions under which scientific data and information are produced, distributed, and managed, and the role of scientists and other actors in these processes have been changing. These changes are in part a result of the revolution in computational capacity and connectivity and advances in hardware and software that together have expanded the quality and quantity of research data and provided scientists with a greatly increased capacity for data gathering, analysis, and dissemination. They are also related to the emergence of new directions in scientific research that demand different types of data, the integration or combination of existing data into new kinds of data sets. The growing involvement of scientists in many countries in joint research projects, together with increased scientific and policy interest in global scale and comparative research activities, further contributes to the changing environment for scientific data and information. Finally, the evolution of the legal concept of intellectual property has transformed how data are viewed. Once considered principally as an input to or result of scientific research, data and the compilation of data are now often considered to be property that can be bought and sold, given or withheld.

The international research community must address a series of new challenges related to data and their management, dissemination, and preservation if it is to take full advantage of data and information resources available today. Equally, if not more important, the research community must assume responsibility for building a robust data and information infrastructure to provide for and preserve today's data and information for future generations of scientists.

The ICSU Priority Area Assessment (PAA) Panel on Scientific Data and Information strongly recommends that ICSU assume a leadership role internationally in identifying and addressing critical policy and management issues related to scientific data and information [1] and that it serve as a bridge between the scientific community and those outside it whose participation is essential to the definition and implementation of this infrastructure. ICSU is uniquely positioned to assemble relevant members of the scientific and policy communities across fields of science and national research systems to discuss data and information collection, management, access, and dissemination for scientific research and to foster agreement on common approaches to these issues. ICSU can also represent the interests of the international scientific community in discussions with governments, commercial entities and non-governmental organizations that are engaged in activities related to current and future production, management, and dissemination of scientific data and information.

We recognize that just as there are some roles that ICSU is uniquely able to fill, there are others that it cannot or should not assume. ICSU should be engaged in activities fostering the adoption of common data policies that will advance science, not data operations. Its strength is its capacity to unify and represent the scientific community and to foster discussions about critical issues and help identify answers to problems. ICSU cannot assume responsibility for the activities of individual scientists, research organizations, publishers, national science agencies, or observation or monitoring efforts.

ICSU played a seminal role in the 1980's and 1990's in establishing an interdisciplinary and internationally coordinated research program on global environmental change. Our current scientific understanding of environmental change and capacity to advance the scientific agenda in this field could not have been accomplished without strong and capable ICSU leadership and judicious use of limited ICSU resources. Today there is a need for an equally strong ICSU role in establishing an international infrastructure and capacity for scientific data and information management and access that meets the needs of scientists in all countries and protects the interests of future generations of scientists.

1. Why Examine the Current State of Scientific Data and Information?

Few aspects of scientific activity have changed as rapidly over the past half century as the production, dissemination, and use of scientific data and information. For centuries, scientists systematically made and recorded observations in the course of their research. Although they often published these data in the scientific literature, there was little sustained effort or even presumed need to preserve data or to re-disseminate them to other scientists, particularly those in other disciplines, countries, or research fields. In the past, it was often deemed sufficient to analyze data and publish results, leaving it to others to test the validity of the research results by further analysis or reproducing the data. The construction and maintenance of scientific databases was a responsibility generally assumed by individuals or small groups of scientists within the context of focused research projects.

As science has become more interdisciplinary and international in scope, the inadequacies of this traditional 'closed' model of scientific data production and management have become apparent, and it has been superceded by greater openness in sharing data. Environmental and global change research is a good example of a field that has been at the leading edge of this transformation. Spurred in part by growing scientific interest in monitoring global scale environmental changes during the 1957 International Geophysical Year (IGY), and in part by advances in observational capacity and sensor development, scientists and scientific policy makers in national science agencies in the last half of the twentieth century began to emphasize the development of data resources, archives, and repositories that would serve the needs of multiple scientific research groups regardless of their geopolitical location. The rapid expansion of computing capacity through the development of more powerful hardware and software in the 1980's and 1990's gave added impetus to these developments. For example, the definition of what constitutes scientific data and information was broadened with the advent of geospatial software. In addition, the availability of multiple-user, large scale data resources meant that scientific research could be more data intensive, and as a result, scientists began to employ data mining and meta-analysis techniques in their research. Advances in science and instrumentation created entirely new types of data for research. For instance, the expansion of civil remote sensing has permitted scientists to conduct new types of geophysical, environmental, ecological, and even social science research.

In part because of the broad influence of global change research and in part because of the new availability of data from multiple time periods, there has been a growing interest in the analysis of change over time. Research of this type depends on a scientist's ability to create time series data from multiple sources and to create comparable data sets across space. There is also an emphasis across fields of science upon the integration of multiple sources of data into new time series or geographically comparable data.

Other areas of science have undergone equally dramatic transitions. Advanced computing technologies have led to rapid expansion in the capacity of scientists to analyze, mine, integrate, disseminate, and model vast amounts of data. The development of the World Wide Web was a direct result of the need of particle physicists at CERN to share documents and data analyses among scientists in different physical locations. The GRID—a new form of distributed computing—is now being developed by computer scientists to facilitate the next generation of distributed data access, integration, and analysis. In biology, new analytic technologies and computing power have permitted the sequencing of the human genome and led to the development of an entirely new discipline of bioinformatics. There is a general move across many fields of science away from reductionist approaches to complex systems analyses. This is characterized by the computerized integration and modeling of large amounts of data from diverse sources.

New Information and Communication Technologies are also revolutionizing scientific publishing. A plethora of new e-journals have appeared in the last three years and new dissemination models, including open access publications and electronic journal archives, are being considered. The long established business model of scientific publishing is being questioned, posing a direct threat to private sector publishers and many scientific societies that have used the income from publishing to subsidize other activities. It is no longer clear where the responsibility and the financial support for maintaining the scientific record of publications lie. New ways of sustaining current practices as well as many new economic models for scientific publishing are being considered and tried.

In this rapidly evolving arena, the necessary institutional structures and legal and regulatory frameworks are not always in place to ensure that the full benefit of scientific data and information can be obtained for science and society. These problems are exacerbated in the least economically developed countries where both the institutional infrastructure and experienced human resources to take advantage of these advances may be lacking. The phrase "digital divide" has been coined to describe this situation at a societal level but despite the expanding access to digital scientific data and information, there is a serious "digital divide" within the international science community as well.

The growing importance of scientific data and information can be seen not only in their significance for research, but also in their changing legal status. Instead of being considered a byproduct of the research process and a possession of the scientist who generated them, scientific data and information are now viewed in new ways. Within the scientific community, they are recognized as part of the broader infrastructure of science. Many science agencies now require scientists to agree to redistribute and share any data collected under a research grant. However, as the production of large scale scientific data and information increasingly serves purposes that extend beyond the originating research project and even takes place outside the structure of publicly-funded research projects, those who produce the data can, in some circumstances, exert intellectual property controls over what they have created in order to obtain economic benefits from the activity. The growing commercialization of scientific data and information raises important legal and ethical, as well as economic issues for science. These issues have implications for data distribution, use, and access and for the role of science in society.

2. Scope of This Priority Area Assessment

This Priority Area Assessment consists of an examination of issues related to the production, management, preservation, and dissemination of scientific data and information and a review of ICSU advisory and interdisciplinary bodies that have a special responsibility for these issues. That is, the assessment examines priorities for scientific data and information policy and practice and also examines the bodies in the ICSU family that are in place to address these issues. The assessment was commissioned by the ICSU Committee for Scientific Planning and Review (CSPR) at the request of the 27th ICSU General Assembly and is part of a broader strategy development process for ICSU. The terms of reference for the assessment, together with list of bodies consulted are attached to this report (annexes 1 and 2).

The assessment has been conducted by an independent *ad hoc* panel of experts, drawn from different disciplines and regions, representative of the breadth of this topic and its universal importance to science. The panel itself was selected by CSPR after solicitation of nominations from the ICSU members (national science academies and international science unions). Members of the panel are listed in annex 3. The panel met on three occasions in Paris between October 2003 and April 2004 and completed its report for submission to CSPR in June 2004. The report represents the advice of the *ad hoc* panel to ICSU.

Although the remit for this assessment is defined in the terms of reference in annex 1, it was clear to the panel at the outset that scientific data and information is an extremely broad field and its own expertise was inevitably limited. The panel had to make choices initially as to what topics it should or should not consider in this assessment and where the principle focus should be so as to assist ICSU in its planning and priority setting. Thus, at the outset of the review, the panel carried out a brief scoping exercise.

The Panel decided first that no specific issues of importance for scientific data and information in an international and multi-disciplinary context should be excluded *a priori*. Second, it decided that its focus should not be on the data and information needs of single scientific disciplines or specific groups of scientists, but rather on generic data and information management and policy issues of relevance to the international science community. However, given ICSU's strong interest in environmental research and sustainable development, the Panel also recognized that when identifying examples to illustrate management and policy issues there might be a particular focus towards this research area. Third, the Panel concluded that both scientific publishing and intellectual property issues should be included in the assessment, as both areas had been highlighted previously by CSPR as worthy of special attention.

Finally, wherever feasible, the panel agreed that this assessment should not only deal with pressing immediate problems, but should take a long-term perspective. The panel believes strongly that decisions taken at present related to scientific data and information will have a long range impact on scientific research for at least the next century. Technological innovations and the use of new types of data and information in the future will inevitably require changes in the data policies and practices adopted today. Nonetheless, members of the international scientific community and those in the public and private sectors who produce or use data for scientific research must recognize that the quality of future scientific research will be dependent on the availability, security, and integrity of data over long time periods. This will be directly dependent upon the adequacy of the data and information policies and commitments put in place today. Data policies are already being developed outside the scientific community for commercial, governmental and other types of data and these policies will govern scientists' access to and use of data in the future. In this situation, it is essential that the scientific community be cognizant of its needs and that it be an active participant in discussions of the broader economic, political, legal, foreign policy, and civil approaches to data and information policy. Only ICSU has the international standing to foster such discussions on behalf of the global scientific community.

BOX
<u>Definitions of Data and Information</u>

For the purposes of the assessment, the panel considered data and information (DI) as a continuum ranging from raw research data through to published papers. "Data" includes, at a minimum, digital observations, scientific monitoring, data from sensors, metadata, model output and scenarios, qualitative or observed behavioral data, visualizations, and statistical data collected for administrative or commercial purposes. Data are generally viewed as input to the research process. "Information" generally refers to conclusions obtained from analysis of data and the results of research. But the distinction between them is flexible and will vary according to the situation. Increasingly, the output of research (traditionally viewed as "information") includes data and has become input to other research, rendering the output-input distinction between data and information meaningless. In this report, both "data" and "data and information" are used interchangeably to refer to the entire DI continuum because the continuum as a whole has been affected by

changes in information technology and is subject to many of the same issues. Where appropriate, the distinction is made in the text between "data" or "DI" and "scientific publications", which are a specific sub-set of scientific information that raise particular issues.

3. ICSU Activities in Relation to Data and Information

Data and information are critical to all scientific endeavors and, as such, are of concern to the entire ICSU family – international science unions, national academies and research funding agencies, interdisciplinary bodies, joint initiatives, and associates. The current ICSU structure for addressing issues of data and information was developed over the past half century in response to opportunities, needs, and problems as they emerged and is based on the work of decentralized and often independent units. Across the entire DI continuum, ICSU's dedicated scientific data and information structure consists of an advisory committee and several interdisciplinary bodies and joint initiatives, with a preponderance of the activities within this structure focused on environmental data.

The single ICSU advisory committee related to data and information is the Committee on the Dissemination of Scientific Information (CDSI), whose special focus is scientific publishing and access to the scientific literature. The International Network for the Availability of Scientific Publications (INASP) is an ICSU interdisciplinary body that was originally established by CDSI and whose role is to bridge the North/South divide in scientific information. The interest in scientific publication goes beyond these two groups, however, for many of the ICSU unions and academies publish their own scientific journals. For some, this role provides a major source of income. The International Council for Scientific and Technical Information (ICSTI), which also focuses on scientific publications, is an associate member of ICSU.

There are several interdisciplinary ICSU bodies whose principal focus is the management and use of large scientific data sets: the Committee on Data for Science and Technology (CODATA), the Panel of the World Data Centers (WDC), and the Federation of Astronomical and Geophysical Data Analysis Services (FAGS). ICSU is also a co-sponsor of the Global Observing Systems (GOS). Both individually and working together, many members of ICSU are very actively involved in issues related to scientific data. For example, an important policy initiative was the creation of an Inter-Union Bioinformatics Group, which produced an authoritative report and recommendations regarding biological databases in 2002. Similarly, several ICSU bodies and Unions have joined forces in an *ad hoc* Group on Data and Information to develop guidelines for data access.

In summary, the ICSU family has an array of disciplinary or topic focused activities in which scientific data and information use are a major component. It also has several overarching committees with special responsibility for data and information. The future role and function of these latter committees in light of the changing scientific and technical data and information environment and the evolving needs of the scientific community is a key aspect of this review.

4. The UN World Summit on the Information Society (WSIS)

The World Summit on the Information Society (WSIS) is a major United Nations inter-governmental summit taking place in two phases, the first culminated in a summit meeting in Geneva in December 2003, and there will be a second major meeting in Tunis in December 2005. Even before the establishment of the current PAA, ICSU had been playing a very active role in the negotiations leading up to Geneva. In particular, an agenda for action – *Science in the Information Society* – had been developed as the result of an on-line discussion forum and international workshop in March 2003. This workshop was co-organized by ICSU and CODATA and hosted by UNESCO. The agenda for action is attached as annex 4 to this report and the majority of its recommendations have been included in the formal declaration and plan of action that were agreed by Heads of State in the Geneva summit meeting. These recommendations had also been formally endorsed in 2003 by many of the ICSU members. They represented a significant contribution to the overall priority area assessment.

II. The Expansion and Diversity of Scientific Data and Information: Opportunities and Challenges

The data and information base for scientific research has expanded rapidly in recent years because of advances in information technology and the invention of new technologies for obtaining observations and data, new types of analysis being used in scientific research, and new ways of dissemination via the Internet. Scientists now routinely employ many types of data and information that were not previously utilized. New data are being generated through, for example, governmental monitoring and observation programs, the generation of model outcomes, the development of new types of sensors, and genetic epidemiological studies. There is a diversity of sources of research data, including traditional knowledge, films, videos, and photographs as well as 'classical' numerical databases. Large amounts of data are also being obtained through new procedures, such as high resolution remote sensing, protein microarray analysis and/or the next generation of nuclear fusion experiments. Not only is the amount of data and information expanding, but also the tools for mining and distributing these resources are rapidly developing. Software and visualization tools, whether open source or proprietary, provide new analytic options to scientists. Powerful search engines permit researchers to mine diverse publications for relevant abstracts or data, without having to move from their desks.
.

Scientific demands for data are also expanding, raising new types of data management and access issues. Research emphasizing change and the dynamics of change, whether at the level of a single mammalian cell or a specific geographical ecosystem, requires time series and comparative data across time and space. The production and use of these data is dependent not only upon standardization and continuity of measurement, but also upon the preservation of data over time. It also may require the "rescue" of previously obtained data that are stored on decaying media or in formats that need to be transformed into digital formats before the data can be available for research. A more 'holistic' approach to research in areas as diverse as sustainable development and cognitive neuroscience necessitates the integration or transformation and integration of data from multiple sources and disciplines into new and increasingly complex databases. Interoperability of data is becoming a critical issue for scientific research.

Scientific publications have traditionally been used for the dissemination of the results of research. Increasingly, however, they also serve as data for further research. Data published in the scientific literature can be combined with other sources of data in meta-analysis or to create new databases. When publications are digitally available on line, they can be electronically linked back to the cited publications and the data used in the analysis, as well as linked forward to citing articles. The data can also be linked to other data to create dynamic databases.

1. The Public Sector Role in the Production of Scientific Data and Information

A characteristic of much of the data now used in scientific research is that their collection is financially supported and/or directly undertaken by the public sector. The "public sector," in this context, includes national, local, or multilateral governments and non-governmental but tax exempt organizations, such as foundations. Using a variety of funding instruments, these groups contribute directly to science through different mechanisms, including research grants for specific projects and the direct support of data collection for research. In addition, many organizations in the public sector collect data and information for their own use that can also be used in scientific research. Direct support for science is often from national research funding agencies and foundations. Many of these agencies support data collection as a resource for multiple scientific users. Scientists also rely heavily on operational data collected by agencies in the public sector or by not-for-profit organizations that have some public support through tax benefits. These include, for example, the data collected by various national meteorological services, economic and statistical data obtained by national statistical agencies and the United Nations, spatial data obtained by land ordnance or geological surveys and mapping agencies, public health and medical data. These operational data are expected to serve multiple users, including, but not exclusively, nor even primarily, scientists.

Public sector support for data collection has been a major factor in the advance of science over the past half century. Without this support, the scientific research enterprise could no longer function as it does. This raises two critical but interrelated problems: First, the international research community is dependent for a large part of its data upon financial decisions centralized in public sector agencies that are not primarily focused on science and second, these bodies are generally in high income countries. Although scientists have in some instances developed successful partnerships with public bodies to promote and facilitate data collection, final control over levels of financial support, data coverage, and continuity of data collection is exercised in the course of broad-based national and international decision making in the context of policy issues that have little relationship to science. In extreme cases, scientifically valuable time series data can be interrupted, discontinued, or altered for administrative, political, or budgetary reasons, regardless of their scientific importance. Even where data resources are directly funded by science agencies, continuity cannot be guaranteed. As the cost of data management increases over the next five to ten years, the scientific community will become even more dependent upon public sector, and ultimately political, structures and processes for financial support for its data collection, management, and archiving.

Recommendations:

3. **ICSU should be a strong advocate to multilateral organizations for the data and information interests and needs of the international scientific community.**

4. **ICSU unions and national members should emphasize the benefits of scientific research and data management to national and multilateral governments and the need for regular scientific consultation on policy and decision making involving data collection and management.**

5. **ICSU should work with its members and key partners to establish guidelines on good practice in public sector data management.**

2. The Private Sector Role in Scientific Data and Information

Commercial data suppliers are increasingly important contributors to many areas of science. This is most evident in the field of Earth observation, where private sector remote sensing firms have existed alongside public sector remote sensing operations since the launch of Earth observation satellites by SPOT Image in France in the 1980's. At present, private sector firms or public-private partnerships for Earth observation exist in Canada, France, the United States, and Russia. In some cases, commercial data production or public-private partnerships compete with public sector data activities; in others, national policy favors the replacement of public sector remote sensing by private sector firms wherever feasible. This is the case in the United States, where high spatial resolution civil remote sensing is, *de facto*, the exclusive preserve of the private sector. Another example of private sector involvement in scientific data collection is in genome sequencing, where the activities of companies such as Celera have complemented as well as competed with public sector efforts. A third example of the commercial role in scientific data and information is in the field of scientific publishing in which commercial publishers have existed alongside non-for-profit enterprises for several decades.

When data are initially produced or financially supported by public sector agencies, they are frequently made available to scientists at no cost, and the cost of production is subsumed into larger budgets. In this situation, individual scientists are generally not responsible for paying for data. Data produced in the private sector, however, are normally licensed or sold to potential users, both to recover the firm's initial investment and to make a profit for its investors. Folding the cost of data purchase or licensing into already tight research budgets can lead to overstretched budgets and ultimately place limits on the research that can be done.

Another issue in the scientific use of commercially produced data is that data collection priorities are determined by market demand. In the case of satellite remote sensing, this means that commercial firms will not continuously monitor the Earth for potential current or future uses of the data because the process is too expensive. Instead, they will obtain remotely sensed images when they have a specific order for the data. Even if all the data obtained at the request of customers are preserved for future scientific use, this means that the surviving Earth observation record will consist of discrete observations that are distinguished only by the fact of their having been ordered, not by scientific priorities or needs. Under private sector financial models of data acquisition, there will not be the continuous monitoring of Earth that has been the hallmark of government civil remote sensing. When public-private partnerships are formed, the public sector partner is more likely to provide for continuous monitoring under the initial agreement between the partners.

In some cases, the commercial desire for confidentiality can also impose restrictions on access to data for research purposes. A good illustration of this is in the area of clinical trials, where meta-analyses of combined data from multiple trials can lead to more accurate assessments of the benefits and side-effects of pharmaceutical interventions. There are now many examples where such analyses have indicated critical changes in drug administration and clinical practice. Although there have been positive steps from some pharmaceutical companies to register their trials and make the data more openly available for secondary analyses, this is still not common practice.

Finally, if data produced in the private sector are to be used by scientists in the future, standard scientific data management practices must be used in data preparation and preservation. Professional data management to meet scientific standards is expensive, but it is also necessary to maintain the integrity of the data for future use. Provision must also be made to deposit private sector data in a public access facility when they are no longer deemed to have commercial value. In the United States, the licensing process for commercial remote sensing requires the private sector entity to agree to turn the data over to a public sector archive when the firm no longer wants to keep the data. This provision is valuable, but its utility depends of the quality of the data management between the acquisition of the data and its disposition in the public sector.

Recommendations:

6. **The scientific community, through ICSU national and union members, should seek to persuade governments and private sector data providers that data produced commercially or through public-private partnerships should be made available for free or for the cost of reproduction for purposes of research and education.**

7. Government authorities who grant licenses to private sector firms for data production should ensure that data originating in the private sector have long-term research utility, are managed according to the same high professional

standards as other types of scientific data, and are ultimately offered to a publicly supported archive for future use.

8. **Governments involved in public-private partnerships to produce data should ensure that commercially produced data are made available for research and education use in a timely manner.**

3. Data and Information rescue

Many types of data, including extant historical data, which have newly appreciated scientific importance for the analysis of changes over time, are not being used for research because they are not available in digital formats, they are in danger of being lost because the media on which they are recorded may decay (*e.g*., paper) or become corrupted (*e.g.*, electronic media), or the software in which they are embedded may be superceded by new types of software. Often, the data are inadequately housed, catalogued, or preserved. This is a particularly acute problem in developing and transitional countries where significant quantities of valuable data, currently only available on paper, need to be digitized before they disappear, but even digital data can be rendered inaccessible if they are imprisoned in outdated software or hardware and not periodically updated.

A similar problem exists with traditional knowledge, which is passed down orally from generation to generation and is rarely recorded. Traditional knowledge about, for example, the healing effects of certain plants or sources of food have proven to be invaluable in pharmacology and agricultural research. However, as many of the communities that possess this knowledge are changing and dispersing, so there is a danger that their knowledge will be lost forever. Similarly, the information may be recorded in languages that are dying or have a declining number of speakers. Recording and validating traditional knowledge raises issues of intellectual property and ethics, which are beyond the scope of the current assessment, but 'rescuing' this knowledge before it is lost altogether is clearly important. The procedures and requirements for doing this have much in common with the rescue of other data of scientific value.

Digitization, data rescue, transcribing, and improved management of traditional or historical data are necessary to preserve these types of data for current and future scientific research. However, the process of data recovery is expensive and often labor intensive, and it requires trained personnel. The digitization of paper records, particularly those that have begun to deteriorate, requires not only technological training, hardware, and software, but also institutional resources to maintain the newly digitized data permanently. There is a need for new and inexpensive methods of data digitization and rescue. This is an area where the public and private sectors could work together. Public-private sector partnerships, could advance the state of the art of data rescue among commercial firms that produce hardware and software for digitization, while at the same time contributing to increased use of and training in information technology in the public or scientific sector. Transcribing and preserving traditional knowledge requires individuals with local language skills, cultural knowledge and the capacity to work with indigenous peoples with full respect for their cultural traditions.

***Recommendations*:**

9. Scientists should inventory major collections of extant data and information and should set priorities for the rescue and permanent preservation of the data and information that are most valuable and at greatest risk.

10. ICSU and its members should draw the attention of scientists, public policy makers, and research foundations to the issue of data at risk and ways to deal with the issue.

11. Scientists or countries that undertake significant efforts in the rescue of data at risk should consider the advantages of public/private partnerships in this effort.

4. Scientific Publications

For most of the last century, the publication of science journals was dominated by learned societies. Increasingly during the past three to four decades, commercial publishers have captured a majority share of the market for scientific publications. With the advent of electronic publishing over the last ten years, new modes of publishing and new organizations have appeared that are changing the landscape and outlook for authors and readers, publishers, and libraries. It is not yet clear which economic and technological models will prove most beneficial to science or which will be economically sustainable.

Scientific publications are increasingly being produced and disseminated electronically. Readers with sufficient Internet and/or institutional support have immediate access to past and current scientific articles, and as a result, despite strains on library and publisher budgets, more people now have access to more scientific information than at any previous time in history. On-line scientific publications are also becoming part of the database of science, with articles providing electronic links both to data used in the analysis and to cited and citing publications. Published data are being re-aggregated into electronic data bases, such as the Petrology Data Base or PetDB (http://petdb.ldeo.columbia.edu/petdb/enterdatabase.htm).

However, at the same time as scientific publications are assuming new roles in research, the economic foundations of scientific publishing are being chipped away. The publishing industry has long been dependent on the scientific and educational infrastructure to provide support for aspects of scientific publication. This included traditional physical libraries, generally but not always at universities, that financially supported publishers by subscribing to publications and storing published journals for future use as a service to the scientific/educational community. In recent years, library budgets have not kept up with journal price increases and the growth of the research literature.

Various elements in the traditional publishing model are being undermined economically by the ubiquity, ease, low cost, and speed of on-line publishing and information access. Individual scientists and institutions are beginning to publish their own work on line – either as a precursor to conventional journal publication (pre-print or e-print) or even at times without recourse to traditional scientific publication. In some scientific communities, such as physics, this has become routine practice and organized central electronic archives are maintained by and for scientists. As a result of advances in information technology, commercial publishers of scientific journals are under increasing pressure to produce their journals and make them available on line at little or no cost. Subscribers are requesting on-line access to new publications and libraries are seeking to purchase collective access to journals. This sometimes undermines and sometimes supplements the traditional financial basis of the scientific publishing industry.

New electronic publishing models, including moves toward 'open access' publishing for readers, raise fundamental questions about all aspects of the traditional scientific publishing process. At the production end, the key issues include the value and need for peer-review, the editorial function, information quality and control, branding, the creation of new journals, and the cost of providing these services. Other issues such as archiving and access are considered in later sections of this report. However, whatever the publishing model, it is clear that, as with any other scientific DI, there are costs associated with both the production and management process. The scientific community is rightly concerned about the escalating costs of scientific publications but there is a danger that, in the transition towards new publishing models, some of the merits of the traditional journal production process might also be lost.

Aside from the information that is normally published in scientific journals, Internet and digital library technologies are facilitating the creation of and online access to other forms of scientific and educational materials. These include digital versions of theses and dissertations, technical reports and courseware. If properly managed, they can provide a valuable resource for research and education. They also have major implications for the role of libraries and the development of virtual institutional repositories.

The transition to new types of scientific publication offers particular opportunities to scientists working in poorer countries. Issues relating to access to data and publications in low income countries are considered in section IV of this report. However, it should be recognized that scientists in the developing world not only need to have access to the scientific literature, but also need publication outlets for their own work and for research that is regionally important.

***Recommendations*:**

12. **Many ICSU Members are directly involved in scientific publishing. The panel recommends that ICSU continue to work with a broad-based organization such as the International Council for Scientific and Technical Information (ICSTI) that includes both the publishing industry and scientists to promote new opportunities provided by information and communication technologies**

and address key issues related to the transition in the scientific publication process.

13. **Journal publishers should encourage authors to make the source data for their articles available in electronic repositories that are stable, widely accessible and professionally managed.**

14. **ICSU should work with its members and relevant bodies in encouraging the coordinated development of digital libraries and their integration with journal publishing and data systems.**

15. **Because of the importance of extending the benefits of digital publications to all scientists worldwide, ICSU should encourage its member organizations to work with the International Network on the Availability of Scientific Publications (INASP) to build cost-efficient and sustainable publishing capacity and journal access in developing countries.**

III. Managing Data and Information for Current and Future Research

Scientific data serve dual purposes. They are a critical component in scientific research and education and they are the foundation upon which future generations of scientists will conduct research. Decisions made—or not made—about data management and preservation today will either ensure that scientists in the future have the data they need for research or will cripple future research because essential baseline and time series data and documentation and monitoring of unique events are no longer available. Once lost, scientific data and observations cannot be obtained again and the research community is permanently impoverished. The growing analytic re-use of data through data mining, meta-analysis, and interoperability or the integration and fusion of data from diverse sources into new data sets suggests that scientists in all countries should give significantly more attention to data management for purposes of research than they have in the past.

A further reason for giving greater attention to data management at this time is the rapid pace of innovation in information and observation technologies and the need to avoid losing data through corruption of the digital record or by imprisoning data in outmoded software or hardware. Because of the need to ensure data quality and integrity over long time periods, data management procedures must take data security, integrity, and routine technological updating into account. Taken together, these various changes in the data environment place a growing responsibility for maintaining the infrastructure of scientific research on data managers.

The panel recommends that ICSU play a major role in promoting professional data management and that it foster greater attention to consistency, quality, permanent preservation of the scientific data record, and the use of common data management standards throughout the global scientific community [16].

1. Professional Data and Information Management

Scientific data and information management can no longer be viewed as a task for untrained amateurs or as part of the routine "clean up" conducted hurriedly by scientists at the completion of a research project. It remains a responsibility of all scientists and should be valued accordingly, but it is also an increasingly important professional activity, one that is essential to the scientific enterprise. Because data centers and permanent archives are now among the most critical components in the infrastructure of science and constitute the legacy that the current generation of scientists will leave to its successors, working partnerships between scientists and data managers will increasingly be required. All scientists need to have some data management awareness, but the use of advanced information technology in scientific data management and dissemination makes it essential that data management be the responsibility of professionals. In some international projects, such as sequencing the human genome, data management is clearly recognized as a fundamental component of the research activity. In other fields, data management can be haphazard at best.

Responsibility for the management of scientific data and information also lies with a variety of institutions. Governments operate data centers and archives, as do libraries, universities and research centers, and increasingly the leaders of research projects make their data available on line to other scientists. Institutional repositories of various kinds have a growing role to play in ensuring good data management. It is notable that in the environmental research arena, operational data collected for applied rather than research purposes are often better managed than data collected in research projects. This is because the operational data are needed for specific purposes and therefore they are maintained in professionally directed data centers. Research data are frequently less consistently managed and preserved and as a result are at greater risk of being lost. This is particularly true in many developing and transitional countries, where scientific resources are more limited and there is little guidance in data management for scientists or data managers.

There is a need for improved management of data in research projects and for dedicated individuals and institutions to disseminate, manage, and archive scientific data and information. Within research projects, data management must be recognized as an essential component in overall project management that takes place in parallel with other research activities. Because of the rapid pace of change in data management technologies over the past two decades and the absence of widespread access to professional courses, most training for scientific data management takes place on the job. Although this can produce good data managers, on-the-job training is of varying quality, and even this is often outside the reach of potential data managers in developing and transitional countries.

Because of the importance of data to scientific research, scientists should be recognized and given credit for the scientific contribution of the data sets that they produce as well as for their own analysis of those data. In some fields, such as bioinformatics, this is already partially the case, but it is not widespread throughout the scientific community. Financial support for data management must become a normal component in all research budgets and part of proposal evaluation criteria.

One consequence of this in a situation where the overall funding is limited is that financial support for other research activities may decline relative to data management costs. However, by providing widespread access to well-documented and managed research data, improved data management practices will provide economies of scale for the scientific enterprise as a whole, now and in the future. In particular, good DI management should significantly improve access to useable data among scientists in developing and transitional countries.

Recommendations:

17. **Recognizing that scientific data management is undergoing rapid innovation and change, information technology specialists, librarians, research scientists, government data producers, donors, and others should be involved in a concerted effort to develop standards and curricula for professional training for scientific data managers.**

18. **Financial support for data and information management should become a routine component in all research budgets and the evaluation criteria for assessing research funding proposals should include evaluation of data management.**

19. **All scientists should receive training in data management as part of their graduate education. ICSU should encourage the development of guidelines for data management by working scientists and their institutions.**

20. **Scientists should be recognized and given credit for the scientific contribution of the data sets that they produce as well as for the analysis of those data.**

21. **ICSU, its members and associated bodies, should raise awareness of the increasingly important role that institutional repositories play in relation to scientific information management and preservation and the need to ensure that such repositories are properly resourced, developed and maintained.**

2. Metadata

If the scientific data and information available today are to serve as a foundation for future research, the data must be adequately documented using metadata based on open and common standards. Metadata are electronic records describing data. At a minimum, they serve as the electronic equivalent of a library's card catalogue ("data about data"), but they are generally far more informative, providing information on the nature and extent of the data, conditions under which they were obtained, and changes that have been made to the data. The use of metadata is a defining characteristic of digital libraries and distinguishes them from other collections of on-line data and information. The use of common standards for metadata across fields of science facilitates the use of distributed data resources both within and across fields of science and provides information that scientists can use in the future to evaluate the data. Metadata should be used as the principal vehicle for documenting known data quality issues.

Metadata standards vary across fields of science. However, in a period in which there is increasing interdisciplinary research directed at problems outside the purview of the traditional disciplines, many scientists look to metadata to obtain information about data in fields in which they have little direct training. It is important that they be able to obtain the basic information they need through on-line, publicly accessible metadata catalogues.

Unfortunately, metadata are not routinely produced for scientific data in many projects, countries, and research fields. International metadata standards can be expensive to obtain and implement (*e.g.,* ISO standards). Because many standards are complicated, preparing metadata can be a difficult and time-consuming process. Once they are prepared, metadata should be archived and made freely available electronically to the larger scientific community in multidisciplinary metadata catalogues on an ongoing basis.

***Recommendations**:*

22. **ICSU should work with its members to promote the development and use of flexible, open and easy to use community standards for metadata. These standards should be interoperable and independent of specific hardware and software platforms. Guidelines for their use should be widely circulated and incorporated into data management training courses.**

23. **Data repositories and publishers should ensure that standard metadata are available for all databases and records.**

24. **Metadata should be archived and made freely available electronically in multidisciplinary metadata catalogues.**

25. **To foster the efficient production of metadata, ICSU should encourage the development of software for writing metadata that can be made available to scientists throughout the world.**

3. Archiving

Permanent archiving is essential for scientific data and information. In some fields, there are institutions that have a clear responsibility for data archiving, but this is not always the case and varies from one scientific field to another. Although it is possible to archive and maintain all digital scientific data and observations, it may be neither practical nor economically feasible to do so. The cost of archiving, which includes the preservation of data integrity and technologically upgrading databases as the software and hardware technologies in which they are embedded are superceded by more advanced versions, is significant and is rarely seen as a high priority activity. However, the adverse effects on future scientific research of not paying due attention to long-term data preservation can be dramatic. For example, data from the first International Polar Year in 1882 would have been invaluable for refining models of global environmental change by today's researchers, but the data have been lost and cannot be recovered.

In some fields, it may be most efficient to have a two-stage structure for data archiving, with active data centers established to distribute data and information and separate data archives set up for permanent preservation of data. This is the approach planned for management of satellite observations in the United States. In other fields, such as with genomic data, the data distribution and archiving responsibilities are shared in the same center(s), which function like virtual data libraries. Because of the expense of data management, active data centers need to examine their holdings periodically and remove unused data and information. Data that are removed from data centers or are to be preserved for longer periods of time should be placed in data archives, which may be at physically distinct sites or co-located. Transitioning data from data centers to data archives will require broad consultation with the scientific community, possibly similar to the peer review of research proposals.

The issues of what to archive, when, and for how long need to be addressed from the perspective of science requirements and should not be driven by financial exigencies. Acquisition of data and information for any centralized data collection activity, whether it be a data center, institutional repository, or virtual library, should be based on defined selection and data quality criteria. Metadata should be freely available for all collected data.

'Physical' libraries have historically been considered as the repositories and long-term guardians of scientific publications. Commercial publishers themselves have also accepted responsibility for archiving publications, although in a competitive commercial world, publishing houses cannot guarantee long-term continuity. The complementarity between the short to medium term role of publishers and the longer-term function of libraries has been a reasonable guarantee of the preservation of the scientific record to date. However, with the advent of new electronic publishing paradigms and the consequent upheaval in scientific information exchange, it is now less clear where the ultimate responsibility for the archiving of scientific information will lie. Many centralized and de-centralized models are being developed but their long-term viability is difficult to assess. In the meantime, there is a crucial role for traditional libraries to play in ensuring continuity in relation to electronic scientific publications. Libraries have a key role to play in setting up and managing institutional repositories that can organize and preserve institutional research output in digital form.

Recommendations:

26. **ICSU and its members should raise awareness of the need for long-term institutional support for data archives both at the national and international level.**

27. **ISCU should foster discussion within the scientific community, including its members and interdisciplinary bodies, on criteria, institutional structures, and models for decision making related to the permanent preservation of scientific data and information.**

28. Ways to reduce the costs of archiving, such as sampling from extant data bases or establishing multiple classifications to prioritize levels of archival support, should be examined by the scientific community.

4. Privacy and Security: Ensuring the Integrity of Data and Information

The ease of combining or integrating electronic research data obtained from multiple sources heightens the need for protection of individual, national, and corporate privacy and confidentiality in scientific databases. By combining data from multiple sources, the integration of various data sets for research may reveal previously private information on individuals, states, and companies. Both research and management of data on human subjects require special procedures so that individual privacy and confidentiality are not violated. In many countries, the privacy and confidentiality of scientific data and information are subject to legal regulation and restrictions. But national legislation or regulation of data privacy differs from country to country and scientists conducting research projects that involve residents of multiple countries may have to respond to multiple types of privacy and confidentiality restrictions covering the same data set. Developing countries often have no, or very limited, regulations on personal data use, which can make the data particularly vulnerable to abuse. There is a need for more widespread discussion and understanding throughout the scientific community of issues related to privacy and confidentiality of research data.

The Internet, the very technology that permits broad and inexpensive dissemination of scientific data, is susceptible to manipulation if network security and data integrity are not maintained. The management of scientific data and information necessarily involves explicit procedures for system security so as to maintain data integrity and prevent the leakage, falsification, or elimination of data in databases that are accessible via the Internet.

Data security and integrity are not only threatened by external manipulation. They can also be compromised by errors or viruses in the system configuration files or in the commands used to check the operating system. Database or system corruption may be the result of malicious intervention in the operating system or its software by an outside force. It can also be the result of non-intentional deletion or substitution of records. But whether the corruption is inadvertent or intentional, it can cause serious problems for the integrity of scientific data, particularly when the data are maintained over long periods of time.

***Recommendations*:**

***29.* ICSU should work with its member unions and other partners to conduct a substantive review of the new ethical and regulatory issues related to use of individual and population data, including the integration of data from multiple sources and variations in national legislation.**

30. **Data security and integrity must be addressed in the context of procedures for data management. A professional information technology staff is required for both archives and data centers if they are to maintain adequate system and database security and integrity.**

31. **Data that are disseminated by scientists or as part of research projects should have metadata that describe the security and integrity measures employed in the collection and management of the data.**

5. Interoperability

The assumption that future scientists will be able to use data collected over long time periods and to integrate data from disparate sources to create new datasets is dependent upon interoperability of the data, software (including both data base and analytic software), and hardware. Coordination in standards development and the use of commonly accepted standards are needed to promote data interoperability, so that data collected in different countries, in different time periods, using different software and hardware configurations, and across different disciplines can be integrated.

Recommendations

32. **ICSU and its partners should "promote interoperability principles and metadata standards to facilitate cooperation and effective use of collected data and information" as recommended in the Science in the Information Society *Agenda for Action*.**

IV. Data and Information Access and Dissemination for Scientific Research

Science has advanced over the years in line with continued improvements in the dissemination of data and information. If science is to further advance throughout the world, scientists must have equitable and affordable access to all the scientific data and information they need for their research. Stable systems for providing universal access to quality data must be developed and maintained.

Limitations on access to scientific data and information can be legal, technological, financial, or institutional, but whatever their origin, they are borne most heavily by scientists in the developing world.

1. Systems for Dissemination of Data and Information

There is growing concern about data quality, security and ongoing integrity of the electronic data base in distributed personal or project web pages maintained outside the institutional structure of a professionally managed data centre or repository.

The ease of disseminating data via the Internet encourages individual scientists to make their research data available to others through personal or project web pages. Many distributed or federated data systems have sprung up that direct scientists to these personal, project, or departmental data servers and provide information on how related data can be found. These data confederations appear to be a valuable means of dividing data management responsibility and assigning it to the individuals who are most knowledgeable about the data, yet little is known about the long term viability of such alliances or their constituent parts. There is no means of ensuring that data managed by individuals or research groups in voluntary distributed systems on the Internet will continue to be available and well-managed over time. Data that depend on a single scientist or a research project, or even a single server for their availability will not be permanently available without periodic expenditures for management and upgrading. Scientists (or their graduate students) inevitably move on to other projects or they retire and cannot maintain support for on-line data from earlier projects. In this regard, the web, for all its advantages, becomes merely a temporary dissemination mechanism; it can only be used effectively if there is an institution or series of institutions that provide continuity in DI provision.

Recommendations:

33. **ICSU should bring together representatives of voluntary data confederations to discuss what has been learned over the past ten years about what contributes to the success of voluntary data confederations, what undermines them, and what must be done to preserve and enhance access to the data in the future.**

2. The digital divide

There are both resource and technical limitations to data access in many parts of the world that not only make it difficult to conduct research, but also interfere with the collection of new data. This problem, which has been labeled the "digital divide," is most evident in low- and medium-income countries, but it has an impact on scientists throughout the world because it limits potential research and access to data in many regions. All scientists who want to use the data and information resources currently available on line require hardware, software, Internet access, and technical capabilities, yet because of the expense of this infrastructure, these resources are not routinely available to scientists in all countries. This limits research, data management, and participation in scientific collaborations. But even where technological resources are available, many scientists are not aware of the utility of on-line sources of scientific research data or the research potential in secondary analysis of extant data because of previous scientific isolation. Unnecessary duplication and consequent waste of valuable resources could be avoided by better communication and information management within the international scientific community.

A major problem for scientists in low-income countries is their lack of access to scientific publications, both as a means of learning about research in other parts of the world and as an outlet for their own research results. Without the opportunity to publish their work, scientists in low-income countries have few opportunities to obtain feedback from others in their field. Even scientists in countries with good access to international scientific publications have only limited access to research findings of scientists in the developing world. The fact that most scientific publications are in English exacerbates this divide, and there is a danger that new electronic publishing paradigms such as "author pays" will extend it even further.

***Recommendations*:**

34. Key issues relating to the digital divide in scientific data and information are addressed in the Science in the Information Society *Agenda for Action*. The panel strongly endorses the actions relating to the provision of high-speed internet access, human capacity building, electronic publishing and initiatives to make scientific information more accessible. ICSU should work with its members and partners to encourage the implementation of these actions.

35. Greater emphasis should be placed on the need for good data and information management in those countries where the scientific data infrastructure is being constructed. ICSU bodies, such as the International Network for Access to Scientific Publications (INASP) and CODATA, have a key education and training role to play in these countries.

3. Equitable Access to Data and Information

Scientific progress relies on full and open access to data and on the open disclosure of research results in the scientific literature. A strong public domain for scientific data and information promotes greater return from the public investment in research by stimulating innovation and more-informed decision making. Principles of open access to scientific data and information can be applied to research data, metadata, or scientific publications, although the specific issues vary with each.

<u>BOX – "Open Access"</u>

["Open access" in relation to scientific data and information means different things to different people. In this regard, an important distinction can be made between data (excluding scientific journals and books) and scientific publications.

"Full and Open Access" to data implies equitable, non-discriminatory access to all data that are of value for science. It does not necessarily equate to immediate access or 'free of cost' at the point of delivery, although this is certainly the ideal in many situations, particularly with regard to publicly funded data. Data should be made available with minimal delay but a short 'privileged access' period for

original data producers may be justified in some situations. Excessive charging for data that is by definition discriminatory against some scientists is clearly contrary to the principle of full and open access but some cost-recovery is not necessarily excluded.

"Open Access" as used in relation to scientific publications implies online access without access charges to readers or libraries, i.e. free of any cost at the point of delivery. "Open access" publishing models necessitate that all production, management and dissemination costs for publications are paid for 'upstream' with no direct subvention from the user. In some cases this can be equated to an 'author pays' as opposed to 'reader pays' model.

Universal and equitable access to scientific publications means ensuring equal opportunities both to publish and to obtain scientific information for all scientists wherever they are located. It does not necessarily imply without any cost at the point of delivery and 'open access' publishing is only one of several models that can be employed to promote universal and equitable access.] **end box**

Because of the scientific and economic benefits of being able to use or adopt another scientist's data for research, and the need to be able to test and retest hypotheses using the same body of data, science has long been best served by a system of minimal constraints on the availability of data and information. However, recent trends towards the appropriation and privatization of scientific data and databases (such as genetic databases), as well as limitations in access to digitized data (e.g., anti-circumvention measures) pose a serious challenge to science.

There is considerable consensus at the international level behind the concept of providing open access to research data and many governments have endorsed the principal of open access in different international fora, including the World Summit on the Information Society (December, 2003). Recently, both the multilateral Group on Earth Observation (July, 2003) and the OECD (January, 2004) have endorsed the concept of open access to research data. However, at the same time as governments are endorsing the principle of open access, limitations on access and fair use of data for scientific purposes are being negotiated in bilateral and multilateral agreements. For example, the recent EC Database Directive (96/9/EC) was developed without due consideration of scientific needs and may potentially deprive scientists of access to valuable data.

The scientific publishing process, unlike public funding for scientific data collection and management, has traditionally been based on a variety of not-for-profit and commercial business models. Scientific journals have not been freely and openly available to all potential consumers, but were restricted to those who could afford them or had access to institutions (such as libraries) that could afford to acquire, store, and preserve them.. The move toward commercialization and profit generation from scientific data is in many ways the opposite of what is happening in relation to scientific publications.

New 'Open Access' publishing models are based on the concept of providing free electronic publications at the point of delivery. This does not mean that there is no cost associated with them nor that they are necessarily 'not for profit.' Instead, it means that any costs or profit margins are transferred to the author or some other subsidizing institution or body at the point of production. The economic models for open access publishing are still under development and the roles within them of existing traditional publishing houses, scientific societies, and individual societies are all subject to speculation. However, the move towards open access publishing has considerable impetus and representatives of a number of major research institutes in Europe and the United States have issued declarations supporting open access to scientific publications (the Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities, 2003, and the Bethesda Statement on Open-Access Publishing, 2003)

One of the major drivers toward "open access" publishing models is the prohibitive cost of printing, mailing, and storing paper journals. These costs are such that they are currently limiting access to journals in even the richest Western research institutions and are completely inhibiting access in poorer countries. However, many 'traditional' commercial publishers and scientific societies are sensitive to these issues and are themselves developing new approaches such as differential pricing and free electronic access to publications after a limited time period to enable broader access.

There is currently a 'mixed market' for scientific publications and some uncertainty as to where this will eventually lead us. ICSU itself has been promoting universal and equitable access to scientific information rather than exclusively focusing on 'open access' publishing. In this regard, what is crucial is that the scientific communities in the developing world are fully included in the discussions on new publishing paradigms and, even more important than the economic issues, that the responsibilities for quality control, archiving and meta-standards of new electronic publications are clearly defined.

Recommendations:

36. **ICSU should continue to stand firmly behind the principle of full and open access to scientific data.**

37. **With regard to scientific publishing, ICSU should ensure that the principle of universal and equitable access to scientific publications is upheld.**

4. Who Pays?

Obtaining and preserving the scientific database for research by current and future generations of scientists will be expensive. The growing capacity of scientists to analyze complex data sets and to redistribute their data to others does not mean that the costs involved will be negligible, particularly when data are to be maintained for decades to centuries. Collection of data, preparation of metadata, institutional assumption of responsibilities for data dissemination, and permanent archiving will add to the expense of conducting research projects and maintaining the larger research infrastructure. At the same time, the possibility of using existing data for research instead of collecting new data for every project could reduce the cost of research for many investigators.

There are several economic models for providing scientists with access to data for research and education. They include, among others, (1) free and open access to research data by scientists, with financial support for data dissemination and preservation assumed by others, including government science agencies and private foundations; (2) open access to scientific data for research and education for the cost of reproduction (that is, recovering the operational costs of data dissemination); (3) free and open access to metadata, and cost-recovery pricing for data (or data licenses) in order to support the full data infrastructure. When the this last approach is employed by a private sector firm, the financial charges for data must be sufficient to recover all investment costs and to make a profit for investors. An important variation on this includes licensing for scientists to use specific bodies of data at reduced cost.

Data production and management are costly. However, the practice of charging for data access discriminates between scientists with and those without adequate funding to purchase or license data. In practice this inequality exists even within national scientific communities where scientists in well funded fields of science or institutions have better prospects of obtaining support for data production and management than those who are in fields in which funding is more difficult to obtain or institutions that are chronically under-funded. Even more importantly, it severely limits access to research data by scientists in low-income countries.

Scientific research has long been viewed as an open process that requires scientists to make their data available to others for testing and retesting hypotheses and conducing new research based on recent findings. With purchased and especially licensed data, scientists may be prohibited from redistributing their data to others. This restricts the use of the data in the short term and impoverishes the science in the long term because of the difficulty in retesting research findings and conducting related research.

As suggested earlier, there is considerable uncertainty about the economic viability of open access publishing models. Although electronic publishing and a 'not for profit' philosophy should reduce the overall cost of publications, there will still be significant costs associated with the production and management of scientific publications. Some 'open access' publishing models can be equated to 'author pays' and for many scientists, particularly in resource poor countries, such a model is not currently viable. The principle of affordable universal and equitable access to scientific information must be embedded in the economic models that are being developed for 'open access' and other e-publishing initiatives.

Recommendations

38. Ensuring the long-term accessibility of increasing quantities of scientific data and information will necessitate increased public (and private) investment in data management and long-term institutional support. ICSU and its members should explore various solutions to meet the financial challenge of providing full and open access to scientific data and universal and equitable access to publications.

5. Intellectual Property Rights (IPR)

As data and information have increasingly become digital and electronic and consequently easier to manipulate, they have become more important not only to scientific research but also to a host of other applications and uses. They are considered both as input to research, policy, decision-making, and management, but also, like digital data in entertainment and commerce, as property in and of themselves. As discussed throughout this report, full and open access to data has been crucial for the development of science. Science is best served by a system of minimal restraints (*e.g.,* those based on privacy considerations) on access to and use of data. Recent trends towards the appropriation of data, such as genetic information, and the protection of databases under *sui generis* regimes, as well as limitations to the fair use of digitized data (*e.g.,* anti-circumvention measures) pose serious obstacles to full and open access to data for scientific purposes.

Intellectual property rights can promote innovation by providing incentives and are, in some circumstances, a useful stimulus to responsibility and data quality. They do not necessarily have to conflict with access. An owner of a data base can permit full and open access via generous licensing agreements. However, restrictive use of IPR can have extremely negative effects on scientific research. Appropriation of scientific data and information for private gain can also raise ethical issues, as, for example, when it leads to control of the human genome sequences.

As with many other issues relating to scientific data, the restrictive use of IPR has a disproportionately adverse effect on science in developing countries. Influential treaties are being negotiated in international arena such as the World Intellectual Property Organization (WIPO) and World Trade Organization (WTO), which have a direct effect on the international exchange of data. Such treaties subsequently become embedded in national legislation. The needs of science must be considered at both the international and national levels in order to ensure full and open access to data. Even in richer countries, the scientific community is ill-prepared to participate in these negotiations; in poorer countries with more limited scientific activity this is an even greater challenge.

In collaboration with member organizations, ICSU should investigate what mechanisms it might institute to ensure that science is fully represented in international treaty negotiations that might have an impact on access to data for scientific purposes. This may be facilitated by the fact that ICSU has consultative status with key UN bodies. Several national academies have already expressed concern and conducted enquiries on the restrictive use of IPR in science.

Recommendations:

39. **Governments and other bodies concerned with international and national policy development, should ensure that IPR legislation recognizes the value of ensuring full and open access to data for scientific research and education purposes.**

40. ICSU should investigate, with its members, appropriate mechanisms to ensure that science is fully represented in international treaty negotiations that might have an impact on access to data for scientific purposes.

V. Review of ICSU bodies

1. The review process

It was not within the remit of this Priority Area Assessment to carry out an in-depth review of the past performance of individual ICSU bodies. However the Terms of Reference charged the panel to examine current activities, identify gaps, overlaps, and synergies of existing activities and propose coordination and responsibilities for specific bodies, now and in the future. In considering the recommendations relating to specific bodies below, it is important to note that these recommendations do not constitute a judgment on the past performance of those bodies but are, rather, an assessment of their future role and potential relative to the strategic priorities identified in the previous sections of this report.

With this forward-looking perspective in mind, the key ICSU bodies that deal with scientific data and information were asked to provide written responses to specific questions defined by the panel. The three bodies that are exclusively concerned with data and information issues were also invited to nominate a representative to meet with the assessment panel. A full list of the bodies consulted in this assessment is attached as annex 2.

2. Specialized data and information bodies

ICSU has established three bodies, the Committee on Dissemination of Scientific Information (CDSI), the International Network on the Availability of Scientific Publications (INASP), and the Committee on Data for Science and Technology (CODATA), that specialize in scientific data and information issues. Each of these bodies were invited to give both written and oral evidence to the panel. The panel appreciated the cooperation of each of these bodies in responding very positively to this invitation and felt that they made a valuable contribution to the overall report, beyond the specific issues detailed in this chapter.

Each of the three bodies was asked to provide a brief written statement on current and planned activities, their relationships with other ICSU bodies, their activities within and outside the ICSU family, and their strategy for the future. This statement was considered by the panel prior to its meeting with representatives of the individual bodies. On the basis of the written statement, the panel formulated four questions around which the subsequent discussions with each representative were focused:

1. How is your strategy evolving to meet future data management and policy needs?
2. How do you prioritize (i.e. choose among) potential actions/activities?
3. What is the value of ICSU to your activities? How do you collaborate with other bodies in the ICSU family?
4. Is there a major responsibility that ICSU should assume that no one else can or will do?

The Committee on the Dissemination of Scientific Information (CDSI)

CDSI is an advisory committee to ICSU and is charged with providing advice to the ICSU family about scientific publication, new developments in information technology, access to data and information, and related legal issues. It has seven individual members (two of whom are ex-officio) appointed by ICSU, and its core funding (~$40,000 p.a.) and administrative support are provided by ICSU. The committee superseded the ICSU Press Publishing Service in 1989**,** which itself grew out of the ICSU Press that had been created as the publishing house of ICSU in 1983.

The CDSI Chairman is Professor Eric Sandewall (Sweden), appointed to this position in 2002. Professor Sandewall attended the second meeting of the PAA panel.

Activities and strategy

CDSI's recent activities included organizing conferences on electronic publications in partnership with UNESCO and working with CODATA to oppose an European Union database directive that could have a negative impact on science. Its future plan is to focus on four main areas:

1. Research knowledge management
2. Encouraging and facilitating the use of multimedia in scientific publications
3. Encouraging the production of 'lay summaries' in association with scientific publications
4. Monitoring restrictions on access to publications, including intellectual property issues

Topics two and three entail limited pilot activities in Latin America. Topic one represents the most substantive activity of the committee. This theme incorporates a number of topics, such as quality assurance and peer-review mechanisms, arising out of changing publication practices related to open access initiatives. The goal of CDSI is to promote a clear understanding of these issues and informed debate at the policy level rather than to advocate or promote particular options.

Assessment

The panel considered the knowledge management theme to encompass a number of interesting topics in which CDSI's advisory role could potentially be beneficial. The combination of foresight and timely communication with the ICSU family was attractive. However, there was some concern at the overall breadth and ambition of the committee's proposals, particularly in the light of its limited resources. Although

CDSI is formally an ICSU advisory body, its proposed activities appeared to have been developed in isolation from the ICSU membership and Executive Board.

It was recognized that many ICSU member organizations are dependent on their publication activities for income and that there is a need for the international scientific community to discuss the issues raised by the transition to electronic publishing. However, it was clear that neither CDSI nor ICSU had the mechanisms in place to discuss the costs and benefits with ICSU members, and it was recognized that there were other bodies with broader and more representative membership, such as the International Council for Scientific and Technical Information (ICSTI), that might be better positioned to take on this role. In this regard, the broader membership and better resources of ICSTI relative to CDSI are an important consideration.

Recommendations

41. CDSI should be disbanded. ICSU should look to ICSTI for the advice formerly expected from CDSI.

42. ICSU should consult with ICSTI to ensure that it is prepared to include the ICSU membership in the debate on these issues. The current Chair of CDSI should participate in this discussion.

43. Channels for regular communication among ICSU advisory structures, such as CDSI, the ICSU Executive Board and the membership need to be improved.

The International Network on the Availability of Scientific Publications (INASP)

INASP was created in 1992 with the goal to "improve world wide access to scientific information". It was originally a program of ICSU Press, but at the time of this assessment, INASP had established itself as an independent company limited by guarantee and was in the process of being registered as a Charity in the UK, which could be formally recognized as an ICSU interdisciplinary body. It operates out of a Secretariat based in Oxford, which includes nine full-time and several part-time staff. The total annual budget is approximately £2.2 million of which a very small amount (~£6,000 to 20,000 p.a.) comes directly from ICSU, with the rest coming from more than 20 different foundations or government development agencies.

The Executive Director of INASP is Dr Carol Priestley, who attended the second meeting of the PAA panel. She is supported in her direction by an international scientific advisory board.

Activities and Strategy

INASP has established an extensive co-operative network of over 3000 partners, including science information professionals, librarians, science managers and academics in the developing world. It responds to requests for assistance in all aspects of scientific literature publication and dissemination, and it assists development and funding agencies in the establishment and implementation of information-related programs in low income countries. Although its original focus was on the compilation of directories and the provision of scientific publications at affordable cost, mainly from the North to the South, it has extended its work considerably beyond these initial activities. INASP defines its work as needs-driven, with priorities being set by scientific communities in low income countries and annual regional coordinator meetings. This has led to an increasing emphasis on training in the use of ICT's and information resources and to the development of local publishing capacity using electronic media.

INASP reported collaborations with 17 out of the 26 ICSU unions and is developing closer links with ICSTI and CODATA in regard to policy issues. INASP itself is very much an operational as opposed to policy body and, for the immediate future, its focus will continue to be on facilitating access to scientific information and on capacity building. Although the demand for INASP's services currently exceeds its capacity to deliver, it would like to expand its activities into the former Soviet newly independent states, where there are relatively strong scientific communities that are currently excluded from the international scientific information networks.

The INASP Executive Director proposed that ICSU consider establishing a "community of practice" among its members. In its simplest form this would be a virtual network of individuals responsible for scientific publications/information management within member organizations. Such a network could provide a focus for communication with INASP (and other ICSU bodies) and the exchange of examples of best practice.

Assessment

INASP is clearly performing an extremely valuable service to a very high standard in a highly efficient manner. Its rapid growth since its establishment in 1992, based on continuous assessment of customer needs and clarity of purpose, was very impressive, as was its capacity to raise the necessary financial support for its program. INASP had adapted well to the rapidly changing context for scientific information exchange and it would need to continue to do so in the future. The establishment of INASP as an independent legal entity was very timely.

Given the historically close relationship between CDSI and INASP, the operational activities proposed by CDSI for Latin American might be continued under the auspices of INASP.

Recommendations

44. INASP should be formally recognized as an ICSU interdisciplinary body.

45. ICSU members who do not already collaborate with INASP should be encouraged to do so.

46. INASP should be encouraged to continue its collaboration with CODATA and ICSTI in a way that complements, but does not distract from, its established aims.

47. The INASP proposal that a "community of practice" be developed among information professionals from ICSU member organizations should be encouraged, with the context of the proposal for an international Scientific Data and Information Forum.

Committee on Data for Science and Technology (CODATA)

CODATA was established as an ICSU interdisciplinary body in 1966. Its principle objectives are improvement of the quality and accessibility of scientific data, as well as the methods by which data are acquired, managed and analyzed; the facilitation of international cooperation on data issues; the promotion of awareness of data issues in the science and technology community; and consideration of data access and intellectual property issues. Its core funding (~200,000 euros p.a.) comes from member subscriptions. There are 23 national members and 15 international union members. A mixture of policy and operational activities are performed mainly by special task forces, which are established and/or renewed at biennial general conferences. Additional funding for *ad hoc* activities, such as workshops, is obtained via the ICSU grants program ($190,000, 2002-2004) and other sources, such as UNESCO and the US National Academy of Sciences. These activities are coordinated by an Executive Director based in Paris, who is supported in her direction by an Executive Board.

The President of CODATA, Professor Shuichi Iwata (Japan), and the Executive Director, Kathleen Cass, attended the second meeting of the PAA panel.

Activities and Strategy

There are currently eight CODATA task forces, ranging from a long-established group on fundamental constants to new groups on preservation and archiving in developing countries and global species data networks (see www.codata.org). In 2002 CODATA launched a peer-reviewed electronic journal specifically focused on data management issues. At the policy level, CODATA participated with several other ICSU bodies in an *ad hoc* group on data and information that produced a document on principles for dissemination of scientific data (June 2000) and was involved with CDSI in lobbying against the EC database directive. More recently it worked closely with the ICSU secretariat on the World Summit on the Information Society (WSIS), including the production of the ICSU/CODATA *Agenda for Action*, Science in the Information Society (March, 2003, see annex 4). CODATA was also involved in the development of an OECD ministerial declaration on access to data from public funding (January, 2004).

In terms of future directions, CODATA plans to continue its work in relation to WSIS, particularly on data access and public domain issues, including economic analyses of different access models. It also hopes to develop closer links with the

ICSU unions on multi-disciplinary data issues and standards and to develop training courses on data management for developing countries. The overall emphasis will be on scientific data management and policies for the benefit of society.

Assessment

CODATA has a mixed portfolio of policy, research, and data management activities, some of which are long established, others of which are newly developing. Many of these activities appear to fill a long term need and some of the newer activities, such as the electronic journal are innovative and forward looking. The recent direction of CODATA is to be complemented for its energy and foresight. However, the overall impression remains one of a collection of opportunistic activities. The future CODATA focus and strategy are not clear. Although the panel recognized that the mechanism for selection of CODATA task forces is responsive to 'bottom up' initiatives from members, it felt that this approach needs to be joined to a long-term strategic plan. Given its limited resources, some of the historical CODATA activities should be terminated or principal responsibility for them transferred elsewhere, if the organization is to initiate new activities. The CODATA strategy should address, membership, funding and other resource issues, which are currently very limiting, with an emphasis on funding work with low-income countries. Current attempts to diversify CODATA funding sources are strongly encouraged.

CODATA has a crucial role to play in advising the scientific community. CODATA policy activities in partnership with ICSU members and other organizations are commendable and in some instances, such as the WSIS and the OECD declaration, they have clearly had a valuable impact. There was some concern that the potentially valuable CODATA/ICSU guidelines on dissemination of data had apparently not been formally considered by ICSU or adopted by the broader ICSU membership. This suggests that ICSU needs to improve the lines of communication with CODATA and the ICSU membership on data policy issues.

There is a valuable and expanded role for CODATA to play in the future in relation to scientific data management and policy issues, particularly in developing countries. CODATA should develop a clear long-term strategy, focusing on the key international data management and policy issues, with a strong emphasis on the needs of low income countries. It could serve a valuable role in data policy and management in low income countries that is parallel to that of INASP in the field of scientific publications.

CODATA is making good efforts to develop its links with other ICSU bodies, such as INASP and ICSTI, and with ICSU members, and is attempting to diversify its funding sources. Both of these are essential if CODATA is to play a major role in data management and policy in developing countries in the future. It will also be necessary to expand CODATA membership, and in this regard, the almost complete absence of European members, needs to be addressed.

Recommendations:

In the context of the development of a long-term strategic framework and international Scientific Data and Information Forum:

48. **CODATA should develop a clear long-term strategy that focuses on key international data management and policy issues and should place a strong emphasis on eliminating the digital divide.**

49. **The lines of communication between CODATA and ICSU need to be improved. CODATA should continue to develop a closer working relationship with ICSU bodies such as INASP and ICSTI in areas where there are complementarities and clear added value.**

50. **CODATA needs a more inclusive worldwide membership. ICSU should encourage those of its members who are not currently affiliated to CODATA to reconsider this position.**

51. **While developing its long-term strategy, a short-term CODATA focus on implementation of relevant aspects of the Science in the Information Society *Agenda for Action* and preparation for the World Summit on the Information Society II is both appropriate and valuable.**

3. Environmental data bodies

ICSU, either alone or in close cooperation with other partners, has been involved in the establishment of a number of bodies that have a major role to play in the production and management of environmental data. All of these were reviewed immediately prior to the current assessment as part of as separate Priority Area Assessment focusing on "Environment in Relation to Sustainable Development." The earlier PAA considered these bodies from the perspective of the scientific user, that is, are they serving the needs of the environmental research community? This panel focused on issues of scientific data management and policy rather than scientific research. However, because of obvious overlaps in the remit of the two panels, all the relevant information collected and produced in relation to the PAA on environment and sustainable development was made available for the current assessment. In addition the chairpersons of the two panels met at the outset of the current assessment in order to ensure continuity and minimize overlap for the two assessments.

A sub-group of the current PAA panel with environment expertise analyzed all the information that was transferred from the earlier assessment and defined where further input might be useful. With regard to the terms of reference for the current exercise, written responses were invited to five specific questions as follows:

1. What does your organization do to promote the adoption and use of common standards and formats for data and metadata in the scientific community?
2. What is your organization's view of the cost of environmental data and of pricing policies? What is the influence (if any) of the changing balance of public and private sector sources of scientific data and information?

3. What is your view of science data archiving policy, for example the time period for data archives and the responsibility for data archives and data rescue policy?
4. What is your organization doing to assist scientists in all countries to gain access to scientific data?
5. In your view how active should ICSU be in developing and promoting policies on the management of data and information? Can improvements be made to the organization or structure of ICSU to enable it to be more active and effective in data and information policies?

The relevant bodies were also invited to describe their future strategies and discuss how they related to other ICSU bodies and members. The written responses, together with the additional information from the earlier PAA exercise, were considered by the entire PAA panel to develop the recommendations detailed below.

Monitoring and Observation bodies: The Global Ocean, Terrestrial and Climate Observing Systems (GOOS, GTOS, GCOS) and the Integrated Global Observing Strategy-Partnership (IGOS-P)

ICSU is a sponsor of the three Global Observing Systems and a partner of the Integrated Global Observing Strategy. In each case ICSU is one of several partners, the others of which are all UN intergovernmental organizations. ICSU currently contributes $60,000 p.a. to each of the three observing systems and participates actively in the IGOS-Partnership meetings.

In their particular areas of activity, the three global observing systems are responsible for coordinating existing and planned operational activities of observing systems with the needs of scientific research programs. IGOS-P has been developing plans for focused and integrated observing systems in several thematic areas, including water, carbon and geohazards. In July 2003 an Earth Observation Summit growing out of a meeting of G-8 nations was held with the purpose of identifying a group of countries committed to developing and maintaining a global observation system that is international, comprehensive and sustainable. The summit established an *ad hoc* Group on Earth Observations (GEO), of which ICSU is a partner, to prepare a 10-year implementation plan for a global observation system. This plan is to be presented at a ministerial meeting in early 2005.

In its coordinating role, the IGOS-Partnership promotes the use of shared data standards and full data access. It also advocates the production and maintenance of standardized metadata and recommends that suitable archive facilities be developed for long-term data preservation. The three observing systems have adopted similar policies and are involved at the management level in standardization and ensuring access to the data. They also encourage data archiving through international data centers, although the ultimate responsibility for this is not clear. The panel agrees with the conclusions of the PAA on the Environment in Relation to Sustainable Development that coordination of observing system activities with the scientific research community and exchange of best practices between the different systems could be improved.

As a founder and sponsor of the World Data Centers as well as the major international scientific research programs that use Earth observation data, ICSU was considered by the observing systems to have a leadership role to play in ensuring good data management and policies in the future.

Recommendations

52. ICSU should play a stronger role in the Global Observing Systems by fostering the development and implementation of appropriate policies and data management procedures and representing scientific data user needs. There is a need for both operational archives for dissemination of data and a strategy for long-term data preservation.

53. ICSU should be actively involved in the *ad hoc* Group on Earth Observations (GEO) as a representative of the international scientific community, advocating appropriate policies for scientific consultation, data collection, data access, and professional management of Earth observation data.

Federation of Astronomical and Geophysical Data Services (FAGS)

FAGS was established in 1956 and includes 12 permanent data services, each operating under the authority of one or more of the three sponsoring unions: Astronomy, Geodesy and Geophysics and Radio Science. The services are maintained nationally and their role is to collect, analyze, interpret, and disseminate observations, information and data related to astronomy and geophysics. The services are independent, but ICSU and the union co-sponsors contribute to the overarching coordinating function which is performed by the Council of the Federation. The Council meets annually and considers reports from each of the services and small requests for funding. When FAGS was created, the funding from ICSU was a major source of income for the services, however, ICSU funding in 2003 ($25,000) now represents only a tiny fraction of the overall funding for each service.

FAGS has a relatively 'hands off' approach to data exchange standards and formats, with each data service using its own scientific standard for formats. The principal goal of the services is to prepare scientific analyses and interpretations of change. These interpretations often form the basis for the definition of global geophysical standards. The services also make data freely available for scientific purposes and retain data permanently. Some of the services run training courses.

Aside from the ICSU financial contribution, the major role proposed for ICSU in relation to FAGS activities was related to data policy and the establishment of an international framework for long-term data production and management

FAGS was developed about 50 years ago when data issues were of a very different scale. It has performed an important coordination role in relation to specific fields of data collection and analysis. However, FAGS is not active with regard to key data policy and management issues, such as development of shared standards, which are identified in this report. It does not interact with other ICSU bodies, such as CODATA and the WDC, on these issues.

Recommendations

54. FAGS should be no longer be an ICSU interdisciplinary body. The responsibility for FAGS should be devolved to its user constituency, the three co-sponsoring Unions.

This is in accordance with the recommendations concerning FAGS in the PAA of the Environment in Relation to Sustainable Development. In addition, it was noted that the individual FAGS services may be important contributors to the proposed Scientific Data and Information Forum.

Panel on World Data Centers (WDC)

The World Data Center system consists of over forty designated World Data Centers (WDC), which collect, manage, and distribute a wide range of defined geophysical, solar and environmental data. The World Data Center program was created during the International Geophysical Year of 1957-1958, and in 1968, ICSU established a Panel on World Data Centers to coordinate and monitor the activities of the centers. Funding for the panel and its small secretariat is largely provided by the US Government, with specific projects occasionally supported by the ICSU grants program. Financial support for specific WDC's is obtained from a variety of sources, usually national governments where the centers are located.

The World Data Center system is decentralized and consists of largely discipline- or field-specific data centers. It is the policy of the WDC's to make data freely open and available for scientific research. All World Data Centers are committed to the long-term retention of their holdings, although there are no WDC standards for permanent data management and they do not necessarily manage permanent archives. New WDC's are nominated by national academies of science or other ICSU members and approved by the ICSU WDC panel. Although neither the system nor the panel promotes the adoption and use of specific standards and data formats, individual World Data Centers generally comply with national and international standards and work with suppliers and users to define preferred formats.

In terms of its future strategies, the WDC system has recognized the need to adapt to changing technical requirements for data management and dissemination. The WDC panel plans to establish mirror sites in developing countries to replicate the paleoclimate, geophysical and solar-terrestrial holdings of specific World Data Centers elsewhere, thereby building capacity and improving the speed of access to data in countries of the mirror sites. It also proposes to continue its interactions with other ICSU bodies, such as CODATA, in order to monitor international policies on data access.

From the perspective of the WDC system, it was proposed that ICSU should play a lead role in data policy development and implementation both in relation to ICSUsponsored research programs and more widely. The WDC system also stressed that ICSU should be promoting full and open access to data in international, inter-governmental fora, such as WSIS.

Assessment

Given the significant changes in research and in data access, availability, and management since the World Data Centers were first established in the 1950's, the panel believes that it is time for ICSU to reexamine its entire data center infrastructure. The current WDC program is based on the concept of nationally supported, field-specific data centers that are physically distributed and not linked electronically. This approach, with its emphasis on broad participation in international data management activity, should be continued, but it should be expanded geographically. The WDC panel has played a valuable role in 'lightly' coordinating the activities of the data centers and in fostering good data management practices. It has also been a strong advocate for full and open data access in international policy discussion.

Looking to the future, however, the panel believes that there are policy and management issues that the WDC's environmental data management activities are not addressing. For example, as identified in the PAA on Environment in Relation to Sustainable Development, there is no obvious strategy for depositing and managing data in the World Data Centers from either ICSU's own global environmental change programs or the Global Observing Systems. There is also no overarching strategy for the development and implementation of a permanent archive and retrieval system for globally relevant international scientific data sets. The existing distributed system of WDC's could be used to fulfill these functions, but planning to assume this responsibility needs to be driven by an updated and widely shared strategic vision that goes beyond environmental data and information and has broader international and disciplinary participation than the WDC's currently afford. Such a system would require a strong central policy and monitoring body that had effective working relationships with other major data and information stakeholders (scientists, governmental and non-governmental organizations, software producers, leading data repositories, and others). Some activities, such as the maintenance of permanent archives, the digitization of existing data, and the establishment of a common global distributed data capacity for scientists, are also likely to require substantial new resources and personnel. Some of these issues were identified in a report on modernization commissioned by the WDC panel in 2002, but the report does not address the larger issues of re-structuring the system.

Recommendations

55. **ICSU should re-examine its entire data center infrastructure in light of technological and scientific changes in data collection, use, and management. Planning for an updated system should build upon the successful accomplishments of the World Data Centers, but should go beyond current practice to take advantages of new technologies and capabilities. This effort should be integrated into the development of a long-term strategic framework**

and international Scientific Data and Information Forum (SciDIF). It must take account of the needs of existing ICSU programs and other new initiatives, such as the Group on Earth Observations (GEO), the forthcoming Polar Year in 2007, and the electronic Geophysical Year (eGY).

4. Other interdisciplinary bodies, Associates and ICSU members

Other Interdisciplinary bodies

In addition to the ICSU interdisciplinary bodies concerned with environmental data, the role of one other interdisciplinary body was considered in detail as part of this assessment. The Scientific Committee on Frequency Allocations for Radio Astronomy and Space Science (IUCAF) was asked to submit written input with reference to the terms of reference for the review. The co-sponsoring bodies for IUCAF, that is, the unions of astronomy (IAU) and radio science (URSI) and the Committee on Space Research (COSPAR), were also asked to comment on the value of IUCAF.

All other ICSU interdisciplinary bodies were invited to give general input to the review. Two bodies--the International Geosphere Biosphere Program (IGBP) and the Scientific Committee on Oceanic Research--responded to this invitation. Their valuable input has been incorporated into this report where appropriate. Both bodies cited a joint meeting that was held in December 2003 on Integrated Data Management for International Research Marine Projects. This initiative was considered by the assessment panel to be a good example of best practice related to the development of community accepted data management policies, and the meeting and its outcome should be publicized.

Scientific Committee on Frequency Allocations for Radio Astronomy and Space Science (IUCAF)

IUCAF was established as an inter-union commission by URSI, IAU and COSPAR in 1960. It was subsequently recognized as an ICSU interdisciplinary body and the name was changed (although the original acronym was maintained). The current membership of IUCAF consists of 13 scientists who represent its three sponsoring organizations and are elected at the General Assemblies of these bodies. ICSU provides no core financial support to IUCAF, although the Committee has successfully applied for ICSU grants in recent years.

IUCAF is a specialized body that fills a particular niche to study and coordinate the requirements of radio frequency allocations for passive radio sciences, such as radio astronomy, space research and remote sensing, in order to make these requirements known to the national and international bodies that allocate frequencies. It is formally recognized as a Sector Member by the International Telecommunications Union (ITU), where it represents the interests of radio astronomy. It also has active ties with key regional bodies of radio astronomers engaged in frequency protection.

Although the current membership of IUCAF is focused on radio astronomy, the committee is interested in advancing its expertise and links with space science and remote sensing in the future. It is expanding its geographical representation and planning capacity-building activities such as workshops and summer schools. IUCAF is performing a valuable specialized service, for which its status as an ICSU interdisciplinary body is appropriate. It also has an appropriate future strategy in relation to expanding its disciplinary base and representative ness, and it is building capacity in an arena that will continue to be important into the foreseeable future.

Recommendations

56. ICSU should encourage the three main partners of IUCAF - IAU, URSI and COSPAR - to organize a joint meeting to define specific procedures and actions that will ensure that IUCAF becomes a truly interdisciplinary committee that has the necessary expertise to coordinate the efforts of *all* the passive radio sciences in frequency management matters.

International Associates

ICSU has two international associates that have a particular focus on data and information issues and from whom written input related to the Terms of Reference of the panel was invited for this assessment. As associates, these bodies were not themselves subject to review but their input was considered by the panel as valuable additional information to inform the overall assessment.

The International Council for Scientific and Technical Information (ICSTI)

ICSTI was established in June 1984 as the successor to the ICSU Abstracting Board. Its aim is to increase accessibility to, and awareness of, scientific and technical information (STI). The ICSTI membership includes some 50 organizations, who represent the full spectrum of professionals concerned with scientific and technical information, from commercial publishers through abstracting and indexing organizations to user representatives, scientific associations and libraries. It was noted that ICSTI has already established links with CODATA, and to a lesser extent with INASP, and has been involved in recent global policy initiatives, such as the World Summit on the Information Society. The multi-stakeholder ICSTI membership was thought appropriate for addressing many of the key issues relating to scientific and technical information in general, and scientific publishing in particular, in ways that are relevant to the operational activities of many ICSU national and union members. Closer links between ICSTI and the ICSU membership could potentially replace much of the advisory function formerly assigned to CDSI (see V.2).

International Federation of Information Processing (IFIP)

IFIP was created in 1960 under the auspices of UNESCO, initially as an organization representing universities and academic organizations but increasingly also representing institutions outside of academia. Its basic aims are to promote information science and technology and to advance international cooperation in the field of information processing. IFIP and ICSU worked together in the early stages of the preparatory process for the World Summit on the Information Society.

IFIP has expertise that is complementary to that of ICSU in the rapidly developing area of computer science and informatics. The panel noted the willingness of IFIP to work with ICSU in relevant areas and the proposals for working together in the future in areas such as "Understanding Information and Knowledge" and ethics. It is suggested that further discussions be held between the relevant ICSU bodies and IFIP with a view to taking forward some of the recommendations laid out in the current priority area assessment, for example, in relation to data security.

ICSU Members and 'Emerging Issues'

There was no special consultation with the entire ICSU membership (national members and unions) for this assessment because of the overlap with parallel consultation exercises such as the PAA on the Environment in Relation to Sustainable Development and a general consultation with all members to identify new emerging scientific issues. However the input to these other exercises was available to the assessment panel and was integrated into its considerations. For example, in the emerging issues exercise, "Data, Information and the Digital Divide" was identified as a key theme with an emphasis on the following specific topics:

- Intellectual property issues arising from data and information
- Database accessibility and quality
- Promoting equitable access to ICT between countries and within countries
- Ethical issues arising from databases on human personal data
- Costs of hard copy publications, in developed and developing countries.
- The need to increase dissemination, access, abstracting/indexing, of publications and traditional knowledge from non-western countries, while ensuring the quality of such publications
- Management and use of data from satellite remote sensors (developing databases; web delivery of spatial information through images and maps)
- Ethical implications of our growing capability to use remote sensing / GIS for tracking things and people in real time.

There are recommendations relating to many of these issues throughout this report. Although specialized ICSU interdisciplinary committees have a key role to play in implementing these recommendations, as described in this chapter, the panel strongly emphasized that this could only be achieved with the active support and collaboration of all ICSU members and other partners.

VI Conclusion: the Need for a long-term Strategic Framework and establishment of an International Scientific Data and Information Forum (SciDIF)

Just as ICSU, in partnership with the international scientific community, met the successive challenges posed by scientists' evolving data and information needs over the past 50 years, so too it needs to address the new challenges and opportunities posed today by advanced information technologies, the ubiquity and power of the Internet, and the expansion of the types of data and information available for research. The remedies for the problems faced in the past were, on the whole, successful, and ICSU is to be commended for its sustained attention to data and information policies and practices. However, the current complex of ICSU data and information activities was developed on an *ad hoc* basis and was largely directed at problems identified within specific scientific specialties. This overall structure and approach are now outdated.

The Terms of Reference given to this Priority Area Assessment Panel were to define an overarching mission and role for ICSU in Scientific Data and Information and to propose a strategic framework for its activities over the next five to ten years. After reviewing the state of the field and current activities within the ICSU family, the first conclusion of the panel was that ICSU and the international scientific community need to develop a strategy for data management across the disciplines and interdisciplinary research areas represented within ICSU, including data access, documentation, dissemination, and preservation, for a period of decades to at least a century. This time frame is recommended because of the panel's conviction that the research data obtained today will be a valuable legacy to future generations of scientists. ICSU already coordinates and plans for research on similar time scales. It should extend this long-term approach to the research data infrastructure. **The Panel recommends that, on the basis of the many specific recommendations in this report, ICSU develop a long-term strategic framework for scientific data and information (policies, practices and infrastructure). An essential part of the development of this framework should be the closer coordination, and in some instances the transformation, of ICSU's current data and information activities [57].** The framework should build on existing data and information structures and services where it is advantageous to do so, but ICSU should be prepared to rethink, reorient, and replace existing structures and bodies where it is necessary

The second conclusion of the panel was that the management of scientific data demands professional attention. It is no longer adequate for data management to be relegated to last minute or end-of-project clean-up activities by scientists whose primary interest is research, not data. Instead, the management of scientific data demands professional care and attention, based on standards, policies, and practices that are common across fields of science.

The third conclusion of the panel was that ICSU should foster greater communication, coordination, and collaboration within and across members of the ICSU family and other partners on issues, policies, practices, and structures for scientific data management. More specifically, there should be widespread participation in the development of the long-term strategic framework for scientific data and information. **The panel recommends that, in parallel to the development of the long-term strategic framework, ICSU establish an international Scientific Data and Information Forum (SciDIF) involving all the key stakeholders: ICSU members, interdisciplinary bodies, science funding bodies and other data providers and users. Through SciDIF, ICSU should aim to ensure that the full benefits of new data and information technologies and capabilities are extended to scientists throughout the world [58]**.

Because of its international scope, its relationships to major scientific institutions, and its capacity to enlist the energies of leading scientists throughout the world, ICSU is uniquely positioned to foster improved scientific data and information policies and management both within the research community and across the digital divide. ICSU's role in SciDIF should be to provide the intellectual and organizational leadership, to use its convening power to bring together representatives of the leading institutions in both high income and low income countries, various fields of science, and multiple sectors (*e.g*., the academic, public, and private sectors) to plan new integrated data and information systems, and to ensure that they are governed on behalf of the international scientific community.

What would SciDIF look like? SciDIF would be both a face-to-face forum and a virtual forum for the discussion and implementation of a distributed and coordinated framework for data management that combines the strengths of the existing ICSU data bodies and adds value to existing data systems in specific countries and internationally. Through SciDIF, the long term integrated strategic framework would promote the use of advanced information technology, professional data management, and affordable access to data and information to meet the needs of scientists in both low income and high income countries. It would provide a forum for discussing data policy and examples of good practice in data management; it would promote interoperability of data, software, and hardware within the community of science; and it would maintain focus on scientific data management and its financial support for long periods of time. Finally, SciDIF would provide opportunities for advanced training and provide information on canons of best practice for professional data managers.

The cost of establishing new mechanisms for data management and preservation, particularly in fields of science or parts of the world where they do not currently exist, will be high. Maintaining these over long periods of time will require a significant financial commitment. Because scientific research resources are shrinking in many countries and have never been available in others, this creates a problem of priorities for scientific community. Under the aegis of SciDIF, ICSU should identify ways to reduce the costs of data collection and management rather than reducing data services. Representatives of the public and private sectors, donor agencies, and the scientific community should be included in detailed consideration of funding issues. Technological innovations that could reduce costs should be examined, as should potential economies of scale through dual use of scientific data in such areas as disaster recovery and management. Over the long term, data preservation and good data management will reduce the total cost of research by adding to the data base for science, but in the short term, it will require new financial commitments.

The panel sees at least three non-exclusive options or stages for ICSU to consider in developing and implementing a long-term integrated strategic framework and SciDIF:

Option I. Re-examine the World Data Center System with a view to extending its disciplinary and geographical coverage and developing a more strategic approach to data collection and management and the introduction of new technologies. This, together with the refocusing of CODATA, modest expansion of INASP, forging a closer relationship with ICSTI, and realigning ICSU with other monitoring and observational bodies (as recommended in this report) would constitute a minimal response to the need for the adoption of professional data management in science, and coordination of data access, dissemination, and preservation strategies in a single Framework. This response would constitute an updating of the current structure and could be categorized as a minimal necessary step. It would require minimal resources and could be achieved fairly rapidly.

Option II. **Establish a Strategic Data and Information Committee to oversee the development of a long-term integrated framework for data and information and a scientific data and information forum (SciDIF). Membership of this committee should include representatives of relevant ICSU bodies and unions, and experts in information technology and professional data management. This should be an *ad hoc* committee, with members (10-12max.) appointed for a period of three years [2]**. The remit for the committee should be to oversee the implementation of the recommendations in this report; to develop a long-term integrated strategic framework; to establish a Scientific Data and Information Forum; and to work with the Forum to address the policy, financial, and operational issues involved in implementing the strategic framework.

Option III. Establish an operational data and information system in ICSU that not only fosters new approaches to data and information management and preservation but also collects data for research, conducts training in data management, and builds software and finding aids for scientists. This option would give ICSU an operational function within the field of scientific data and information and would require operational funding. The need and location for any new operational functions should first be defined in Option II. ICSU is currently developing regional offices, which might have a role in this regard.

The Priority Area Assessment Panel believes that *Option II* is the way to start and that the development process for ICSU programs such as IGBP, which have had a far reaching effect on the practice of science internationally, might provide a model for this effort. The thoughtful and strategic reorganization of the data and information activities of ICSU under the aegis of a new ICSU data and information framework is an essential first step. The establishment of SciDIF would further enable collaboration with ongoing national and multilateral data and information activities and help extend full data and information services to scientists in low income countries. By focusing attention on data and information management for the long term, ICSU will be providing a valuable service to the scientific community now and building a lasting foundation for improvements in scientific research and education.

Annexes

Annex 1 – terms of reference
Annex 2 – ICSU bodies/members consulted
Annex 3 – members of the review panel
Annex 4 – Science in the Information Society- Agenda for Action
(ICSU/CODATA/UNESCO, March 2003)
Annex 5 – table of recommendations and principal audiences and responsibilities
Annex 6 – list of acronyms and abbreviations – *to be prepared*

Annex 1

Terms of Reference

1) Define an overarching "mission" and role for ICSU in the area *of "Scientific Data and Information"*, taking into account relevant activities outside of ICSU;
2) Propose a strategic framework for ICSU to take this area forward for the next 5-10 years;
3) Examine current activities within the ICSU family; identify gaps, overlaps and synergies of existing activities, and; propose responsibilities for individual bodies;
4) Propose modalities for promoting collaboration and co-ordination within the ICSU family when necessary and propose potential partnerships with bodies outside ICSU;
5) Examine and propose, if appropriate, changes either in the future direction of individual bodies or the way they operate including relationships with other bodies/organizations.
6) To consider the ethical issues related to scientific data and information and, where necessary, propose how ICSU might develop policies in response to these issues.
7) To identify policy issues of particular importance to science and society, which should be highlighted in the World Summit on the Information Society.

Annex 2

List of bodies consulted during the course of the assessment [4]

1. Written and oral in put

Committee on Dissemination of Scientific Information (CDSI)

International Network on the Availability of Scientific Publications (INASP)

Committee on Data for Science and Technology (CODATA)

2. Written in put only

Scientific Committee on Frequency Allocations for Radio Astronomy and Space Science (IUCAF)

World Data Centres (WDC)

Federation of Astronomical and Geophysical Data Analysis Services (FAGS)

Global Ocean Observing System (GOOS)

Global Terrestrial Observing Systems (GTOS)

Global Climate Observing System (GCOS)

Integrated Global Observing Strategy- Partnership (IGOS-P)

Additional Consultation

International Council for Scientific and Technical Information (ICSTI)

International Federation of Information Processing (IFIP)

International Union of Radio Science (URSI), re: IUCAF

International Astronomical Union (IAU), re: IUCAF

Committee on Space Research (COSPAR), re: IUCAF

All other ICSU Interdisciplinary Bodies and Joint Initiatives were also invited to provide input relative to the terms of reference for the review, although only two – the Scientific Committee on Oceanic Research (SCOR) and the International Geosphere Biosphere Programme (IGBP), responded.

[4] In parallel to this PAA a number of other consultation exercises were being performed with the ICSU family and the information form these exercises was fed into the assessment where appropriate (see full report for details). Further to this the draft report was circulated to all ICSU members and bodies for consideration before it was published.

Annex 3

PAA panel membership

Chair: Roberta Balstad Miller, Columbia University, USA, is Director of the Center for International Earth Science Information Network (CIESIN). She is currently Chair of the U.S. National Committee on Science and Technology Data (CODATA) and former Chair of the National Research Council's Steering Committee on Space Applications and Commercialization. Dr. Balstad Miller previously headed the Division of Social and Economic Science of the National Science Foundation and was the founding Executive Director of the Consortium of Social Science Associations (COSSA). She has published widely on the human role in the environment, science and technology policy, and South African history.

Jean Bonnin is Professor of Solid Earth Geophysics at Institut de Physique du Globe, Louis Pasteur University, Strasbourg, France. Fellow of the École Normale Supérieure, he has previously held posts at the University of Montpellier, Lamont-Doherty Geological Observatory (Columbia University, USA), Centre National pour l'Exploitation des Océans and the Institut National d'Astronomie et de Géophysique (France). His fields of interest include : plate tectonics (in particular in the Atlantic Ocean and the Mediterranean Sea), deformation of plates and earthquakes as markers of this deformation. He has headed the European-Mediterranean Seismological Centre, in charge of quasi real-time earthquake source parameters determination and dissemination. He is currently active in the analysis of the actual informational content of geophysical data (earthquake parameters), and in the area of decision-making for disaster management.

Marc H. Brodsky is the Executive Director and CEO of the American Institute of Physics (AIP). AIP is a nonprofit member corporation of ten physics and astronomy related Societies with a total of 130,000 members. AIP publishes and distributes journals, magazines, and conference proceedings. Dr Brodsky is Chair of the Executive Council of the Professional and Scholarly Publishing Division of American Association of Publishers. He is also on the Executive Committee and Board of Directors of Publishers International Linking Associates, Inc (PILA is the corporate entity of the CrossRef linking service between online publications). He is also a member of the Steering Committee of a liaison group between IFLA (the International Federation of Library Associations and Institutions) and IPA (the International Publishers Association). Dr Brodsky was at the IBM Watson Research Center for 25 years as a researcher (in semiconductor physics and devices), manager and executive.

Liu Chuang is Professor and Director of Global Change Information and Research Center, Institute of Geography and Natural Resources, Chinese Academy of Sciences Beijing, China. She currently serves the Committee on Data for Science and Technology (CODATA) as Co-Chair of Task Group on Preserving and Archiving for Scientific and Technology Data in Developing Countries and CEOS (Committee of Earth Observation Satellites) as the User Co-Chair of the Working Group of Information Systems and Services (WGISS). She is Chair of Spatial Data Committee of Chinese Associate of Geographical Information Systems (CAGIS) and Secretary General of a Working Group of Remote Sensing and Data Information Systems

(RS/DIS), China National Committee for IGBP. Dr. Liu is active in research on the broad issues regarding strategy, technology and capacity building for open access to environmental data. She is geographer in GIS and Asia Land Ecosystems studies.

Carlos A Correa is Professor of economics and law and Director of the Masters Program on Science and Technology Policy and Management, and of the Post-graduate Courses on Intellectual Property of the University of Buenos Aires. He is also Director of the Center for Interdisciplinary Studies of Industrial Property Law and Economics of the same University. He is currently in charge of the "Innovation, Development and Intellectual Property Policy" project at the South Centre, and chairs the Genetics Resources Policy Committee of the CGIAR. He has been a consultant to several UN agencies and other regional and international organizations in different areas of law and economics, including investment, science and technology and intellectual property. At different times he has advised the governments of Canada, Spain, Ecuador, Guatemala, Dominican Republic, Uruguay, Jordan, South Africa, Indonesia on these issues. He was a member of the UK International Commission on Intellectual Property, established in 2001.

Norihisa Doi is currently a professor of the Faculty of Science and Engineering, Chuo University, and also professor emeritus of Keio University. He is a member of the Council for Science and Technology and the Council for Information and Communication, which are advisory to the Japanese Governement, President of the not-for-profit Japan Information Security Audit Association and Chair of the Japan Chapter of the Association for Computing Machinery (ACM). He was until recently a member of the Science Council of Japan (1994-2003).

Ray Harris is Professor of Remote Sensing and Vice-Dean of the Faculty of Social and Historical Sciences at University College London. He has been a lecturer at the University of Durham and has worked in management and consultancy positions in Software Sciences Limited and Logica plc. He has been active in satellite remote sensing research since 1972 and has produced two books on Earth observation data policy. He has worked with the European Commission, the European Space Agency and the US National Oceanic and Atmospheric Administration on data policy. He has also worked on the use of Earth observation data for agricultural monitoring in the UK and the Middle East.

Andrew M. Kaniki, National Research Foundation (NRF), RSA, is Executive Director of Knowledge Management and Strategy. He was, until October 2002, Professor: Information Studies, and Pro-Vice Chancellor, University of Natal, South Africa. Dr. Kaniki also taught at the University of Zambia, and worked as Science Information Specialist: Engineering and Science Library, Carnegie Mellon University, USA. He holds a Bachelors degree in politics and library science (University of Zambia); Master of Science (University of Illinois), and PhD and Agricultural Information Specialist Certificate (University of Pittsburgh). He has published and presented numerous conference papers on information needs and use; and knowledge management. He is currently President: Southern African Research and Innovation Management Association (SARIMA)

Vitaly A. Nechitailenko, Geophysical Center RAS, Russian Federation, is the Head of Dept. for Information Processing and Electronic Publishing. He is currently a member of Russian National CODATA Committee and former member of ICSU Panel on World Data Centers and Committee on Dissemination of Scientific Information. Dr. Nechitailenko previously was Vice President of the Soviet Geophysical Committee and was for many years involved in International Geophysical Programs, developing research equipment and software for data processing. He has published widely on meteor radar observations, computer network simulations, data processing and electronic publishing technology

Pierre L.-J. Ritchie, is Full Professor, University of Ottawa, Canada where he is responsible for practicum and internship training in the School of Psychology. He is the Secretary-General of the International Union of Psychological Science and psychology's main representative to the World Health Organization. Dr. Ritchie is also Executive Director of the Canadian Register of Health Service Providers in Psychology. His clinical expertise focuses on differential diagnosis. His scholarly interests address ethics in health practice and in research involving human participants as well as health policy. He has published widely on these subjects and been extensively involved in national and international professional and scientific organizations.

T.B. Rajashekar is Principal Research Scientist and Associate Chairman of the National Centre for Science Information (NCSI) in Indian Institute of Science (IISc), Bangalore, India. Dr. Rajashekar develops and manages digital library and e-information services for the IISc research community. His primary areas of teaching and research interest are Digital Libraries and Information Organization. He has held several R&D project grants from government and private agencies and has also organized national and international training workshops in these areas. Dr. Rajashekar is currently coordinating a UNESCO/FAO project on development of self-learning courses on creation and sharing of digital library collections. He has published widely and has served as resource person in several national and international programmes.

Annex 4

Science in the Information Society [1]

Scientific research is one of the key factors underpinning the development of the Information Society. All of the fundamental technological components of the Information Society were discovered or invented in academic laboratories: electricity, radio waves, lasers, the World Wide Web (www) and the web browser. Ensuring equitable access to scientific knowledge is essential in order to achieve the Millennium goals and the use of Information and Communication technologies (ICTs) now offers incredible opportunities in this regard. Scientific research leads to the development of new technologies themselves and to the production of data and information that, when combined with these technologies, can be of huge benefit to society as a whole. **The essential role of science and scientists in building the Information Society should be clearly acknowledged in the declaration of principles and reflected in the plan of action from WSIS.**

Principles

Scientific knowledge and data are of enormous importance in a global Information Society:

- To foster innovation and promote economic development
- For efficient and transparent decision-making, particularly at the governmental level
- For education and training

Scientific data and information should be as widely available and affordable as possible: the more people that are able to share them, the greater the positive effects and returns to society. Scientific knowledge is a "public good".

The development of new ICTs opens up **unprecedented opportunities** to ensure universal and equitable access to scientific data and information and to enhance the global knowledge pool. However, **excessive privatization and commercialization of scientific data and information** is a serious threat to the realization of these opportunities for the benefit of society as a whole.

Agenda for Action:

1. Ensure that all universities and research institutions have affordable and reliable high-speed Internet connections to support their critical role in information and knowledge production, education and training.

2. Promote sustainable capacity building and education initiatives to ensure that all countries can benefit from the new opportunities offered by information and communication technologies (ICTs) for the production and sharing of scientific information and data.

3. Ensure that any legislation on database protection guarantees full and open access to data created with public funding. In addition, restrictions on proprietary data should be designed to maximize availability for academic research and teaching purposes.

4. Promote interoperability principles and metadata standards to facilitate cooperation and effective use of collected information and data.

5. Provide long-term support for the systematic collection, preservation, and provision of essential digital data in all countries.

6. Promote electronic publishing, differential pricing schemes, and appropriate open source initiatives to make scientific information accessible on an equitable basis.

7. Encourage initiatives to increase scientific literacy and awareness of how to interpret web-based scientific information.

8. Support urgently needed research on the use of information technologies in key areas, such as geographical information systems and telemedicine, and on the socio-economic value of public domain information and open access systems.

9. Recognize the important role for science in developing and implementing the new governance mechanisms that are necessary in the information society.

---

1. This statement is the product of a workshop "Science in the Information Society", that was organised by ICSU and the ICSU Committee on Data for Science and Technology (Codata) in partnership with UNESCO. The workshop took place in Paris on 12th March 2003 and involved over 60 scientists, science managers and representatives of international agencies from all over the world. A full report of the workshop and other ICSU activities related to WSIS can be found at **www.icsu.org**.

Annex 5

Recommendations and principal audiences and responsibilities

The recommendations in this report are addressed to multiple audiences within and outside the ICSU family. These include ICSU interdisciplinary bodies, national and union members and institutions responsible for science policy and practice throughout the world. In the table below, the assessment panel indicates which of these various audiences was envisioned as being responsible for taking forward which recommendations. The designations here should be considered as a guide, rather than a definitive assignment of responsibilities. Moreover, many of these recommendations also need to be considered by the proposed new *ad hoc* strategic committee and taken forward under the umbrella of the Scientific Data and Information Forum (SciDIF).

In the table below, the panel recommendations are assigned to four categories of actors:

ICSU Executive refers to the ICSU Executive Board, CSPR, and the Secretariat;
ICSU Bodies includes interdisciplinary bodies, joint initiatives and advisory/policy committees;
ICSU Members/scientists refers to the national and union members of ICSU and the scientific communities and individual scientists that they represent;
Other Stakeholders includes national governments, inter-governmental organizations, scientific funding bodies, scientific publishers, and other data producers, managers, and users.

Although responsibilities for data and information policy and management and the recommendations of this report are normally shared, the category where greatest responsibility is expected is noted as "primary;" (**P**); others with substantial responsibilities for the recommendation are designated as "secondary" (**S**). Where a particular ICSU body is expected to assume leadership, that body is specifically noted.

Rec. no	**ICSU Executive**	**ICSU bodies**	**ICSU members/scientists**	**Other Stakeholders**
1	P	S	S	
2	P	S	S	
3	P	S	S	
4		S	P	S
5	S	P	S	S
6		S	P	S
7			S	P
8				P
9		S	P	S
10	S	S	P	S
11			S	P
12	S	ICSTI	S	
13		S	S	P
14	S	P	S	S

15	S	INASP	P	S
16	P	S	S	
17		S	S	P
18			S	P
19			S	P
20			S	P
21	S	S	P	S
22	S	S	P	S
23		S	S	P
24			S	P
25	S	S	S	P
26	P	S	S	S
27	P	S	S	S
28			P	S
29	S		P	
30			S	P
31			P	S
32	S	S	P	
33	P	S	S	
34	S	P	S	S
35		S	S	P
36	P	S	S	S
37	P	S	S	S
38	S	S	P	S
39		S	S	P
40	P	S	S	S
41	P	S		
42	P	S	S	
43	P	S	S	
44	P		S	
45		S	P	
46		P		
47		P	S	
48		CODATA		
49	S	P	S	
50	S	S	P	
51		CODATA		
52	P	S		S
53	P			S
54		FAGS	P	
55	P	S	S	S
56	S	IUCAF	P	
57	P	S	S	
58	P	S	S	S

N.B. Where boxes are left blank or specific bodies are identified, it should not be assumed that a recommendation is of no relevance to that audience or other bodies.