AD HOC STRATEGIC COMMITTEE ON INFORMATION AND DATA

Final Report to the ICSU Committee on Scientific Planning and Review

June 2008

Table of contents

Executive summary

The Strategic Committee on Information and Data (SCID) was established by ICSU to advise on the future organisation and direction of its activities in relation to scientific data and information. Following an earlier priority area assessment exercise in this area, ICSU's declared strategic goal is: *to facilitate a new, coordinated global approach to scientific data and information that ensures equitable access to quality data and information for research, education and informed decision-making.* The role of SCID was to assess how this goal might best be achieved.

Taking the Priority Area Assessment on Data and Information (ICSU, 2004) as its starting point, SCID met on three occasions in 2007-2008 and considered input from the following ICSU Interdisciplinary Bodies: the World Data Centres (WDC), the Federation of Astronomical and Geophysical Data analysis Services (FAGS) and the Committee on Data for Science and Technology (CODATA). The SCID members themselves also collected and presented information on major international data and information initiatives of strategic importance to ICSU.

The major recommendations of SCID are that:

1) ICSU assert a much-needed strategic leadership role on behalf of the global scientific community in relation to the policies, management and stewardship of scientific data and information;

In order to achieve this, ICSU must reform some of its current interdisciplinary bodies and establish a new committee that will provide overall strategic direction and advice.

2) a new ICSU World Data System be created (as an ICSU Interdisciplinary Body), incorporating the WDCs and FAGS as well as other 'state of the art' data centres and services;

This new structure or system must be designed clearly to support ICSU's mission and objectives, ensuring the long-term stewardship and provision of quality-assessed data and data services to the international science community and other stakeholders.

3) CODATA focus its activities on the three main initiatives identified in its draft strategy and extend its links to other organisations and networks to play a more prominent role within ICSU and within the wider scientific community;

This will require the close alignment of implementation mechanisms, e.g. working groups and task groups, with the 3 main initiatives identified in the draft CODATA strategic plan (appendix G). The bi-annual CODATA conference should also be modified to provide closer links to ICSU priorities and the new ICSU World Data System.

4) a new *ad hoc* ICSU Strategic Coordinating Committee for Information and Data be established to provide broad expertise and advice to ICSU in this area;

This Strategic Coordinating Committee will act as an interface between scientists and data and information professionals that can advise on the data needs and possible solutions for existing and new ICSU programmes and other international initiatives. It should be established for three years in the first instance, potentially renewable for a further three-year term. During which period it will be expected to establish visible

and effective leadership for ICSU and ensure proper coordination among ICSU activities.

5) ICSU National Members and Unions be strongly encouraged to establish committees or commissions, where these do not already exist, focussing on data and information issues.

Where national committees or liaison structures already exist for CODATA and/or the WDCs, consideration should be given to amalgamating and expanding these to integrate data policy, management and stewardship issues. Professional data services must be recognised and supported at the national level as part of the long-term infrastructure of science.

1. The context

The ICSU Vision

The long-term ICSU vision is for a world where science is used for the benefit of all, excellence in science is valued and scientific knowledge is effectively linked to policy making. In such a world, universal and equitable access to high quality scientific data and information is a reality and all countries have the scientific capacity to use these and to contribute to generating new knowledge that is necessary to establish their own development pathways in a sustainable manner[1].

It is implicit in this mission that ICSU has an important responsibility on behalf of the global scientific community for promoting the optimal stewardship[2] and policy development for scientific data and information.

1.1. Background

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 rapidly in recent years. These changes are partly a result of the rapid evolution in computational capability and connectivity that together have expanded the variety and quantity of research data. They are also related to the emergence of new questions in scientific research that require different types of data. Taken together, these changes are providing scientists throughout the world with 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. At the same time, as the quantity and diversity of data expands, the challenges in assuring the quality, accessibility and long-term preservation of this precious resource are also amplified. And the capacity to deal with these challenges is very unevenly distributed across the globe.

In 2004 ICSU conducted a Priority Area Assessment (PAA) on Scientific Data and Information[3]. This was part of an overall strategic planning exercise and it laid out over 50 recommendations on future needs and priorities. The first and broadest recommendation was that ICSU should:

assume a leadership role internationally in identifying and addressing critical policy and management issues related to scientific data and information.

The PAA report highlighted the importance of professional data and information management and the need to build capacity in this area in all countries, the

[1] ICSU Strategic Plan 2006-2011

[2] Stewardship is used throughout this report as a broad term encompassing all aspects of data collection, management, archiving and distribution, including publication. For a summary of the key aspects of data stewardship, see http://www.egy.org/declaration.php

[3] ICSU (2004) Report of the CSPR Panel Area Assessment on Scientific Data and Information, International Council for Science, Paris, ISBN 0-930357-60-4, 42pp

importance of coordination within the ICSU family and beyond, and the need to modernize or replace existing structures.

1.2. ICSU strategy

On the basis of the PAA recommendations, the ICSU Strategic Plan 2006-2011 (pp 40-41) includes the following goal,

> *to facilitate a new, coordinated global approach to scientific data and information that ensures equitable access to quality data and information for research, education and informed decision-making.*

ICSU has made a commitment to re-focusing its own data and information structures, in line with the PAA recommendations:

> *The World Data Centre (WDC) system and the Federation of Astronomical and Geophysical data Services (FAGS) will be reformed taking account of user needs, including those of existing and new ICSU programmes. This will form part of the development of a broader strategic framework for data and information.*
>
> *The Committee on Data for Science and Technology (CODATA) will be encouraged to develop a long-term strategy, giving special attention to the needs of less economically developed countries.*

In order to achieve this, the 28th ICSU General Assembly agreed to establish an *ad hoc* Strategic Committee on Information and Data (SCID).

BOX 1. Definitions of data and information

For the purposes of the work of SCID the definitions of data and information were adopted from those of the earlier ICSU PAA report:

Data and information (DI) can be considered 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.

Scientific information is sometimes considered as being synonymous with text publications in scientific journals. Such journals are certainly an important part of scientific information. However, they are not the major focus of the current report, for which a broader definition of 'information' has been adopted.

1.3. *Ad hoc* Strategic Committee on Information and Data

The remit of the Scientific Committee on Information and Data (SCID), given by ICSU, was to oversee the implementation of the key recommendations in the PAA report and in particular those that concern ICSU interdisciplinary bodies (WDCs, FAGS, CODATA). A key conclusion of the PAA was that ICSU should foster greater communication, coordination and collaboration within and across members of the ICSU community and with other partners on issues, practices and structures for

scientific data management. A multi-stakeholder Scientific Data and Information Forum (SciDIF) was proposed by the PAA as a mechanism to achieve this (PAA recommendation 58). Since the publication of the PAA report a number of significant multi-stakeholder forums concerned with data and information had already been established. These included the planning exercise for a Global Earth Observation System of Systems (GEOSS) and the launching of a Global Information Commons for Science Initiative, both of which were supported by ICSU. The Electronic Geophysical Year (eGY) was also acting as a focus for coordination and collaboration on data issues and the International Polar Year (IPY) was having a federating effect on all those involved in data and information management as regards polar research. These developments provided a new context for SCID to reconsider the original PAA recommendation for a forum.

1.4. Terms of reference

Taking the PAA report and the ICSU Strategic Plan 2006-2011 as points of departure, and adding recent developments and programmes in scientific data and information management, the terms of reference given by ICSU to the SCID were as follows:

1. to *guide* and *oversee* the reform of the World Data Centre (WDC) system and Federation of Astronomical and Geophysical data analysis Services (FAGS);
2. to *liaise* with CODATA in the development of its strategic plan;
3. to *advise* the Committee on Scientific Planning and Review (CSPR) on any other actions that might be appropriate for ICSU to consider in order to facilitate a coordinated global approach to scientific data and information, including the potential need for a Scientific Data and Information Forum (SciDIF).

1.5. Working practices

The terms of reference of the SCID were approved by the Committee on Scientific Planning and Review (CSPR) in September 2006. After consultation with relevant bodies, the membership of the SCID was agreed by CSPR at its meeting in February 2007. The members of the SCID are listed in Appendix A.

Three meetings of the committee were held in Paris during the period July 2007 to February 2008. Critical inputs to these meetings included a white paper from FAGS and a report from a meeting of WDC directors (2007), both of which focussed on the strengths and weaknesses of the existing structures. The CODATA draft strategic plan was also made available and the CODATA Executive Director attended part of the second SCID meeting. All three bodies were also well represented on the committee itself. Hence, the discussion and final recommendations of SCID have not been made in isolation but after consultation with the bodies concerned.

The draft report of the SCID was circulated to all ICSU Members and Interdisciplinary Bodies for comment and the final report submitted to the CSPR in April 2008. The main SCID proposals are to be presented to the ICSU General Assembly in October 2008.

2. The case

2.1. The challenge

Science is increasingly problem-oriented, not discipline-bound. While scientists do often work in subject disciplines, exciting and important scientific (and societal) challenges do not necessarily follow disciplinary boundaries. Scientific data and information are increasingly used in a multi-disciplinary or inter-disciplinary context, increasingly used to provide information services rather than simply raw data and increasingly characterised by model output as well as source data. Scientists use data and information from many different sources; these include government and commercial sources as well as traditional scientific sources. Just as there is a market place for ideas there is a market place for data and information suitable for testing ideas and solving scientific problems.

Three major trends in data and information management are dramatically changing science. The first is the major step change in the sheer volume and diversity of data suitable for science. Many fields from geo-demographics to particle physics are witnessing dramatic increases in data and information volumes (see Box 2). The second is the availability of new information and communication technologies, such as Grid computing or Sensor Web[4], which means that very ambitious modelling and data processing are within the scope of an increasing number of scientists. The third is the increasing need for scientific datasets to be properly identified, quality-assured, tracked and accredited (for example, through assignment of digital object identifiers or DOIs). This requires professional data management and, in some areas, may involve review and publication of datasets. Publication and accreditation can also act as an important incentive for primary data producers to make their data available.

Virtual observatories, seen particularly in astronomy and geophysics[5], use these three major trends to tailor massive data and information capabilities to their scientific user communities (Box 3). In some countries, government science agencies are beginning to establish new organizations and structures to respond to this new context for scientific data and information. Alongside these major trends and responses, and of particular importance from the ICSU perspective, is the overarching requirement to improve access to data and information for scientific research, education and informed decision-making[6].

Box 2. **Challenges in managing astronomical data**

The volume of data collected by current astronomical instruments is estimated to double every 12-18 months. For example, the cutting-edge IRAS space telescope in 1984 produced a total data volume of about one gigabyte (10^9 bytes) whilst 2MASS in 1998 generated a total of 10 terabytes (10×10^{12} bytes). In 2012, the Australian Square Kilometre Array Pathfinder radio telescope will produce 10 terabytes every hour, with a storage capacity of about 60 petabytes (60×10^{15} bytes)

[4] An emerging new paradigm is the concept of a dynamic network of intelligent sensors; a sensor web, in which devices may be controlled remotely via the web. "Data" from many such devices could automatically be collected, aggregated and styled for easy comprehension. See (http://www.geospatial-solutions.com/geospatialsolutions/article/articleDetail.jsp?id=52681 and http://www.opengeospatial.org/projects/groups/sensorweb

[5] Dalton R (2007) Geophysicists combine forces, *Nature* 28 June 2007, 1037

[6] ICSU strategic plan 2006 - 2011

and the Large Synoptic Survey Telescope will produce about 5 petabytes per year. Future telescopes, such as the Square Kilometre Array, due for completion in 2020, will produce volumes of data orders of magnitude greater. This huge growth in data volume is accompanied by a corresponding increase in data complexity. There is a major initiative underway in the international astronomical community to develop a Virtual Observatory to provide the user with the specialised suite of tools and processes, which will be essential for those scientists wishing to access these data.

Box 3 **Virtual observatories**.

Virtual Observatory (VO) definitions differ depending on the application domain.

In astronomy, a VO is a collection of integrated astronomical data archives and software tools that utilize computer networks to create an environment in which research can be conducted. Thus, data from all the world's major observatories at many electromagnetic wavelengths are available to all users and to the public.

In solar and space physics, a VO is a suite of software applications on a set of computers that allows users to uniformly find, access, and use resources (data, software, documents, and image products and services using these) from a collection of distributed product repositories and service providers. A VO is a service that unites services and/or multiple repositories.

In geosciences, a VO should increase efficiency, and enable new science by greatly enhancing access to data, services and computing resources. VOs include tools: to locate and retrieve data from many sources, for analysis, simulation, and visualization, to compare observations with results obtained from varied sources with a spectrum of time availability, and supporting information. At a fundamental level VOs provide interoperability: services that can be used regardless of the clients computing platform, operating system and software capabilities.

2.2. Why are data and information systems necessary?

One of the characteristics of the new information and communication technologies and the World Wide Web is that they encourage innovation and permit individual scientists and institutions to make data and information easily available. This is one of the reasons that there is more scientific information available at the touch of a keyboard than ever before. However, one of the other characteristics of the web is that it is constantly changing and somewhat chaotic; URLs disappear and previously available information can be lost without trace overnight. Web-searching or automated crawling software and data storage facilities are attracting major commercial investment but the quality and reliability of the available data and information is less readily assured (Box 4). There is a trend towards data being managed by individuals or groups in voluntary distributed systems on the internet but the quality assurance and long-term accessibility of this data is frequently neglected.

One way to address these concerns and ensure the long-term stewardship and availability of critical scientific data and information is via the development of a more stable system or systems that have explicit responsibility for collecting, managing and distributing critical data and ensuring their transition through inevitable future technological developments. As asserted in the PAA report, there is a need for global federations of professional state of the art data management institutions, working together and exchanging practices. Such federations can provide quality assurance and promote data publishing, providing the backbone for the development of a global virtual library for scientific data. They can also complement and assist the multitude

of very worthy voluntary initiatives that flourish alongside them by helping to develop and disseminate good practices and standards.

In some areas of science, such data systems, services, federations or virtual libraries already exist. For example, the bioinformatics community has developed a very effective 'hub and spokes' network for genomic and proteomic data. This is built around a small number of major centres in the USA, Europe and Asia that act as a virtual global library and is supported by a large number of ancillary centres. Other examples include the development of virtual observatories by different scientific communities (see Box 3). Oceanographic and meteorological data, which are routinely collected by government agencies primarily for operational reasons but also used by scientists, are managed via well established networks of national data centres and data transmission systems. In other areas such as satellite observations, responsibility for the long-term stewardship of data and information is still being established (see Box 5).

Box 4 – **Commercial publishers, the dotcom industry and scientific data**

As a response to the growing demand for primary data related to journal publications, commercial publishers have opened up on-line spaces for the deposition of supplementary material. Similarly, Google with its 'Google Base' (http://base.google.com/) has offered an open platform for exchange of digital objects that is sometimes proposed as a possible solution for long term archiving and dissemination of scientific data. More recently, with its 'Palimpsest' project (http://research.google.com/) the same company has started an initiative that specifically focuses on scientific data. In fact, 'Palimpsest' - enhanced by an offline shipping service and visualisation services - might turn out to be useful, in particular for the dissemination of large published data sets. However, all of these approaches are lacking homogenous, reliable, and acknowledged quality management structures and procedures for the hosted data. In addition - as storage of data and metadata does not follow any common standards - none of the data providers is interoperable in the sense of GEOSS (Box 5). Thus, usage of such data in a serious scientific context is not only disputable but, due to the effort required to fully re-constitute the data, also inefficient. Whilst companies such as Google have an enormous amount to offer, there remains a large question mark over devolving long term data stewardship responsibility to commercial domains.

Box 5 – **The Global Earth Observations System of Systems**

The Global Earth Observation System of Systems (GEOSS) is an inter-governmental initiative to achieve comprehensive, coordinated and sustained observations of the Earth system. It aims to improve monitoring of the changing state of the planet, increase understanding of complex Earth processes and enhance prediction of the Earth system. GEOSS, collectively, has several functional components: to address identified common user requirements; to acquire observational data; to process data into useful products; to exchange, disseminate and archive shared data, metadata, and products; and to monitor performance against the defined requirements and intended benefits. With a focus on access and sharing of Earth observation data and products, there has been significant advancement in the definition of interoperability standards and mechanisms for the allocation and use of data and information products, and in the synergetic system development resulting in improved data access and data sharing. The GEO Web Portal and Clearinghouse aim to provide a single interface for access to GEOSS data and information ((see http://www.earthobservations.org/index.html). GEONETCast, a satellite-based dissemination system, allows users to access real-time, global, Earth observation data and derived information

2.3. The key principles and characteristics of an ideal system

What might constitute an ideal system for scientific data and information provision and management to help deliver ICSU's vision and strategic priorities for global science?

The essential principles and characteristics of such a system include:

- Enabling universal and equitable access (accessible to everyone everywhere without discrimination, see Box 6)
- Ensuring reliable and efficient access
- Facilitating improved data deposition and retrieval
- Maintaining and validating the quality and authenticity of data and products and ensuring adherence to standards
- Ensuring long-term sustainability
- Enabling and encouraging interdisciplinary research
- Maintaining flexibility in response to changing demands, changing science and changing technology

In addition to these, from an ICSU perspective, the ideal system would:

- Enact a common vision for the stewardship of data and information on behalf of the global science community
- Provide a federation of active participating organisations in which internal communication is highly valued
- Provide a forum to identify, articulate and advocate the common needs and interests of the components of the system
- Promote data publication and accreditation
- Encourage complementary and linked provision of data and information

From a scientist's point of view the ideal system must be built on trust in the supply of data and information; trust not only in data, but also in the many steps required for the management of this data and information. The ideal system is also therefore one that scientists, scientific organisations and other bodies wish to support and/or join.

Disciplinary coverage

There is no obvious limit to the disciplines or areas that might be included in an ICSU system and indeed a major overarching principle for such a system should be to enable interdisciplinary science. This can only be achieved by ensuring both deep disciplinary support and interoperability between data from different fields. At the same time, it is recognised that many parts of the scientific community are already developing their own data and information management systems and the immediate benefit to them of joining a new ICSU structure may not be obvious. One of ICSU's goals should be to ensure effective communication and good interoperability among different systems and communities.

In the first instance a logical starting point for ICSU would be the development of a system that effectively fulfils the needs of its own interdisciplinary programmes. A development and expansion of ICSU's established role in relation to geoscience,

space, astronomical and environmental (including socio-economic) data would provide an important service to the global science community and to the multiple UN conventions which depend on scientific data and information in these areas. Over time such a system might expand to include other disciplines or fields of science, as indeed ICSU's own programmes are becoming even more inter-disciplinary.

Box 6 **Data access policies**

It is the declared policy of ICSU's current world data centres and data analysis services to make their data freely open and available for research. In practice many of these centres have considerable additional datasets and/or data products for which access is restricted for a variety of reasons and these are not considered to be part of the WDC or FAGS holdings.

ICSU's formal policy is to promote "full and open access" to data, as discussed in the PAA report:

"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, which is by definition discriminatory against some scientists, is clearly contrary to the principle of full and open access but some cost-recovery is not excluded. [ICSU PAA, 2004]

More recently these principles have been built upon in developing a data policy for the International Polar Year, which also explicitly recognises that restrictions on openness may be necessary when dealing with some data types, e.g. personal medical data or data that identify the nesting sites of rare species. (see http://classic.ipy.org/Subcommittees/final_ipy_data_policy.pdf). CODATA has taken the lead in developing a detailed white paper on GEOSS data sharing principles, which are also focussed on "full and open access".

With regards to scientific information, and scientific journals in particular, ICSU's formal policy is to promote "universal and equitable access" [ICSU PAA, 2004]. This wording had been adopted partially to avoid any confusion with open-access journal publishing, which is a narrower term linked to a particular publishing model. Open-access publishing is one of several models that can be used to promote universal and equitable access.

2.4. Functions of the ideal system

On the basis of the ascribed principles and characteristics for an ideal ICSU system, SCID proposes that the functions of such a system fall under three main headings: mission, coordination and execution.

Mission

- Enable and encourage the advancement of science through the open provision of high quality data and information services.
- Increase global knowledge and reduce the knowledge divide between richer and poorer countries by providing universal and equitable access to scientific data and products.
- Identify structural gaps in data and information provision and seek solutions to fill these gaps.
- Develop further the structure for long term stewardship of scientific data, including in the form of formal public libraries for data.

Coordination

- Foster multi-disciplinary, large scale, complex science.
  - Encourage the transfer of knowledge and expertise across discipline boundaries
  - Foster systems interoperability and develop common registries, directories and portals.
- Lead and champion professional data management.
  - Identify and facilitate training in professional data management in an international, multi-disciplinary context
  - Promote best practice operational data and information management processes.
  - Identify and organise data backup and archival services.
  - Encourage and adhere to accepted scientific standards and conventions.
  - Develop common or shared processes, for example shared licences for data access.
- Inform discussions on data policy from a science perspective.

Execution

- Take a lead role in developing, testing and implementing standards for data access to provide services for all scientists.
- Promote the publication of data and data products, with the associated recognition and accreditation that are common to peer-reviewed science publications.
- Provide reliable and trustworthy science-reviewed data and derived products.
- Serve discipline-based science communities with exemplary data repositories and data products.
- Integrate data sets using community-consensus algorithms.
- Enable seamless access to data.

Such a system should go a long-way towards meeting the needs of major international interdisciplinary research programmes such as the International Polar Year (see Box 7)

Box 7 **International Polar Year and data stewardship**

The year 2007 marked the beginning of the International Polar Year (IPY), which is co-sponsored by ICSU and WMO. IPY is an ambitious program of globally coordinated research, comprising over 200 multi-disciplinary projects centred on the poles and their connections with global phenomena. This modern IPY will build on knowledge derived from three previous scientific campaigns, spanning the last 125 years. Historically, each of the former campaigns has led to step shifts in scientific understanding of the polar regions and provided new tools and techniques to better model the role of the poles and oceans in global climate systems. The 2007/2008 IPY intends to leave as its legacy a coordinated set of global observing and monitoring systems, supported by a robust scientific data management framework. Progress towards realising this

framework is happening through the development of the IPYDIS (IPY Data and Information Service), a global partnership coordinated by the US National Snow and Ice Data Centre. IPYDIS (http://ipydis.org/index.html) will eventually be a federation of data centres, archives and networks that provides access to, and long-term management of, data produced from IPY projects and successive observing programs. IPYDIS of necessity must build upon existing data networks and institutional activities but participants are cognisant that much of the current global data management infrastructure cannot adequately meet the coordination and functional aspects required of a truly integrated and interoperable data access system. Significant collaborative work therefore lies ahead for participating IPY related data centres, IPY research scientists, data scientists in general and existing global data networks in order to create the requisite infrastructure.

3. The solution

3.1. Current situation

The current data and information organisations in ICSU provide part but not all of the ideal system requirements listed above. They provide the foundation on which a broader and more effective system can be built

WDCs and FAGS

The World Data Centre (WDC) network was created during the International Geophysical Year (IGY) of 1957-58, originally consisting of 27 centres in the US, Russia, Europe and Japan. There are now 50 WDCs in 12 countries and their holdings include a range of solar, geophysical, environmental and human dimensions data (see appendix B). Individual centres are hosted and funded nationally in a variety of different institutions and they serve mainly a national mandate, with their global WDC-role being a secondary function. This function is frequently unsupported, although being recognised as a WDC does enable some centres to attract additional finances. All data held in WDCs are available on a full and open access basis, either directly on-line or for no more than the cost of copying and sending the requested information.

The principal strengths of the WDC network include its durability and associated 'good name'. In some disciplines or areas of science, World Data Centres are recognised as providing a good service. Since their inception the WDCs have been able to arrange data access for scientists in regions where governmental restrictions would otherwise have prevented exchange. Importantly, the WDC principle of making data freely open and available for research (Box 6) has inspired other organisations to adopt similar principles. The WDC panel, which is made up of 8 people selected by ICSU, oversees the WDCs. It has been hampered by a lack of resources but has, over time, made some efforts to identify areas where the network needs expanding and to fill these gaps. Over the past three years, the panel has made considerable efforts to try and improve electronic data access and exchange and networking amongst centres, including the promotion of 'mirror sites'. When the WDC Directors have met together there has been a lot of interest in exchanging experiences and identifying common actions.

When they reviewed the whole WDC network in 2007 the WDC Directors themselves identified three principal weaknesses with the current WDC arrangements[7]. First, there is no real WDC 'system'. The WDCs have not collectively responded in a coordinated way to evolving user needs and increasing expectations for online access to data and supporting services that cut across disciplinary boundaries. Second, there has been no active management of the WDCs in a collective manner. The concept of data interoperability has been only minimally or superficially implemented in the WDCs. Directories, broadband communications and data set interfaces are only dealt with at the local level of an individual WDC. Third, the WDCs have not yet found a way to provide effective science support to less economically developed countries (LEDCs). The success in building East-West bridges during the Cold War has not yet been matched by building extensive South-North bridges. Of the 50 WDCs, nine are in China, one in India and all the others are in OECD countries.

[7] *World Data Centres. Status and priorities for the future*, 11 July 2007. See also appendix D

In 1956 the Federation of Astronomical and Geophysical data analysis Services (FAGS) was established also in the framework of the International Geophysical Year (IGY). The principal purpose of FAGS has been to encourage the analysis of long-term data sets and produce data products for the scientific community and other end-users. There are now 12 permanent services (see appendix C), each operating under the authority of one or more of the scientific unions, namely the International Astronomical Union (IAU), the International Union of Geodesy and Geophysics (IUGG) and the Union Radio-Scientifique Internationale (URSI). The services, or components of services, are maintained nationally and their role is to collect, analyse, interpret and disseminate observations, data and information related to astronomy and geophysics. FAGS is a loose federation of diverse services, some of which themselves act as the hubs for their own distributed networks of data centres. The Federation is overseen by a Council of eight people, including representatives from the relevant Unions, which also provide a small amount of funding that is distributed to the services, via the panel. The Unions provide a quality assurance function for the services and also identify gaps, where new services are needed, within their disciplinary scope.

The acknowledged weaknesses of FAGS are very similar to those of the WDC network (appendix E). It is notable that some key services, even for the disciplinary areas within the FAGS remit, are not currently part of the Federation. The Council has overseen the merger of some services but has not played a strong leadership role in actively developing the system.

The FAGS structure is very much a product of history rather than strategy and, to a lesser extent, the same can be said of the WDC network. Both these networks recognise the need to change and the potential advantages of being part of a broader and better coordinated system (see appendices D and E). The WDC panel has already initiated a number of significant steps to reinvigorate its centres, including an accreditation exercise for all existing centres and site visits and in-depth reviews of specific centres (Chinese centres in 2005, with Russia and Japan being planned). A proposal that all WDCs should make a small financial contribution to the coordination activities of the network as a whole has also been aired. Following the publication of the PAA report in 2004, the FAGS Council took steps to strengthen its interaction with the services. Council members have participated on the directing boards of individual services and service directors have been invited to annual Council meetings. These meetings have been appreciated by the directors as providing a useful forum for discussion of common issues and concerns. Both the WDCs and FAGS responded very positively to the opportunity to provide input to SCID, emphasising the importance that they attach to being associated with ICSU and being forthright in their identification of current problems and weaknesses.

CODATA

The Committee on Data for Science and Technology (CODATA) was established as an ICSU inter-disciplinary body in 1966. Its principal objectives are as follows.

1. Improvement of the quality and accessibility of scientific data, as well as the methods by which data are acquired, managed and analysed.
2. Facilitation of international cooperation on data issues.
3. Promotion of awareness of data issues in the science community.

4. Consideration of data access and intellectual property rights.

In addressing these objectives, it has a particular focus on policy development.

CODATA has 23 national members and 15 scientific union members (appendix F).

CODATA's strengths lie in its cross-disciplinary focus on scientific data, its broad membership and the active participation of individual data experts and, to a lesser extent, scientists. CODATA's core activities include a bi-annual conference on data issues, at which various task-groups are proposed and selected, and the publication of an on-line *Data Science Journal*. However, beyond its immediate membership, CODATA's work is not as widely known as it should be, even amongst those working in scientific data and research centres. National Members from Western Europe and from LEDCs are limited in number and the active links to other organisations and networks with major interests in scientific data are, on the whole, poorly developed. These apparent weaknesses must be considered within the overall context of the limited resources available to CODATA, which has core funding of ~€200k per annum, a full-time director and a single part-time support person.

In response to the PAA on data and information, CODATA developed a draft strategy (see appendix G for Exec. Summary), which focussed on three major initiatives: 1. global information commons; 2. the digital divide; and 3. advanced data methods and information technologies (including virtual observatories). It was notable that each of these topics had considerable overlap with the interests of the WDCs and FAGS and yet until very recently, these three ICSU bodies barely interacted. CODATA had also recently provided support to the data subcommittee for the International Polar Year and had taken the lead in developing a data policy document for the Global Earth Observation System of Systems (GEOSS). Both of these latter data policy activities were in areas of high strategic priority for ICSU.

Links among existing ICSU structures, ICSU research programmes and other stakeholders

As part of its strategic planning exercise, ICSU carried out a PAA on "Environment in relation to Sustainable Development" (ICSU, 2003). This was critical of the WDCs and FAGS for not responding strategically to the data needs of ICSU's own interdisciplinary programmes. This included having poorly developed links with the four global environmental change programmes – in climate, geosphere-biosphere, biodiversity, and human dimensions – as well as the Global Observing Systems. Since that time, a major new ICSU programme – the International Polar Year – had been launched, for which the lack of a mechanism for developing and implementing a coordinated data management strategy presented a major problem. Ensuring the potentially unique data legacy of IPY was now the single major concern of the Joint ICSU-WMO Committee that had responsibility for overseeing this programme [See Box 7].

Other international initiatives, such as the planning for a Global Earth Observations System of Systems (GEOSS, Box 5) have illustrated the need for a forum in which coordinated data management strategies frameworks can be developed and the role for ICSU structures identified in an effective and timely manner. New ICSU programmes are currently under development in areas such as natural hazards, urban health or ecosystem services, which will all present their own complex challenges for data management. Some of these challenges should be met by a new ICSU World

Data System (see ahead) and some by CODATA but it will be crucial that this is coordinated with other stakeholders, including the ICSU Membership – many of whom are very active in data management. As asserted in the PAA there is an opportunity, and a need, for ICSU to take a leadership role in this area on behalf of the global scientific community. In order to achieve this it is proposed that, in addition to restructuring and refocusing existing structures, a new *ad hoc* Strategic Coordinating Committee for Information and Data be established[8].

3.2. Recommended future structures and directions

Individually the WDCs, FAGS and CODATA continue to make a positive contribution to data and information management and policy development as asserted in the PAA in 2004. However, they could have a much greater impact if they strengthened their existing networks and worked more closely together. Taking into account the changing science demands and needs, it is recommended that FAGS and WDCs be expanded into a new unified system that also incorporates other 'state of the art' scientific data and information centres. This new ICSU World Data System (ICSU WDS) should develop closer links with CODATA in areas of common interest. In order to ensure a coordinated and strategic response from the global scientific community to international initiatives such as IPY and GEOSS, a new *ad hoc* ICSU Strategic Coordinating Committee for Information and Data should be established. This Committee would be advisory to ICSU and work closely with the WDS, CODATA and other international data and information bodies – both within and outside the immediate ICSU community. Figure 1 provides a summary of the proposed new structures.

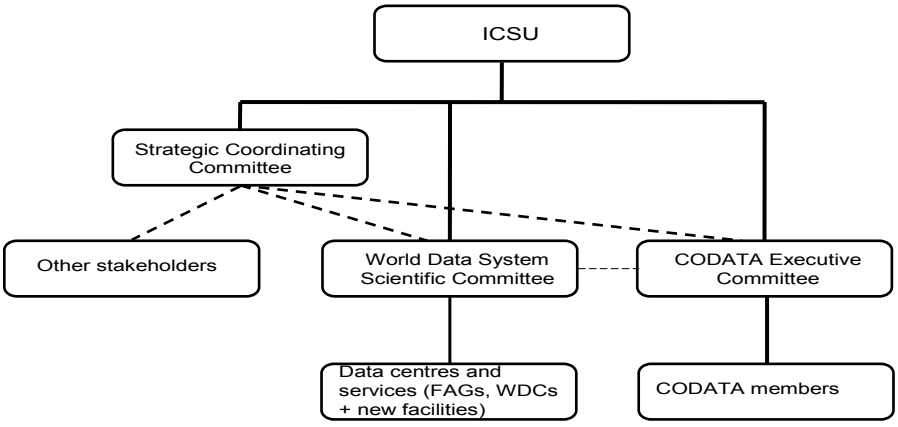


Figure 1: The relationship between the proposed new ICSU structures.

Solid lines indicate structural links and/or direct reporting responsibilities. Broken lines indicate strategic coordination and cooperation functions. As Interdisciplinary Bodies, both CODATA and the ICSU World Data System have direct reporting lines to ICSU, in addition to strong cooperative links with the Strategic Coordinating Committee

[8] In the PAA report on Data and Information (ICSU 2004) it was proposed that ICSU establish a broad-based Scientific Data and Information Forum (SciDIF). SCID considered that a single large forum was no longer appropriate but that more-focussed multi-stakeholder fora on specific strategic topics should be organised by the new *ad hoc* Strategic Coordinating Committee making use of existing mechanisms, such as the bi-annual CODATA Conference.

In summary, it is recommended that:

1) **ICSU assert a much-needed strategic leadership role on behalf of the global scientific community in relation to the policies, management and stewardship of scientific data and information;**

This is a reassertion of one of the primary recommendations in the earlier ICSU Priority Area Assessment on Data and Information. In order to achieve this, ICSU must reform its current interdisciplinary bodies and establish a new committee that will provide overall strategic direction and advise and ensure proper coordination among ICSU activities.

2) **a new ICSU World Data System be created (as an Interdisciplinary Body), incorporating the WDCs and FAGS as well as other 'state of the art' centres;**

This new structure or system will:

- support ICSU's mission and objectives, ensuring the long-term stewardship and provision of quality-assessed data and data services to the international science community and other stakeholders;
- have a broader disciplinary and geographic base than the current ICSU networks and be recognised as a world-side 'community of excellence' for data issues;
- work closely with CODATA and with the new ICSU *ad hoc* Strategic Coordinating committee.

It will require its own governance system and dedicated secretarial support (see ahead)

3) **CODATA focus its activities on the three major initiatives identified in its draft strategy and also extend its links to other organisations and networks to play a more prominent role within ICSU and within the wider scientific community;**

This will require:

- the close alignment of implementation mechanisms, e.g. working groups and task groups, with the 3 major initiatives identified in the strategic plan;
- close working links with the new ICSU World Data System in areas of common interest, e g data policies and data publishing;
- close links with ICSU and the new *ad hoc* Strategic Coordinating Committee, e.g. in organising sessions at the bi-annual Conference.

In order to be effective, CODATA needs more financial support and, as previously proposed in the Priority Area Assessment on Data and Information, ICSU should strongly encourage its National Members and Unions to become Members of CODATA.

4) **a new *ad hoc* ICSU Strategic Coordinating Committee for Information and Data be established to provide broad expertise and advice to ICSU in this area;**

The new ICSU committee will:

- act as an interface between scientists and data and information professionals that can advise on the data needs and possible solutions for existing and new ICSU programmes and international initiatives such as GEOSS;
- act as an expert advisory and monitoring committee for the implementation of the new ICSU WDS structure and the continued strategic development of CODATA;
- establish an agenda for international discussions on key data and information issues, which could be largely implemented via existing mechanisms, for example special sessions at the bi-annual conference of CODATA.

The membership of this committee (~12 persons) should include both scientists and data and information professionals from a range of disciplines and countries. It is proposed that this committee be established for three years in the first instance, potentially renewable for a further three year term. It should have strong links with all the relevant existing ICSU bodies (CSPR, WDS, INASP, ICSTI).

5) **ICSU National Members and Unions be strongly encouraged to establish committees or commissions, where these do not already exist, focussing on data and information issues;**

Such structures should:

- ensure an interface with ICSU and between scientists and data and information professionals;
- raise the profile of data and information issues at the national or disciplinary level;
- promote training and capacity building for data and information management and recognition of data services as an essential part of the long-term infrastructure of science.

Where national committees or liaison structures already exist for CODATA and/or the WDCs, consideration should be given to amalgamating and expanding these to provide an integrated approach to data policy, management and stewardship issues.

3.3. Expected benefits of the proposed new structures and directions

A number of benefits, arising from the successful implementation of recommendations 1-5 can be expected for the organisations themselves and for a variety of key stakeholders.

ICSU

ICSU has not been in a position to easily provide a coordinated response to global data and information initiatives such as GEOSS or the World Summit on the Information Society. Nor has ICSU been well-positioned to give the necessary attention to data issues arising from the development of new interdisciplinary programmes, such as the International Polar Year. The new *ad hoc* Strategic Coordinating Committee will provide a focus for discussion of such activities. This should enable ICSU to establish leadership in data and information for science, permit a more effective response and avoid duplication of activities.

Data services and centres

By combining the efforts of its component parts, the ICSU World Data System will be able to promote their individual and collective capabilities more widely. A coordinated focus on topics such as virtual observatories, building on the interests of CODATA and the new ICSU World Data System, as well as ICSU Members, will help provide better data and information services to scientists. Sharing of experiences and practices in challenging areas such as the management of real-time data streams should benefit both individual facilities and the system as a whole.

CODATA

CODATA will benefit from its closer contact with both ICSU and the new ICSU WDS. CODATA's strategic activities will be strengthened by its close links with the new *ad hoc* ICSU committee. CODATA's expertise on data policy and best practice can inform the World Data System. CODATA itself will be strengthened by exposure to the operational challenges facing the data centres and services.

Scientific Unions

The new *ad hoc* committee will provide a focus and forum within ICSU for the strategic consideration of data and information issues of importance to the Unions. This will complement and add value to the Unions' own structures for considering data issues.

The scientific unions have a critical role to play in ensuring the scientific quality of both the services and their products within the new ICSU WDS. The unions should be involved in the identification and validation of new data services and in the overall governance of the system.

Closer involvement of the Unions with CODATA, for example in specific sessions at the bi-annual conference, will ensure that scientific needs and data policies are developed simultaneously in a coordinated manner.

Interdisciplinary Programmes

The new *ad hoc* committee will provide a forum for the timely discussion of, and response to, the data requirements of existing and new international interdisciplinary programmes.

The expanded ICSU WDS will be specifically implemented to ensure the interoperability of data from different sources.

Individual Scientists and Users

Scientists will have a simpler entry into the ICSU World Data System network to access quality-assessed scientific data and information. This should be the case for scientists in all countries, including those working in less economically developed and transition countries.

Data Providers

The establishment of a 'state of the art' ICSU WDS that has specific responsibility for the long-term stewardship of scientific data and information can help take the burden off data providers, such as observing systems, whose main mission does not encompass long-term data stewardship. The development of data tracking and publishing mechanisms within the WDS will ensure that data providers are accredited for the data that they provide.

Other data and information networks

There are a number of data centre networks that have links to the current FAGS and WDC but are not part of these ICSU structures. Some of these, for example the national oceanographic data centres or the national meteorological services, are very well structured with their own governance bodies, such as the International Oceanographic Data Exchange (IODE), and common portals. Data interoperability and the development of common standards are a major interest that these networks share with the new WDS. The data policy activities of CODATA should also be of interest to all these networks. The establishment of the *ad hoc* ICSU strategic coordinating committee will provide a focus within ICSU for bringing these interests together, identifying best practices and developing common approaches. It should ensure better liaison and coordination with key data management bodies outside of ICSU, including the rapidly developing Global Earth Observation System of Systems.

Developing Country perspectives

The overarching aim of the restructuring and refocusing of ICSU's data and information bodies is to ensure universal and equitable access to scientific data and information. It is crucial that data services from the South (for example, see Box 8) be included in the new ICSU WDS. The enhancement of capacity for professional data management in all countries should be a major focus for the new ICSU *ad hoc* committee and the WDS. As recommended in the PAA, reducing the digital divide is one of the three main strategic priorities for CODATA and a greater focus on this area can have considerable benefits, for example by improving data policies.

Other stakeholders

The visible assertion of a leadership role for ICSU on data and information issues and the establishment of an *ad hoc* committee to ensure expert advice and strategic coordination among ICSU activities, will make ICSU a more desirable and effective partner.

The envisaged benefits of re-structuring of ICSU's data and information bodies, as listed above, can also be considered as indicators of success. It will be important that ICSU monitors these indicators and consults all stakeholders on a regular basis.

Box 8 **The Network of Data and Information Curation Centres (NeDICC): a stimulus to science in Africa**.

In South Africa, a number of institutions and domain-specific networks have established data curation activity. There are no national practices and policies in place yet. A proposed 'eResearch for South Africa' initiative has thus far failed to capture the attention of key stakeholders in the system, but a group of stakeholders have initiated a feasibility study for a National Data and Information Curation Centre (NaDICC).

In parallel with this, the publication of the Principles & Guidelines for Access to Research Data from Public Funding, designed to give shape to the OECD Declaration on this topic, gave rise to a study conducted for the Department of Science and Technology. This led up to a Workshop in September 2007, to review the policy implications of South Africa's commitments in terms of this declaration.

The NaDICC concept has been refined via this process and is now expressed as the Network for Data and Information Curation Center (NeDICC), a digital commons designed to provide thought leadership and practical support to researchers, research institutions and government departments that make, or wish to make, digital data and information available to researchers and other users in Southern Africa, and possibly, Africa as a whole.

4. Governance and finance

ICSU and the new *ad hoc* Strategic Coordinating Committee on Data and Information

As has been discussed, ICSU itself, via the Committee on Scientific Planning and Review (CSPR), should play a more proactive role in developing the global agenda for scientific data and information and in guiding and monitoring the activities of its data and information bodies.

The membership of this new *ad hoc* committee (~12 persons) should include both scientists and data and information professionals from a range of disciplines and countries. It should have strong links with all the existing ICSU structures (CSPR, WDS, CODATA, INASP, ICSTI) but members should be selected for their individual qualities and not simply as representatives. This new *ad hoc* committee should be established as soon as possible and be directly supported by ICSU. Its performance and achievements, including the need for its continuity in the light of the development of other ICSU structures, should be reviewed by CSPR after three years.

ICSU World Data System

Concerns have been raised by both the WDC and FAGS directors, in relation to the existing governance bodies, the FAGS Council and WDC Panel. Despite the earnest efforts of devoted individuals, these bodies have been perceived as making very little contribution to the networks as a whole. This appears to be in part related to both the remit and composition of these committees but mostly to their lack of resources. These issues need to be explicitly addressed in establishing an effective governance structure for the new ICSU World Data System.

SCID proposes that a new scientific committee of the World Data System will be responsible for developing an overall strategy for the development of the system, ensuring active participation by all WDS facilities and services in the network, adherence to agreed standards, review and accreditation of new facilities or services. This committee should be composed of leading scientists and data centre/service directors (maximum 10 people), and be balanced for geographical representation and gender. The length of service of individual committee members should be limited, in line with other ICSU committees. In order to be effective the committee will need dedicated personnel support (minimum 0.5 FTE), which might be located within one of the centres or services.

Individual World Data System facilities will continue to be self-governed with oversight and input from the ICSU Unions where appropriate. However, a General Assembly of all facilities should be convened at regular intervals, possibly in association with the CODATA bi-annual congress.

During the initial 3-year period of the development of the WDS, ICSU should provide financial support for the scientific steering committee. The scientific Unions, which currently provide funding for FAGS, should also be invited to transfer this support to the WDS scientific committee. In the longer-term the main mechanism of support for the WDS system might be a small levy/membership fee for each of its facilities; this would increase the onus on the system to provide ‘added value’ to what can be achieved by the centres acting in isolation. It is recognised that many of the current

data centres and services are opposed to such a membership fee and indeed the FAGS services currently receive a small amount of money annually via the FAGS panel. However, it is also clear that ensuring adequate support to the WDS scientific committee in the longer term will be critical to the viability of the system as a whole.

Funding for individual facilities within the ICSU-WDS cannot be provided by ICSU. These facilities should be recognised as an essential part of the core infrastructure of science and must be funded accordingly by national and international funding bodies. Within the existing WDC and FAGS facilities there already exists a wide variety of business models and the scientific committee for the WDS should facilitate comparison of these and identification of best practices.

CODATA

It is recognised that CODATA has its own Membership, governance and dues structure independent of ICSU. At the same time, it is an ICSU Interdisciplinary Body and should be an important component of the implementation of ICSU's strategy. It is envisioned that the existing CODATA Members will embrace the idea of extending the CODATA networks and working closely with the new ICSU *ad hoc* committee and the WDS system. In particular, the bi-annual CODATA conference provides an ideal mechanism for ensuring close interchange between these structures. It is suggested that this conference be modified to incorporate special forum sessions, which would be co-organised with the new ICSU *ad hoc* committee and address both data policy and management issues in an integrated way. For example, a forum session could consider both data policy and stewardship issues for a developing ICSU programme. It is also suggested that the General Assembly of the WDS system might on occasions be held in association with the CODATA conference, which would help in developing areas of common interest.

For CODATA to realise its full potential it requires more resources and ICSU should strongly encourage its own Members to consider joining and supporting CODATA. The establishment of national or disciplinary data and information committees or commissions by ICSU Members could provide a new focus for promoting and supporting CODATA.

5. A model implementation plan

5.1 Approach to implementation

This section concentrates on the implementation of the main recommendations contained in section 3 of this report. It should be considered mainly as illustrative rather than prescriptive. There are broadly two categories of implementation required at this critical stage of the process of developing ICSU's leadership role in scientific data and information management. First, to create or modify the structures and associated governance and membership arrangements of the various bodies and organisations involved. Second, to enable the new or revised structures to develop plans and implement actions that are relevant to their revised missions. The former aspect, in terms of process, including the development of the detailed terms of reference for the new committees, is largely the responsibility of the ICSU Executive Board and Secretariat. Clear indications on the role, remit and composition of the new structures are given throughout this report.

With regards to developing plans and implementing key actions for the new structures, SCID did not want to be too prescriptive but has identified a number of key milestones, which it considers should be achieved and which could serve as indicators of progress. The dates given in parentheses are provisional target achievement dates but it is recognised that some of these may need to be adjusted once the new structures are in place.

5.2 Confirming ICSU leadership

I. Following key recommendations 1 and 4 earlier in this report, the ICSU Board to establish an *ad hoc* Strategic Coordinating Committee for Information and Data, which will include representation from CSPR.

II. In line with key recommendation 2, the ICSU Board to establish a new World Data System with its own scientific committee as a new ICSU Interdisciplinary Body.

III. The ICSU Board and Secretariat to use the PAA and SCID reports to promote widely in the scientific community and elsewhere, ICSU's new leadership role and aims in scientific data and information and the importance of partnering with other key stakeholders

Action on all three of the above should begin as soon as possible after the publication of the SCID report, although it is recognised that the formal establishment of new structures cannot be done until Members have been fully consulted and the General Assembly has given its approval in October 2008. Both of the new committees (actions I and II above) should be in place and functioning early in 2009.

Once established, **the new *ad hoc* Strategic Coordinating Committee for Information and Data** will:

IV. Develop a strategy for data stewardship and coordinated ICSU participation in relation to the planned Global Earth Observation System of Systems and existing global observing systems [2009].

V. Consider and advise, as necessary, on the data and information issues (needs and solutions) of new ICSU programmes, including natural hazards, urban health, Millennium Ecosystem Assessment follow-up [2009-2010].

VI. Develop a coordinated strategy for training and capacity enhancement in data stewardship, involving the activities of CODATA, ICSU WDS and other relevant Interdisciplinary Bodies.[2009-2010]

VII. Organise multi-stakeholder forum sessions on key strategic issues in data and information (e.g. V above) for inclusion in the CODATA bi-annual conference in 2010.

VIII. Monitor and assess the progress made in implementing the new ICSU World Data System and revised CODATA strategy, reporting on these to CSPR in 2011.

IX. Develop a sustainability plan for maintaining the established strategic coordination and leadership role of ICSU for consideration by the General Assembly in 2011.

5.3 Developing the ICSU World Data System

I. The existing WDC panel and FAGS council to be replaced by a new WDS scientific committee. Dedicated secretarial support (minimum 0.5FTE) to be identified for this committee from within one of the accredited centres or services. [mid 2009]

II. WDS scientific committee to define its operating procedures, by-laws and rules, including a model consortium agreement/MoU for new and existing centres and services. Such an MoU should enshrine the long-term commitment of the responsible national funder(s) as well as the relevant facility authorities and ICSU itself. [mid 2009]

III. Development and implementation of a single data portal for locating and accessing all WDS data and services on line [2009]

IV. Accreditation process for all existing data centres and services to be completed [end 2009]

V. After consultation with ICSU Members and Interdisciplinary Bodies new facilities and services to be introduced into WDS, including those in LEDCs and transition countries [begin in 2009 and establish annual targets]

VI. Define and achieve minimal interoperability across the large majority of centres and services [2010]

VII. Define and achieve good interoperability across all centres and services [2012]

VIII. Stimulate the development of formal, peer-reviewed data publications accompanied by formal citation procedures and author recognition [2009-2011]

IX. Work with the new ICSU *ad hoc* committee and CODATA to organise forum sessions in the CODATA bi-annual conference in 2010.

X. Convene a General Assembly of WDS Directors in 2010

5.4 Focusing and strengthening CODATA

I. The CODATA Executive Committee and General Assembly to adapt its strategic plan and focus its implementation mechanisms, such as task forces, in line with the SCID recommendations [2008 and 2010]

II. CODATA to work with the new *ad hoc* ICSU committee and the ICSU WDS scientific committee to organise forum sessions at its bi-annual conference in 2010.

III. ICSU to encourage its own Members to become members of CODATA.[2008-2009]

6. Appendix A Members of SCID

Roberta Balstad
Earth Institute, Columbia University
New York
USA

Nicole Capitaine
Observatoire de Paris
Paris
France

Michael Diepenbroek
Institute for Marine Environmental Sciences (MARUM)
University of Bremen
Germany

Kim Finney
Australian Antarctic Division, Department of Environment & Heritage
Kingston
Australia

Peter Fox
High Altitude Observatory
Earth Sun Systems Lab, National Centre for Atmospheric Research
Boulder
USA

Alexei D. Gvishiani
Director, Geophysical Center, Russian Academy of Sciences
Moscow
Russia

Ray Harris [Chair]
University College London, Department of Geography
London
UK

Toshio Koike
Department of Civil Engineering,
University of Tokyo
Japan

Jean-Bernard Minster
Scripps Institution of Oceanography
University of California,
USA

Ruth Neilan
Jet Propulsion Laboratory
Pasadena
USA

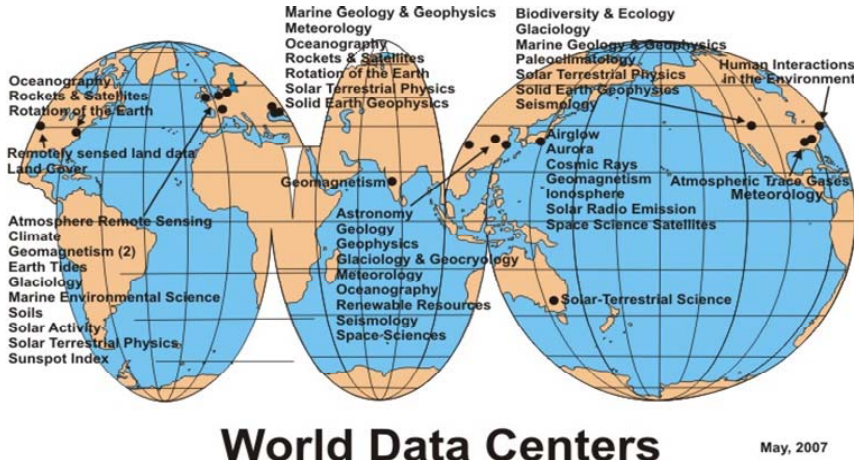
Ray Norris
CSIRO Australia Telescope National Facility,
Australia

*Alejandro Pisanty
Director General de Servicios de Computo
Academico, Universidad Nacional Autonoma de Mexico (UNAM)
Mexico

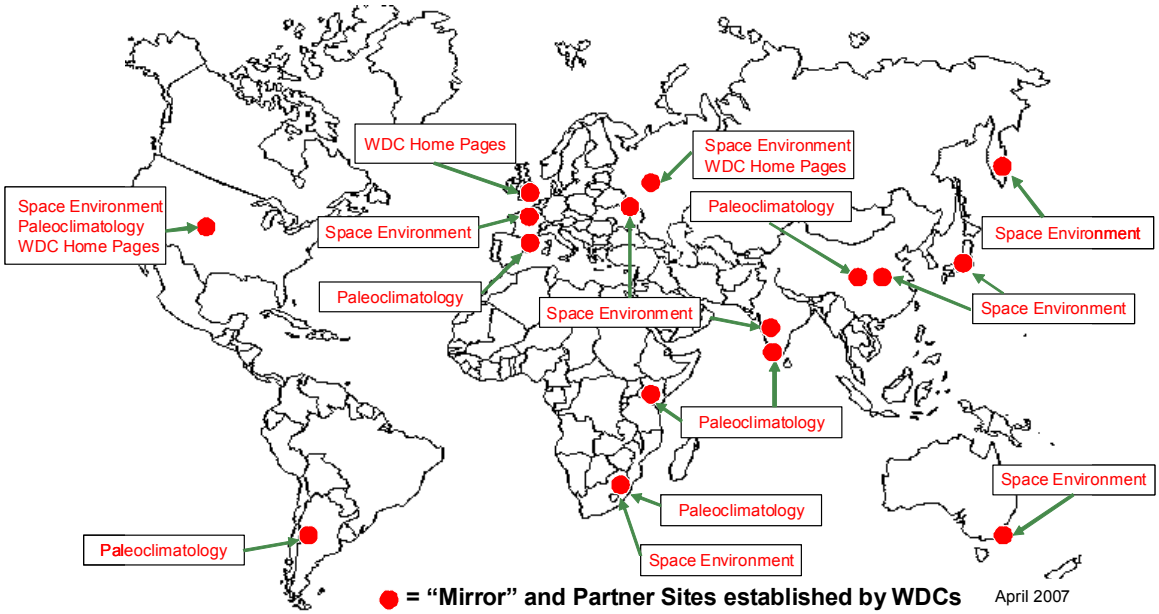
Daisy Selematsela
Knowledge Management & Strategy Directorate, National Research Foundation
Pretoria
South Africa

*Alejandro Pisanty was unable to attend the SCID meetings in person but provided extensive input and comments on the various issues under discussion and on the text of the final report.

7. Appendix B



The WDC system was created 50 years ago with sites in Europe, The Soviet Union, Japan and North America The system was subsequently enhanced in a variety of disciplines in response to international programs and extended to other countries, most notably China. Recent additions are more broadly environmental. A new WDC on "Biodiversity and Human Health" is currently under consideration in South Africa. However, the large majority of WDCs are still located in northern hemisphere countries, an imbalance that the Panel has undertaken to remedy by focusing on electronic technologies for data access and exchange, including the development of 'mirror sites'.



List of World Data Centers with Hyperlinks to WDC Guide page and WDC home page				
	World Data Center (Hyperlink to WDC Guide page)	**Host Institution**	**World Data Centre Web site** (Hyperlinked, as taken from the WDC Guide Web pages)	**Date established**
1	Airglow	National Astronomical Observatory, Tokyo, Japan	http://solarwww.mtk.nao.ac.jp/wdc.html	1957
2	Astronomy	Beijing Astronomical Observatory, CAS, Beijing, China	http://badc.lamost.org	1988
3	Atmospheric Trace Gases	Oak Ridge National Laboratory, USA	http://mercury.ornl.gov/cdiac/ No WDC Web page	
4	Aurora	National Institute of Polar Research, Tokyo, Japan	http://polaris.nipr.ac.jp/~aurora/	1981
5	Biodiversity and Ecology	USGS Center for Biological Informatics, Denver, CO, USA	http://wdc.nbii.gov	2002
6	Climate	Max-Planck-Institute for Meteorology and German Climate Computing Centre, Hamburg, Germany	http://wdc-climate.de/	2002
7	Cosmic Rays	Solar-Terrestrial Environment Laboratory, Nagoya University, Toyokawa, Japan	http://www.env.sci.ibaraki.ac.jp/database/html/WDCCR/index.html	1957
8	Earth Tides	University of French Polynesia, BP 6570, 98702 Faaa, Tahiti	http://www.astro.oma.be/ICET/	
9	Geology	Chinese Academy of Geological Sciences, Beijing, China	http://www.wdcgeo.net	1988
10	Geomagnetism, Copenhagen	Danish Meteorological Institute Copenhagen, Denmark	http://dmiweb.dmi.dk/fsweb/projects/wdcc1/	
11	Geomagnetism, Edinburgh	British Geological Survey, Edinburgh, UK	http://www.wdc.bgs.ac.uk/	
12	Geomagnetism, Kyoto	Kyoto University, Kyoto, Japan	http://swdcwww.kugi.kyoto-u.ac.jp/index.html	1957
13	Geomagnetism, Mumbai	Indian Institute of Geomagnetism, Mumbai, India	http://www.wdciig.res.in/	
14	Glaciology, Boulder	CIRES, University of Colorado, Boulder, USA	http://nsidc.org/data/wdc.html	1976
15	Glaciology, Cambridge	Scott Polar Research Institute, Cambridge, UK	http://wdcgc.spri.cam.ac.uk/	
16	Glaciology and Geocryology, Lanzhou	Cold and Arid Regions Environmental and Engineering Research Institute, Chinese Academy of Sciences, Lanzhou, China	http://wdcdgg.westgis.ac.cn/	1988
17	Human Interactions in the Environment	CIESIN, Columbia University, Palisades, NY, USA	http://sedac.ciesin.columbia.edu/wdc/	1995
18	Ionosphere	National Institute of Information and Communications Technology, Tokyo, Japan	http://wdc.nict.go.jp/index_eng.html	1957
19	Land Cover Data	USGS, EROS Data Center, Sioux Falls, USA	http://landcover.usgs.gov/ No WDC web page	2002
20	Marine Environmental Sciences	Centre for Marine Environmental Sciences and Alfred Wegener Institute for Polar and Marine Research, Bremen, Germany	http://www.wdc-mare.org/	2001
21	Geophysics and Marine Geology, Boulder	NOAA NGDC, Boulder, CO, USA	http://www.ngdc.noaa.gov/mgg/aboutmgg/aboutwdcmgg.html	1975
22	Marine Geology & Geophysics, Moscow	Glav NIVC MNR RF, Moscow, Russia	Bad URL	
23	Meteorology, Asheville	National Climatic Data Center, Asheville, NC, USA	http://www.ncdc.noaa.gov/oa/wdc/index.php	1957
24	Meteorology, Beijing	National Meteorological Information Center, Beijing, China	http://data.cma.gov.cn/index.jsp	1988
25	Meteorology, Obninsk	All-Russian Research Institute of Hydrometeorological Information, Obninsk, Russia	wdcb@meteo.ru No WDC Web Page	
26	Oceanography, Obninsk	All-Russian Research Institute of Hydrometeorological Information, Obninsk, Russia	wdcb@meteo.ru No WDC Web page	
27	Oceanography, Silver Spring	NOAA/NODC, Silver Spring, MD, USA	http://www.nodc.noaa.gov/General/NODC-dataexch/NODC-wdca.html	

28	Oceanography, Tianjin	National Marine Data & Information Service State Oceanic Administration, Tianjin, China	http://wdc-d.coi.gov.cn/english/eindex.html	1988
29	Paleoclimatology	NOAA/NGDC, Boulder, USA	http://www.ncdc.noaa.gov/paleo/datalist.html	1989
30	Remotely Sensed Land Data	USGS, EROS Data Center, Sioux Falls, SD, USA	http://edc.usgs.gov/ No WDC web page	
31	Remote Sensing of the Atmosphere	German Aerospace Centre, DLR, German Remote Sensing Data Centre (DFD), Oberpfaffenhofen, Germany	http://wdc.dlr.de/	2002
32	Renewable Resources and Environment	Institute of Geographical Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing, China	http://eng.wdc.cn:8080/Metadata/index.jsp	1988
33	Rockets and Satellites	All-Russian Research Institute of Hydrometeorological Information, Obninsk, Russia	http://meteo.ru/english/ No WDC web page	
34	Satellite Information	NASA Goddard Space Flight Center, Greenbelt, USA	http://nssdc.gsfc.nasa.gov/about/about_wdc-a.html	
35	Space Science Satellites	Inst. Space & Astronautical Science, Kanagawa, Japan	http://darts.isas.jaxa.jp/index.html.en No WDC web page	1969
36	Rotation of the Earth, Obninsk	All-Russian Research Institute of Hydrometeorological Information, Obninsk, Russia	http://meteo.ru/english/ No WDC Web page	
37	Rotation of the Earth, Washington	U.S. Naval Observatory, Washington, DC, USA	http://maia.usno.navy.mil/ No WDC Web page	
38	Seismology, Denver	USGS, Denver Federal Center, Denver, USA	http://neic.usgs.gov/ No WDC Web page	1986
39	Seismology, Beijing	China Earthquake Networks Center, Beijing	http://210.72.96.21:8080/wdc/home-1.html	2004
40	Soils	ISRIC - World Soil Information, Wageningen, The Netherlands	http://www.isric.org/UK/About+Soils/WDC+for+Soils/	1989
41	Solar Activity	Observatoire de Meudon, Meudon, France	http://bass2000.obspm.fr/home.php No WDC Web page	1978
42	Solar Radio Emissions	Nobeyama Solar Radio Observatory, National Astronomical Observatory, Nagano, Japan	http://solar.nro.nao.ac.jp/norp/archive.html	1969
43	Solar Terrestrial Physics, Boulder	NOAA/NGDC, Boulder, USA	http://www.ngdc.noaa.gov/stp/WDC/wdcstp.html	1957
44	Solar Terrestrial Physics, Chilton	Rutherford Appleton Lab, UK	http://www.ukssdc.ac.uk/wdcc1/data_menu.html	
45	Solar-Terrestrial Physics, Moscow	Geophysical Center, RAS Moscow, Russia	http://www.wdcb.ru/stp/index.en.html	1956
46	Solar-Terrestrial Science, Sydney	IPS Radio and Space Services, Bureau of Meteorology, Sydney, Australia	http://www.ips.gov.au/World_Data_Centre	2000
47	Geophysics, Beijing	Institute of Geology and Geophysics, CAS, Beijing, China	http://gp.wdc.cn/wdc/english/indexnew.htm	1988
48	Solid Earth Physics, Moscow	Geophysical Center, RAS, Moscow, Russia	http://www.wdcb.ru/sep/	1971
49	Space Science	Center for Space Science and Applied Research, Chinese Academy of Sciences, Beijing, China	http://www.cssdc.ac.cn/	1988
50	Sunspot Index	Royal Observatory of Belgium, Brussels, Belgium	http://sidc.oma.be/	

8. Appendix C FAGS services

Status of Services, March 2008

Service	Title	Parent Unions	Hosting Institution	International cooperative network#	Date of association
BGI	Bureau Gravimétrique International	IUGG	CNES &Observatoire Midi-Pyrénées, Toulouse, France		1956
CDS	Centre de Données astronomiques de Strasbourg	IAU	Strasbourg Observatory, France		1985
ICET	International Center for Earth Tides	IUGG	Observatoire géodésique de Tahiti, Université de Polynésie Française*		1966
IERS	International Earth Rotation and Reference system Services	IAU, IUGG	Central Bureau at BKG, Frankfurt am Main, Germany	Yes	1988
IGS	International GNSS Service	IUGG,	Central Bureau at Jet Propulsion Laboratory, Pasadena, USA	Yes	1994
ISES	International Space Environment Service	IAU, URSI IUGG	Central Bureau at Geomagnetic Laboratory, Natural Resources, Ottawa, Canada	Yes	1956
ISGI	International Service of Geomagnetic Indices	IUGG, URSI	Coordinating Centre at CETP, Saint-Maur-les-fosses, France	Yes	1956
IVS	International VLBI Service for Geodesy and Astrometry	IAU, IUGG	Coordinating Centre at GSFC, Greenbelt, MD, USA	Yes	2001
PSMSL	Permanent Service for Mean Sea Level	IUGG	Proudman Oceanographic Laboratory, Liverpool, UK		1957
SIDC	Solar Influences Data Analysis Centre	IAU, URSI IUGG,	Royal Observatory of Belgium, Bruxelles, Belgium		1985
QBSA**	Quarterly Bulletin on Solar Activity	IAU	National Astronomical Observatory, Minamimaki, Japan		1956
WGMS	World Glacier Monitoring Service	IUGG	Central Office at Universität Zürich, Switzerland		1986

Several of the FAGS services are themselves 'hub and spokes' networks, with data and information from ancillary facilities across the globe being essential for the products of the central bureau or coordinating centre. In some cases, eg IERS and IGS, the number of networked agencies are in the hundreds.

*Previously hosted at Royal Observatory of Belgium, until January 2008

** Disbanded as FAGS service in April 2008

9. Appendix D Recommendations by the WDC

World Data Centre recommendations, compiled as input to SCID, in a paper, dated 11 July 2007, arising out of a meeting of WDC Directors in Bremen in May 2007.

- The WDC system must respond robustly and effectively to ICSU's program data management needs. ICSU will not support the *status quo ante*. The WDCs will work with ICSU to meet this requirement.
- The WDCs must strongly and actively support the data management needs of the IPY, a major new program of ICSU.
- The WDCs must become an active partner in the planning of the GEOSS data activities.
- The WDCs must implement network links between WDCs utilizing interoperable data systems to support current scientific programs. The WDCs must begin to promote the adoption and promotion of standards within the systems in order to achieve this interoperability.
- The WDCs must establish a baseline of IT capabilities that will form the backbone of a Global Science Data Network of WDCs.
- The WDCs and the WDC Panel must identify and secure funding for managing the System.
- Data publishing is an effective way of making historical data available and should be widely implemented in the WDC System.
- The WDC system and FAGS should discuss common operations and a possible merger of their activities.
- Since the some of the National Oceanographic Data Centers of the IODE are WDCs, closer interactions and coordination between the two systems are logical and necessary. IODE relies on the WDCs for data archival and dissemination.
- WDCs need to integrate their future IT activities with new state-of-the-art technologies like Virtual Observatories and the activities of electronic Geophysical Year.
- The WDC system needs to expand its discipline structure and its geographic distribution to better serve ICSU programs—while perhaps at the same time consolidating in some areas. This includes a concerted push to expand to developing countries and the Southern Hemisphere. The WDC Panel and ICSU need to identify mechanisms for the establishment of new WDCs within national structures.
- The WDCs should be open to look at a new paradigm in restructuring the WDC System (virtual WDCs?).
- As the world transitions into a digital environment, the WDCs must address the fate of analog data and make it a priority to rescue the analog data by converting them into digital data.

10. Appendix E Recommendations by FAGS

FAGS recommendations to SCID in a 'white paper', dated 10 July 2007 and compiled by the FAGS panel after consultation with Service Directors

- FAGS should be restructured to include more data products and to serve more useful coordinating and interfacing functions than at present.
- The word "Service" may be changed in order to adapt to the range of information available.
- It was suggested that a new ICSU data/information entity could provide "one-stop shopping" – a way that scientists could discover world resources and obtain guidance on how to deposit and retrieve the data and information; and policy makers could ask relevant questions and be led to relevant and credible answers. The Unions could be enlisted to ensure that information was reliable.
- Coordinating the Services with a renewed set of WDCs within a new ICSU data/information entity.
- Introducing other services in the federation in order to fill existing gaps between services and initiating new required ones.
- Facilitating as necessary small data analysis services to merge in order to address a more inclusive field (e.g. atmosphere, gravity, solar environment)

11. Appendix F Members of CODATA

List of CODATA members, March 2008

Co-opted Organizations (no dues)
WDC
FAGS
ICSTI
WFCC

Scientific Unions (no dues)
IAU
IUPAC
IUPAP
IUBS
IGU
IUCr
IUBMB
IUGS
IUPsyS
IUPAB
IUNS
IUPHAR
IUIS
IUMS
IUSS
IUGG

Countries (pay dues)
Australia
Brazil
Cameroon
Canada
Czech republic
China, Beijing
China, Taipei
India
Indonesia
Ireland
Israel
Japan
Korea
Nigeria
Poland
Russia
Senegal
South Africa
Thailand
Ukraine
United States
France (Associate member)
Germany (Associate member)

12. Appendix G Executive summary from the draft CODATA Strategic Plan[9]

Executive Summary

CODATA, as an interdisciplinary body of ICSU focused on scientific and technical data, affirms its commitment to the long-term vision articulated by ICSU of a "world where science is used for the benefit of all, excellence in science is valued and scientific knowledge is effectively linked to policy-making." CODATA is committed to the principle of "universal and equitable access to high quality scientific data and information" and in particular to the goal to "facilitate a coordinated global approach to scientific data and information that ensures equitable access to quality data and information for research, education and informed decision-making" (ICSU, 2006). CODATA is committed to tackle issues related to the *Digital Divide*.

This *CODATA Strategic Plan* articulates CODATA's overall approach and specific plans to meet this goal during the period 2006-2012. The Plan reviews the major obstacles to universal and equitable access to data and assesses the potential role of CODATA in overcoming these obstacles. The Plan proposes a new CODATA mission statement, identifies key priorities for CODATA's scientific agenda, and also recommends organizational changes to improve CODATA's capacity to carry out its agenda.

Specifically, this Plan recommends the following new CODATA mission statement:

> *The mission of CODATA is to strengthen international science for the benefit of society by promoting improved scientific and technical data management and use.*[10]

The Plan also recommends that CODATA pursue three major initiatives over the next 6 years:

1) ***The Global Information Commons for Science Initiative (GICSI).*** Launched by CODATA and several partner organizations at the second phase of WSIS in Tunis, GICSI represents an innovative effort to accelerate the development and "scaling up" of global open-access scientific data and information resources. GICSI will promote full and equitable access to scientific data in key policy arenas and among major stakeholders in the world's diverse scientific community. Through both "bottom up" and "top down" efforts, GICSI will help create a tangible, shared *information commons for science* containing valuable scientific data, information, tools, and other resources accessible to all.
2) ***The Scientific Data across the Digital Divide (SD³) Program.*** To address the pressing needs of developing country scientists, students, and applied users for scientific data related to sustainable development, CODATA will develop a specific program of activities aimed at making critical scientific data and associated tools and resources related to sustainable development widely accessible in developing countries. As part of this effort, CODATA will work with several major international scientific data

[9] Version: Public Draft, July 2007

[10] This new mission statement has been approved at the 25th CODATA General Assembly in Beijing in October 2006.

management activities such as the Global Earth Observing System of Systems (GEOSS), the International Polar Year (IPY), the electronic Geophysical Year (eGY), the Global Risk Identification Program (GRIP), the ICSU World Data Centers, and the Global Biodiversity Information Facility (GBIF) to make their data more accessible and usable for sustainable development. CODATA will develop selected partnerships with key development agencies, nongovernmental organizations, universities, research institutes, and other groups to further this effort. A new opportunity in this regard is the United Nations Global Alliance for Information and Communications Technologies and Development (GAID), an open, multi-stakeholder forum that brings together governments, international organizations, civil society, the private sector, media and other stakeholder constituencies in a common effort to better harness ICT for advancing development.

3) ***Advanced Data Methods and Information technologies for Research and Education (ADMIRE)***. Another key area where CODATA could have both a significant technological and institutional impact is in the application of advanced data mining and integration techniques in research, education, and other applications. ADMIRE will seek to strengthen linkages between the computer science community involved in data mining, data integration, artificial intelligence, and other techniques with particular scientific areas where such approaches could be especially valuable, including materials science, the geosciences, astronomy, ecology, and genetics. One activity currently under development is participation in a project to address the multicultural and multilingual aspects of accessing, exploiting, using, and re-using digital content in Europe. ADMIRE will also address both technical and institutional issues related to long-term stewardship and accessibility of data. The new framework programme (FP7) for research of the European Commission could provide additional funding opportunities for this initiative**.**

In order to successfully carry out these three initiatives, CODATA will need to expand its own scientific, technical, and institutional capacity in several ways:

1) Strengthening of its national and union membership, both by expanding membership to new countries, unions, and interdisciplinary bodies and by helping to energize existing members and help and support of ICSU would be sought, whenever necessary;

2) Expansion of the number and breadth of Supporting Organizations and other partners to include the key data and research centers, organizations, and networks that engage many data-oriented scientists and data professionals, especially those focused on areas critical to sustainable development and those located in developing countries;

3) Development of an "Associates Program" to encourage individual scientists and data professionals from around the world to become active, long-term contributors to CODATA activities;

4) Establishment of an International Data Academy to provide a select expert pool of data information and knowledge scientists who can be called upon for advice on data issues;

5) Expansion of externally funded activities that permit CODATA to develop concrete products and services, involve key stakeholders, hire additional staff or consultants when needed, and increase its visibility and impact;

6) Establishment of a Gift and Endowment Fund to provide CODATA with a stable and flexible source of income; and

7) Strengthening of the CODATA Secretariat.

Hand-in-hand with these efforts, CODATA must also focus and improve its existing portfolio of activities, coordinate its activities with ICSU and other key partners, and increase its flexibility and responsiveness to ongoing, rapid changes in data management, technology, and policy. In particular, CODATA will:

1) Encourage the CODATA Task Groups and Working Groups and the editors of the *CODATA Data Science Journal* to make substantial contributions to GICSI, SD^3, and ADMIRE in their areas of activity;

2) Participate actively in the planned ICSU *ad Hoc* Strategic Committee on Information and Data (SCID) and possible follow-on Scientific Data and Information Forum (SciDIF) and develop cooperative agreements and reciprocal memberships with key partners;

3) Appoint a new Data Policy Committee or Working Group of the Executive Committee charged with monitoring of international data policy issues and recommending CODATA responses in a timely manner;

4) Establish a new Technology Committee or Working Group of the Executive Committee charged with developing a plan for introducing new technologies that can facilitate CODATA's work and its interactions with the broader scientific community;

5) Establish an *ad hoc* Committee of the Executive Committee charged with reevaluating CODATA's dues structure and suggesting modifications or alternative approaches for consideration at the 2008 General Assembly; and

6) Improve the CODATA's outreach to the broader scientific community through a coherent program of publications, Internet-based services, selective participation in key scientific activities, and interactions with key scientific publications.

A number of these actions have already been initiated by the CODATA General Assembly and the CODATA Executive Committee.