



Australian Academy of Science

A Submission to the
Prime Minister's Science, Engineering and Innovation Council
Working Group on Data for Science

11 August 2006

Contact: Professor Philip Kuchel
Secretary, Science Policy
Australian Academy of Science
GPO Box 783
Canberra ACT 2601
Tel: (02) 6201 9401
Fax: (02) 6201 9494
Email: es@science.org.au

A Submission from the Australian Academy of Science to the Prime Minister's Science, Engineering and Innovation Council Working Group on Data for Science

The Australian Academy of Science does not support the notion of a *national* approach for managing data for science. Science is an *international* activity with international, discipline-based databases and is underpinned by publication of scientific findings in prestigious journals, or as patent applications. Editorial policies that support the professional maintenance and use of international databases are the most cost-effective way to ensure access to publicly and privately funded research.

Recent International Activities of Relevance to the Working Group on Data for Science

The Australian Academy of Science funds an annual Elizabeth and Frederick White Conference on a topic considered by the Academy to be of strategic importance. In 2006, the conference was on the topic of *Mastering the data explosion in the Earth and environmental sciences*, and was held at the Shine Dome on 19–21 April. The conference organisers were Dr Malcolm Sambridge, Australian National University; Dr Louis Moresi, Monash University; and Dr Fabio Boschetti, CSIRO Exploration and Mining.

The conference focused on the Earth and environmental sciences and brought together Australian and international experts in spatial statistics, data mining, inversion, grid computing, computer visualisation and numerical simulation of natural processes, together with those solving real world problems using these tools.

A key issue discussed throughout the conference was how to deal with the enormous amount of data that is being collected and generated as researchers try to make sense of new and complex observations about the world around us.

The conference papers are available at <http://rses.anu.edu.au/cadi/Whiteconference>.

Key presentations of relevance to the PMSEIC Working Group for Data for Science included:

Inversion and Imaging for the Solid Earth
Prof. B.L.N. Kennett, FAA, FRS,
Research School of Earth Sciences,.

Fixed Rank Kriging for Massive Datasets
Prof. N. Cressie,
Dept. of Statistics, Ohio State Univ., USA.

Data Explosion: The challenges for Geoscience Australia
Dr. P. McFadden, FAA
Chief Scientist, Geoscience Australia.

Data Mining and its Environmental Applications
Prof. S. Dzeroski,
Dept. of Knowledge Technologies, Jozef Stefan Institute, Slovenia.

Towards a Geoscience Information Commons: the Electronic Geophysical Year, 2007-2008 and the Global Earth Observing System of Systems
C. Barton, A. Held
Australian National University, CSIRO

Computational Frameworks enabling multi-scale multi-physics models
Bill Appelbe, S. Quenette
Victorian Partnership for Advanced Computing

Desperately Trying to Cope with the Data Explosion in Astronomical Sciences
Prof. Ray P. Norris
CSIRO Australia Telescope National Facility

The working group may wish to explore whether Australia is taking full advantage of its links with the international scientific community, parts of which (e.g., the International Council for Science (ICSU), and The Committee on Data for Science and Technology (CODATA)) are actively addressing these questions.

CODATA

Recognizing a world-wide demand for useful, reliable and readily available scientific and technological data, ICSU established in 1966 the Committee on Data for Science and Technology (CODATA). The aim of CODATA is to promote throughout the world the evaluation, compilation and dissemination of data for science and technology and to foster international collaboration in this field.

CODATA's 20th International Conference, *Scientific Data and Knowledge within the Information Society*, sponsored in part by the InterAcademy Panel, is to be held in Beijing, 23-25 October, 2006. CODATA says:

“The 20th International CODATA Conference marks the 40th Anniversary of CODATA. The world is entering a new era of enormous potential for everyone, that of the Information Society and expanded human communication.

CODATA 2006 will be a great opportunity to promote the use of scientific data and knowledge and harness the benefits of the Information Society to reach scientific and developmental targets.

These targets build upon strides the scientific community has taken during this information explosion, particularly in the areas of data and technology. Data have become both more accessible and more necessary to maintain; more abundant and more complex to manipulate; more global and more expected to share.

The 20th International CODATA Conference, aptly titled "Scientific Data and Knowledge within the Information Society," will address issues such as fair and equitable access to data, data archiving, data protection and data security– to name but a few of the issues that will be touched upon.

Science has never relied more heavily on the collection, management and dissemination of data, and CODATA's task to contribute to all aspects of data, from preservation to sharing and value extraction and addition, will be addressed at this International Conference.

CODATA 2006 will feature advanced topic sessions such as:

- Digital Divide
- Internet-based international science communication
- Data IT and Applications
- Global Information Commons for Science Initiative
- Crossing the Social-Natural Science data Boundaries
- Role of Scientific Data in Natural Disaster Management
- I*Ys (Planet Earth, IPY, IHY, and eGY) data management & policy
- Scientific Data Sharing and Application in China
- Infoscience
- CODATA's New Vision”

Declaration for a Geoscience Information Commons

The Australian Academy of Science, through the National Committee for Earth Science, has endorsed the Declaration for a Geoscience Information Commons. This is an initiative of the Electronic Geophysical Year (eGY) 2007-2008), ICSU and the World Summit on the Information Society.

Organisation for Economic Co-operation and Development (OECD)

The working group may wish to examine the extent to which Australian facilities are complying with the OECD Declaration on Access to Research Data from Public Funding, arising from the meeting chaired by Minister Peter McGauran in Paris on 30 January 2004.

The OECD Committee for Scientific and Technological Policy met at Ministerial level on 29, 30 January 2004. Mr. Peter McGauran, Minister for Science of Australia, chaired the meeting, with Ms. Claudie Haigneré, Minister for Research and New Technologies of France, and Mr. Jaime Parada Ávila, Director General of the National Commission of Science and Technology (CONACyT) of Mexico, as Vice-Chairs.

The meeting was preceded by a High-level Forum on "Key Challenges for Science and Innovation Policy", in which prominent representatives of research institutions and business participated.

The Ministers concluded that:

- **Changing innovation processes and the evolution of the relative contribution made by the private and public sectors have emphasised the need for strong industry-science linkages. A well-functioning interface between the innovation and science systems is more necessary than ever to reap the economic and social benefits from public and private investments in research, ensure the vitality and quality of the science system, and improve public understanding and acceptance of science and technology and the importance of innovation.**
- Patent regimes play an increasingly complex role in encouraging innovation, diffusing scientific and technical knowledge, and enhancing market entry and firm creation. As such, they should be subject to closer scrutiny by science, technology and innovation policy makers.
- Increasing participation and maintaining quality standards in tertiary education in science and technology are imperative to meet growing demand for workers with scientific and technological knowledge and skills. Complementary efforts are needed to improve mobility and the attractiveness of research careers in the public and private sectors.
- Improving the accountability of science and technology policy should usefully be addressed through more systematic evaluation exercises. Additional effort is needed to identify and disseminate good practices in this area.
- **Co-ordinated efforts at national and international levels are needed to broaden access to data from publicly funded research and contribute to the advancement of scientific research and innovation. To this effect, Ministers adopted a Declaration entrusting the OECD to work towards commonly agreed Principles and Guidelines on Access to Research Data from Public Funding.**
- Greater international co-operation in science and technology is vital to meet a broad range of global challenges related to economic growth, better health, sustainable development, and enhanced safety and security, as well as for implementing large science projects in a growing range of disciplines. In this regard, Ministers adopted a Declaration aiming at strengthening international S&T co-operation for sustainable development. They endorsed efforts to establish a framework for a Global Biological Resource Centre Network and gave their support to promote scientific co-operation in the fields of high-energy physics and neuroinformatics.

See also **ANNEX 1**

Editorial Practices for International Scientific Journals

Unrestricted access to publicly funded research data is most satisfactorily enforced by international scientific journals with editorial policies that insist that materials and data underlying any publication are available to readers on request. For example, in the areas of biochemistry and molecular biology, papers reporting protein or DNA sequences and crystallographic structures will not be accepted for publication without proof (i.e., an accession number) from a recognised database in general use in the discipline.

The editorial policies in place in this regard for *Nature Genetics* are typical for biomedical research and are reproduced here:

Editorial Policy for *Nature Genetics*

AVAILABILITY OF MATERIALS AND DATA

An inherent principle of publication is that others should be able to replicate and build upon the authors' published claims. Therefore, a condition of publication in Nature Genetics is that authors are required to make materials, data and associated protocols available to readers on request. Any restrictions on the availability of materials or information must be disclosed at the time of submission of the manuscript, and the methods section of the manuscript itself should include details of how materials and information may be obtained, including any restrictions that may apply. One preferred form of disclosure is a link from the methods section to a copy of the relevant Material Transfer Agreement (MTA) form, which will be hosted as supplementary information on the journal's web site. Authors may charge a reasonable fee to cover the costs of producing and distributing materials. If materials are to be distributed by a for-profit company, this should be stated.

For materials such as mutant and genetically modified organisms and cells, authors are expected to use established public repositories (for example, Jackson Laboratory, Mutant Mouse Regional Resource Centers, American Type Culture Collection, UK Stem Cell Bank, and so on) wherever possible.

Referees may be asked to comment on the terms of access to materials, methods and/or datasets, and we reserve the right to refuse publication in cases where authors are unable to provide adequate assurances that essential resources will be made freely available to the community.

*Papers reporting protein or DNA sequences and crystallographic structures will not be accepted without an accession number to Genbank/EMBL/DDBJ, PDB, SWISS-PROT or other appropriate, identified, publicly available database in general use in the field that gives free access to researchers from the date of initial publication (normally the date of online publication); see *Nature* [394, 105; 1998](#) and [404, 317; 2000](#). All novel sequence or structure data must be made available to editors and referees either as Supplementary Information (5 copies if provided on disk) or by an accession number to an appropriate publicly accessible database which can be accessed before publication. Please note this policy includes even short stretches of novel sequence information such as epitopes, functional domains, genetic markers or haplotypes. Short novel sequences must include surrounding sequence information to provide context. An author's web site is not acceptable for providing this type of information.*

Structures:

Papers must state that atomic coordinates have been deposited in the Protein Data Bank (or Nucleic Acids Database, as appropriate), and must list the accession code(s). Accessibility must be designated "for immediate release upon publication". For X-ray crystal structures, structural factor files must also be submitted to PDB and released upon publication. Authors must provide atomic coordinates and structure factor files upon request of referees or editors for the purposes of evaluating the manuscript, if they are not already freely accessible in a publicly available and recognized database. Please note that a table containing data collection and refinement statistics (see templates for [X-ray crystallography](#) and [NMR](#)), as well as a figure showing experimental electron density map (for X-ray crystallography) or superposition of the lowest energy ensemble (for NMR) must be included at the time of submission. At this time, there is no formal requirement for deposition of NMR assignments and constraints, molecular envelope reconstructions from electron microscopy data or coordinates generated from modeling. However, we encourage authors to deposit such data in appropriate, publicly available databases, provide accession codes and release data upon publication.

Microarrays:

Please see the MGED open letter specifying microarray standards at http://www.mged.org/Workgroups/MIAME/miame_checklist.html. Authors submitting manuscripts containing microarray data must supply the data as Supplementary Information on CD at time of submission. The data must be MIAME-compliant and supplied in a form that is widely accessible, with the completed checklist also placed on the CD. Five copies of the CD are required, so they can be sent to referees. We also require submission of microarray data to the GEO (<http://www.ncbi.nlm.nih.gov/geo/>) or ArrayExpress (<http://www.ebi.ac.uk/arrayexpress>) databases, with accession numbers at or before acceptance of the paper for publication. See Nature 419, 323; 2002 for further details and explanation.

Any supporting data sets for which there is no public repository must be made available to any interested reader on and after the publication date from the authors directly, the author providing a URL to be used in the paper on publication. Researchers who encounter a persistent refusal to comply with these guidelines should contact the editors of Nature Genetics by email, with "materials complaint" as part of the subject line. See Nature 416, 1; 2002 for further details. In cases where editors are unable to resolve a complaint, we reserve the right to refer the correspondence to the author's funding institution and/or to publish the fact that readers have been unable to obtain necessary materials or reagents to replicate the findings.

The Australian Academy of Science does not support the notion of a *national* approach for managing data for science. Science is an *international* activity with international, discipline-based databases and is underpinned by publication of scientific findings in prestigious journals, or as patent applications. Editorial policies that support the professional maintenance and use of international databases are the most cost-effective way to ensure access to publicly and privately funded research.

ANNEX 1
DECLARATION ON ACCESS TO RESEARCH DATA FROM PUBLIC FUNDING

Adopted on 30 January 2004 in Paris

The governments (1) of Australia, Austria, Belgium, Canada, China, the Czech Republic, Denmark, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Israel, Italy, Japan, Korea, Luxembourg, Mexico, the Netherlands, New Zealand, Norway, Poland, Portugal, the Russian Federation, the Slovak Republic, the Republic of South Africa, Spain, Sweden, Switzerland, Turkey, the United Kingdom, and the United States

Recognising that an optimum international exchange of data, information and knowledge contributes decisively to the advancement of scientific research and innovation;

Recognising that open access to, and unrestricted use of, data promotes scientific progress and facilitates the training of researchers;

Recognising that open access will maximise the value derived from public investments in data collection efforts;

Recognising that the substantial increase in computing capacity enables vast quantities of digital research data from public funding to be put to use for multiple research purposes by many research institutes of the global science system, thereby substantially increasing the scope and scale of research;

Recognising the substantial benefits that science, the economy and society at large could gain from the opportunities that expanded use of digital data resources have to offer, and recognising the risk that undue restrictions on access to and use of research data from public funding could diminish the quality and efficiency of scientific research and innovation;

Recognising that optimum availability of research data from public funding for developing countries will enhance their participation in the global science system, thereby contributing to their social and economic development;

Recognising that the disclosure of research data from public funding may be constrained by domestic law on national security, the protection of privacy of citizens and the protection of intellectual property rights and trade secrets that may require additional safeguards;

Recognising that on some of the aspects of the accessibility of research data from public funding, additional measures have been taken or will be introduced in OECD countries and that disparities in national regulations could hamper the optimum use of publicly funded data on the national and international scales;

Considering the beneficial impact of the establishment of OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data (1980, 1985 and 1998) and the OECD Guidelines for the Security of Information Systems and Networks (1992, 1997 and 2002) on international policies for access to digital data;

DECLARE THEIR COMMITMENT TO:

Work towards the establishment of access regimes for digital research data from public funding in accordance with the following objectives and principles:

Openness: balancing the interests of open access to data to increase the quality and efficiency of research and innovation with the need for restriction of access in some instances to protect social, scientific and economic interests.

Transparency: making information on data-producing organisations, documentation on the data they produce and specifications of conditions attached to the use of these data, available and accessible internationally.

Legal conformity: paying due attention, in the design of access regimes for digital research data, to national legal requirements concerning national security, privacy and trade secrets.

Formal responsibility: promoting explicit, formal institutional rules on the responsibilities of the various parties involved in data-related activities pertaining to authorship, producer credits, ownership, usage restrictions, financial arrangements, ethical rules, licensing terms, and liability.

Professionalism: building institutional rules for the management of digital research data based on the relevant professional standards and values embodied in the codes of conduct of the scientific communities involved.

Protection of intellectual property: describing ways to obtain open access under the different legal regimes of copyright or other intellectual property law applicable to databases as well as trade secrets.

Interoperability: paying due attention to the relevant international standard requirements for use in multiple ways, in co-operation with other international organisations.

Quality and security: describing good practices for methods, techniques and instruments employed in the collection, dissemination and accessible archiving of data to enable quality control by peer review and other means of safeguarding authenticity, originality, integrity, security and establishing liability.

Efficiency: promoting further cost effectiveness within the global science system by describing good practices in data management and specialised support services.

Accountability: evaluating the performance of data access regimes to maximise the support for open access among the scientific community and society at large.

Seek transparency in regulations and policies related to information, computer and communications services affecting international flows of data for research, and reducing unnecessary barriers to the international exchange of these data;

Take the necessary steps to strengthen existing instruments and - where appropriate - create within the framework of international and national law, new mechanisms and practices supporting international collaboration in access to digital research data;

Support OECD initiatives to promote the development and harmonisation of approaches by governments adhering to this Declaration aimed at maximising the accessibility of digital research data;

Consider the possible implications for other countries, including developing countries and economies in transition, when dealing with issues of access to digital research data.

INVITE THE OECD:

To develop a set of OECD guidelines based on commonly agreed principles to facilitate optimal cost-effective access to digital research data from public funding, to be endorsed by the OECD Council at a later stage.

(1) Including the European Community

ANNEX 2

CASE STUDY IN AUSTRALIAN RESEARCH IN EARTH SYSTEMS SCIENCE

- These following observations relate to data storage, maintenance and acquisition of data that is pertinent to Earth System Science (ESS). ESS is the global study of the dynamically interacting systems of atmosphere, waters, ice, soils, biosphere and human life support systems that are impacted by and have an impact on human welfare and sustainability of human existence on Earth.
- To conduct its research, ESS has an underpinning need for long uninterrupted coherent records of the state of the planet over many timescales and space scales in such areas as diverse as atmospheric composition, climate variables, extreme events, natural energy fluxes, water and land resources, ice extent, agricultural, forestry and fisheries production and potential, satellite imagery of surface properties of land and ocean, fossil fuel use and resources and measures of human health and wellbeing. As these are trans-jurisdictional attributes, the data needs of ESS do not always sit comfortably with pressures for proprietary and commercial-in-confidence requirements of many data streams.
- Ready, affordable access to such data is critical to effective understanding of the interactive Earth System. Accordingly the principle that seems to be adopted in the USA (in many areas of environment monitoring at least) that data that has been paid for from the public purse should be freely available to everybody even those in other countries, needs to be promoted and adhered to in Australia and elsewhere. The Working Group should consider the proposition that public good benefits of the free flow of data far outweigh the small amounts of money that could be made from taxing information flow, especially when the cost of such taxation is taken into account.
- Every type of information is unique and has special qualities and requirements. It would therefore be a daunting, perhaps impossible, task to devise a universal system of data storage, maintenance and acquisition. It is probably not worth considering.
- However, a flexible structured system culminating at the highest level in a meta-document that points the way for how to find data on specific matters, not just in Australia but also globally, is worth consideration. The facilities of the world wide web and its search engines have already pointed the way in how this might be achieved.
- In a specific area, which is critical in a world experiencing global climate change, the lamentable state in Australia of meteorological and climatic monitoring and old record maintenance needs strong emphasis. Historical data are not being preserved, and being lost, are scattered among various agencies and generally have not been digitized. Weather stations are often not well maintained, have been or are being closed “all over the place”, and new automatic weather stations are not being funded for deployment. It seriously hampers research in weather and climate related sciences in areas as diverse as aviation safety, climate change, and fire management.