

**Strategies for Permanent Access to Scientific Information in Southern Africa:  
Focus on Health and Environmental Information for Sustainable Development**

*An International Workshop*

**EXECUTIVE SUMMARY**

**AND**

**RECOMMENDATIONS**

**Editors: William Anderson  
Paul Uhler  
Robyn Arnold**

**5–7 September 2005**

**CSIR Convention Centre, Pretoria, South Africa**

Jointly organised by:

CODATA Task Group on the Preservation of and Access to S&T Data  
in Developing Countries

and

South African National Committee for CODATA

and

United States National Committee for CODATA

and

with support of

South African National Research Foundation

Publication date: April 17, 2006

## ACKNOWLEDGMENTS

A workshop of CODATA's calibre requires a concerted effort from a number of partners for it to be successful. The National Research Foundation would like to heartily thank all our partners and sponsors who generously provided material and financial support for the workshop. These valuable partners are:

- The International Development Research Centre (IDRC), Canada
- The South African Department of Science and Technology (DST)
- Statistics South Africa (StatsSA)
- International Council for Science, Regional Office for Africa (ICSU)
- Committee for Data on Science and Technology (CODATA)
- International Network for the Availability of Scientific Publications (INASP)
- Open Society Institute (OSI)
- US National Institutes of Health (NIH)
- US National Science Foundation (NSF)
- US National Weather Service
- The National Academies, US
- Chinese Association for Science and Technology
- South African National Research Foundation (NRF)
- *Soetkaroo* and *Distell* for a generous donation of refreshments for social events.

This work was carried out with the aid of a grant from the International Development Research Centre (IDRC), Ottawa, Canada.

We would also like to acknowledge the contribution of session rapporteurs in writing the sections that report on their sessions:

- Mr LeRoy Charles, George Washington University, United States – Session 7.1a: Health and Biomedical Data and Information
- Dr William Anderson, Praxis101, United States – Session 7.2a: Earth and Environmental Data and Information
- Mr Paul Uhlir, US National Academies, United States – Session 7.3a: Biodiversity Data and Information
- Mr Roy Page-Shipp, SARIS – Session 7.4a: Scientific, Technical and Medical Journals.

We are also indebted to members of the South African National Committee for CODATA under the leadership of Prof. Steve Rossouw. We wish to thank Prof. Liu Chuang, Co-chair of the CODATA Task Group, for facilitating Asian participation. The unwavering support and guidance of our North American partners: Dr Fraser Taylor, Dr William Anderson, Paul Uhlir and Amy Franklin, is also deeply appreciated. Their critical inputs and guidance throughout the entire planning and execution of the workshop provided valuable insights. We would also like to acknowledge the contributions made by Ms Henda van der Berg of the National Research Foundation, who was responsible for overseeing all logistical arrangements for the workshop. Finally, we would like to thank all our regional and international speakers who travelled to South Africa to share their experiences with their peers. Without their well-researched papers and participation, the workshop objectives would not have been met.

Dr Eddy Maepa  
Local Organising Committee  
NRF

Prof. Steve Rossouw  
Chair: SANC CODATA

Dr William Anderson  
Prof. Liu Chuang  
Co-Chairs: CODATA  
Preservation Task Group

“Digital resources will not survive or remain accessible by accident.”

Bernard Smith, European Commission,  
ICSTI/ICSU/CODATA Digital Preservation Workshop,  
15 February 2002, Paris, France

## INTRODUCTION

The Committee on Data for Science and Technology (CODATA) workshop on *Strategies for Permanent Access to Scientific Information in Southern Africa* is one of a series of workshops on the preservation of and access to scientific data and information in developing countries. The initial workshop was organised by CODATA and the National Research Foundation (NRF) and held in Pretoria in 2002. This was followed by workshops in Brazil and China.

The 2002 workshop had a limited focus, aimed at addressing scientific data and information management to advance research. In addition to effective data and information management and access challenges, the 2005 workshop on *Strategies for Permanent Access to Scientific Information in Southern Africa* focused on issues related to sustainable development, which is a primary concern for the Southern African Development Community (SADC) region and Africa as a whole. The theme of the series of workshops was premised on the vision of the World Summit on Sustainable Development (WSSD), held in South Africa in 2002, as well as on declarations of the World Summit on the Information Society (WSIS) in Geneva in 2003, with the aim of addressing some of the developmental issues raised at both summits. To this end, the focus of this workshop was on health and environmental information for sustainable development.

The workshop was attended by nearly 100 people and included representatives from eight of the fourteen SADC countries, Asia, Europe and the United States. The excellent meeting venue and the extensive interactions among the participants generated discussion of many issues and suggestions for solving individual and local challenges. The workshop also provided a valuable opportunity to renew and initiate personal contacts.

## WORKSHOP OBJECTIVES AND REPORT

The workshop, described more fully at <http://stardata.nrf.ac.za/html/workshopCodata.html>, had the following objectives:

1. Review the current status of practices for sharing and archiving scientific information resources related to sustainable development in the SADC region, with specific reference to health and biomedical data, earth and environmental science data, and scientific, technical and medical literature.
2. Identify and discuss scientific, legal and policy, institutional and economic, and management and technical factors relevant to providing permanent access to digital scientific information resources. Examine different models, and their benefits and shortcomings in the SADC region, drawing on examples of related digital archiving and access regimes.
3. Identify follow-up activities that can be taken to improve access and preservation for the major types of digital scientific information resources discussed.
4. Provide a networking opportunity for workshop participants across discipline, institutional and national boundaries.

The format of the workshop report is based on the following set of questions that were used as the focus for each workshop session:

- What are the challenges and barriers?
- What are the existing resources and mechanisms?
- What are some potential realistic projects or collaborations to help address the challenges and barriers that have been identified?
- How can the new or improved initiatives specifically be implemented?

The full workshop report contains summaries of many of the presentations as well as a list of the recommendations that arose during discussions and the breakout sessions. The recommendations are included in this executive summary. The individual presentations are available online at <http://stardata.nrf.ac.za/html/workshopProgramme.html>.

## **RECOMMENDATIONS OF THE WORKSHOP**

The workshop forum and structure enabled participants to (1) share information about regional conditions and projects, and (2) generate recommendations for practical short-term actions and longer-term strategic possibilities. Many of these actions and opportunities are captured in these recommendations.

The recommendations generated discipline-specific actions as well as broader and longer-term actions. The recommendations are directed at two primary audiences. One audience includes the scientific committees, agencies and funding organisations in the SADC region. These institutions and their members and social networks are most able to effect local changes and to follow up on the recommendations. A second audience includes international organisations such as the United Nations, ICSU, CODATA and others. These organisations can help facilitate the broader objectives.

The recommendations of the workshop should also be viewed in the context of the ongoing implementation of the objectives of the CODATA Task Group on Preservation of and Access to Scientific and Technical Data in Developing Countries; namely, to provide:

- A bridging role in reducing the digital divide in the management and use of scientific and technical data, in collaboration with CODATA national committees, ICSU bodies and relevant regional and international organisations.
- A partner role by participating in joint activities, such as workshops, meetings and training programmes. The Task Group will continue to encourage its members to be more actively involved in joint activities.
- An outreach role in developing the Task Group's Web site, publications and information network and making them more timely and effective.
- A leading role in identifying new issues and challenges in scientific and technical data in developing countries, and in organising a country series of international workshops to establish outreach and organisational networks that can help solve the problems of the information society in developing countries.

As workshop conveners and supporters, CODATA and ICSU are seen as two organisations that can help promote and facilitate solutions to the many challenges raised regarding permanent access to scientific resources. Several key messages for these organisations were expressed by participants and are presented here. Workshop participants and organisers realise that while CODATA and ICSU have international visibility, these organisations are primarily facilitators of action, and networking resources. The work of resolving problems and addressing issues remains, for the most part, with regional institutions and research communities.

### **CODATA and ICSU: Organisational recommendations**

Workshop participants look to CODATA and ICSU for leadership on issues that span discipline and regional boundaries through the work of its Task Groups. The publication of the ICSU Priority Area Assessment on Data and Information<sup>1</sup> and the establishment of an ICSU Regional Office for Africa generated several organisation-level recommendations. The following suggestions are directed to these groups to help inform their activities:

- ICSU should work with CODATA to develop a long-term, coordinated framework for data and information policies, practices and infrastructure.

---

<sup>1</sup> [www.icsu.org/1\\_icsuinscience/DATA\\_Paa\\_1.html](http://www.icsu.org/1_icsuinscience/DATA_Paa_1.html)

- The South African National Committee for CODATA is an ICSU presence in southern Africa, which should contribute to the execution of the ICSU strategic plan for 2006–2012 concerning data and information.
- It may be useful for ICSU to have a scientific advisory committee that understands archiving and preservation problems and provides advice on decision-making related to the time periods for which different datasets should be preserved. CODATA could play an influential role in this regard.
- CODATA and ICSU should address the sharing and improved access of data across national boundaries. SADC regional initiatives are important and should be represented in ICSU discussions on policy for standardising metadata or developing common practices for data management and preservation.
- CODATA membership should be promoted. Only four CODATA national committees exist on the African continent. CODATA should establish a regional committee to work with the ICSU Regional Office on outreach and recruitment of African CODATA members.
- Scientific and technical (S&T) data and information policy
  - CODATA outputs should include examples of national policies that establish the record-keeping policies of various nations such as the USA, United Kingdom and Australia. This could help inform the South African National Archives Act, which currently makes no mention of the mandatory collection and preservation of S&T data and information, but refers only to public administrative records.
  - Involve government representatives in forums organised by CODATA:
    - Ensure that workshop objectives describe convincing scenarios about the importance of science and technology, at the appropriate audience level, and convey core data requirements and issues.
    - Encourage funding agencies such as the South African National Research Foundation to promote research on the preservation of digital objects as a major priority.
  - Data sharing: the CODATA Task Group should consider a major action that it can implement in the next few years (involving ICSU, NEPAD, the African Union and the Pan African Council) with respect to governance and regulation of scientific data (for example, to advance the attainment of the Millennium Development Goals).
  - Develop partnerships:
    - CODATA should consider seeking a formal liaison, or contributor status, in TWNSO.
    - Investigate the possibility of a CODATA contribution to data coordination among the many ministries involved in implementing the Indigenous Knowledge Policy in South Africa, and generally encourage such coordination among those ministries.

### **General workshop recommendations**

The following recommendations are not directed specifically at CODATA and ICSU, but rather to the broader S&T policy, funding, and research management communities. They arose from several of the plenary discussions and are more general than the discipline-specific suggestions.

- Data sharing
  - Raise awareness of S&T data and information preservation, access and sharing successes and challenges:
    - Promote awareness of data issues in ministries and universities.
    - Follow up with the participants in this workshop to continue further dialogue.
    - Identify regional conferences and workshops at which the results of this workshop can be presented. Take the conversation to others.
    - Continue to describe and promote workable models for sustainable open access.
  - Foster the development of a list of core datasets of who holds what data. This would facilitate data sharing. The NASA Global Change Master Directory is an example of such a directory and a possible model for action. Leveraging the emerging Global Earth Observation System of Systems (GEOSS) initiative is one such opportunity.

- Training and development of good practices:
  - Create a “Data Management Academy,” building on the presentations and discussion from the workshop, to develop and provide training for researchers, managers and government officials about data management and sharing operations and policy development and implementation. Specifically:
    - Consider a phased implementation of a virtual academy online;
    - Investigate implementing a tool such as ‘Ask an Archivist’ that can field questions; and
    - Develop online training materials.
- Specific data issues
  - Data quality:
    - Document existing regional data quality initiatives with respect to various disciplines.
    - Develop a metadata inventory of information resources to identify gaps and cross-link opportunities. Do not limit dissemination of this information only to online modes.
  - Socio-economic data:
    - Socio-economic data tend to be neglected. Encourage GEOSS/SAEOS to include socio-economic data.
  - Indigenous knowledge data:
    - Promote the recording and translating of oral history data.
  - Measures of data value:
    - Sponsor a study of S&T data in jeopardy of loss as a result of factors such as obsolescence, neglect or lack of funding.
    - Create measures of value for data preservation and sharing.

### **Discipline-specific recommendations**

The workshop included four break-away sessions to explore discipline-specific issues and to generate focused recommendations. Those recommendations are summarised below and identify opportunities for action by individuals and institutions. Some of these recommendations are related to the larger, organisational ones listed above.

#### ***Biomedical data and information***

Action recommendations in this area fall under two broad headings, organising collections of biomedical data and establishing practices that support data preservation policies.

- Organisation of biomedical data and information collections will be facilitated by the following actions:
  - Establish health surveillance mechanisms in individual countries, in the way that Centers for Disease Control (CDC) collects, records and analyses health and disease data in the United States.
  - Establish health information systems at the national level first, and then integrate them internationally. It needs to be recognised that many biomedical data are collected and stored in paper form held by individuals and small organisations.
  - A standard language is required for data sharing; microbiology culture collections are just one example of the need. These standards need to be established and maintained locally, but be able to work regionally and globally. Support for training and establishing these standards is essential.
- Establishing biomedical data collections and their preservation policies and practices will be facilitated by the following actions:
  - Develop a compendium online of the kinds of biomedical information being collected in SADC countries and their sources, so that people know where to look.
  - Collect biomedical information from traditional healers; leverage their contact with people and their knowledge of Africa’s biodiversity and its use during the past.
    - More specifically, follow up on Zambia’s efforts to recruit someone to manage data from traditional healers (e.g., share project stories).

### ***Biodiversity data and information***

Actions with respect to biodiversity data and information recommendations fall into three broad categories.

- Improving data quality:
  - Development and broad implementation of data cleaning tools (for example, see the 2005 Global Biodiversity Information Facility [GBIF] report on this topic).
  - Training to improve data management and related applications:
    - Create online training material.
    - US National Science Foundation's Science Environment for Ecological Knowledge (SEEK) training modules may be useful and could be focused on developing country practitioners and users.
    - Training workshops by GBIF, the Long Term Ecological Research (LTER) program, ODINAfrica (funded by Belgium), and other organisations should be investigated.
  - Sharing and linking many kinds of primary data resources (species, specimen and molecular biology) with other related data and information will improve research and applications.
  - Demonstration projects show the value of biodiversity and ecological databases and research at the local and regional levels to help stimulate greater appreciation of such work by decision-makers and funding sources.
  - Biodiversity and ecosystem data must be coordinated with the GEOSS initiative in South Africa and the Marine Science Remote Sensing Data Centre in South Africa.
  - Federations of culture collections are organised in other regions, but such a federation is needed for the SADC region.
  - Collaborative research projects and programmes will help.
  - The African Ocean Biological Information System regional group is hosted by the Southern African Data Centre for Oceanography. This programme is now being initiated and needs to be coordinated with other existing databases and research institutions and activities. Funding ends in September 2006, however.
  - A metadata inventory of information resources to identify gaps and cross-link opportunities is being developed by SAEON.
- Establishing and nurturing social networks
  - Make efforts to bring African scientists to upcoming meetings on biodiversity and associated topics, for example:
    - The GBIF Governing Board meeting in Cape Town, April 2006.
    - Associated meetings on the "barcode of life" project.
    - New JRS Foundation to bring African scientists together on biodiversity informatics, education and funding initiatives for developing countries.
- Promoting open access funding policies
  - GBIF is developing policy statements for funding agencies that require data-sharing and maintenance plans, similar to the International Long Term Ecological Research programme's data policy model.
  - Various scientific 'information commons' initiatives are being established worldwide, including some specific to the area of conservation commons. The Southern African Millennium Assessment (2001–2005) provides an avenue to promote this work in the SADC region.

### ***Earth and environmental data and information***

The earth and environmental data and information discussions yielded actions in four areas. Some actions are for individual and small institutional actions, whereas other actions are best supported by larger, more global organisations.

- Build social networks: this requires being personally proactive, but does not cost much (for example, establish an e-mail circulation list among the workshop participants).

- Establish practicable data sharing practices (one major issue is that data are not shared adequately, and also may not be easy to share):
  - Practise what we preach to our governments in our individual actions:
    - Facilitate data sharing among colleagues and data organisations.
    - Data-producing organisations need to reward people for creating and sharing data.
    - Be transparent by sharing data, soliciting feedback and transferring knowledge and learning.
  - List what data are available: identify the datasets that exist in the region and build a database of databases to serve as an online directory of data.
    - The Namibian project (presented in Session 7.2A) that canvassed data-holders is a possible model for this work.
    - Identify core datasets in the region and document their consistency and quality.
    - South Africa used to have a database of environmental data, but it has not been maintained. It is proposed that the NRF revive this activity for South Africa, and it could perhaps ultimately grow into a SADC activity. Establish a SADC CODATA working group for this activity.
  - Make data policies explicit and available.
    - NEPAD and SADC structures can be used to work on revisiting sharing regulations.
    - Align data sharing projects with existing regional organisations (avoid establishing new organisations wherever practicable).
- Data management training: presentations and discussions at the workshop highlighted needs for capacity and competency building, for both developing and developed countries and regions. The shared needs can promote collaboration.
- Influence policy decisions that are implemented by policy makers.
  - Develop convincing examples about current data preservation and access conditions and the importance of these to science and technology.
    - These examples need to be crafted to aid understanding in the target audience and to convey core data requirements and issues.
    - Use the scientific method to estimate scientific capacity and competence under different policies (for example, closed versus open access).
  - Leverage the capabilities of international participants in the workshop to find doors into areas that can be changed. CODATA and ICSU have member networks that can be consulted.

### ***Scientific, technical and medical journal information***

The following kinds of actions should be taken with regard to scientific, technical and medical (STM) journals:

- Establish and implement policy interventions by research funders (including governments and institutions) that:
  - Mandate that scholars make pre-prints and e-prints of their research available via an open access medium.
  - Mandate long-term curation of research outputs, both in the interests of the individual researchers who produce the articles, but also in recognition of the shared character of the global research enterprise.
- Promote the value of open-access approaches to the research funding bodies by:
  - Involving researchers and managers in describing the real challenges as well as solutions (using available local success stories).
  - Establishing training programmes for researchers and for journal funders and producers.

- Create high-quality regional information repository facilities where individual publications, or the output of small subgroups of scientists, can be cost-effectively preserved, and openly available.
  - This will support the digitisation of more African material.
  - Promote the establishment of open institutional repositories.
  - Include national repositories to archive national heritage items and provide quality-control functions such as selection, appraisal and retention.