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

An International Workshop

WORKSHOP REPORT

Editors: Robyn Arnold
William Anderson
Paul Uhlir

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

TABLE OF CONTENTS

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[1] 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.

[1] 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.

SESSION 1: OPENING

Chair: Dr Lulama Makhubela, National Development Agency, South Africa

Presentations/ Papers presented

Welcome Address: *Dr Khotso Mokhele, President and CEO: National Research Foundation*

Objectives of the Workshop: *Dr William Anderson, Workshop Co-chair and Co-chair, CODATA Preservation Task Group, United States*

The Role of the ICSU Regional Office for Africa in Sustainable Development of Sub-Saharan Africa: *Sospeter Muhongo, Director, ICSU Regional Office for Africa*

What are the challenges and barriers?

The theme of the workshop – strategies for permanent access to scientific information – is one of the most central issues to the way in which science is conducted and the knowledge generated by science is transmitted, shared and preserved.

The ICSU Committee for Scientific Planning and Review has conducted a series of priority area assessments, one of which was on scientific data and information. The report observed that the nature and use of scientific data and information, the conditions under which they are produced and managed, and the role of scientists and others in the process have been changing rapidly in recent years. The changes have partly been due to the revolution in computational capacity, connectivity and advances in hardware and software, which have provided scientists with greater capacity in terms of data gathering, analysis and dissemination. These advancements have also enhanced the opportunities for scientists to collaborate in scientific work, even from remote locations. These changes hold out hope for scientists in developing countries to collaborate more intensively with their counterparts in the developed world, but at the same time increase the risk of developing world scientists falling even further behind if they are not in a position to employ effective and efficient data and information management techniques and processes to ensure appropriate access to and utilisation of scientific data and information. The challenges within the SADC region of the rapid advances in information and communication technology (ICT) are to properly capacitate research initiatives so as to maximally enjoy the benefits.

Statistics from Africa (2003) paint a far from rosy picture. Africa has a population of some 703 million people, and an annual population growth of 2.1%. Life expectancy at birth is 45.8 years, the mortality rate of infants per 1000 live births is 103.1, and about 8.5% of the most productive section of the population (aged 15–49) has HIV/AIDS. Africa's total public and private foreign debt is estimated at about US$500 billion, against a gross national income (GNI) of only US$400 billion. Almost all the income generated by the continent thus goes to servicing private and public debt. The GNI per capita is only US$490 per annum. The harsh reality is that 40% of sub-Saharan Africa (about 280 million people) lives in abject poverty.

Against these stark realties, Africa is challenged with striving towards the Millennium Development Goals (MDGs):

- To eradicate extreme poverty and hunger
- To achieve universal primary education
- To promote gender equality
- To reduce child mortality
- To improve maternal health
- To combat HIV/AIDS, malaria and other diseases
- To ensure environmental sustainability
- To develop a global partnership for development.

African countries risk falling increasingly further behind developed countries, given their limited resources to invest in science, technology and development. Only about 58% of sub-

Saharan African children are annually enrolled in primary schools. Developed countries have only 21% of the world population but generate 58% of world gross domestic product (GDP) and account for 80% of global gross expenditure on R&D (GERD) (which was US$746 billion in 2000) and 72% of world researchers. In 2000, Africa accounted for only 1% of world GERD. Africa has only 78 researchers per million inhabitants, compared with 5206 in Japan, 4006 in the United States, 554 in China and 143 in India. South Africa is relatively better off in this regard than most other African countries, with 309 scientists per million inhabitants.

Africa needs to generate data so as to be in a position to plan a course of sustainable development and address human-induced hazards, such as water pollution, deforestation, siltation, mining waste disposal, acid drainage, destruction of ecosystems, as well as social problems, including child labour, health care and immigration problems.

What are the existing resources and mechanisms?

In September 2005, the first Regional Office of the International Council for Science (ICSU) anywhere in the world was launched in Pretoria. This marks a milestone in ICSU's efforts to give practical effect to promoting science and technology in developing countries as a central *raison d'être* of the organisation. According to this new conceptual model, ICSU will have four offices in developing regions – in Africa, Latin America, Asia and the Arab world. The South African National Committee for CODATA is another manifestation of an ICSU presence in southern Africa, which will contribute to the execution of ICSU's strategic plan for 2006–2012 concerning data and information.

Africa is well endowed with energy resources, strategic reserves of industrial materials and strategic metals. Opportunities thus exist for Africa to invest in and benefit from exploitation of these resources and reserves, which is dependent on science and technology.

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 priority areas of the ICSU Regional Office are central to addressing the most fundamental problems that Africa faces, namely: health and human well-being, sustainable energy, natural and human-induced hazards and global climate change. Some of the actions in which ICSU will engage in pursuing its objectives are promoting the formation of African networks of scientists, compiling a database of African scientists, including those in the Diaspora, capacity building and maintenance, and promoting constructive indigenous science, knowledge and skills.

As CODATA considers how to carry out its mandate in developing countries, it could find a valuable partner in the Third World Network of Scientific Organisations (TWNSO), which is supported by the Academy of Sciences for the Developing World. TWNSO comprises the representatives of the ministries of science, higher education institutions, academies of science and research councils of forty developing countries and has the potential, if properly supported, to become a powerful non-governmental organisation dealing with science and technology in developing countries.

SESSION 2: INTERNATIONAL PERSPECTIVES ON PERMANENT ACCESS TO PUBLIC SCIENTIFIC INFORMATION

Chair: Mr Emmanuel Mutale, National Science and Technology Council, Zambia

Papers presented

Accessing Scientific and Communication Information: Methodological Underpinnings, *Prof. Radhamany Sooryamoorthy, Associate Professor, University of KwaZulu-Natal, South Africa*

Scientific Data and Information: Priorities for Science, *Dr Roberta Balstad, Director, Center for Earth Science Information Network, and Chair, United States National Committee for CODATA*

Global Earth Observing System of Systems (GEOSS*), Mr Dhesigen Naidoo, South African Department of Science and Technology*

What are the challenges and barriers?

Scientific data and information

In any field of science, major changes will be found as a result of the digitisation and advancement of electronic resources in data and information. Data and information constitute an electronic continuum ranging from raw data to published papers and data products, such as maps.

In the past, collecting data was generally the most time-consuming part of any research initiative. Today, many decisions about data collection are made in the commercial sector by publishers or in the development sector by agencies that are collecting or funding data. The challenges include the fact that scientists increasingly rely on data collected and managed by others, which is both an advantage and a disadvantage. It means that scientists can base their analyses and findings on very large datasets, and that it is more readily possible to test and retest data and in the process discover any errors in the data, thereby improving the quality of data and science. However, when data are collected by others, it takes longer to understand the dimensions related to the way in which the data were collected and the factors that influenced the shape of the database.

Data collection, analysis and management have become separate activities with differing professional requirements and training. In the past, scientists were their own data managers. Today, scientists require professional data managers. The growth in analyses of change over time requires careful preservation of today's databases for use by future generations of scientists.

The cost of data collection, instrumentation and management is rising disproportionately to other research expenses. The cost of research is constantly going up, but the cost of data management is going up faster and will continue to do so as databases, and the size of existing databases, continue to increase. New financial models for current and future data and information management and preservation are needed so as not to deplete current funding levels available for the research activity itself.

Access to data and information (including publications) and ownership of data and information may be subject to costly commercial and legal restrictions. The commoditisation of data is an everyday reality. Data have become a commodity that can be bought, sold or licensed.

The growing emphasis on interdisciplinary research requires more extensive metadata and interoperability of databases, hardware and software. There is insufficient communication and coordination of data and information activities across fields and countries. Science is an international activity and is increasingly a multidisciplinary activity. There is a need for common standards across both national and disciplinary barriers.

There is a lack of clarity on who is ultimately responsible for the data and information on which science is based, how they assume that responsibility and who carries the costs.

The plenary discussion noted that the ideal situation may be to devise data and information management practices and policies that have relevance over decades and centuries, but this poses problems in disciplines that turn increasingly to numerical models for data. Such models can produce gigabytes of data in a few days, which researchers consider as valid data. The quantitative management of such data is becoming a major problem. Modelled data need to be preserved so that anyone that wants to question the assumptions or findings of research based on such data can go back to the original data. It may not be possible to store all data electronically, and scientists will have to become involved in decision-making about which data are to be preserved permanently and which data are to be preserved for a specified period only.

Questions have been raised about the sustainability of open access. Scientists are generally in favour of open access, while commercial publishers are not. This is another case where financial and business models need to be worked out.

As governments become increasingly involved in data collection, preservation, archiving and provision, scientific input and advice may be needed for decision-making about such data. Decisions about data that may be important for subsequent scientific research are too important to be dictated by political criteria or budgetary considerations.

Accessing scientific and communication information

It could be assumed that advances in ICT will enhance the opportunities for collaboration between scientists in the developing and developing worlds, but research does not support this assumption. A longitudinal study has been ongoing since 1994 in Kerala (India), Ghana, Kenya, South Africa, the Philippines and Chile to look at the impact of the Internet on research communication within the developing world and with the international scientific community through an analysis of the conditions associated with interpersonal networking and information search behaviour.

The research came up with the counter-intuitive finding that collaboration is unrelated to productivity. For academics only, collaboration is positively and significantly associated with productivity. For scientists employed in research institutes, collaboration is negatively correlated with productivity. Not only do collaborative efforts fail to improve productivity for government scientists, but they may actually hinder the production of written output.

The reasons may be the limited access to resources available in terms of Internet connectivity, and that the systems available may not be fast enough to establish useful links. There may be practical difficulties; for instance, a researcher in Kenya was found to be accessing the Internet via a police station. If scientists in the developing world take advantage of the Internet to increase collaboration, they may cancel out any productivity benefits by escalating the liabilities of shared work as well. Moreover, the study suggests that where collaborations are introduced by donors from afar, collaboration cannot be expected to enhance productivity as it does in the developed world.

The plenary discussion revealed the research findings about the negative correlation between research collaboration and productivity to be contentious, and contrary views were offered that the most productive researchers in Africa are those that collaborate with advanced laboratories, and that African researchers that have studied aboard have built up collaborative networks, the benefits of which extend also to their research teams and graduate students.

What are the existing resources and mechanisms?

ICSU

An ICSU programme assessment panel on data and information (D&I) was appointed in 2003 with a mandate to propose strategic priorities for ICSU over the next five to ten years in the

area of scientific D&I, to review D&I capabilities in the ICSU family (including CODATA) and to advise national and international agencies. The panel made over fifty recommendations related to D&I production, management, access and dissemination directed at the scientific community, science funding bodies, research and educational institutions, intergovernmental organisations, commercial data producers and publishers, the media, decision-makers and policy-makers. ICSU already has a policy of open access to data and universal and equitable access to publications, which the panel endorsed.

The official government policy in the United States is open access to all data. The United States is unusual in this respect. Some other governments sell data such as weather, climate, demographic and geospatial data. There is a need to reach agreement between nations on common access policy, and at least scientific and education institutions ought to have free access, but the financial model for that is not clear.

ICSU does not replace or supplant existing initiatives, but learns the lessons of collaboration and applies these in areas where they do not exist.

Scientific data and information

The Center for Earth Science Information Network, for example, which is a data-producing and data archiving organisation, has a policy of obtaining funding for data collection in advance and then making data available freely.

Global Earth Observation System of Systems (GEOSS)

GEOSS is a new comprehensive, coordinated and sustainable international partnership initiative with a membership of 58 countries and the European Commission and 43 international organisations. The membership includes 29 developing countries, 14 of which are from Africa. Its governance structure comprises four co-chairs – two from developing countries and two from the developing world. Currently, the co-chairs are from South Africa, the United States, the People's Republic of China and the European Commission.

The key objective is to strengthen cooperation and coordination among global observing systems and research programmes for integrated global observations, taking into account the need for building capacity and sharing of data from ground-based observations, satellite remote sensing and other sources among all countries. There is a general international readiness for this type of coordinated effort, and the world is becoming more uniform in its ability to correlate forecasting between developed and developing countries. In the last ten years, weather forecasting has developed equity in data utilisation, with the differential between the North and the South converging to a point. Earth observation applications exist in crop yields, water and air quality, weather forecasting, climate change, disease patterns (fauna, flora and human) and disaster mitigation and management.

The concept of the system of systems is, in the short term, to integrate the multiple national and regional systems, and eventually to develop a single international system for earth observation.

The work of GEOSS is organised into nine societal benefit areas, within which an examination is conducted of data, user requirements, capacity to run the system and a more informed decision-making model. The information challenges GEOSS is battling to resolve include data accessibility; whether a professional structure is required to be in place to make use of the data; issues of interoperability related to the metadata connection; different software, governance systems and reporting regimes; the capacity required; the increasing divide between countries with better platforms and those without; and the beneficiation loop of increasing the effort and the product when all participants see direct benefit to their own objectives.

GEOSS provides the space and opportunity for implementing practical initiatives around data.

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?

ICSU

ICSU should develop a long-term, coordinated framework for D&I policies, practices and infrastructure that will operate over periods of decades to centuries.

Scientists themselves must be involved in setting data and information policies, priorities, and practices. Consistent D&I access policies across nations could benefit both science and public policy.

ICSU (and members of the ICSU family such as CODATA) should assume a leadership role internationally in identifying and addressing scientific data and information policy issues. ICSU should promote cross-disciplinary and cross-national collaboration and capacity building by holding an international Scientific Data and Information Forum (SciDIF) for data and information stakeholders to discuss broad issues and the needs of current and future scientific research. ICSU should establish an ad hoc committee to organise SciDIF and oversee the implementation of the institutional and policy recommendations of the panel assessment report.

If may be necessary 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.

ICSU is not a research funding or performing agency. ICSU provides the space for discussing common approaches, strategy and policy and fosters a sense of standards that could be used across and within regions. 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.

GEOSS and the South African Earth Observation System (SAEOS)

Being in a leadership position as one of the four co-chairs of GEOSS, South Africa has responded strongly by forming the South African Earth Observation System (SAEOS), the draft of which is being finalised and will be presented to Cabinet in November 2005. The challenge for SAEOS is to bring diverse initiatives into one fold through negotiation. A second national response is the development of international projects located in southern Africa, to be part of the starter group of projects for GEOSS as a whole.

The plenary discussion yielded suggestions that GEOSS could be improved by having a formal scientific advisory structure; by including socio-economic data, since most of the GEOSS goals concern societal benefits; and by finding the right kind of modality for coordination, since the current series of overlays will solve some short-term problems but not address fundamentals. If coordination is to be successful, it is important not to be pulled in one direction or another by the specific objectives of the various players. GEOSS will gain experience in coordination and use that to develop the model that is eventually used.

The local initiatives of GEOSS will have a national focus and regional orientation. The discussion proposed that GEOSS become involved in regional initiatives on the global ocean observation system, especially for the Western Indian Ocean subregion.

SESSION 3: REGIONAL PERSPECTIVES ON PERMANENT ACCESS

Chair: Dr Eddy Maepa, National Research Foundation, South Africa

Papers presented

National/Regional Publication of Original Research in a Globalising World, *Prof. Wieland Gevers, Executive Officer, Academy of Science of South Africa*

The Status of Environmental Data and Information in the SADC Region: Legal and Institutional Frameworks, *Mr Clever Mafuta, Head, Musokotwane Environment Resource Centre for Southern Africa (SARDC IMERCSA,) Harare, Zimbabwe*

What are the challenges and barriers?

Scholarly publishing

Any developing country needs its own science publishing initiatives. Local and regional publishing of research is desirable and valuable for several reasons, including wide local participation in editing and peer reviewing; networking scholars and postgraduate students; fostering disciplinary coherence; facilitating local contributions to local, regional and global knowledge; reflecting a local/regional focus; disseminating the results of local research so that these can influence policy-making; providing depth and strength in particular fields; show-casing the country's/region's scientific achievements in a concerted way; and drawing in wide local/regional audiences in government, higher education science councils, schools, the media and the general public. The value of a national scholarly publishing effort is premised on a system of high quality. If the quality is low, every argument in favour of a local scholarly publishing system becomes a counter argument.

Researchers learn an enormous amount as scholars through the practice of peer review, and that opportunity should be maintained in developing countries. Peer review should remain the bedrock of the science system, even in the electronic age, so that it can be accepted that when a paper has been published it has gone through a universal process. Publication records help indicate what research to fund and reward and assist young people in developing an idea of quality. All these advantages are built into the paper-based system, and the emerging electronic publishing system will have to explore how to adapt or replicate the system that has been tried and trusted over centuries.

A survey in 2005 among 213 editors of research journals in South Africa reveals that the acceptance rate is relatively high, with 60–80% of submitted articles being accepted. This raises doubts about whether the scholarly publishing effort in the country is serving the purpose of quality control.

In the plenary discussion, the issue was raised that international publishing houses are already poaching the better South African research journals and turning them into international journals that become unaffordable for South African libraries and individuals.

The South African Department of Education has decided to accredit only two international indexes of research publications – the International Bibliography of the Social Sciences (IBSS) and the Institute of Scientific Information (ISI) – because they lend themselves to bibliometric analysis (through the completeness of their author records). There may thus be valuable and important papers that the Department of Education subsidy system overlooks. The Academy of Science of South Africa (ASSAf) intends to compare the ISI and ISSB with other databases of publications, feature for feature. One of the difficulties lies in the fact that the modern world wishes to weigh, count and measure the impact of publications. There is a need to look at an alternative way of valuing a scholarly paper, as the value lies not only in the value of the paper itself but also in its functional value to users of the system.

All SADC countries have in place environmental policies that directly or indirectly subscribe to the right to live in a clean and healthy environment, which in turn carries with it the right to be

informed about environmental issues and calls for the development of appropriate information systems. The need for reliable information and data is further emphasised in the region's key development initiatives, including the Millennium Development Goals (MDGs), the New Partnership for Africa's Development (NEPAD) and the Regional Information Strategic Development Plan (RISDP), which require data as indicators for measuring their success.

Status of environmental data and information in the SADC region

Southern Africa suffers from a lack of regionally generated statistics and depends largely on international organisations for data and information. In some cases, the data and information from such international sources is disputed (especially for the purpose of policy-making), as it is based on projections and is not grounded truth. Policy-makers prefer locally referenced data. Many factors force the region to depend on data from international sources, including lack of up-to-date statistics, lack of comparability of data across countries, and limitations in dissemination mechanisms such as databases, networking and connectivity.

Concerns about the protection of intellectual property rights result in a wide range of information materials being classified as confidential, including maps and photographs that tell the story of the region. Data sharing and access are further hampered by the growing shift towards commercialisation, which puts a monetary value on all sorts of products, including information and data, particularly when the data are generated by international sources. Initiatives that use external funding for data collection tend to take the form of projects rather than processes (with data collected just to fulfil the mandate of the project). Data collection generally stops when the project ends, rather than being continuously active. There may also be many players involved, which makes it difficult to source and access the data later.

Government, academic and research institutions dealing with environmental information across the SADC region have important environmental datasets and databases. Most such datasets and databases are available online. However, many of these databases have little spatial reference information, and the datasets and databases are of varying quality and are scattered (and therefore difficult to access) and not well networked.

The institutional set-up for official statistics varies among the different SADC countries. In the majority of countries, statistics are handled by many different institutions, which means that it is common to find overlaps and gaps in their work.

The RISDP strategy highlights challenges such as inadequate resources allocated to statistics in some member states; disparities in capacities among countries; inadequate statistical capacity at the SADC secretariat to coordinate statistics in the region; absence of a legal instrument for regional cooperation in the area of statistics; lack of dissemination mechanisms, including databases and Internet connectivity; and a long time lag between collection of data and the dissemination of statistics.

The large number of networks and initiatives involved in environmental data in SADC can be a barrier to accessing data. Efforts are needed in SADC to optimise resources and improve access and sharing, as there may be duplication, overlap and repetition when agencies or research teams are not aware of what others are doing, or what others have done. There needs to be a way of targeting data and information requirements for beneficiaries.

What are the existing resources and mechanisms?

Scholarly publishing

ASSAf is nearing completion of a DST-sponsored study with the objectives of: developing a strategic approach to publishing research journals in South Africa; promoting and enhancing their national and international standing; improving their productivity and efficiency in the local/regional science system; and ensuring a wide audience for research published locally within both the research communities and society.

Some emerging ideas for strengthening the South African science system through locally published journals include increased numbers of listed journals; an improved peer review system (and the recognition of peer review itself as a legitimate and valuable product of the research process); increased frequency and magnitude of issues; as well as 'on time' publication; disciplinary consolidation; financial viability through revenue mix (including subscriptions, page charges on the strength of the Department of Education subsidy, advertising, sponsorship and subsidisation); world-wide dissemination (both print and especially online); continental link-up and planning within Africa.

ASSAf has recently begun publishing a new science magazine called *Quest*. The magazine showcases South African research, with articles written by research authors themselves, targeting an audience comprising the general public, senior school learners, teachers and researchers. Cooperation contracts are planned with government departments and agencies, as well as distribution to schools and commercial sale and subscriptions.

The plenary discussion revealed that a forum of editors of African medical journals (known as FAME) had been established. The forum is working with the *British Medical Journal* on a training programme for editors. Training has already been conducted for anglophone and francophone countries, and training for lucophone countries is planned. Editorial guidelines are being drawn up. The initiative is intended to improve the quality and regularity of publishing. Many countries do not have a critical mass of peer reviewers, and such countries will have access to international peer reviewers through this initiative.

Status of environmental data and information in the SADC region

The development of southern Africa's environmental information systems (EIS) can be traced to the late 1980s, when the then SADC environment and land management sector (ELMS) organised a series of meetings and workshops, in collaboration with a number of partners such as the United Nations Environmental Programme (UNEP), the World Conservation Union (IUCN), the United Nations Institute for Training and Research (UNITAR) and the Southern African Research and Documentation Centre (SARDC). The region's EIS programme was boosted with the establishment of the training and education sub-programme (implemented by the University of Botswana under an initiative called the SADC EIS training and education [SETES] initiative) and the technical and networking sub-programme (implemented by the then SADC Food Security Technical and Administrative Unit in Harare, through the SADC EIS Technical Unit [SETU]).

SADC SETU took the lead in pursuing the development of the EIS network in the region, through partnerships with, amongst others, the IUCN Regional Office for Southern Africa (ROSA) and SARDC IMERSCA. The collaborative effort coordinated by SADC SETU resulted in the development of the region's initial EIS databases, which included bibliographic databases, contact databases and the geospatial biodiversity metadatabase.

A number of other EIS initiatives were undertaken, including:

- The development of the Biodiversity Information System in the SADC region.
- The Southern African Biodiversity Support Programme (SABSP) Regional Biodiversity Information Systems (RBIS), built on the national environmental information systems through biodiversity information centres.
- The UNEP Global Environmental Information Exchange Network (INFOTERRA), considered one of the world's largest networks for the exchange of environmental information, the objective of which is to stimulate and allow the exchange and flow of technical and scientific environmental information between the countries, is key to the development of the southern African EIS network. The INFOTERRA network is coordinated by UNEP and is a 173-country affiliation, including most SADC countries. The Southern African Sub-regional INFOTERRA Network (SASIN) coordinates activities in southern Africa.
- Another initiative supportive of the region's EIS process is the SADC Water Resource Database (WRD), which evolved from the need to estimate the potential of small water bodies fisheries developments. This database integrates information on surface water

bodies, watersheds, rivers, fish species distribution and related data in a geographic information system (GIS).

- The EIS process in southern Africa also includes the identification, agreement and development of indicators used in trend analysis in state-of-the-environment reporting in the SADC region.
- A regional collaborative initiative is being undertaken by the Communicating the Environment Programme (a partnership of SADC, IUCN ROSA and SARDC I Musokotwane Environment Resource Centre for Southern Africa [IMERCSA]), with support from the South African Department of Environmental Affairs and Tourism (DEAT) and UNEP. Under the initiative, a core set of indicators for state-of-the-environment reporting has been developed, based on internationally accepted criteria for indicators of sustainability, and is being tested for relevance and data availability in the region.
- A process to collect data on socio-economic driving forces and integrate these into a GIS system was also undertaken jointly by the Human Sciences Research Council (HSRC) and SARDC as part of efforts to strengthen the region's EIS.
- Following the realisation that data resources for environmental management are segmented, tend to be application-specific, originate from different organisations that collect and maintain data for their own particular purposes, and are very often designed and operated independently of one another, a comprehensive Africa-wide initiative called the Africa Environment Information Network (AEIN) is being implemented by UNEP through a network of collaborating centres and pilot countries. UNEP's collaborating centre for southern Africa is SARDC IMERCSA, while the two pilot countries that recently concluded the first phase of the AEIN are Lesotho and Zambia. One of the key outputs of the AEIN is the Africa Environment Outlook EIS, through which a comprehensive electronic database of bibliographic information, experts and contacts, statistics and map information is available online.
- SADC, through the RISDP, considers statistics a priority intervention area, and aims to provide relevant, timely, accurate and comparable statistical information for planning, policy formulation, implementation, and monitoring and evaluation of SADC integration activities.
- The development of the SADC statistics is guided by the strategy document approved by the SADC Council of Ministers in 1998, which stipulates that SADC statistics are critical in the monitoring and evaluation of the SADC Programme of Action and the regional integration process. The RISDP strategy document focuses on developing a legal framework for regional cooperation in statistics; harmonisation of statistical information; collating, processing and disseminating official statistics; developing indicators for monitoring and evaluating regional integration; and building capacity for national and regional statistical systems. RISDP targets for statistics include the development of a legal framework (2004–2006); harmonisation of SADC statistics by 2015; development of an integrated regional statistics database in all priority areas by 2015, including poverty, gender, the informal sector, HIV and AIDS; development of indicators for monitoring and evaluation (2004–2005); enhancement of statistics capacity in SADC by 2015; and development of economic models and forecasting mechanisms for statistics (2004–2006).

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?

Scholarly publishing

ASSAf is partnering with the Council on Higher Education (CHE) to develop the role of research publishing in South African journals as part of the quality system. The National Research Foundation, the Medical Research Council (MRC) and other bodies that fund research at South African higher education institutions and museums have a keen interest in the success of the project, as they need to know what value to attach to a local publication when evaluating research. Higher education institutions also want to know the role that can be played by publishing in South African journals as opposed to international journals. The Department of Science and Technology, the National Council on Innovation (NACI) and research groups such as at the Human Sciences Research Council have an interest in knowing the value to attach to publications in South African journals as indicators of research

productivity. The strength of research could be built through close collaboration among the departments of Education and Science and Technology, the National Research Foundation and the Council on Higher Education.

An independent body, such as ASSAf, could be given responsibility for considering submissions of journal titles to be included on the list of accredited journals, since the Department of Education grants a subsidy only for articles published in journals on its accredited list (which includes journals listed on the ISI and IBSS databases, as well as selected locally published journals).

ASSAf intends to carefully examine the Australian model of incentives for publication in local journals, for possible application in South Africa. The plenary discussion suggested that researchers should get bonus points in terms of their subsidy from the Department of Education if they publish in South African ISI-listed journals.

Status of environmental data and information in the SADC region

The plenary discussion revealed that the NEPAD Ministers' Council for Science and Technology, which would be meeting in Senegal in September 2005, would be deliberating the possible adoption of a programme to develop an African observatory for science, technology and innovation, which would be tasked with producing the Africa Innovation Outlook (along Organisation for Economic Cooperation and Development [OECD] lines). Such an initiative would rely heavily on a network of competent national authorities and agencies associated with them to collect the data. A single point, wherever it was based, would not be in a position to collect primary data, and would rely on the provision of excellent data that meet specified standards to populate databases and assist in generating the Innovation Outlook.

RISDP targets for statistics include the establishment of a multi-sectoral forum of users and producers of statistics and the enhancement of the utilisation of cost-effective ICTs in sharing information in the priority areas by 2015.

Discussion revealed the existence of the Environmental Long-term Observatories Network of Southern Africa (ELTOSA), which is a regional body for long-term ecological research, with different degrees of support from the respective governments. It was recommended that RISDP make contact with ELTOSA, as it is a key player in data on the environment.

SESSION 4: POLICY AND LEGAL ISSUES

Chair: Mr Ahmed M. Makbel, National Bureau of Statistics, Tanzania

Papers presented

Access to Data and Information from Publicly Funded Research: A Policy Perspective, *Paul F. Uhlir, Director, Office of International Scientific and Technical Information Programmes, The National Academies, Washington DC, United States*

Compilation and Implementation of Policy Agreements for National Data Providers in the South African Biodiversity Information Facility, *Ms Renee le Roux, National Research Foundation, SABIF Project, South Africa*

Collaboratory Environments in Developing Countries, *Prof. Gary Olson and Paul M. Fits, Professor of Human–Computer Interaction and Associate Dean of Research, University of Michigan, United States+*

Intellectual Property and Indigenous Knowledge Systems: International Development and their Implications for Southern Africa, *Dr Mogege Mosimege, Director: IKS, Department of Science and Technology, South Africa*

What are the challenges and barriers?

Policy on access to data and information from publicly funded research

The digital revolution is fundamentally transforming the conduct of scientific research and virtually all spheres of human endeavour. The extent and rapidity of that change offers unprecedented opportunity for creating, managing, disseminating and using scientific and technical data and information. It also raises new challenges that need to be confronted. Many opportunities arise involving information societies and knowledge economies, whether with respect to data-intensive research, applications and conversion of data into higher levels of information knowledge, or making vast amounts of factual and scholarly information available to a broad spectrum of users. The inherent challenges lie in effectively managing these information resources for optimal access and use, and developing rational rules and structures for that process.

There are compelling reasons for placing government-generated data and information in the public domain and under open access conditions:

- A government entity needs no legal incentives from exclusive property rights to create information. The activities that the government undertakes, as well as the information produced by it in the course of those activities, are a global public good.
- The ethical consideration is that the taxpayer has already paid for the production of the information. The burden of additional access fees falls disproportionately on the individuals least able to pay. This is an important consideration in the transfer of knowledge from developed to developing countries.
- The political consideration is that transparency of governance is undermined by restricting citizens from access to and use of public, including scientific, data and information. Rights of freedom of expression are compromised by restrictions on re-dissemination of public information, particularly factual data.
- Numerous economic and non-economic positive externalities (especially through network effects) can be realised on an exponential basis through the open dissemination of public-domain data and information on the Internet. Conversely, the commercialisation of public data and information on an exclusive basis produces *de facto* public monopolies that have inherent economic inefficiencies and are contrary to the public interest on social, ethical and good governance grounds.

There are economic, legal and socio-economic pressures on the public domain status of publicly funded scientific data and information and for open access to such sources:

- The economic factors relate to the fact that the privatisation of government research and data collection activities has been going on for decades. Where this is inevitable in the interests of efficiency and effectiveness, it needs to be done in a socially responsible way in the public interest. A growing trend over recent decades has been the commercialisation of academic research and publishing (and of government data in many countries). This has resulted in very high costs of academic literature, controlled by a decreasing number of publishers. This trend has had a negative impact on access to scientific literature worldwide, particularly in the developing world, along with the privatisation of data from the source. There are different models for the long-term sustainability of data and information dissemination and archiving activities. Some include public–private partnerships. This is an area that has to be considered carefully in the research policy framework for every country, as well as cooperatively and internationally.
- There are many legal barriers, mostly arising from contingencies outside public science. Intellectual property laws have become broader, and are manifested in copyrights, patents, restrictive licences, statutory database protection and other restrictive laws, especially those based on national security, personal privacy and confidentiality. A balance is needed between the public interest of access and use and the legitimate interests of protecting personal or national interests through restriction.
- Sociological factors include competitive behaviour within the public research community, and the cultures and norms of traditional disciplines that operate to block collaboration and openness (in some cases these have legitimate genesis, while in others they need to be changed over time).

The broad implications of excessive restrictions on access to and re-use of data and information from public research include:

- Disadvantage and marginalisation of developing country researchers through high access costs, legal barriers that do not allow the recombination of information to give new knowledge, and restrictions on sharing information
- Significant lost opportunity costs to the research community, and the related failure to capture maximum value from public investment in public research activities
- Sole-source problems (monopolisation) exacerbated in scientific database and journal markets, both public and private
- Higher transaction costs (not just costs of access, but also the costs of the infrastructure required to make the transactions legally and institutionally accepted, which adds to the overheads associated with the transfer of information)
- Less effective international, inter-institutional and interdisciplinary cooperation using digital networks.

The plenary discussion yielded interesting debate on the open access versus cost recovery model of making data available. Questions were raised about the sustainability of the open access model of the United States, and it was mentioned that many African government institutions sell their data under cost recovery policies, citing the diminishing resources for data collection as the motivation.

It was explained that the United States model, under federal law, does not allow copyright protection of federally generated data, and there are no intellectual property barriers to the use of that information, which is in the public domain. Moreover, federal policy does not allow cost recovery other than for the information dissemination activity. For example, for meteorological data, it would not be possible to recover the costs of the whole programme, nor is this allowed. The preferred pricing policy for information is the marginal cost of dissemination, which is the cost of fulfilling the user request (essentially zero when done online).

The situation in Europe varies between countries. In many, if not most, cases, the government can charge up to full cost recovery for the activity. The prices are therefore quite high, and data acquire intellectual property protection, which inhibits re-use of information and diminishes the amount of subsequent follow-on activity with the public information. There is therefore a difference in the overall economic activity based on a free and open model versus

a restricted and expensive model. The European model takes the approach that has been mentioned with respect to Africa, based on a feeling that data dissemination activities should be self-sustaining or should generate revenue.

South African biodiversity information

Open access to publicly funded data and information is a relatively unfamiliar concept in developing countries. In South Africa, the existence and *modus operandi* of the South African Biodiversity Information Facility (SABIF) has raised concerns with respect to data security, data ownership, data use, data sharing policies, intellectual property, acknowledgements, copyright issues and privacy rights.

The implementation challenges identified by SABIF relate largely to the assumption that publicly funded data should be subject to access restrictions. These identified implementation challenges include the fact that data providers do not have data use and sharing policies; the problems of monitoring data usage (since each acceptance agreement is for using data only once); policing commercial versus non-commercial use; preventing bio-prospecting; and ascertaining that the data provider is the owner of the data.

The plenary discussion raised comments that SABIF is in many ways very advanced within the GBIF stable in terms of the way in which it has developed agreements for data providers and end users. However, it was questioned why SABIF and data providers have concerns with respect to distinguishing between use of the data for profit or non-profit in view of the fact that social gains can be identified from uses such as writing value-added search engines for the kind of data that SABIF serves or using the data to compile useful data products, which could be considered commercial uses.

The response to this probe was that a major concern is that use for profit does not advance science but furthers profit-seeking by consultants. Data providers do not always want to receive compensation, but rather just to be acknowledged as the source of the data. However, where commercial use is concerned, data owners want to be paid. In such cases, the data provider and end user sign a separate agreement for payment (subject to adherence to the Indigenous Knowledge Policy and the Biodiversity Act). Resolution of disputes between the end user and data provider are dealt with in terms of the agreement between these parties, without the intervention of SABIF.

Collaboratory environments in developing countries

A collaboratory is an organisational entity that spans distance, supports rich and recurring human interaction oriented to a common research area, and provides access to data sources, artifacts and tools required to accomplish research tasks. A key reason that many collaboratories have evolved has been the need for data and for understanding and sharing knowledge about data. ‘Collaboratory’ is a hybrid of ‘collaborate’ and ‘laboratory’. The word was first used in the late 1980s as scientists recognised the potential of expanding national and international computer networks. Collaboratories are generally physical spaces where people are co-located. The idea behind the concept of a collaboratory is to use the Internet to create the modern version of laboratories. The definition highlights the core capabilities of a collaboratory, namely, technology linking people to people, technology linking people to information, and technology linking people to facilities such as remote scientific instruments.

The challenge for collaboratories in Africa is that the less developed technical infrastructure makes it difficult to use some of the more sophisticated tools, or even to use *e*-mail or get access to the computer. People have limited or mixed experience with some of the more advanced technologies. However, very high motivation exists to use the collaboratory approach.

The factors that affect the success of remote long-distance collaboration are both technical and social. Few, if any, groups have mastered them all, whether in the United States or elsewhere. Developing countries may have poorer technical structure, but when people are

willing to work together, progress can be made. Developing countries may lack the human and institutional resources to build capacity, but partnering with others can help.

The implications of collaboratories for developing countries hold a mix of challenges, including the fact that such countries are high on collaboration readiness but may be low on technical readiness. The tools are improving, and include Web conferencing and voice over IP (Internet protocol). The impact of collaboratories could potentially be greatest in the developing world, with benefits for both research collaboration and education.

The potential benefit of international technologically mediated scientific collaboration goes beyond the immediate impact of scientific investigation. Scientists may be able to make new connections and should remain open to interesting secondary effects, as they may be the sources of the next great scientific discovery or the wellspring of the solution to an important local problem.

Intellectual property and indigenous knowledge systems

Over the past ten years, there has been a dramatic increase in interest in the role that indigenous knowledge can play in truly participatory approaches to sustainable development. It may not be accidental that the growing interest in the potential contribution of indigenous knowledge to development is becoming manifest at a time when current development models have not proved particularly successful. Hundreds of millions of marginalised people all over the world are still excluded from the mainstream of development. New insights reveal that development interventions have failed to induce people to participate because of the absence of instruments and mechanisms that enable them to use their own knowledge.

Indigenous knowledge is unfortunately vulnerable to exploitation for commercial gain, often without compensation to the communities within which it originated and was nurtured and developed over centuries.

The Pacific Island Forum Secretariat and others involved in indigenous knowledge systems (IKS) have argued that the current intellectual property rights (IPR) system cannot protect indigenous knowledge. Three reasons were advanced: (1) the current system seeks to privatise ownership and is designed for knowledge to be held by individuals or corporations, whereas indigenous knowledge is collectively owned; (2) the protection is time-bound, whereas indigenous knowledge is held in perpetuity; and (3) it adopts a restricted interpretation of invention, which should satisfy the criteria of novelty and be capable of industrial application, whereas innovation in indigenous knowledge is incremental, informal and occurs over time.

Other arguments have been advanced against the use of the intellectual property rights regime in protecting IKS. There is concern that IPR systems encourage the appropriation of indigenous knowledge for commercial use without sharing the benefits with the knowledge holders. As a result, it has been recommended that a *sui generis* or alternative law is necessary to protect indigenous knowledge. A further argument is that current IPR regimes, particularly patents, threaten to worsen the piracy of biological resources and associated traditional and indigenous knowledge. Biopiracy of indigenous knowledge could be regarded as a double theft: it steals creativity and innovation, and patents on stolen knowledge rob owners of economic development.

However, arguments have been advanced for using or adapting current IPR regimes for the protection of IKS, or parts of IKS. Copyrights can be used to protect the artistic manifestations of traditional knowledge holders, especially artists that belong to indigenous communities, against unauthorised reproduction and exploitation. The patent system could be used for the protection of technical solutions that are industrially applicable and universally novel and that involve an inventive step. New plant products, cultivars and varieties of all species of plants may be protected under plant breeders' rights (PBRs). The design and shape of utilitarian craft products such as furniture, receptacles and garments, as well as articles of ceramics, leather, wood and other materials, may qualify for protection as industrial designs. Examples

have also been given of trademarks, trade names, geographical indications and appellations of origin, and repression of unfair competition.

It is clear that there are conflicting views on whether indigenous knowledge can be protected adequately by existing IPR regimes. It is important that different countries continue to explore the possibilities for protecting indigenous knowledge, whether they find existing IPRs appropriate, adapt the existing regimes or opt for *sui generis* legislation.

The World Intellectual Property Organisation (WIPO) acknowledges the difficulties related to debates on indigenous peoples, and regards indigenous knowledge as part of traditional knowledge. WIPO conducted nine fact-finding missions between June 1998 and November 1999 in 28 countries in the South Pacific, southern Africa, East Africa, West Africa, South Asia, North America, Central America, South America, the Arab countries and the Caribbean to enable WIPO to identify the intellectual property needs and expectations of traditional knowledge holders. Arguments were generated for and against the use of intellectual property rights to protect traditional knowledge.

What are the existing resources and mechanisms?

Policy on access to data and information from publicly funded research

The economic value extracted from the use of public sector information is ten times greater in the United States than the European Union (EU), although the investment is only double. The European Union invests 9.5 billion Euro per year in publicly funded data, compared with 19 billion Euro for the United States. The European Union derives a benefit of 68 billion Euro per year from its investment, compared with the benefit of 750 Euro per year from the United States investment (PIRA International 2000). The reasons for this disparity are primarily that it is the policy of the United States to make public sector information openly available as far as possible, which generates economic activity, value added and secondary economic activity based on the exploitation of public information.

South African biodiversity information

SABIF (the South African node of GBIF) has been created as an enabling platform for researchers in South Africa to share information while at the same time having access to global biodiversity information. Through this platform, space is created for researchers to explore new and existing areas for improved knowledge generation (by identifying gaps and new directions), and to develop end-user applications (for the benefit of society and policy-makers). Data providers share information with SABIF either directly through a virtual network or by providing data that can be stored on a database in the SABIF portal.

SABIF's vision is achieved by promoting the sharing of data and information under a common set of standards by developing and maintaining metadatabases; establishing and maintaining a Web portal; providing funding for data digitisation; providing funding for data provider interface; facilitating the development of application tools for data use; encouraging networking and coordination of activities; contributing to education, training and capacity development for promoting national access to data; facilitating marketing and awareness-raising of the benefits and applications for end users; and advising on national and international initiatives.

Collaboratory environments in developing countries

The worldwide AIDS epidemic illustrates the value of bringing scientists from around the world together to collaborate on a problem that none of them can solve on their own. AIDS has hit sub-Saharan Africa particularly hard. The International AIDS Research Collaboratory (IARC) was established to study and support the collaboration needs of projects. It is based primarily at the University of Michigan. The philosophy is to use an iterative user-standards designed philosophy, by working closely with scientists to understand their needs and then introduce and track the use of collaborative tools that might help them in their research. This is done

through visits to their places of work, interviews, surveys and training to support the technology. This requires technical and community structure.

The first collaboratory programme to be supported was the HIV pathogenesis programme (HPP), a three-way collaboration among the Nuffield Department of Clinical Medicine (Oxford University), the AIDS Research Centre (Harvard University) and the Nelson Mandela College of Medicine (University of KwaZulu-Natal). The second programme was the partnership for HIV research and education between the Harvard AIDS Institute, the School of Public Health at Harvard University and the Ministry of Health of Botswana.

IARC supports laboratory meetings, remote lectures, clinical meetings and remote colloquium speakers. In the early days, collaboration was accomplished by visits to sites, which still remain important. Because of the need to get together face-to-face, the need for more real-time communication was identified early on. Data are difficult to interpret and repeat at various sites to get the same findings. There is therefore a need to collaborate closely over data. On the basis of interviews and observations, the Centra System was recommended to support real-time collaboration. Centra has a large suite of tools to support real-time conferencing. The most significant feature is voice over IP, optimised for long bandwidth networks.

IARC is part of a larger initiative, the Science of Collaboratories, which is funded by a five-year grant from the US National Science Foundation (NSF). The purpose of the initiative is to investigate social and technical elements that contribute to successful collaboration across many scientific disciplines. The initiative takes the approach of building bottom-up principles of successful long-distance collaboration.

Intellectual property and indigenous knowledge systems

The South African Cabinet adopted the Indigenous Knowledge Systems (IKS) Policy in November 2004, and the Department of Science and Technology launched the policy in March 2005. The policy addresses key policy drivers; IKS and the national systems of education and innovation; stakeholders and role-players in IKS; the institutional framework; IKS funding and principles; national and international imperatives; and the role of various government departments and the intergovernmental committee on IKS.

With respect to efforts to develop IKS policies for the various countries in the SADC region, a SADC workshop was held in June 2004 as one of a series of regional IKS workshops, the first of which was held in Mozambique in 2002. A recommendation of the June 2004 workshop was that each country in the region should develop its own policy on IKS. A framework was discussed, for which the South African policy would serve as a guide. This has been agreed to by SADC member states.

In July 2005, a science and technology policy workshop was held in Namibia, where it was agreed that the South African IKS Policy should be largely adopted by most SADC states, with adaptations by each country. It is important for the region to work together to protect indigenous knowledge. For example, it is difficult to protect an endangered plant in one country if it also occurs in a neighbouring country where it is not protected.

An important development at the international level has been the establishment of the Intergovernmental Committee (IGC) on Intellectual Property and Genetic Resources, Traditional Knowledge and Folklore at the WIPO general assembly in October 2000. The IGC has met on eight occasions in Geneva, Switzerland. The work of the IGC has included a study of the operational definitions relevant to traditional knowledge; a review of existing national systems of intellectual property for traditional knowledge; an analysis of the elements for a possible *sui generis* system for the protection of traditional knowledge; the use of databases to promote defensive protection; and the development of an intellectual property management toolkit for the documentation of traditional knowledge.

About 90 countries are involved in the IGC, and after eight meetings over four years, there is still a stalemate, as developed countries are not in agreement with developing countries on international instruments that would be legally binding on IKS matters.

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?

Policy on access to data and information from publicly funded research

The following principles should be used to inform the creation of a legal and policy regime that supports the open availability and unfettered use of publicly funded scientific data and information produced by government and government-funded sources:

- Express prohibition of intellectual property protection of information produced directly by the government
- Contractual reinforcement of the sharing norms of science through open data terms in government research grants and contracts, and deposit of scientific literature in open repositories (which would greatly enhance the accessibility of information produced with public funding by academics and non-profit organisations and lead to increased availability)
- Maintenance of a large and robust public domain for non-copyrightable data, as well as other immunities and exceptions under intellectual property law favouring public science and education. (This is more difficult to achieve, because statutory developments are typically outside the scope of the scientific community, but that community does have a role to play in informing policy-makers and legislators about their interests and the importance of not overly restricting access to and use of information for research.)
- The explicit treatment of data and information resources as a major research infrastructural investment by the research community and policy-makers, with appropriate institutional mechanisms and support, including generating the institutional frameworks for protecting information over the long-term and making it available
- Disseminating data and information freely online as the most effective way to maximise the investment in public scientific research.

The social and economic benefits of open access to information, the economic activity that is generated by it, the taxes paid to Treasury based on that activity, and the employment involved surpass the amount recovered from the few users that are willing and able to pay for accessing the information. Even though institutions do recover some of their costs by selling data, the cost to society is much greater. Very few people can afford to pay more than simple distribution costs. This is a policy issue that has to be addressed at a high level, not at the level of the institution that charges the access fees. This leads to the question of sustainability of the activities, but there are different models for funding information dissemination activities that are typically at a much lower cost than the cost of assembling the information. It ought to be part of the cost of the whole infrastructure, and be much lower in cost than the original cost of assembling the data.

Another issue raised in the plenary discussion was that the arguments for open access to scientific information are even more compelling in developing countries than they are in developed countries. Sociological barriers were mentioned in the presentation, including mindset, competitive behaviours and institutional cultures. It was queried whether there is any advice for dealing with this kind of institutional resistance. With respect to sustainability, it was mentioned that the South African Human Sciences Research Council (HSRC), which has adopted an open access publishing policy (which is proving successful, despite a thorny period of implementation), is beginning to learn to measure its impact in ways other than revenue.

The response was that the benefits of open access are greater in developing than developed countries, as even low cost access can be an insurmountable barrier. Open access allows all potential users with access to the Internet to use the information most effectively. Models with respect to sociological barriers should be considered from both the top-down and bottom-up

perspective, with the institution in the middle. Normative behaviour can be changed from the top with laws and polices that provide incentives and disincentives for certain types of activities (including economic, legal and policy issues). From the bottom up, it is possible to demonstrate demand. In Europe and the United States, there are many self-organising activities. There are many examples of bottom-up collaboration without a top-down edict to do so, based on volunteer-type effort and thus at low cost, within the ambit of already funded public research. This does not add much to the cost but clearly has many benefits.

At an institutional level, there are ways of measuring the benefits of greater access, some of which are economic and others of which are non-economic, social and scientific. With open access journals, studies show higher levels of citation and exposure of open-access authors. Open-access journals have risen quite high in terms of the ISI-type impact index. It has already been demonstrated that openness is very effective and lower in cost overall. There is still the problem of sustainability over the long term, and the need to find a mix of funding or volunteer assistance. That is the most difficult problem to address, and the jury is still out on whether the open-access model for journals is sustainable in the long term, especially high-end journals (for example, the Public Library of Science, which has not yet covered its costs).

In the plenary discussion, the possibility of pursuing international efforts to standardise data access through treaties and conventions was raised.

Collaboratory environments in developing countries

Given the benefits of collaboratory initiatives for developing countries, outreach programmes could be launched to promote this type of activity and to reach out to other communities. The University of Michigan has been involved in outreach projects for the last ten years, since the tools became available. The first big project they supported was Newark, the upper atmosphere research collaboratory, which started in the early 1990s when many of the tools were at an early stage. They have been involved in about ten such projects.

Intellectual property and indigenous knowledge systems

Regional and continental collaboration in IKS activities is important, as the following examples show:

- Most of the Hoodia debates and agreements have focused on the San community in the Kalahari, South Africa. However, San communities are also found in Botswana, Namibia, and Angola. Based on the example of the Hoodia in the region, it is clear that collaboration in IKS (within the SADC region and the African continent) is essential to prevent loss of monetary benefit and intellectual property.
- Collaboration at the WIPO IGC meetings has ensured that Africa is able to withstand the pressure of developed countries in negotiations on indigenous knowledge. This collaboration needs to be encouraged and promoted.
- Many institutions have a database of some kind related to medicinal plants and traditional medicines, including the Council for Scientific and Industrial Research (CSIR), the South African National Biodiversity Institute (SANBI), the Medical Research Council (MRC), the Agricultural Research Council (ARC), universities and technikons (universities of technology). Often these databases have no reference at all to one another. These databases can serve as a first step towards the creation of a South African Indigenous Knowledge Systems Digital Library (SAIKSDL). The South African Department of Science and Technology (DST) has embarked on a process of identifying all existing databases so that it can verify what is available and what needs to be done. DST has also approved a proposal by the CSIR to develop a Traditional Medicines Database for South Africa. This is just one of many such projects that would need to be undertaken in developing SAIKSDL.
- The IKS Policy provides for the creation of indigenous knowledge databases, which should be harmonised with international systems so that cross-referencing is possible.
- In fulfilment of their obligations under the Convention on Biological Diversity (CBD), countries have enacted, or are drafting, legislation to regulate access and benefit-sharing arrangements in relation to biodiversity and knowledge of its use. In South Africa, the

Department of Environmental Affairs and Tourism has been involved in drafting the Biodiversity Bill. The Department of Science and Technology (DST) has made inputs to specific chapters dealing with access and benefit-sharing. In addition, DST is leading discussions on benefit-sharing models that need to be in place for the benefit of knowledge holders, who are the most exploited parties in these arrangements.

IKS is a cross-cutting field, involving the collaboration of scientists from different disciplines. The inter-departmental structure for IKS involves forty South African ministries and government departments in operational responsibility for IKS, and the way in which it will function is still in the early stages of development.

The plenary discussion investigated how to involve communities in discussions on IKS and IPR and in related research. It was pointed out that the first audit of indigenous knowledge, conducted by nine South African universities in 1998, recommended that in the interests of ethical practice, IKS research not be conducted without the appropriate involvement of communities, and that such communities be acknowledged when the work is published. The National Research Foundation had been engaged in this regard. To accomplish this requires a change of perspective among researchers, which takes time to effect.

SESSION 5: INSTITUTIONAL AND ECONOMIC ISSUES

Chair: Prof. Michael Kahn*, Human Sciences Research Council, South Africa*

Papers presented

E-research Support Services: a Cost-effective Solution for the South African Research Community*, Ms Susan Veldsman, Coalition of South Africa Library Consortia and the South Africa Site Licensing Initiative*

The Work of the International Network for the Availability of Scientific Publications (INASP) in Access to Information through the Programme for the Enhancement of Research Information (PERI), *Dr Buhle Mbambo, International Network for the Availability of Scientific Publications, PERI Country Coordinator and University Librarian, University of Zimbabwe, Zimbabwe*

A National Data and Information-sharing Platform for the Non-profit Sector: an Initiative by the Department of Social Development, National Development Agency and Nelson Mandela Foundation, *Bok Mapena, Tsholo Mogotsi and Dr Lulama Makhubela, South Africa*

The Role of Archives in Permanent Access to Scientific Data and Information in Southern Africa*, Dr Patrick Ngulube, University of KwaZulu-Natal, School of Sociology and Social Studies, Information Studies Programme, South Africa*

What are the challenges and barriers?

E-*research support services*

Researchers are dependent on access to global research outputs and interactive tools if the quality of research is to be maintained. An increasing proportion of information and knowledge is available online as part of the emerging *e*-world. Two main obstacles are the high and increasing cost of access rights (in terms of negotiated licences) and the cost of connectivity (for example, in India Internet access costs 5c per minute, whereas in South Africa it costs 25c per minute).

There are many disconnected small initiatives across South Africa, making scholarly communication disjointed and inaccessible for the majority. Preliminary assessment points to the need to create a 'Team South Africa' approach, with high-level participation and commitment in the interests of all researchers – not only those at tertiary institutions or research institutions.

Overseas visits indicated a wider scope than online content, known as *e*-research. The fundamental elements of *e*-research are *e*-science, digital curation and access to content, which are pre-competitive and lend themselves to joint investment. There are common infrastructural issues and common researcher support needs related to the various aspects of *e*-research.

The challenges of *e*-research are to integrate IT and *e*-research in the daily work processes of researchers, teachers and students; to take the needs of researchers as the starting point (including access to information, study environment, collaboration tools, personalised information, support with statistical databases and information literacy); integrate the digital library with the digital learning environment; and be actively involved in creating and maintaining institutional repositories and open access as a support tool for researchers. This requires closer connection between libraries, researchers, information technologists, government and other stakeholders to enhance and support a Team South Africa approach.

Online access in South Africa

Expenditure on online access amounts to approximately US$18 million per annum, which is an enormous investment for a developing country. The South Africa Site Licensing Initiative (SASLI) is doing valuable work in negotiating terms and conditions for site licensing, and in

2005 achieved savings of US$50 million for subscriptions to online access for tertiary and research institutions. However, SASLI has no high-level 'national' clout in negotiations. The Coalition of South Africa Library Consortia (COSALC) is an ineffective empowerment structure, and would like to have the backing of the Department of Science and Technology or the Minister in negotiations.

Publishers still strive to make the rules. Many researchers are well served, but others are very poorly off. Savings allow well-resourced libraries to increase their holdings, but poorly resourced libraries remain disadvantaged. The exchange rate of the South African rand is recovering against other currencies, but subscriptions are denominated in euros or dollars, resulting in extensive cancellations when the rand/dollar exchange rate deteriorates.

Open access in South Africa

There are a number of scattered initiatives, including *e*-thesis repositories and *e*-journals, but there is no focus for the development of support systems. Researchers are in bondage to a subsidy system focused on a limited list of peer-reviewed journals. Approximately 90% of global journals now allow publication in institutional repositories, but South Africa lacks incentives to use such repositories (with the HSRC publishing model as a notable exception).

Digital curation in South Africa

The exponential growth of data is a major resource investment. There are moves to make primary research data and findings accessible, but considerable data were lost in the 1980s and 1990s. Most databases are institutionally held (or may even be held at a project level). There is no managed access or promotion of use beyond the original data generators.

E-*science*

Some researchers are battling with global data transfer. For instance, a researcher involved in an international human genome research project is on a sub-network of a major university in South Africa. Each time she tries to submit her data, the system crashes. She therefore has to submit the data on a compact disk, which is starting to marginalise and isolate her. *E*-science is a new growth area.

In the plenary discussion, it was mentioned that moving from paper to electronic versions gives rise to copyright issues when seeking permission to digitise. Many library projects are being delayed or terminated because they cannot obtain such permission. The response was that copyright issues were not a major aim of the project, and operational issues would have to be dealt with in the innovative leg of the model. Copyright issues are equally challenging in the paper environment.

International Network for the Availability of Scientific Publications (INASP)

Typical scenarios of most participants in the Programme for the Enhancement of Research Information (PERI) include limited bandwidth, limited financial resources, pioneering in *e*-resources provision, developing new ways of doing business, growing student populations, limited ICT skills and limited equipment.

The challenges to the access initiatives of PERI include information overload and stretched infrastructure (in terms of both bandwidth and number of computers). The PERI mid-term review identified challenges with respect to the political and institutional will in supporting research; effectively situating *e*-resources within a context of *e*-learning; increasing the use of *e*-library services and resources; the development of evaluation and impact indicators; and supporting strategies towards sustainability.

A national data and information-sharing platform for the non-profit sector

The South African Department of Social Development (DSD) is mandated to administer the Non-profit Organisation Act (Act No. 71 of 1997), for which it registers and manages the

annual reporting of registered non-profit organisations (NPOs) in South Africa, of which there are more than 33 000. The NPO sector in South Africa employs more people than the mining sector, the transport sector or the energy sector, and spent R10 billion in 1998. Globally in 38 countries, expenditure was more than a trillion rand. The sector has not seen much benefit from the advances in IT that have benefited the business and public sectors. The Department of Social Development receives 200 NPO registration applications per week, and the number is increasing. The current NPO database system is not designed to cater to the increased demand for registration of NPOs. This situation presents challenges in dealing with public expectations and requests for information on NPOs, the processing of reports of registered organisations and the unveiling of statistical information.

The South African National Development Agency (NDA) administers the National Development Agency Act (Act No. 108 of 1998), with its mission to eradicate poverty and its causes through partnerships. A comprehensive database of civil society information is a requirement of the Act and is fundamental to NDA in achieving its aims.

The Nelson Mandela Foundation promotes and enables the growth of human fulfilment and the continuous expansion of the frontiers of freedom, and is a promoter and catalyst of the Nelson Mandela legacy. Through strategic networks and partnerships, it directs resources, knowledge and practice to add value and demonstrate new possibilities. In fulfilling its mission, it requires access to a comprehensive civil society information platform to support its work in developing strategic networks and partnerships.

The role of archives in permanent access to scientific data and information in southern Africa

In the highly intensive field of scientific knowledge creation and transfer, preserving data and information into the future is very important. One of the keys to sustainable development and the ability of the scientific community to conduct innovative and valid research are access to and use of long-lived digital data and information. Access to knowledge – primarily in agriculture, medicine and technology – can help to create stronger social, economic and technical infrastructures that are essential in the development process. Furthermore, access to data and information makes research efficient and opens up opportunities for further research.

Digital data collections give researchers access to data from a variety of sources and enable them to integrate data across fields. The relative ease of sharing digital data (compared to data recorded on paper) allows researchers, students and educators from different disciplines, institutions and geographical locations to contribute to the research enterprise. It democratises research by providing the opportunity for all who have access to data collections to make a contribution. However, continued access to digital data collections by the research community in sub-Saharan Africa is not guaranteed. Until recently, the main focus has been on the potential opportunities offered by the new technologies without addressing the challenges they pose in relation to the long-term processibility of and access to digital information. Although ICTs can no longer be seen as a luxury for the elite, they create and store information in a variety of formats, and the technologies are changing at an increasingly rapid rate. This situation presents myriad challenges to the information society, which is largely dependent on digital data collections.

National archival institutions, as the central agencies for official records in their countries, are key players in the long-term preservation of information irrespective of format, but national archival institutions in Africa have tended to concentrate on public records at the expense of private ones.

There is need for vigorous action if the loss of society's electronic memory is to be averted. As record-keeping moves from paper to electronic systems, it is essential for archival professionals to formulate policies to ensure access to comprehensive, authentic and tamper-proof digital collections. Records created today should be readable with tomorrow's technology. In essence, archivists in southern Africa should take an active role in ensuring

that electronic records are protected from technological obsolescence and media impermanence.

The limited capacity of national archival institutions has led sceptics to think that national archival institutions cannot play a significant role in digital preservation. International examples have demonstrated that the capacity of national archival institutions may be developed through partnerships. Partnerships are fundamental to service delivery in the information society and to safeguarding society's memory. The need to collaborate and develop partnerships is underscored by the globalisation of information, which in turn has led to the emergence of an information society. The ideals of the information society will remain a pipe dream if information produced by research agencies is not managed through its life cycle and made accessible when it is needed. National archival institutions have experience in managing data and information into the future. In that regard, national research foundations and councils should work in partnership with their respective national archival institutions in order to ensure that digital data and information are not lost to obsolete technology.

Although electronic records are proliferating throughout southern Africa, many archival repositories have not yet addressed the implications of the preservation of and access to digital materials. Policies and other technical guidance addressing best practices in digital preservation are inadequate. Expressed differently, insufficient attention has been paid to adopting and implementing policies that address the preservation of digital collections. In addition to that, archival institutions have little expertise in the area of preserving electronic data collections. Consequently, electronic data collections in southern Africa are in grave danger of being lost and becoming inaccessible. Very little research is being done on the digital preservation of collections in southern Africa.

In the plenary discussion, the issue was raised that archives do not have the resources to manage all records on their own. One of the debates is whether national archival institutions should take all records into custody or manage some of them according to a distributed system, with institutions managing their own digital objects with the help of the national archive.

It was questioned whether any attention was being paid to capturing non-digitised data and what could be done in that regard. The response was that some paper records had been scanned to create digital records, but that this has to be weighed in terms of the costs, including the costs of managing the records into the future. It is possible to digitise paper records, but that is not a priority. The paper records are already being preserved according to international standards. The focus should rather be on standards for preserving electronic information into the future. Efforts are under way to digitise the records of freedom organisations in southern Africa.

What are the existing resources and mechanisms?

E-*research support services*

The South African Research Information Services (SARIS) project focuses on online content.

Open access in South Africa

The Department of Science and Technology is contemplating legislation to make research results more accessible (following the lead of countries such as the United States where there is open access to publicly funded research). The South Africa Site Licensing Initiative (SASLI) offers regular workshops to create awareness of open access, as well as training in setting up institutional repositories.

Digital curation in South Africa

South African repository initiatives include the South African Integrated Spatial Information System (SA-ISIS), the HSRC Human Resources Development (HRD) Data Warehouse, the

South African Data Archive (SADA) in the NRF, Digital Imaging South Africa (DISA) and some initiatives by the Department of Science and Technology.

E-*science*

The use of data and models across research teams is increasing. The South African National Research Network (SANReN), as envisaged by the Department of Science and Technology, will link South Africa to GéANT, Internet2 and the Australian Academic and Research Network (AARNET).

International Network for the Availability of Scientific Publications (INASP)

The mission of the International Network for the Availability of Scientific Publications (INASP) is to enable worldwide access to information and knowledge with a particular emphasis on the needs of developing and transitional countries. Existing initiatives to provide access to international publications (online) include the Health InterNetwork Access to Research Initiative (HINARI) (managed by the World Health Organisation [WHO] and specialising in health content), Access to Global Online Research in Agriculture (AGORA) (managed by the Food and Agriculture Organisation [FAO] and specialising in agricultural information), Electronic Information for Libraries (EIFL), to provide low-cost access for developing countries (managed by EIFL.net), and PERI (the Programme for the Enhancement of Research Information, managed by INASP).

PERI aims to support the enhancement of capacity in the research sector in developing and transitional countries by strengthening the production, dissemination and access to information and knowledge utilising new information and communication technologies. PERI is mandated to work in countries that fulfil GDP and human development index (HDI) criteria, which in essence comprise the poorest countries.

PERI works with country coordinators and networks of scientists, researchers, institutions and libraries to identify needs. Country coordinators or coordinating teams plan, implement and develop services, and PERI responds and facilitates.

PERI had 22 participating countries in January 2005, seven of which were self-funded, and there were 22 more country requests in the pipeline. A total of 14 500 full-text peer-reviewed journals titles were available. In 2004, there were 360 896 full text downloads.

PERI assists national research publishing with online publishing support (for example, on the www.ajol.info platform, which has been taken over by the National Inquiry Services Centre [NISC], and via other services) and skills and knowledge-building through workshops, study tours, partnerships and published resources.

PERI provides ICT and *e*-resource management training in nine workshop series, including 65 different modules, ranging from 'Using the Internet' to 'Bandwidth Management and Optimisation'. PERI also provides 'travelling' training materials, 'cascading' methodology (by training trainers within the countries), local facilitation and in-country delivery.

PERI supports bandwidth optimisation and management for research and development (R&D), development and training in managing available resources, and development and implementation of local search mechanisms.

INASP is a non-profit organisation. PERI was initially funded by donors, through which it built up its resources. PERI has a model of self-finance by participating countries, with self-sustaining mechanisms.

The PERI programme has been critical in the provision of multidisciplinary digital information. It has worked in the participating countries to strengthen local journal *e*-publishing, has strengthened the dissemination of local information electronically, and is working to ensure the sustainability of the provision of and access to digital information.

A national data and information-sharing platform for the non-profit sector

The Department of Social Development (DSD), the National Development Agency (NDA) and the Nelson Mandela Foundation have formed a partnership to establish a single national data and information-sharing platform for the non-profit sector in South Africa. The memorandum of agreement, activity plan and budget are currently being finalised.

The DSD (the regulator of NPOs in South Africa) will provide its extensive NPO database as well as business processes for the sustainable collection of data and the provision of public access to information, since registered NPOs have to submit an extensive annual report to DSD. The NDA will facilitate the convening of civil society stakeholders and provide support in knowledge management, R&D and policy development. The Nelson Mandela Foundation will direct international resources and knowledge to add value to the development of the platform, and will provide the facilities to convene stakeholders.

The anticipated benefits to civil society organisations (CSOs) of the national platform are to offer all CSOs, regardless of their size or means, a free opportunity to explain their work and demonstrate their accountability to the public; to increase the visibility of South African CSOs and their access to a significant range of local and international grant-makers, donors and other potential givers and supporters; to streamline reporting to grant-makers, donors, intermediaries and any individuals, researchers, institutions and agencies requesting information about their work; and to allow CSOs to understand the work of peer organisations and access benchmarking data.

The anticipated benefits to donors and grant-makers are to satisfy the first level of due diligence for grant applicants; to allow convenient mapping of grant programme areas; to permit grant-makers to identify other non-profit organisations, for comparison purposes, that are similar to grant applicants; and to identify other donors that have committed resources to specific CSOs and causes.

The anticipated benefits to individual donors and society in general are to provide the tools to identify, compare and track the records of non-profit organisations undertaking activities that they wish to support, which promotes confidence in CSOs among the general public; to allow individuals to confirm the status of any fundraising entity and assess its strategy, programmes, organisational capacity and impact; to establish transparency as the accepted practice in civil society; and to increase transparency, which enables the public to gain greater confidence in and satisfaction from giving and volunteering.

In the plenary discussion, the issue was raised that the annual narrative and financial report to the Department of Social Development (DSD) is an additional cost and an onerous task for small non-governmental organisations (NGOs). Experience so far has shown that there has been no value added from DSD, and it is therefore gratifying that the database intends to benefit NGOs. NGOs would like to see another benefit, namely negotiation with the South African Revenue Service (SARS), so that if in the opinion of the DSD an organisation is non-profit, it does not have to go through the process of persuading SARS of this.

An organisation has to be registered with the NPO directorate in order to be registered as a public benefit organisation for tax purposes. SARS and DSD use roughly the same information to establish whether an institution qualifies. The platform is intended to provide an opportunity for an organisation to indicate, when it submits an application to register as an NPO, whether it wants automatic application to SARS for public benefit status. The data used are the same, but the criteria of SARS and the DSD are different. SARS also looks at the cause being pursued. Some NPOs are legally non-profit but not public benefit organisations.

The project aimed to address the cost of compliance. The benefits to non-profit organisations are both direct and indirect. The direct benefit is marketing and making information about NPOs accessible on the Internet, rather than stored in a filing room. The indirect benefit is to provide an opportunity for CSOs to improve their practices by gaining access to the reports of other organisations.

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?

E-*science*

Some potential users will cope, while others need help by means of a national help desk.

The proposed structure for an *e*-research support service for South Africa includes an *e*-research board that would be responsible for *e*-research development, innovation and service delivery. It is not yet clear who would take ownership of the *e*-research board. Roll-out possibilities include the Department of Science and Technology (DST) or a consortium of stakeholders (including the Ford Foundation), and funding possibilities include the Carnegie Foundation or DST.

International Network for the Availability of Scientific Publications (INASP)

In the plenary discussion, it was mentioned that increased bandwidth and improved access to digital information has also helped proliferate viruses in many institutions. It is up to institutions to manage their firewalls and protect themselves. There are several affordable packages on the market, and PERI would consider an application from a university for consortium access to Norton Anti-virus. The security of digital records is problematic, but should be considered in terms of the cost–benefit relationship and risk management. Even firewalls are not foolproof, and there is also a need for a disaster management plan so that when disaster strikes, it does not affect all repositories or vital digital documents.

The plenary discussion suggested that it would be useful for PERI to collect statistics on the impact of the programme on the quantity and quality of published research, the quality of journals and the frequency of use, as such statistics would facilitate negotiating for electronic journals in the future. Current feedback is anecdotal, and there has been no objective assessment of the programme.

It was also suggested that making PERI more widely available to individuals, in addition to making it available through institutions, would help increase the critical mass of users and make the sharing of costs more manageable, as well as facilitate access to information for larger numbers of researchers.

A national data and information-sharing platform for the non-profit sector

The major steps forward for the national data and information-sharing platform for the non-profit sector are to convene stakeholders and understand their needs; leverage international experience, knowledge and resources; develop an implementation business plan for a sustainable national platform; direct resources; and implement the national data and information-sharing platform for the non-profit sector.

The role of archives in permanent access to scientific data and information in southern Africa

Preserving digital collections is becoming urgent if the impending digital data and information 'tsunami' is to be averted. Some of the strategies for preserving digital data and information into the future include that national archival institutions should spearhead programmes such as the National Digital Information Infrastructure Preservation Programme; provide leadership in formulating policies for digital preservation; help institutions design data management plans; influence national research-funding agencies to allocate financial resources for research in digital preservation; provide training in accessing and using digital collections; provide guidance on the most suitable formats in which to create and preserve digital collections; help facilitate access to publicly funded research outputs and assist in building institutional archives that are compliant with the Open Archive Initiative for Metadata Harvesting; develop partnerships with research agencies; develop national strategies for digital preservation; assume a proactive role in promoting standards and best practices in

digital preservation; help organisations to define policies for the retention of digital data and information over time; assist organisations in selecting file formats that are suitable for archiving digital data; support the Open Archive Initiative and promote open access and long-term archiving of information; and lobby funding agencies such as the South African National Research Foundation to promote research on the preservation of digital objects as one of its major priorities.

In the plenary discussion, it was stressed that archivists have to preserve and protect valuable digital information, but an almost equally important function is to market the availability of the data and ensure a wider audience. The critical issue is answering the question: 'Why preserve?'

Conclusion

The title of the session could have been 'Government as an enabler or a disabler of information sharing'. One of the themes that came through strongly, perhaps because all the speakers were talking from an African perspective, was the enormous cost of catching up. In the South African environment, one of the reasons for the high cost of telecommunications is that South Africa has adopted a universal service agenda, which means that those that can afford to pay subsidise those that cannot. Bandwidth is therefore at extortionist levels even for schools. The cost of compliance has been another theme. Being very well informed and identifying champions in government is perhaps one of the ways of making progress.

SESSION 6: MANAGEMENT AND TECHNICAL ISSUES

Chair: Dr Roberta Balstad, Director, Center for Earth Science Information Network, and Chair of the United States National Committee for CODATA

Papers presented

Managing the Impacts of Programmatic Scale and Enhancing Incentives for Data Archiving, *Dr Raymond McCord, Oak Ridge National Laboratory, United States*

Selection, Appraisal, and Retention of Digital Scientific Data, M*r John Faundeen, United States Geological Survey's EROS Data Center, United States*

Oceanographic Data and Information Network for Africa (ODINAfrica), *Mr Mika Odido, Intergovernmental Oceanographic Commission of UNESCO, Nairobi Office, Kenya*

National Research and Education Networks: the Potential for SADC Participation, *Dr Duncan Martin, Tertiary Education Network (TENET), South Africa*

What are the challenges and barriers?

Managing the impacts of programmatic scale and enhancing incentives for data archiving

We are at the doorway of many opportunities for data archiving, and we need to plan now how to best leverage the future. Despite concerns about funding models, we have the time to think through the issues thoroughly. Many of the issues have to do with human behaviour, rather than money. The future limits will not be technology but our minds.

For the 400 years during which the scientific method has been in use, it has been based on original observations, which are analysed. Now, research is increasingly based on secondary observations, which are selected and extracted from data archives.

Storing data is easy. Finding and using data later is not. Systematically and consistently organised data do not occur without cost. Consider the results from previous science projects that had no extra effort for data archiving. The natural tendency over time for data and information is chaos. Effort must be exerted to overcome this, although successfully managed data produced by research projects may not be ready to be archived.

Certain assumptions are made about archiving, namely: information-sharing is important; multidisciplinary data access will foster more robust scientific discoveries; archiving can be improved; the 'neurons' of archives are metadata; the current number of permanent data archives will increase; and the Internet creates certain expectations.

Archives need structure and standards. The reasons for archiving include that archiving supports better science. The metadata required for archiving will improve data quality. Archiving thus improves science. Archiving extends data usefulness. Archived data increase the information base for doing research, increasing both data volume and diversity. Proper archives permit the replication of results.

The reasons that researchers do not archive their data include lack of incentives; lack of acknowledgement that producing datasets is a legitimate research output; fear of giving up publication rights, with the associated fear of being scooped by other researchers mining the data; not making provision for data archiving in the research plan; lack of resources to pay for archiving; lack of clarity on who will support the management of the dataset in the future; lack of training; and uncertainties about metadata content and how much metadata is enough.

Most of the information technology in use was built not for research but for business, which does not deal with fundamental and changing datasets over long periods of time.

Designing for change is a significant issue. Many issues will change with respect to archiving in the future, including: access expectations; removal or addition of access restrictions; scope and logical hierarchy of the information; and new parameters, disciplines, study sites, data sources or methods. There will be revisions and additions to metadata codes for parameters, sites and measurements; data values from reprocessing; and updates of hardware and software.

The initial view of data was in terms of measurement from the single experiment view, entailing parameter name, date, sample identification and location. This has evolved through the research project view and the long-term or multidisciplinary view to the integrated system and archive view.

Planning for and decisions about archiving need to be intentional and not accidental. It is easier and more effective to develop metadata before rather than after data measurement. Measurement takes place near the end of the information process, and the only information that is added after the measurement is from the quality assurance results and possibly some detailed records.

Perspectives on metadata depend on one's worldview. Investigators tend to take a project focus and believe that extensive formal metadata are not required. The criteria for metadata differ depending on whether the perspective is that of individual investigators, groups or science teams, project information systems, or data distribution and archive centres of master data directories. Establishing metadata practices that work for individuals, projects, institutions and disciplines remains a challenge.

Selection, appraisal and retention of digital scientific data

Selection, appraisal and retention of records work together. Appraisal is subjective, but must be done by putting a value on the data. Retention leads to dispositions, and may result in transfers, not just destruction. It is critical to institutionalise these processes and imperative to make a start, even if that start is initially very small.

Selection of the records to be preserved is necessary, but also daunting and time-consuming, and may become political.

Appraisal is the bridge between selection and retention. Appraisal is one of the most critical elements of the records management process and the most subjective and judgemental. Alignment with the mission of the organisation must be taken into account, along with agreements with other agencies, organisations or legislative mandates. A collection policy is useful, as a brief statement by the organisation on the collections in which it is interested (related to its mission), as well as what it is not interested in. This may make it easier for an organisation to turn down collections that are irrelevant to its mission. Appraisal must consider the continuing value of the data in terms of secondary uses, science data interdependencies and possible future research use (by predicting the ways in which the data may be valuable to generations to come).

The criteria of different agencies may differ according to particular needs, but the criteria generally relate to the following factors: uniqueness, adequacy of documentation (metadata), availability of hardware to read the data records, cost of replacement, and evaluation by peer review. Complete metadata should define the content, format or representation, structure and context of a dataset. Another issue to address is whether the data really need to be digital. Some agencies consider printing out data, storing the hard copy and destroying the digitised original. Digital data are useful in that words can be indexed and used to search the data, but in the event of a dilemma about the affordability of maintaining the digital archive, reducing it to paper may be the only solution.

In the United States, every federal agency has to have a scientific records schedule, in which records are described and valued as permanent or temporary – if temporary, the time period for retention has to be stipulated. The value of 'temporary' or 'permanent' has to be agreed on with the national archives, which may work to change that value. Once agreement has been

reached, a legal document to that effect is signed by the heads of the agency and the national archives. The data will be destroyed after the set time period, and the cost of destruction has to be budgeted for. There is a communication process prior to destruction, but a strong justification would be needed for revaluation.

The United States Geological Survey (USGS) strategy for digital archiving is to manage three copies of the data, stored respectively near line or online, off-line and off-site, but it currently meets only the first two of these objectives. It is intended to meet the off-site storage objective in 2006.

Questions were raised in the plenary discussion about the cost of data management as a percentage of operating costs. It seemed that 10% was a realistic percentage (or up to 20% if the costs of quality assurance and analysis are included). It was stressed that the cost of data management is a continuous rather than once-off cost.

Oceanographic Data and Information Network for Africa

The challenges for the Oceanographic Data and Information Network (ODINAfrica) include language differences (English, French and Portuguese), severe communication problems, differences in focus of the designated institutions (ranging between education, research and resource management, sometimes with a narrow focus such as oceanography or fisheries, and sometimes with broad coverage as in a university), differences in size and personnel numbers, and differences in skills levels (ranging from basic skills to advanced proficiency).

National research and education networks

Bandwidth is expensive in Africa because Africans, as latecomers to the Internet, bear the full cost of the long-haul connectivity to Europe and the United States. Moreover, many countries still have a single incumbent operator protected by restrictive licence regulations. SADC lacks European Union-type regional authority. Regional bodies that do exist tend to be associations of vested interests, such as telecommunication companies and regulators. Consumer interests are not represented. Cross-border connectivity is tightly regulated for the benefit of the incumbents.

A ‘shareholders' club’ model for financing submarine cables prevents competitive access. Such clubs do not operate as companies: they are private, are not incorporated as accounting entities and are very secretive about costs. South Africa’s second national operator has no landing rights for traffic from the SAT-3 cable. There is healthy competition to provide international connectivity via satellite (V-SAT), but V-SAT bandwidth is inherently much more costly to provide and requires each site to install a large dish (at a cost of US$45 000). Operating prices vary between $2.50 and $4.00 per kb/s (one-way) per month. The system is not feasible for high bandwidths (over 100 Mb/s). The result is that pricing for cable connectivity is not cost-related, and the price is pitched just to compete with V-SAT.

What are the existing resources and mechanisms?

Selection, appraisal and retention of digital scientific data

The United States Geological Survey (USGS) uses the USGS EROS tool for appraising records. The tool uses over 70 categorised questions relating to mission relevancy, general policy (ISO standard), physical media, metadata and cost–benefit.

Oceanographic Data and Information Network for Africa

ODINAfrica (comprising 40 institutions in 25 countries) falls under the International Oceanographic Data and Information Exchange programme (IODE), under the auspices of the United Nations Educational, Scientific and Cultural Organisation (UNESCO) Intergovernmental Oceanographic Commission (IOC). As such, ODINAfrica adheres to the IOC data policy, approved in 2003, which entails timely, free and unrestricted access to all data, associated metadata and products generated under the auspices of IOC programmes. ‘Timely’ relates to the need to be sufficiently rapid to be of value for a given application. ‘Free and unrestricted’ means non-discriminatory and without charge (and

'without charge' means at no more than the cost of reproduction and delivery). For non-IOC programmes, this policy encourages timely, free and unrestricted exchange where this is essential for the preservation of life, or relates to the beneficial public use and protection of the ocean environment, forecasting of weather, operational forecasting of the marine environment, monitoring and modelling of climate, or sustainable development in the marine environment. In relation to the research and education community, this policy encourages timely, free and unrestricted exchange on condition that the products or results of such use shall be published in open literature without delay or restriction.
ODINAfrica operates with the purpose of getting access to data available in other data centres, developing infrastructure for collection, archiving and analysis of data, and developing skills for the manipulation of data and preparation of data and information products. The goal is that each country should have its own priorities, which ODINAfrica will support.

National research and education networks

In the developed world, national research and education networks (NRENs) have arisen to ensure that advanced networking traffic is not disabled by congestion from commodity-type traffic and to develop next-generation networking and applications in research and higher education. In SADC, there are other roles as well. Bandwidth is very expensive. In South Africa, the Tertiary Education Network (TENET) has negotiated a favourable deal with Telkom SA, and is essentially a bandwidth consortium comprising some forty institutions. It negotiates volume discounts and special deals, and lobbies government and regulators (for example, for relaxed V-SAT licence conditions). The TENET network has a direct connection to GéANT, but the bandwidth is too costly and limited to significantly enable participation in collaborative international research projects. TENET also runs a donor-funded programme called Development of IT Capacity in Higher Education (DITCHE), which is aimed at developing ICT knowledge and skills in the local universities.

TENET is a section 21 company, owned by universities, which, together with the public research institutions, are the parties it exists to serve. Each university or research institution appoints TENET as the agent for procurement of its Internet access. TENET is authorised by each of the participating institutions to represent that institution in negotiation for bandwidth and to commit it contractually.

In South Africa, TENET is the bandwidth consortium that arranges general Internet access. The Department of Science and Technology of the South African government is creating an NREN (SANReN) (with the assistance of TENET), which will be a parallel network to TENET, particularly for connection to the global network, driven by the needs of large *e*-science projects. This will be in place in the first quarter of 2006. All TENET institutions will connect, either directly or via a SANReN/TENET gateway. 'Connectable' institutions in other countries are welcome to join.

The Southern African Regional Universities Association (SARUA) initiative, launched in March 2005, comprises 46 public universities in the SADC region (26 have access to SAT-3 and 20 use V-SAT connectivity). SARUA envisages adopting a TENET-like agency model; negotiating a common V-SAT deal for SARUA universities; lobbying, where necessary, for relaxation of V-SAT licence restrictions and/or fees; achieving a shared connection to GéANT/Internet2; and pursuing capacity development programmes throughout the region.

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?

Managing the impacts of programmatic scale and enhancing incentives for data archiving

Incentives to archive data are a management responsibility, which management in many cases may still need to assume. Management must understand the impacts of scale on archiving.

Management needs to promote data archiving in relation to career advancement by giving researchers recognition for archiving, considering scientific journals that also provide companion data publications, and encouraging the notion that archiving will help in conducting research from a broader perspective. Peer pressure should be developed around archiving as good scientific practice.

Professional development is required through training in data management. Daily interaction is needed between scientific and information specialists. Support is needed for long-term stewardship, so that clear responsibility is assigned for the management of the data after the project has been completed.

Other institutional incentives could include requiring data archiving by the project sponsor and ensuring that data archiving is included in the research plan and resources are available to support it, and that archival preparation (metadata development) is interwoven with the scientific planning and publication processes.

Managing change is a significant issue, and consistent 'self-discipline' is still challenging.

The goal is to minimise changes and maximise documentation. Design considerations to be taken into account include: creating 'extensible standards' for metadata; having a process for implementing new standard metadata codes; recording the effective dates of changes; building databases and applications software for change; putting labels in lookup tables (outside the software code); not letting the flexibility needed to store the information become constrained by software that is too complex to be changed; including notification procedures to data users about changes; distributing information to previous data users; and allocating resources for reprocessing.

Graduate schools worldwide still do not teach data management. In the plenary discussion, several opportunities for developing countries to study data management were mentioned. Developing countries could benefit from participating in a two-week training course, soon to be certified, offered by the National Archives of the United States. International candidates are accepted on this course. The National Archives also offers multiple one-, two- and three-day courses. Training is also available from the Society of American Archivists, the International Archives Institute and the American Records Management Institute. Some of the training is via a distance model, which would not require travel to the United States and would be more cost effective than a residential course in the United States.

In the plenary discussion, it was proposed that there is a need to influence policy and legislation, and that the role of CODATA in this regard should be explored. The South African National Archives Act makes no mention of the mandatory collection and preservation of scientific data and information, but refers only to public administrative records. South Africa at least has an Act on national archives, but some other SADC countries bemoan the fact that they have no legislation to control archiving. In such cases, even if people are trained in data preservation, governments are not committed to the collection, appraisal and retention of archives.

The plenary discussion proposed several other ways of influencing attitudes and policy-making. One way is to make known the value of the data and what it would mean to lose the data resource. Another is to showcase examples of records that have been lost that would

have been of scientific value today (for example, at the National Aeronautics and Space Administration [NASA], a manager took it upon himself to maintain records slated for destruction, and they proved a critical factor for later research). Another suggestion was to place a monetary value on the cost of not doing the particular data management activity.

In the plenary discussion, means for dealing with parties that would not share data were investigated, and it was suggested that a 'carrot and stick' approach was needed. For example, a condition of US National Science Foundation grants is that all grantees provide data to repositories. Another way is to get researchers to understand why data management is essential, using scientific logical arguments, namely, that researchers in the modern world use increasing amounts of data that are not theirs, and scientific results based on larger amounts of data will be more robust. Researchers need to understand that if data management planning and scientific planning are interwoven, data archiving is not a huge burden. Researchers need to be able to plan both for collecting their own data and using that of others.

Selection, appraisal and retention of digital scientific data

The USGS EROS Data Center is willing to extend its data migration effort related to the preservation of digital scientific data. The USGS knows land satellite data best, but is willing to work with other data on a case-by-case basis.

National research and education networks

'Connectable' institutions in other countries are welcome to join the South African NREN (SANReN).

Prospects for a SADC regional research and education network (RREN) include two possible options. The first would be a V-SAT bandwidth consortium, with a common V-SAT service provider. Participation would be contingent upon V-SAT licence conditions in each country. The RREN would offer shared connections to the general Internet and to GéANT (or Internet2) from a common base-station in Israel, Europe or the United States. The second option would rely on terrestrial connectivity, which would be more difficult to realise because multiple cross-border links would have to be established; multiple licences in multiple administrations would be required to land and transit traffic; multiple operators would be involved, with no uniform prices in the region; and competitive submarine cable connectivity would be required, which is still some years away. Both options require important capacity development work.

Even V-SAT is not a viable option in some countries, such as South Africa, because of regulations that disallow it or make it very expensive through licence charges. Such issues can only be solved on a country-by-country basis, through lobbying. The SARUA initiative has strong blessing from NEPAD, and could bring considerable influence to bear in discussions with national governments in negotiating favourable terms for research and education.

In 2004, the Partnership for Higher Education in Africa, comprising four United States donor foundations (Carnegie, Ford, MacArthur and Rockefeller) convened a meeting in Dar es Salaam. The message was that bandwidth providers (V-SAT providers) are tired of receiving special requests from universities, as are donors, and that efforts must be coordinated. An important way to do so might be through NREN formation. The only functioning NRENs in East and southern Africa are KENET in Kenya and TENET in South Africa.

Some publishers are connected to their local NREN, while others are not. Publishers could consider lobbying research and education networks for access in order to facilitate the distribution of scholarly publishing.

SESSION 7.1A: HEALTH AND BIOMEDICAL DATA AND INFORMATION

Chair: Prof. Wieland Gevers, Academy of Science of South Africa
Rapporteur: Mr LeRoy Charles, George Washington University, United States

Papers presented

Utilisation of Health Information System (HIS) in Namibia: Challenges and Opportunities Faced by the Health Care Delivery System/Health Sector, *Dr L. Haoses-Gorases, Faculty of Medical and Health Sciences, University of Namibia, Namibia*

Ensuring Access to Data and Information: Capacity Building for Tropical Diseases, *Dr Lester Chitsulo, Tropical Diseases Research, World Health Organisation, Geneva, Switzerland*

The Role of Indigenous Knowledge in Primary Health Care and Sustainable Utilisation of Natural Resources in Malawi: Opportunities and Challenges for Scientific Information Access, *Prof. S. Sosten Chiotha, University of Malawi, Malawi*

Data Sharing: Perspective from the National Institutes of Health, *Dr Belinda Seto, Deputy Director, National Institutes of Biomedical Imaging and Engineering, National Institutes of Health, United States*

Considerations in Designing a National or Regional Microbiological Data Archiving System, *Dr Micah Krichevsky, Bionomics International, United States*

What are the challenges and barriers?

Utilisation of the Health Information System (HIS) in Namibia

The challenge in Namibia is that health information is located in different directorates. The problems include turnover of staff, with implications for training, timeliness at the national level, lack of designated staff at a district level, lack of computer literacy, unavailability of private-sector information, the agendas of development partners, the coordination of the systems, and the involvement of top-level management.

The role of indigenous knowledge in primary health care and sustainable utilisation of natural resources in Malawi

The challenges are related to the limited research in indigenous knowledge; the fact that much valuable information is locked in the oral medium; the rapid loss of biodiversity before characterisation; the loss of local control over and benefits of indigenous knowledge and biodiversity; the impact of HIV/AIDS; and the erosion of the indigenous knowledge base and biodiversity through the harvesting of natural resources.

Challenges to sharing and practice include toxicity from malpractice; the fact that published indigenous knowledge lacks detail about drug combination and administration; and that the validation of indigenous knowledge is mostly at the inventory and basic chemical analyses level (in other words, how to use the plant), but the active ingredients remain unknown.

Data sharing: perspective from the National Institutes of Health

The challenges to data sharing include data sensitivity, the protection of confidentiality and the danger of deductive disclosure. A further challenge is that timely release of data is essential. Data sharing is made more difficult by an adverse environment and the challenges of interoperable technologies.

What are the existing resources and mechanisms?

Utilisation of the Health Information System (HIS) in Namibia

The objectives of the Health Information System are to improve the performance of health care providers, measure the quality and efficiency of government strategies, compare performance over time, provide support to health facilities and monitor trends. The system covers indicators on human resources, population, health facilities and financing. Information comes only from public institutions, rather than the private sector. The legal framework is provided by the Facility Act, and a Health Act is being drafted.

The Health Information System was strengthened and revised in 1994. The initiative included a new design for forms and procedures, the adoption of international standards, computerisation, training and the introduction of standard registers.

International partners such as the United States Agency for International Development (USAID) and United Nations agencies invest in specific programmes. One of the programmes is the prevention of mother-to-child transmission of disease.

Strategies for building up health indicators include surveys and the household census, and information on births and deaths. Health indicators are used for planning, resources allocation, and monitoring and evaluation.

Ensuring access to data and information: capacity building for tropical diseases

The objectives of the Special Programme for Research and Training in Tropical Diseases (TDR) are to support research on and the development of new approaches and tools for preventing, diagnosing, treating and controlling neglected diseases and to strengthen the capacity of countries where the diseases are endemic to undertake the research required. The TDR Disease Strategic Emphases Matrix provides a grid of priorities organised by disease.

TDR awards grants to individuals and institutions, and training grants to young scientists, establishes research networks, and ensures access to scientific literature for the researchers they support through HINARI.

Capacity building initiatives include training in project planning and management; capacity building in research ethics; quality practice in basic biomedical research; good clinical practices; and clinical data management (for which eight centres have been established, one of which is in Ethiopia).

To ensure that data generated from research are widely disseminated, there is support for scientific writing and editing of scientific journals in endemic countries.

TDR creates networks, such as the Multilateral Initiative on Malaria, and enters into partnerships with universities to initiate epidemiology classes and programmes (for instance, with the University of the Witwatersrand).

TDR is establishing data centres to support the collection, analysis and publication of data generated from research so as to have a portal with all information (resources, news and funding) in one place.

The role of indigenous knowledge in primary health care and sustainable utilisation of natural resources in Malawi

The importance of indigenous knowledge is evident where local communities have turned to it to solve problems of public health, food security and preservation of biodiversity. Indigenous practices enhance agro-biodiversity and associated ecosystem health.

Initiatives include enhancing research and publication on indigenous knowledge to improve access to and sharing of knowledge; capacity building in taxonomy; mechanisms for local researchers to access indigenous knowledge publications abroad; utilising advances in digital technology; overcoming the stigma/inferior perception of indigenous knowledge; and popularising indigenous vegetables, fruits and cereals.

Data sharing: perspective from the National Institutes of Health

Researchers comply with the NIH Data Sharing Policy because funding is revoked if they do not. Funding applications must have a data sharing plan, which is reviewed by external peer reviewers. Data must be shared as soon as publication takes place. This policy ensures the timely release of final data for use by other researchers. The NIH pays for data sharing.

Considerations in designing a national or regional microbiological data archiving system

This presentation outlined a planned project (which has not yet been initiated) to design a generic model for a distributed system for capturing, managing and archiving national microbiological data. The components would include commercial laboratories and university collections. The services offered would include a repository of data, a microbial diversity database, data tools, a database of material transfer agreements, access to data, data producer education and user training. The system is aimed at a wide audience comprising the general public, the media, non-governmental organisations, regulators, the commercial sector, students, government laboratories, investigators and educators. Final delivery of information and services must become, and remain, a local responsibility. It would be up to data producers to say what restrictions they want on the data.

Resources for the model would include communication with other stakeholders, the National Federation for Culture Collections and other national federations, as well as the World Federation for Culture Collections.

Implementation would be through a steering committee (with a user as the chair). Other committee members would represent data users and providers, as well as government representatives. The repository would need a secretariat, a mechanism for user feedback, a panel of consultants, coding of data with a controlled vocabulary, user system training, education and hands-on training in microbial coding and data analysis. Both external and internal people would be involved in these functions.

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?

Utilisation of the Health Information System (HIS) in Namibia

Opportunities exist to strengthen and coordinate the system, designate staff at the district level, mobilise the resources of stakeholders, involve policy-makers, match indicators with the National Development Plan, finalise the Facility Act and Health Act, ensure proper utilisation of the system, and make information available in a user-friendly manner.

The role of indigenous knowledge in primary health care and sustainable utilisation of natural resources in Malawi

Opportunities exist for ongoing research on plant and animal taxonomy; building indigenous knowledge; research on medicinal properties; multi-stakeholder dialogue among researchers, doctors and traditional healers; training on conservation, sanitation and dosage; and the establishment of formal associations.

Data sharing: perspective from the National Institutes of Health

The resources of the NIH are available through the National Library of Medicine as a world resource, and the NIH Data Sharing Policy serves as a model for other institutions worldwide.

Considerations in designing a national or regional microbiological data archiving system

The potential exists to apply the model for national microbiological data to other sciences.

Plenary discussion

What are the challenges and barriers?

The challenges exist at many levels, and are related to policy, privacy issues, intellectual property rights, resources, systems and capacity. There are problems of the interoperability of systems between institutions and countries.

There is a general lack of policies for data preservation in SADC. There is a need to create awareness of data issues among high levels of ministries and universities. An added problem at the national level is the heterogeneity of data recording and systems. Moreover, there is fragmentation and lack of clarity on how information is archived and shared across national boundaries.

Data sensitivity relates to protecting patient rights. Part of the knowledge is not necessarily open access for reasons of privacy, and sometimes as a result of intellectual property rights. Reliability of information is often not verified: data may be recorded without laboratory follow-up to confirm the diagnosis. Intellectual property rights of healers that use indigenous knowledge are not always protected. There is a lack of documentation of biodiversity. Such documentation is mostly done by traditional healers. Dilemmas exist with respect to documenting the biodiversity without losing it to international pharmaceutical companies. A large part of the health care system is left out if traditional knowledge is omitted.

There is a general lack of awareness of existing resources, with the associated risk of reinventing the wheel through not knowing what already exists.

Capacity building requires the training and retaining of people.

What are the existing resources and mechanisms?

South Africa publishes a journal of research on indigenous knowledge and traditional healing, namely, *Indilinga.* Examples from other countries could serve as models for developing countries (for example, the United States NIH).

There are very few resources for data management in SADC (or poor visibility for them). Internet resources exist, but there is a lack of awareness about them.

What are some potential realistic projects or collaborations to help address the challenges and barriers that have been identified?

Countries become sentinels for collecting health data, in the way that Centers for Disease Control (CDC) sentinels in the United States must have some health surveillance and recording. Health information systems should be established initially at national level, and then integrated internationally, but there would be challenges to combining these systems. A codified language for collections would be useful. With respect to the institutionalisation of health records, it was noted that in some countries people hold on to their own data in books they keep at home.

The needs are many and varied. Policy needs to be translated into practice. A compendium of the kinds of information being collected in SADC countries is required, as well as a

compendium of sources of knowledge, which are widely distributed, so that people know where to look. Common resourcing would help. Information should also be collected from traditional healers, as they have more direct contact with people and know Africa's biodiversity and its use during the past. Specifically, a data manager is needed in Zambia to manage data from traditional healers.

How can the new or improved initiatives specifically be implemented?

Whom Africa would report health information to is one of the dilemmas to be resolved. Africa could learn from the Chinese model for the integration of traditional medicine information in the national health information system. The aim must be a comprehensive system and a regional network. An international health information system could start between just two countries to work out methods and interoperability issues as well as quality control. This would constitute a pilot test between two counties with mutual interest in such a system, to leverage each other's strengths while fostering a sense of joint ownership. The model could then be instituted across many countries.

SESSION 7.2A: EARTH AND ENVIRONMENTAL DATA AND INFORMATION

Chair: Mr Robert Kriger, National Research Foundation, South Africa
Rapporteur: Dr William Anderson, Praxis 101, United States

Papers presented

The Status of Geospatial Data for Natural Resource Management and Food Security Assessment in the SADC Region, *Mr Tamuka Magadzire, United States Geological Survey/Famine Early Warning System Network and the SADC Regional Remote Sensing Unit, Botswana*

The Southern African Data Centre for Oceanography (SADCO): Regional Facility for Storing and Sharing Marine Environmental Data, *Dr Marten L. Grundlingh, CSIR, Manager of SADCO*

Oceanographic Data and Information: the Need for Regional Collaboration*, Dr Alfonse Dubi, Director, Institute of Marine Sciences, University of Dar es Salaam, Zanzibar, Tanzania*

Building African Capacity to Use Geographic Information for Development Planning and to Influence Sustainable Development Policy Debate, *Mr Jacob Gyamfi-Aidoo and Sivas Govender, Executive Director, EIS-Africa, South Africa*

Remote Sensing, *Mr Andrew Terhorst, Satellite Application Centre, CSIR, South Africa*

Metadata Clearinghouse and Open Access to Geographic Data in Namibia, *Ms Emma Noongo, Ministry of Environment and Tourism, Namibia*

What are the challenges and barriers?

Internet access and network infrastructure generally

Many Internet service providers are not sensitive to the volume of data associated with science and technology. Bandwidth tends to be a serious barrier in many places, even when connectivity is available. The Internet is not the only option for data transfer: consideration should also be given to physical media, but the dependability of the mail is questionable. New portable disks are not very durable (and cannot be viewed as primary storage devices). Sending out physical copies requires version control (and this adds cost and work).

Another primary challenge is capacity. Human capability to use the Internet effectively is varied, and ways need to be found to address this. Education in Internet use is the most useful response. The first step is to ensure that all libraries have Internet access and to help decision-makers understand what technology can do. Internet infrastructure differs substantially among countries (for instance, there are big differences between Tanzania and South Africa). A remark was made that, “Sometimes we’re working as if we’re in the dark.”

Bandwidth and access cannot wait. If we do not make the right connections with politicians for action, then it remains frustrating. Governments know very well that access is needed. Many governments run the networks in their countries, with monopoly-like control of the network, which keeps access costs high.

Political buy-in and political will

Consideration needs to be given to how politicians can be influenced. CODATA and the NRF made the workshop meeting possible, but very few government representatives attended the workshop. CODATA needs to get this discussion into forums in which government representatives are involved.

A language barrier exists between scientists and politicians. Politicians do not have space for uncertainty, while scientists always see the uncertainty. Scientists and data managers need to

learn to talk to politicians in a language they can hear. It is possible to learn to unify the voice and tell a good story.

The political and scientific communities have different priorities:

- The scientific community needs to share data. Sharing requires a certain ethos and can begin when both parties gain (for example, the original requestor makes available data that the provider might want). The political community is interested in improved research and information (and for politicians, their own careers may be their foremost concern). The message to politicians needs to be that improvement in infrastructure is necessary and in their interests. In this regard, initiatives should be regional and local, and should build on existing institutions and capacity.
- Politicians need to see the results and be educated as to the use of data. Initiatives should start by involving politicians known to be open to scientific and technical data sharing in order to build momentum. Politicians need to be guided to consider data as a national resource. This could be fostered by citing and using illustrative studies that quantify the economic benefits (for example, the Chinese data sharing initiative). Issues that already have the attention of politicians should be used – climate change and disaster management are two such examples (and these issues sometimes require access to the same types of data).

Scientists need to sort out the issues and engage the political community. It was mentioned that the South African marine science experience has generally been positive in expressing a success story to the public. It is necessary to capitalise on such experiences.

Data sharing

One issue is not knowing what exists, and the possibility of establishing indices or directories of who holds what data needs to be explored. The NASA Global Change Master Directory is an example of such a directory and a possible model for action.

Sometimes data sharing is best left as a person-to-person interaction. Rules often preclude sharing. Civil servants are bound by regulations, although this is sometimes just used as an excuse.

What matters is the purpose of specific data sharing projects. Custodians need to know the purpose for which data will be used before sharing. Clarity in data needs can help manage restrictions. Policy and guidelines on how to use statistics already exist.

It needs to be considered what it would take to help people share. For instance, would a request for data analysis be treated differently from a request for the data? Data sharing may thus be enhanced by asking for the desired result.

Governance has not kept pace with technology. This is true more generally as Internet use grows, and the challenge is not restricted to developing countries. ICSU and CODATA have a civic duty that impacts on policy and regulation. Science impacts on policy-making. The Task Group should consider a major action that it could implement in the next few years (involving and impacting ICSU, NEPAD, the African Union and the Pan African Council) with respect to governance and regulation (for example, so as to advance the attainment of the Millennium Development Goals).

Data sharing includes knowledge-sharing (What are the data? How to use the data?) and technology-sharing (What analyses and tools are available?). A relationship of reciprocity is required for sharing, and openness is essential with respect to the purpose and use to which shared data will be put.

Data sharing rationales include that the issue is not science for the sake of science but science for society (the ICSU view). Changes of view are needed on the roles of science.

There is frustration at the pace of change. This is a general issue and is not limited to developing countries. The challenge is to make visible the changes that have occurred, as well as the recurring barriers.

What are the existing resources and mechanisms?

Resources

Resources include organisations and information that can be utilised to answer challenges and take action.

Availability of current data

A major issue regarding resources is that we often do not know of the availability of current resources. In this respect, data-producing and data-maintaining institutions within the SADC region are, overall, falling short of working together as a region, and consideration will have to be given to addressing this by probing what is holding things back.

Organisational resources

Numerous organisational resources exist, including CODATA as a resource, the NEPAD Secretariat, the International Oceanographic Data (IOD) ocean science data management forum in Belgium, the ICSU Regional Office, the EIS Africa newsletter and network, the NEPAD 'Mapping Africa for Africa' issues, including the articulation of the core datasets that are needed to address the issues, and GEOSS.

Spatial Data Infrastructure Act

The South African Spatial Data Infrastructure Act makes spatial data freely available, although not necessarily free, as the cost of distribution has to be covered. There is a clearing-house, but it seems that there is a general lack of awareness of availability.

Economic data use example

An example of the use of economic data is that several years ago, the government of Angola lost track of US$2 billion, which was an accounting problem, and was covered up by the government. The source of the problem was agencies from the North and local NGOs in Angola. Sometimes the actual potential partners (and resources) can be overlooked. International agencies may have data that can support specific data requests, and government and big institutions are not always adversaries.

Internet connectivity and access resources

Wireless broadband access will help distribute Internet connectivity. The Sensor Webs initiative is working with GEOSS and EIS.

Initiatives

Mauritius: IT sector development experience

In surveying the current state of ICT capability, the government of Mauritius found a huge gap. The government set up a committee led by top-level government officials and educated the population as to the need.

The government was especially interested in ICT in the education sector and investigated the minimum criteria for IT connections in primary schools. High schools all have IT laboratories. Even housewives were introduced to information technology. Mauritius turned to India for help with cyber infrastructure. The example of Mauritius shows what is possible if there is national commitment. Ultimately, competition took hold.

African marine and coastal management

A series of meetings was held, which served to drive the focus. The initiative kept building on previous results and involved scientists and resource managers.

Tanzania: national Web site

The initiative engaged the President himself and put his speech on the Web site. The result was that the President is now interested in what is on the Web site.

Namibia

Namibia provides an example of a participatory approach in eliciting information regarding data sources and encouraging data sharing.

Data preservation

The United States Geological Survey has offered to copy old media containing scientific and technical data (preferably satellite Earth observation) to newer media on a case-by-case basis.

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?

Building social networks

Building social networks requires being personally proactive, which does not cost much. The participants in the CODATA workshop have already made an investment simply by attending. A possible initiative is to start an *e*-mail circulation among workshop attendees to establish what data we have oversight over. It might be necessary to identify a person or organisation to drive this initiative after the workshop.

Data sharing

As individuals, we need to practice what we preach to our governments with respect to social network actions. There is a need to identify the datasets that exist in the region and build a database of databases, to serve as an Internet portal. The Namibian presentation of visiting data-holders offers a possible model for this work. South Africa used to have a database of environmental databases, 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. It was suggested that a SADC CODATA working group be established around this activity.

Collaborative generation of data will facilitate sharing. The core datasets in the region need to be identified and investigated for consistency. Data policies must be explicit and available. NEPAD and SADC structures could be used to work on revisiting sharing regulations. Projects need to be aligned with existing regional organisations (to avoid establishing new organisations wherever practicable). Knowledge-producing organisations need to reward people for creating and sharing knowledge. It is imperative to be transparent by sharing data, soliciting feedback, and transferring knowledge and learning.

CODATA specific actions

Only four CODATA national committees exist on the African continent. A CODATA or ICSU regional committee should be established to work with the ICSU Regional Office. A working group needs to be formed to take this forward.

Data management academy

The proposal for a data management academy builds on material from the workshop presentations and discussion. The function of such an academy would be to provide training for researchers, managers and government officials, including data management operations training and policy outreach. This is something that CODATA and/or ICSU must promote.

Influence policy decisions

Those with an interest in promoting the cause must learn to lobby politicians so that recommendations are implemented, and help politicians to hear what we want them to hear. This can be facilitated by developing convincing stories about current situations and the importance of science and technology. We must learn to tell the story about sharing to people who do not always understand us, at the appropriate audience level, to convey core data requirements and core issues. The list of core datasets needs to be built. The scientific method should be used to estimate scientific capacity under the current policy versus the different policies (for example, closed versus open access).

We need to learn to express measures of value and performance for data preservation and sharing. Data management measures of performance are associated with distribution (collecting the distribution data) and document successes and failures (both of which are useful). We need to find doors into areas we can impact. The workshop participants are an international group whose potential ought not to be underestimated. CODATA must be a participant, and it needs to be established whether there are any legal questions about how CODATA could do this.

SESSION 7.3A: BIODIVERSITY DATA AND INFORMATION

Chair: Prof. H. Baijnath, University of KwaZulu-Natal and SANC CODATA
Rapporteur: Mr Paul Uhlir, US National Academies, United States

Papers presented

Open Access to Scientific Data on Biological Diversity, *Dr James Edwards, Global Biodiversity Information Facility, Denmark*

SAFRINET to SABIF: Initiatives to Coordinate and Realise Southern Africa's Biodiversity Heritage, *Dr Mervyn Mansell, International Services, Animal and Plant Health Inspection Service, United States of America Department of Agriculture, South Africa*

SIBIS: SANBI's Integrated Biodiversity Information System, *Mr Reuben Roberts, South African National Biodiversity Institute, South Africa*

The South African Environmental Observation Network (SAEON): Delivering Long-term Data for Research and Policy*, Mr Avinash Chuntharpursat, National Research Foundation, South Africa*

The United States Long Term Ecological Research Program: Site and Network Level Information Management, *Prof. Kristin Vanderbilt, Assistant Research Professor and Information Manager, Sevilleta Long Term Ecological Research Site, Department of Biology, University of New Mexico, United States*

What are the challenges and barriers?

Vast amounts of specimen-associated legacy data are incarcerated in museums and herbaria in analogue formats. There are major problems associated with digitising those original data and making them broadly available online by different organisations, including the problems posed by amateur collections. The issues relate to funding, efficient methods, legal concerns with respect to liability, and subsequent protection. Repatriation of information to countries of origin is another issue, mostly from developed countries and institutions to developing countries where the specimens were originally collected.

Data quality assurance needs to be carried out. There is a lack of a comprehensive taxonomic (and common name) authority file or catalogue, including in particular in southern Africa.

Data are disjointed and duplicated, and insufficient interoperability exists (both technical and semantic) among organisations. There is also insufficient prioritisation to identify data to be digitised, catalogued or made freely available. There is a general lack of primary data resources to support practical applications and policy activities.

A shortage of skilled taxonomists and systematists is a worldwide reality, including in southern Africa, and 'ivory tower' attitudes of practitioners do not help matters. The training for these professions also is not very good, but moreover there is great inequality of capacity between countries in the SADC region.

Questions exist about the sustainability of data activities over the long term, given the shortage of funds in the southern African region. This situation is exacerbated by insufficient information about SADC country institutions and people working in these subject areas online, which makes it difficult for foreign scientists and institutions to identify opportunities and targets for collaboration in the region.

What are the existing resources and mechanisms?

Southern Africa is the third most biologically diverse region in the world, which presents many potential opportunities. The biodiversity and ecological characteristics of the region are a

major source of economic and other benefits (such as ecotourism, bioprospecting and advancing basic biological knowledge). Biodiversity and ecological databases are essential for supporting these important applications.

Several existing organisations that work on biodiversity and ecological data and information preservation and access activities were represented or discussed at the workshop. At the international level, these include the Global Biodiversity Information Facility (GBIF) and the International Long-Term Ecological Research Network (ILTER). At the African regional and national levels, the workshop included representatives from the following organisations and initiatives: the Southern African Network of BioNET-International (SAFRINET) and several South African national initiatives – the South African Biosystematics Initiative (SABI), the South African National Biodiversity Institute (SANBI) and its Integrated Biodiversity Information System (SIBIS), the Southern African Botanical Diversity Network (SABONET), the South African Biodiversity Initiative (SABIF), which is a national data node of GBIF, the South African Environmental Observation Network (SAEON), which is linked closely to the ILTER, the South African Integrated Spatial Information System, the Marine Ecosystem Programme, BioMove, the Collaborative Spatial Analysis and Modelling Platform (CoSAMP), and the Environmental Long-term Observatories Network of Southern Africa (ELTOSA).

Other important organisations that have relevant programmes or that might provide funding include:

- At the inter-governmental level, the Convention on Biological Diversity Global Strategy Plan for Conservation (to develop a complete plant checklist by 2010) and the Global Taxonomy Initiative (which is just getting started), the World Bank/GEF Botanical and Zoological Network in East Africa, as well as the International African Fruit Fly Initiative.
- Other United Nations specialised agencies that could play a role include the Food and Agriculture Organisation (FAO), the United Nations Environment Programme (UNEP) and the United Nations Educational, Scientific and Cultural Organisation (UNESCO).
- South African government organisations include the Department of Science and Technology, the National Research Foundation, the Department of Environmental Affairs and Tourism, the Agricultural Research Council, the nature conservation agencies, the CSIR, SAEOS, and the Technology and Human Resources for Industry Programme (THRIP) for public–private partnerships in commercialising research.
- Government organisations in other SADC countries would need to be explored, and include other African organisations at regional and national levels, non-governmental organisations (which generally play a development role rather than a research role, and may have relevant data holdings and possible support) and numerous universities.
- Private foundations include the Mellon Foundation African Plant Type Initiative, the Sloan Foundation (African Ocean Biological Information System [AfOBIS]) and many other foundations that can potentially provide funding.
- United States government agencies include the National Science Foundation Partnership for Enhancing Expertise in Taxonomy and various cooperative research programmes, as well as the Agency for International Development (which in past years formed groups on International Cooperation in Biodiversity, although this is no longer a focus).

What are some potential realistic projects or collaborations to help address the challenges and barriers that have been identified?

Several initiatives and inputs are required to improve data quality:

- There is a need to develop and broadly implement data cleaning tools (for example, see the 2005 GBIF report). Training is required to improve data management and related applications. Online training material would be useful. The 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, Long Term Ecological Research (LTER) and other organisations are possible. ODINAfrica (funded by Belgium) also organises workshops on data management.

- Sharing and linking many kinds of primary data resources (species, specimen and molecular biology) with other related data and information would improve research and applications (see, for example, a report by Arthur Chapman).
- Demonstration projects would show the value of biodiversity and ecological databases and research at the local and regional level to help stimulate greater appreciation of such work by decision-makers and funding sources.
- Collaborative research projects and programmes would help as well.
- Biodiversity and ecosystem data must be coordinated with the Global Earth Observation System of Systems (GEOSS) in South Africa and the Marine Science Remote Sensing Data Centre in South Africa.
- Federations of cultural collections are organised in other regions, but such a federation is needed for the SADC region.
- 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.
- A metadata inventory of information resources to identify gaps and cross-link opportunities is being developed by SAEON.

The GBIF Governing Board is meeting in Cape Town early in April 2006 and convening a symposium on the use of biodiversity data for various applications, with a focus on southern Africa. Other meetings will probably be held in conjunction with the GBIF meeting, possibly on bar-coding of life and a new JRS Foundation to bring African scientists together on biodiversity informatics, education and funding initiatives for developing countries. Funding will be available to bring African scientists to these meetings, and individuals should be nominated. GBIF will have information about these meetings online later in 2005. GBIF is also developing a language for funding agencies to require plans for data sharing and maintenance, similar to the LTER data policy model.

Finally, various scientific 'information commons' initiatives are being established worldwide, including some specific to the topical area of conservation commons. The Southern African Millennium Assessment (2001–2005) provides another opportunity for promoting work in this area.

SESSION 7.4A: SCIENTIFIC, TECHNICAL AND MEDICAL JOURNALS

Chair: Prof. Theo Bothma, University of Pretoria, South Africa
Rapporteur: Mr Roy Page-Shipp, SARIS, South Africa

Papers presented

African Journals Online (AJOL), *Ms Margaret Crampton, National Information Services Corporation (NISC), South Africa*

HINARI and AGORA Enhancing Access to STM Literature in Southern Africa, *Mr Gracian Chimwaza, HINARI, TEEAL, AGORA Office, Zimbabwe*

Open Access and Capacity Building in ICT4D, *Ms Ann Tothill, Association for Progressive Communications (APC), Australia and South Africa*

Making a Difference: Open Access Publishing and Effective African Research Dissemination, *Ms Eve Gray, Eve Gray and Associates, Strategic Publishing Solutions, South Africa*

Towards the Implementation of Open Access Policies in Southern Africa, *Ms Heather Ford, Creative Commons South Africa*

What are the challenges and barriers?

The challenges and barriers were seen to fall into two categories:

Category 1 – Exogenous factors

These are little subject to research community influence, and include:

- General resource deficiencies in Africa, arising from poverty, where too little has been spent on research infrastructure and resources for too long. This is frequently seen as mainly an issue of ICT infrastructure and connectivity, but well-staffed libraries and publishing facilities are also lacking.
- The global information imbalance, where African output has difficulty in achieving global visibility, and African researchers are not readily able to tap into the stream of global outputs. There are even examples where African journals, devoted to African research, are published abroad. This leaves Africa as a disadvantaged partner.

Category 2 – Endogenous factors

These can be influenced by the research community and include:

- A common need for visibility of African outputs, by strategically managed publication and improved access to resources through intelligent marshalling of negotiation power, whether political or financial. (Strategic management of these issues can ameliorate the global information imbalance.)
- All players in the research system, from researchers to administrators, government and funding agencies, display a lack of awareness of the realities of the dynamics relating to both permanent access and one of the emerging alternative modalities of publication, namely, open access**.** The main issues are:
    - Recognition that the dissemination of research outputs, including re-usable data, is a line function of the institution, and that in a developing country, the social science interpretation of more technologically focused outputs is essential if assimilation and implementation are to be achieved.
    - Acknowledgement that the cost of dissemination, including original publication, is a real research cost that must be budgeted for.
    - Unrealistic expectations with respect to the potential income from institutional publishing, which can rarely break even, let alone turn a profit.

  - The importance of persistent archiving of digital outputs (documents, data, models and graphics) beyond the lifetime of the project that created them. Where this also includes the marketing of the data and training of potential users in accessing them, this integrated activity is sometimes called 'digital curation'. The ongoing nature of this expense, which recurs as archiving software and hardware evolve, must also be recognised.
  - Editorial and peer review functions and the skills and processes that go towards efficiency and effectiveness, not to mention the costs.
  - Publishing costs, which arise in open access, albeit in different form from print and online commercial publishing.
  - Legal issues, especially those relating to copyright. In this respect, the researchers' naïve expectations of the commercial value of the outputs and ignorance of the realities, especially of such aids as creative commons licensing, are a major problem.
- These factors create resistance to change.

What are the existing resources and mechanisms?

Always, at the top of the list, must be the many committed and enterprising individuals in the system. They have created many convincing case studies (HSRC, Council for the Development of Social Science Research In Africa [CODESRIA]), and others are available from overseas sources (arXiv).

In principle, the online journals from African sources (for example, the Southern African Bibliographic Information Network [SABINET] *e*-journals and African Journals Online [AJOL]) can achieve global visibility if they fall within the reach of indexing and harvesting systems.

Access support systems, including INASP and various national site-licensing initiatives, such as the South Africa Site Licensing Initiative (SASLI), make an ongoing contribution, but more effort is required to increase usage by marketing and training, especially in institutions where specialised staff are at a premium. Pay per view options exist, where the individual researcher can track down an item via widespread publication of abstracts and purchase any particular article by download. (The British Library is a supplier of last resort.) Open-source software, open platforms and open standards all have a role to play, as do creative commons licences.

Parallel investment in ICT, which is not owned by the publishing activity, but which must be seen as part of a synergistic whole, especially considering huge investments in ICT, is currently being made in Africa.

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?

Policy interventions among funders (including governments) and institutions should mandate that scholars make pre-prints and *e*-prints of their research available via an open access medium, and report accordingly. Curation of outputs should also be mandatory – in the interests of the producers, as individual researchers, but also in recognition of the shared character of the global research enterprise.

Marketing and promotion is needed by researchers and managers of the real challenges as well as solutions, preferably via the burgeoning store of success stories. This should be supported by training programmes that could also use an open-access approach.

Creation of aggregation facilities is needed, where individual publications, or the output of small subgroups of scientists, could be cost-effectively produced to high quality and curated. This would support the need for more digitisation of African material. The facilities could include institutional and national repositories, the latter absorbing the function of archiving national heritage items as well as being the publisher of last resort for isolated researchers. They would provide quality support functions such as selection, appraisal and retention, as well as widespread publication of abstracts.

SESSION 8: VISIONS OF THE FUTURE OF ICSU AND CODATA

Chair: Prof. Steve Rossouw, South African National Committee for CODATA, South Africa

Paper presented

CODATA's Future Role in Permanent Access to Scientific Information Resources, *Prof. Liu Chuang, Chinese Academy of Sciences, and Co-chair, CODATA Preservation Task Group*

What are the existing resources and mechanisms?

The CODATA Task Group on Preservation of and Access to Scientific and Technical Data in Developing Countries was approved by the CODATA 23rd General Assembly in Montreal in 2002, and renewal was approved by the CODATA General Assembly in Berlin in 2004. The strategies and critical issues that the Task Group should focus on were identified as promoting a deeper understanding of the needs of developing countries with regard to long-term preservation, archiving and access to scientific and technical data; advancing the development and adoption of improved scientific and technical data archiving procedures, technologies, standards and policies, with special attention to such efforts in developing countries; providing an interdisciplinary forum and mechanisms for exchanging information about scientific and technical data archiving requirements and activities, with particular focus on the concerns and needs of developing countries; and publishing and disseminating broadly the results of such efforts.

In the implementation phase from 2002–2005, the Task Group cooperated closely with international organisations, the National Committees for CODATA, the sponsors and decision-makers, as well as the community of users. The organisations include CODATA, ICSU, UNESCO, the International Council for Scientific and Technical Information (ICSTI), INASP, OECD, the International Development Research Centre (IDRC), the US National Academies, the Chinese Academy of Science, the South African National Research Foundation (NRF), the Ministry of Science and Technology of China, the National Natural Science Foundation of China, the United States National Science Foundation (NSF), the United States National Institutes of Health (NIH), the South African Department of Science and Technology, the Open Society Institute, Statistics South Africa, the United States National Weather Service and the Chinese Association for Science and Technology.

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 most important progress in the implementation of the activities of the Task Group is to identify the gaps between the scientific data-holders and users, scientific communities and decision-makers, national and international organisations, and industry and developing countries, and to create bridges between them through a series of workshops.

The near and mid-term objectives of the CODATA Task Group on Preservation of and Access to Scientific and Technical Data in Developing Countries are as follows:

- Continue to play a bridging role in reducing the digital divide of scientific and technical data in collaboration with CODATA national committees, ICSU bodies and relevant regional and international organisations
- Continue to play 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.
- Continue to play a special role in developing the Task Group's Web site, publications and information network and making them more timely and effective
- Play 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.

APPENDIX 1: WORKSHOP PARTICIPANTS

Anderson, Dr William	Praxis101, United States - Rye, New York	band@praxis101.com
Arnold, Mrs Robyn	Write Connection CC	writeconnect@intekom.co.za
Arnold, Mr Trevor	South African National Biodiversity Institute	Arnold@sanbi.org
Baijnath, Prof. Himansu	University of KwaZulu-Natal	baijnathh@ukzn.ac.za
Balstad, Dr Roberta	Center for Earth Science Information Network (CIESIN), United States - Columbia University, New York	roberta@ciesin.columbia.edu
Becker, Ms Michelle	National Research Foundation	michelleb@nrf.ac.za
Bothma, Prof. Theo	University of Pretoria	Theo.bothma@up.ac.za
Busgeeth, Ms Karishma	Medical Research Council	Kbusgeeth@mrc.ac.za
Charles, Prof. LeRoy	George Washington University	lcharles@gwu.edu
Chimwaza, Mr Gracian	Essential Electronic Agricultural Library (TEEAL), Zimbabwe	gracian@cyberplexafrica.com
Chiotha, Prof. Sosten	University of Malawi, Malawi	schiotha@chanco.unima.mw
Chitsulo, Dr Lester	World Health Organisation, Switzerland	chitsulol@who.int
Chuntharpursat, Mr Avinash	SAEON, South Africa	avinash@saeon.ac.za
Coetzee, Ms Laurette	Open Spatial Solutions	lcoetzee@openspatial.co.za
Crampton, Mrs Margaret	National Information Services Corporation, South Africa	info@nisc.co.za; liza@nisc.co.za
De Beer, Ms Jennifer*	University of Stellenbosch, South Africa	jenniferdebeer@gmail.com
Dookhit, Mr Rambrich	Mauritius Ministry of Education and Science Research	rdookhit@mail.gov.mu
Dubi, Dr Allfonse	Institute of Marine Sciences, Tanzania	director@ims.udsm.ac.tz
Edwards, Dr James	Global Biodiversity Information Facility, Denmark	jedwards@gbif.org
Jiyane, Ms Velile	University of Zululand	gvjiyane@pan.uzulu.ac.za
Faundeen, Mr John	United States Geological Survey's EROS Data Center (USGS/EROS), United States - Sioux Falls, South Dakota	faundeen@usgs.gov
Fatoyinbo, Ms Temilola	University of Virginia	Tef3p@virginia.edu
Ford, Ms Heather	Creative Commons South Africa, South Africa	hfordsa@gmail.com
Franklin, Ms Amy	The National Academies, U.S.	afranklin@nas.edu
Gevers, Prof. Wieland	Academy of Science of South Africa, South Africa	wieland@telkomsa.net
Glaser, Ms Robyn	SA Department of Science and Technology	
Govender, Mr Sives	EIS-Africa, South Africa	SGovender@eis-africa.org
Gray, Ms Eve	Eve Gray and Associates, South Africa	egray@icon.co.za
Grundlingh, Dr Marten	CSIR, South Africa	MGrundli@csir.co.za
Gyamfi-Aidoo, Mr Jacob	EIS-Africa, South Africa	gyamfi@worldonline.co.za
Haoses-Gorases, Dr L.	University of Namibia, Namibia	lhaoses@unam.na; iisaacs@unam.na
Hammes, Mrs Monica	University of Pretoria	monica.hammes@up.ac.za
Heunis, Ms Marilene	SEFGIS	marilene@sefsa.co.za
Kahn, Prof. Michael	Human Sciences Research Council	mkahn@hsrc.ac.za
Khaole, Mr Gaabi	National Development Agency	
Krichevsky, Dr Micah	Bionomics International, United States - Rockville, Maryland	micahk@earthlink.net; micahk@bioint.org
Kriger, Ms Robert	National Research Foundation	rskriger@nrf.ac.za
Le Roux, Ms Renee	National Research Foundation, South Africa	renee@nrf.ac.za
Liu, Prof. Chuang	Chinese Academy of Sciences, China	lchuang@igsnrr.ac.cn
Maepa, Dr Eddy	National Research Foundation	eddy@nrf.ac.za

Mafuta, Mr Clever	Southern African Research and Documentation Centre, Zimbabwe	cmafuta@sardc.net
Magadzire, Mr Tamuka	United States Geological Survey/Famine Early Warning System Network, SADC Remote Sensing Unit, Botswana	TMagadzire@fews.net
Magubane, Mr Phumelele	SA Department of Science and Technology	Phumelele.magubane@dst.gov.za
Makbel, Mr Ahmed	National Bureau of Statistics, Tanzania	amakbel@yahoo.com
Makhubela, Dr Lulama	National Development Agency, South Africa	lulamaM@nda.org.za
Makhura, Mr Moss	University of Johannesburg	mossm@sanparks.org
Malherbe, Ms Carina	Transvaal Museum	tersiap@yahoo.com
Mansell, Dr Mervyn	United States Animal and Plant Health Inspection Service, South Africa	Mervyn.W.Mansell@aphis.usda.gov
Martin, Dr Duncan	Tertiary Education Network, South Africa	ceo@tenet.ac.za
Masamba, Dr Wellington*	Openheimer Research Institute, Okavango Delta, Botswana	wmasamba@orc.ub.bw; hoorc@orc.info.bw
Mati, Dr Xola	Academy of Science of South Africa	xola@assaf.nrf.ac.za
Maubane, Mr Huddleston	National Development Agency	
Mbambo, Dr Buhle	University of Zimbabwe, Zimbabwe	bmbambo@uzlib.uz.ac.zw; psmart@inasp.info
McCord, Dr Raymond	Oak Ridge National Laboratory (ORNL), United States - Oak Ridge, Tennessee	mccordra@ornl.gov
Mecenero, Dr Silvia	University of Cape Town	smecener@adu.uct.ac.za
Mladenov, Ms Natalie	University of Virginia	Nm6w@virginia.edu
Mniki, Ms Noxolo	National Information Services Corporation	info@nisc.co.za
Mohoto, Mr Themba	Statistics South Africa	thembaM@statssa.gov.za
Mokhele, Dr Khotso	National Research Foundation	khotso@nrf.ac.za
Morris, Mr Chris	CSIR	cmorris@csir.co.za
Mosimege, Dr Mogege	Department of Science and Technology, South Africa	Mogege.Mosimege@dst.gov.za
Muhongo, Prof. Sospeter M.	ISCU - Regional Office for Africa, South Africa	s.muhongo@icsu-africa.org
Musungwa, Flt Lt Washington	Research Council Zimbabwe	blandina@rcz.ac.zw
Mutale, Mr Emmanuel	National Science and Technology Council, Zambia	nstc@zamnet.zam
Naidoo, Mr Dhesigen	Department of Science and Technology, South Africa	Dhesigen.Naidoo@dst.gov.za
Ngulube, Dr Patrick	University of KwaZulu-Natal, Pietermaritzburg	ngulubep@ukzn.ac.za
Nicholson, Ms Denise	University of the Witwatersrand	Nicholson.d@library.wits.ac.za
Noongo, Ms Emma N	Ministry of Environment and Tourism, Namibia	noongo@cc.joensuu.fi
Odido, Mr Mika	Intergovernmental Oceanographic Commission of UNESCO, Tanzania	m.odido@odinafrica.net; m.odido@unesco.org
Olsen, Prof. Gary	University of Michigan (UMICH), United States - Ann Arbor, Michigan	gmo@umich.edu
Page-Shipp, Dr Roy	Academy of Science of South Africa	royconps@yebo.co.za
Patanapanich, Dr Warunee	National Center for Genetic Engineering and Biotechnology, Thailand	warunee@biotec.or.th
Perregil, Ms Tersia	Transvaal Museum	tersiap@yahoo.com
Pienaar, Dr Heila	University of Pretoria	heila.pienaar@up.ac.za
Pillay, Mr Dechlan	CSIR, South Africa	dlpillay@csir.co.za
Roberts, Mr Reuben	South African National Biodiversity Institute, South Africa	Roberts@SANBI.org
Rong, Dr Isabel	Agricultural Research Council	rongl@arc.agric.za

Rossouw, Prof. Steve	CODATA, South African National Committee	steveros@iafrica.com
Salojee, Mr Imraan	Department of Science and Technology, South Africa	imraan.Salojee@dst.gov.za; chanel.isaacs@dst.gov.za
Selematsela, Ms Diasy	University of South Africa	selemsd@unisa.ac.za
Seto, Dr Belinda	National Institutes of Health (NIH), United States - Bethesda, Maryland	setoB@mail.nih.gov
Sharif, Mr Raed	Syracuse University	rmalshar@mailbox.syr.edu
Smit, Mr Arnoldus	Council of Geoscience	leslies@geoscience.org.za
Smith, Ms Michelle	Council of Geoscience	michelle@geoscience.org.za
Steenkamp, Ms Yolande	South African National Biodiversity Institute	Steenkamp@sanbi.org
Strachen, Mr Leslie	Council of Geoscience	leslies@geoscience.org.za
Sooryamoorthy, Prof. Radhamany	University of KwaZulu-Natal, Durban	Sooryamoorthyr@ukzn.ac.za
Suddhoo, Dr Arjoon*	Mauritius Research Council, Mauritius	mrc@intnet.mu
Teng, Mr Mianzhen	China Ministry of Science and Technology	lchuang@igsnrr.ac.cn
Teng-Zeng, Dr Frank	University of Stellenbosch	fktz@sun.ac.za
Terblanche, Mrs Hettie	National Research Foundation	hterblanche@nrf.ac.za
Terhorst, Mr Andrew	CSIR, South Africa	aterhorst@csir.co.za
Tothill, Ms Ann	Association for Progressive Communications, South Africa	AATothill@apc.org; Ann.Tothill@pixie.co.za
Tsebe, Ms Phindiwe	National Research Foundation	Phindiwe@nrf.ac.za
Uhlir, Mr Paul	The National Academies, United States	puhlir@nas.edu
Van der Berg, Mrs Henda	National Research Foundation	henda@nrf.ac.za
Vanderbilt, Prof. Kristen	University of New Mexico (UNM), United States - Albuquerque, New Mexico	vanderbi@sevilleta.unm.edu
Van Rooyen, Ms Marna	National Research Foundation	marna@nrf.ac.za
Veldsman, Mrs Susan	COSALC, South Africa	sasli@cosalc.ac.za
Walwyn, Dr Dave	CSIR	dwalwyn@csir.ac.za
Zimmerman, Dr Ann	University of Michigan (UMICH), United States - Ann Arbor, Michigan	asz@si.umich.edu

* Paper forms part of the proceedings, although the author did not attend the workshop.

APPENDIX 2: LIST OF ACRONYMS

AEIN	Africa Environment Information Network
AfOBIS	African Ocean Biological Information System
AGORA	Access to Global Online Research in Agriculture
AIDS	Acquired immune deficiency syndrome
ARC	Agricultural Research Council
AREN	Australian Research and Education Network
ASSAf	Academy of Science of South Africa
BioNET	BioNET International
CBD	Convention on Biological Diversity
CDC	Centers for Disease Control
CEO	Chief executive officer
CHE	Council on Higher Education
CODATA	Committee on Data for Science and Technology
CODESRIA	Council for the Development of Social Science Research In Africa
COSALC	Coalition of South Africa Library Consortia
CoSAMP	Collaborative Spatial Analysis and Modelling Platform
CSIR	Council for Scientific and Industrial Research
CSO	Civil society organisation
D&I	Data and information
DEAT	Department of Environmental Affairs and Tourism
DISA	Digital Imaging South Africa
DITCHE	Development of IT Capacity in Higher Education
DSD	Department of Social Development
DST	Department of Science and Technology
EIFL	Electronic Information for Libraries
EIS	Environmental information system
ELMS	Environment and land management sector
ELTOSA	Environmental Long-term Observatories Network of Southern Africa
EU	European Union
FAME	Forum of Editors of African Medical Journals
FAO	Food and Agriculture Organisation
GBIF	Global Biodiversity Information Facility
GDP	Gross domestic product
GEF	Global Environment Facility
GEOSS	Global Earth Observation System of Systems
GERD	Gross expenditure on R&D
GIS	Geographical information system
GNI	Gross national income
HDI	Human development index
HINARI	Health InterNetwork Access to Research Initiative
HIS	Health Information System
HIV	Human immunodeficiency virus
HPP	HIV pathogenesis programme
HRD	Human resources development
HSRC	Human Sciences Research Council
IARC	International Aids Research Collaboratory
IBSS	International Bibliography of the Social Sciences
ICSTI	International Council for Scientific and Technical Information
ICSU	International Council for Science
ICT	Information and communication technology
IDRC	International Development Research Centre
IGC	Intergovernmental committee
IKS	Indigenous knowledge systems
ILTER	International Long-Term Ecological Research Network
IMERCSA	I Musokotwane Environment Resource Centre for Southern Africa
INASP	International Network for the Availability of Scientific Publications
INFOTERRA	Global Environmental Information Exchange Network
IOC	Intergovernmental Oceanographic Commission

IOD	International Oceanographic Data
IODE	International Oceanographic Data and Information Exchange programme
IP	Intellectual property
IP	Internet protocol
IPR	Intellectual property rights
ISI	Institute of Scientific Information
ISO	International Organisation for Standardisation
IT	Information technology
IUCN	World Conservation Union
kb/s	kilobytes per second
LTER	Long Term Ecological Research
Mb/s	Megabytes per second
MDG	Millennium Development Goal
MRC	Medical Research Council
NACI	National Council on Innovation
NASA	National Aeronautics and Space Administration
NDA	National Development Agency
NEPAD	New Partnership for Africa's Development
NGO	Non-governmental organisation
NIH	National Institutes for Health
NISC	National Inquiry Services Centre
NPO	Non-profit organisation
NREN	National research and education network
NRF	National Research Foundation
NSF	National Science Foundation
ODIN	Oceanographic Data and Information Network
OECD	Organisation for Economic Cooperation and Development
PBR	Plant breeders' rights
PERI	Programme for the Enhancement of Research Information
R&D	Research and development
RBIS	Regional Biodiversity Information Systems
RISDP	Regional Information Strategic Development Plan
ROSA	Regional Office for Southern Africa
RREN	Regional research and education network
SABI	South African Biosystematics Initiative
SABIF	South African Biodiversity Information Facility
SABINET	Southern African Bibliographic Information Network
SABONET	Southern African Botanical Diversity Network
SABSP	Southern African Biodiversity Support Programme
SADA	South African Data Archive
SADC	Southern African Development Community
SAEON	South African Environmental Observation Network
SAEOS	South African Earth Observation System
SAFRINET	Southern African network of BioNET-International
SAIKSDL	South African Indigenous Knowledge Systems Digital Library
SA-ISIS	South African Integrated Spatial Information System
SANBI	South African National Biodiversity Institute
SANReN	South African National Research Network
SANC CODATA	South African National Committee for CODATA
SARDC	Southern African Research and Documentation Centre
SARIS	South African Research Information Services
SARS	South African Revenue Service
SARUA	Southern African Regional Universities Association
SASIN	Southern African Sub-regional INFOTERRA Network
SASLI	South Africa Site Licensing Initiative
SciDIF	Scientific Data and Information Forum
SEEK	Science Environment for Ecological Knowledge
SETES	SADC EIS Training and Education Initiative
SETU	SADC EIS Technical Unit
SIBIS	South African Integrated Biodiversity Information System

TCE	Traditional cultural expression
TDR	Special Programme for Research and Training in Tropical Diseases
TENET	Tertiary Education Network
THRIP	Technology and Human Resources for Industry Programme
TWNSO	Third World Network of Scientific Organisations
UN	United Nations
UNEP	United Nations Environmental Programme
UNESCO	United Nations Educational, Scientific and Cultural Organisation
UNITAR	United Nations Institute for Training and Research
US/USA	United States of America
USAID	United States Agency for International Development
USGS	United States Geological Survey
WHO	World Health Organisation
WIPO	World Intellectual Property Organisation
WRD	SADC Water Resource Database
WSIS	World Summit on the Information Society
WSSD	World Summit on Sustainable Development