Scientific Committee On Antarctic Research (SCAR)

Data and Information Management Strategy (DIMS)

2009 - 2013

Prepared by
Kim Finney (Manager, Australian Antarctic Data Centre)
Chief Officer, SCAR Standing Committee on Antarctic Data Management
Australian Antarctic Division
kim.finney@aad.gov.au

Edited by
SCAR Ad-hoc Group On Data Management
Colin Summerhayes (SCAR Executive Director)
Chuck Kennicutt (SCAR President)

June 2009

Table of Contents

Page No.

Executive Summary 2

1.0 Strategic Vision 4

2.0 How Can This Vision Be Realised 9

2.1 Data Management Imperatives 9

2.2 SCAR Institutional Arrangements 11

2.2.1 SCADM 11
2.2.2 SC-AGI 13

2.3 Current Approaches To SCAR Data Management 14

2.3.1 SCADM and SC-AGI Lead Data Management Activities 14
2.3.2 SCAR Science Lead Data Management Activities 18
2.3.3 SCAR Data Management and the ATS 19

2.4 Strategic Recommendations 20

2.4.1 Policy, Leadership, Coordination and Governance 20
2.4.2 Cultural Change and Incentives 24
2.4.3 Leveraging Resources and Systems 25
2.4.4 Standards and Interoperability 27
2.4.5 Outreach and Guidance 30

3.0 Next Steps 33

References 34

Glossary of Terms 35

Appendix A - Draft SCAR Data Policy 36

Executive Summary

The Scientific Committee on Antarctic Research (SCAR) - an inter-disciplinary scientific body of the International Council of Science (ICSU) - initiates, develops and coordinates high quality, international scientific research in the Antarctic and on the role of the Antarctic region in the Earth system. SCAR also provides objective and independent scientific advice to the Antarctic Treaty System Consultative Parties and other organizations on issues of science and conservation affecting the management of Antarctica and the Southern Ocean. In support of this dual mission, SCAR has been developing capacity for international data management amongst its member nations since 1992. In order to effectively implement programs, policies and procedures, and activities to fulfil these roles, SCAR must clearly map out a Data and Information Management Strategy for the future. This document is the next logical step in this process.

The committee in SCAR responsible for all aspects of data and information management is the Standing Committee on Antarctic Data management (SCADM). Data and information are valuable and irreplaceable resources. In the pursuit of many science objectives (especially those of a pan-Antarctic nature) it is necessary to use data and information collected by scientists from many countries. SCAR recognizes the critical and essential importance of the stewardship of data and information within national and international programs and of its accessibility by the international Antarctic scientific community. This management is not an "add-on" or an additional task. It is a fundamental aspect of modern earth system science and essential to addressing complex questions about how our planet works and how it will respond in the future.

This Strategy's vision is to build an Antarctic Data Management System (ADMS), capable of supporting inter-disciplinary Antarctic science and SCAR activities within the Antarctic Treaty System. The ADMS should be viewed as a science enabler. Through a range of individual activities SCAR is already making progress towards achieving this vision. But much more can be achieved. The likelihood of realising the desired goal will be greater if appropriate strategic foundations are put in place to enable better coordination of individual and often disconnected efforts. These strategic foundations should encompass:

a. **Policy, Leadership, Coordination and Governance:** better articulated governance arrangements and strong leadership, suitable for driving the development of a distributed, but loosely federated, shared infrastructure. This requires development of a SCAR Data Policy that stipulates the norms that SCAR members should adopt with respect to data sharing and access; data management planning; and establishment of National Antarctic Data Centres (NADCs). Recognising that dedicated leadership is essential for driving development of any shared infrastructure, SCAR members should consider seconding appropriately trained professionals to the SCAR Secretariat and/or assist with raising external funds to support infrastructure development positions. To strengthen existing components of the ADMS, opportunities for partnering arrangements should be explored between SCAR data management groups and those institutions involved in the reformation of the Intergovernmental Oceanographic Data Exchange (IODE) and ICSU World

Data Centre Systems. If the ADMS ultimately expands more through partnerships with these types of global systems than through an expansion of the SCAR NADC network, it may then be prudent to review the role, membership and function of SCADM.

b. **Cultural Change and Incentives:** fostering a culture willing to share and collaborate on data management related activities. Data sharing between SCAR scientists is highly patchy both within and between member countries. Data citation systems are being touted as a mechanism to foster improved data sharing practices between scientists. The Scientific Committee on Oceanic Research (SCOR) has been trialling approaches to data citation. SCAR could formally partner with SCOR in piloting such a system within its NADCs. More could also be done to build an ADMS and to change cultural practices if SCAR's peak data groups harnessed their collective capabilities to garner funding from external sources. Additionally, more money would be available for scientific data management if SCAR educated funding sources about the need for data management to be an explicitly funded component of supported projects.
c. **Leveraging Resources and Systems:** leveraging existing SCAR and non-SCAR systems, capabilities and resources and supplementing these where there are obvious deficiencies (the primary purpose of such leveraging is to create a network of designated permanent data archives capable of the long-term management and publication of all types of SCAR related data). The number of NADCs is low relative to the number of national SCAR Members. Of the NADCs that do exist, only a few have significant capabilities. A functional ADMS will be difficult to develop solely through an expansion of the NADC network. SCAR should identify a small number of existing and complementary data access networks with which to affiliate and then promote NADC involvement in these networks. By "affiliating", rather than building from scratch, SCAR can expand its ADMS at minimal cost and at the same time achieve greater interoperability with other networks. It is also important that SCAR's peak data management groups (ie. SCADM and the Standing Committee on Antarctic Geographic Information - SCAGI) work more closely together in pursuing common goals. Now that the distinction between managing and publishing spatial and non-spatial data is disappearing, consideration might be given in the future to amalgamating SCADM and SCAGI.
d. **Standards and Interoperability:** agreement on, and implementation of, standards that support the interoperation of technology platforms and data transport protocols. In particular, development or adoption of standards to describe and encode data objects, equipment, processing techniques and instruments that ultimately function to permit data integration and aggregation. A key component of the ADMS is the Antarctic Master Directory (AMD) metadata system. It is therefore crucial that SCAR works closely with the AMD host organisation (i.e. the GCMD) to help determine the functionality of future iterations of this technology platform. Equally important is the need to recognise that SCAR science covers highly diverse data types and data management requirements. The ADMS must be geared to meeting this diversity of needs. To achieve this goal, further enhancement of the ADMS should be under-pinned by developing an implementation roadmap.

e. **Outreach and Guidance:** education, outreach and guidance on all facets of the systems operation, protocols and functions. Growing the number of NADCs and improving the capabilities of those that exist could be achieved using a more formalised training and mentoring campaign. Both SCADM and SCAGI should improve their communication mechanisms and mediums.

Much of the data management that currently occurs within SCAR science projects is conducted under circumstances outside of the influence of either of SCAR's peak data management coordinating groups. The network of NADCs on which the SCAR ADMS should be founded therefore needs to be expanded and become interdependent with other, successful thematic and global data networks, that are currently being patronised by SCAR research programs or which have the potential to add value to SCAR science. Several opportunities exist to more closely align SCAR data management with large international data management facilities and networks (notably the ICSU WDCs, IODE, the WMO Information System [WIS], the IPY Data and Information Service [IPYDIS] and the Polar Information Commons[PIC] initiative), all of which conversely need to align themselves with scientific data sources (such as SCAR).

To realise its strategic data management vision SCAR needs to develop a roadmap to action recommendations in this report in the form of a Data and Information Strategy Implementation Plan.

1.0 Strategic Vision

Consider the following future testimonials from two scientists operating within the SCAR network. Their experiences paint a picture of a well-patronised, coordinated, technically robust, flexible and functional Antarctic Data Management System (ADMS). Development of such a system is the vision of this Strategy.

Testimonial From A Young Research Scientist – Recently Recruited Into A New SCAR Endorsed Science Project (A Data User's Perspective)

"The SCAR Antarctic Data Management System has significantly reduced the amount of time I have had to spend familiarising myself with research relevant to the project that I'm currently undertaking. The SCAR metadata system, and its affiliated, inter-linked core web-based systems permit me to readily discover, and in most cases access almost immediately: raw data; derived data; publications; products; and model output, regardless of when, where and how those data were originally collected. There must be a lot of redundancy, intelligence and flexibility built into the system because I never have to wait very long for a response, even for complicated requests. The hits I get are usually spot on and I don't have to spend much time going through material that isn't relevant, particularly if I use some of the high level visual discrimination tools that let me browse datasets quickly before even downloading them. Most resources I do access are really well described, making it very easy for me to judge the quality of the material for research purposes. I'm amazed at how simple it is for me to grab data from different sources and rapidly integrate and manipulate them in a meaningful way, using very intuitive utility tools that are advertised and available within the system as public services from a wide variety of institutions and individuals. I've developed a few utility tools of my own recently and I'm just about to register them with the system.

The great thing about the SCAR systems is that they don't confine you to Antarctic themed resources. If you need to search for resources more broadly, SCAR systems interoperate with many other relevant systems. I'm also able to get access to large volumes of high resolution spatial data as a backdrop for much of my scientific work, no matter whether I'm on my way down south on a ship, back at the lab, or on station. I don't know if I could ever move into another research field that wasn't supported by such a fantastic data management system."

Testimonial From A Seasoned Research Scientist – Participating In A SCAR-Sponsored Ocean Monitoring Project (A Data Provider's Perspective)

"I must admit that when development of the SCAR Antarctic Data Management System began in earnest I was very sceptical about using it, or making any sort of contribution. I just couldn't see the value in it for me personally, or for my colleagues. We had our own systems and they worked fine for us. But now I'm a total convert. Last year we put in for funding for a series of voyages around the Antarctic continent and because funding agencies now insist on data management being costed into project proposals and expect a project plan to be developed at an early stage in the research, for the first time we had sufficient resources allocated to do basic data management which meant it was also much easier for us to do our analyses.

The new SCAR dataset citation system is now widely accepted as a legitimate research performance indicator by SCAR participants, and has made a big difference to how my work is being received in my home agency. I used to struggle in selling the value of my observational research to senior management, but now I'm being lauded because of the wide range of uses that my data is being put to. The systems that are now available for registering my datasets are very easy to use and I actually use them even in the early stages of my work to organise my data as I work up results. When I'm ready to make my data public it only requires the push of a button.

Perhaps one of the main things that changed my mind about the usefulness of the SCAR system was when the virtual observatories really began to expand and came on-line. This has considerably changed the way I conduct my science. I can now control a number of sensors from my desk-top, and have just added my sensors to a cluster of sensor networks which really broadens the types of measurements that I am able to get access to. Since we have dynamic control of many aspects of sensor operations and we can collaboratively analyse our data in real-time, thanks to sensor network technologies and agreed standards, we can rapidly change configurations to examine interesting phenomena as we detect them. Because we now have such high-resolution spatial data for so much more of the region, coming together daily from all collaborating SCAR Members, we are finding numerous niche environments to explore with our sensors and making many new scientific discoveries as a result.

I used to be worried that I would get gazumped and people would publish before me if I put my data out there too early, but now I have access to such a wide range of data, in real or near real-time, that there is more data to interpret than any of us can reasonably keep up with.

The sheer volume and variety of the data has even provided some of us with new fields of enquiry focussed purely on mining the data. Of course I can see now that without the foresight that was put into standardising and specifying many aspects of the system, like data, metadata and communication standards, none of what we now enjoy would have been possible. We might still be capturing these large volumes of data – but sharing in its use and using it efficiently like we do today just wouldn't have happened."

2.0 How Can This Vision Be Realised?

Whilst the vision espoused in the testimonials is aspirational it can be achieved through an incremental change in how individuals, institutions and nations work together to solve common data management problems that require a globally co-ordinated response. It is unlikely that this vision will be achieved in totality during the life-time of this Strategy (2009-2013), but the strategic directions set out in this document will lay the foundation for its realisation within the next 10 years. Before explaining how to achieve the vision, an argument is made in the following section, as to why SCAR needs to act now to embrace a more strategic direction for conducting its data and information management.

To appreciate the strategic recommendations made in relation to SCAR data and information management, an overview of SCAR and the SCAR operating environment is provided. Current weaknesses in the SCAR data and information management system are also highlighted before a series of high-level recommendations are made, designed to address these weaknesses and to set SCAR on a path towards achieving the vision.

2.1 Data Management Imperatives

In 1952, the International Council of Scientific Unions (ICSU) proposed a comprehensive series of global geophysical activities to span the period July 1957-December 1958. The International Geophysical Year (IGY), as it was called, was modelled on the International Polar Years of 1882-1883 and 1932-1933 and was intended to allow scientists from around the world to take part in a series of coordinated observations of various geophysical phenomena. Given the state of science in the late 1950s, the timing of the IGY was highly opportune. Research technologies and tools had advanced greatly since the 1930s, allowing scientists a scope of investigation without precedent. Cosmic ray recorders, spectroscopes, and radiosonde balloons had opened the upper atmosphere to detailed exploration, while newly developed electronic computers facilitated the analysis of large data sets (NAS, 2005). It is now 50 years since the IGY, and we have concluded another International Polar Year (IPY 2007-2008). It is argued that the timing, yet again, is opportune to fundamentally change how we collect, archive, access, manipulate, and preserve data given the enormous advances in technology since the IGY.

From the mid-1990s we have witnessed rapid improvements in the sophistication, availability and accuracy of sensor technologies, importantly coupled with a decrease in sensor size. We can readily deploy these tools in inaccessible and often harsh environments, using remote control and 'set and forget' techniques. It is now possible to use marine mammals as opportunistic sensor platforms, capturing physical oceanographic data in remote oceanic regions from small body-mounted sensors as these animals perform their daily foraging activities. Data capture rates are beyond the comprehension of our 1950s IGY colleagues and the rich types of data we are now able to acquire has

led to an explosion of information. Taxonomists can now classify biota by bar-coding their DNA, and satellite-based sensors regularly sweep the earth recording parameters from which we can derive ice thickness and ocean biological productivity at repeated, and often very high spatial and temporal resolutions.

Unlike in 1957, computers are now ubiquitous. They are highly configured, reasonable in price and researchers can readily build their own low-cost processing infrastructure or easily access and harness high performance computing platforms that routinely run very complex and data rich simulation models. This in turn generates even more data. Advances in communication technologies, particularly the Internet, are providing new models for highly distributed, networked collaboration. These changes have lead to newly emergent scientific disciplines such as data mining and genetic sequencing, which are able to exploit the massive volumes of data and the more sophisticated data processing tools now at our disposal. Since the last IGY there has been a quantum shift in scientific capabilities.

IGY was significant for many achievements but importantly it left a global legacy, the World Data Centre System. This system, still in operation today, is an international system of nationally based Centres, dedicated to the management and publication of multi-disciplinary scientific data. Although now due for renewal, this system has provided many scientific disciplines with a stable framework for exchanging data and has been a mechanism for the compilation of a large range of comprehensive data products that are global in coverage. One possible legacy of IPY 2007-2008 is a system of long-term observing networks to support Polar research for decades to come. These networks will among other things enable establishment of a baseline against which to detect and forecast future environmental, biological and climate change (ICSU, 2007). Significantly this IPY goal will also require substantial enhancement of the existing Polar data management infrastructure if we want to effectively discover, access, use, store, archive and protect the data being captured and add value to these data many times over through their re-use in cross-jurisdictional, inter-disciplinary research.

This strategy hopes to stimulate the creation of a sustainable, network-based infrastructure for the future that is capable of meeting the information management challenges that accompany the technological advances which we now enjoy at the beginning of 21st Century. In this Century it is obligatory for any Antarctic Data Management System (ADMS) to be interoperable with other existing global infrastructures and initiatives, particularly with the reform of the ICSU World Data Centre System, and to this end it must leverage existing and emerging global standards and protocols whenever possible.

If we are to leave a credible data management infrastructure as a legacy of IPY, we must take action now to plan and coordinate its deployment.

2.2 SCAR Institutional Arrangements (for details on SCAR see www.scar.org)

SCAR is charged with initiating, developing and coordinating high quality international scientific research in the Antarctic region, and advising on the role of the Antarctic region in the Earth system. The scientific business of SCAR is conducted by its Standing Scientific Groups in the Physical-, Life- and Geo- Sciences which represent the scientific disciplines active in Antarctic research. These groups share information on disciplinary scientific research being conducted by national Antarctic programmes; identify research areas or fields where current research is lacking; coordinate proposals for future research by national Antarctic programmes to achieve maximum scientific and logistical effectiveness; identify research areas or fields that might be best investigated by a major SCAR Scientific Research Programme; and establish Action and Expert Groups to address specific research topics within the discipline.

The management of data and information on behalf of SCAR's scientific community is carried out by SCAR's national members, with coordination and leadership from two Standing Committees:

- Standing Committee on Antarctic Data Management (SC-ADM), and
- Standing Committee on Antarctic Geographic Information (SC-AGI).

In addition to carrying out its primary scientific role, SCAR also provides objective and independent scientific advice to the Antarctic Treaty Consultative Meetings (ATCM) and other organizations on issues of science and conservation affecting the management of Antarctica and the Southern Ocean. Article III 1c of the Antarctic Treaty states that "Scientific observations and results from Antarctica shall be exchanged and made freely available." Aside from meeting the needs of SCAR's scientists, SCAR's data and information management activities should assist Treaty Parties in meeting that aspiration. However, it should be borne in mind that SCAR is a body of the International Council for Science (ICSU) and as such its membership is made up of representatives of national academies, whereas the Treaty Parties are governments.

2.2.1 SC-ADM

SCAR's involvement in data and information management began in the late 1980s, in response to Recommendation 5 of the 13th ATCM (1985), which asked SCAR for advice to improve the comparability and accessibility of Antarctic scientific data. In 1989 SCAR formed the *ad hoc* Committee on the Coordination of Antarctic Data (CCAD). Recommendation 16 of the 15th ATCM (1989) asked Treaty Parties to assist the work of the CCAD and the development of an Antarctic scientific data directory. At the XXII SCAR meeting in 1992 it was agreed that SCAR and COMNAP should replaced CCAD with a joint SCAR-COMNAP *ad hoc* Planning Group on Antarctic Data Management whose terms of reference included developing a plan for the coordination and management of Antarctic data, taking into account SCAR's programmes along with requirements under the Antarctic Treaty System, especially with respect to the needs of the

Treaty's Protocol on Environmental Protection. The Planning Group met in October 1992 and proposed that a committee for Antarctic data management be established to manage the development of an Antarctic Data Directory System (ADDS), comprising National Antarctic Data Centres (NADCs) linked to an Antarctic Master Directory (AMD). The proposal was accepted by the Antarctic Treaty Consultative Meeting in November 1992, and by SCAR and the Council of Managers of National Antarctic Programs (COMNAP) Executives in April 1993. In due course the Planning Group was replaced by the Joint SCAR/COMNAP Committee on Antarctic Data Management (JCADM), which held its first meeting in Christchurch, NZ, during the 21st Antarctic Treaty Consultative Meeting. In response to Information Paper 85, submitted by SCAR and COMNAP, the 22nd ATCM adopted Resolution 4 (1998) on Antarctic Data Management, which recalled the commitment of Parties under Article III (1)(c) of the Treaty to promote international cooperation in scientific investigation by exchanging and making freely available scientific observations and results from Antarctic: welcomed the establishment by SCAR and COMNAP of JCADM and the ADDS; recognised the enhanced efficiency for Antarctic research to be gained from effective data management; and recommended that: (i) Parties who have not yet done so should establish NADCs and link these to the ADDS managed by JCADM; (ii) Parties and their NADCs should encourage their scientists, through a process of education, support, and the development of policies and procedures, to provide in a timely manner appropriate information to their NADCs for distribution through the ADDS; and (iii) Parties give priority consideration as to how the requirement for freedom of access to scientific information, in accordance with Article III(1)(c) is achieved within their national data management systems.

Some ten years on from its formation, in 2008 JCADM comprised 30 national representatives reponsible for Antarctic data management within their respective jurisdictions. As is evident from the description above, JCADM's remit was narrowly focused. To ensure that it was functioning effectively, JCADM's performance was reviewed by an external group in 2005 and 2008, and several improvements were put in place to ensure that JCADM met the needs of its primary users, the SCAR science groups. The 2005 review recognised the need for JCADM to widen its remit and to provide advice on developing a data and information strategy for the future, as recommended in the SCAR strategic plan (2004-2010).

In 2009, JCADM was replaced by SCADM after COMNAP withdrew its support for the group, citing a need to sharpen its focus around core business issues. SCADM's purpose is to advise SCAR on all key facets of the management of Antarctic data. A primary role is to provide strategic guidance on the development of a broadly encompassing Antarctic Data Management System (formerly referred to as the Antarctic Data Directory System), which is founded on the activities of National Antarctic Data Centres (NADCs). SCADM is also responsible for the recruitment of NADCs into the System and for encouraging scientists to submit metadata to the SCAR endorsed metadata system, directly, or via the NADCs. SCADM, in conjunction with the NASA-based, Global Change Master Directory (GCMD), has developed the Antarctic Metadata Directory (AMD) System to manage these metadata records. Given the remit to provide

advice on a data and information management strategy for the future, SCADM must now work to compliment the existing system with other components that can provide a more holistic framework for Antarctic data and information management.

The SCADM Chief Officer reports anually to the SCAR Executive Committee (EXCOM). The Chief Offier is elected by the SCADM member parties, generally for a period of four-years, which can be on a rolling basis. The role is currently part-time and is supported by the home institution of the elected official. SCADM also appoints 4 Liaison Officers whose roles are to forge links with the SCAR Science Standing Groups and with the CEP, and has representatives on each of SCAR's major Scientific Research Programmes, in order to facilitate closer communication with the user community.

2.2.2 SC-AGI

At the XXIX SCAR meeting in Hobart 2006, the Expert Group on Geographic Information (EGGI) was repositioned from within the Standing Scientific Group for the Geosciences to become the Standing Committee on Antarctic Geographic Information (SC-AGI). SC-AGI is the direct descendant of the SCAR Working Group on Cartography, formed near SCAR's beginnings, in 1958.

Most scientific and logistical work carried out in Antarctica relies on a consistent geographic framework, and the main function of the new SC-AGI is to manage and improve this framework. This Group's primary focus is aimed at developing an Antarctic Spatial Data Infrastructure (AntSDI). An idealised AntSDI would be characterised by:

- Easy and efficient mechanisms for sharing and exchanging topographic and other spatial datasets (e.g. management zones, protected areas, infrastructure footprints, bathymetry) using agreed datums, reference systems, place names and features either at the point of data capture or through a process of mapping to commonly agreed standards at the point of exchange,
- Complete and regularly updated seamless map coverage of the Antarctica and surrounding ocean at a series of scales commonly determined as useful for research, territory and environmental management purposes,
- Globally shared services and infrastructure that permit access to these data from anywhere on the Earth, and which provide users with a variety of data portrayal methods dependent upon their needs.

Building on the work done by its predecessor, SC-AGI continues to deliver and manage a range of Geographic Information products that have been developed to address some of the AntSDI goals outlined above. These products include the SCAR Composite Gazetteer of Antarctica, the SCAR Antarctic Digital Database, the SCAR King George Island GIS Database, the SCAR Feature Catalogue and Symbology Editor and the SCAR Map Catalogue. SC-AGI also seeks to integrate and coordinate, where feasible, national Antarctic mapping and GIS programs of the SCAR member countries. Using its GIS expertise SC-AGI has forged liaisons with other key non-polar bodies that have an influence on the spatial information framework for Antarctica. These bodies include the geographic standards group within the

International Standards Organisation (ISO), the International Steering Committee For Global Mapping (ISCGM), the International Hydrographic Organisation (IHO) and the International Society for Photogrammetry and Remote Sensing (ISPRS).

The SC-AGI Chief Officer reports annually to the SCAR Executive Committee (EXCOM). As with SCADM, the Chief Officer is elected by the membership and appointed by EXCOM for a period of four years, which may be on a rolling basis. The position of Chief Officer is resourced by the elected official's home agency and is a part-time appointment. The group meets face-to-face annually to discuss progress with its work programme.

There are obvious overlaps between the roles of SCADM and SC-AGI and it is one of the key goals of this Strategy to ensure that both groups work in unison to develop a coherent and integrated framework for Antarctic data management, which meets the requirements of Antarctic research, logistic operations and environmental management now and into the future. In practical terms there is very little difference between the stated goals of the two groups, as both aim to initiate a comprehensive system for managing and utilising data and information derived from Antarctic research and national Antarctic operations.

2.3 Current Approaches To SCAR Data Management

2.3.1 SCADM & SCAGI Lead Data Management Activities

SCADM and SC-AGI have been founded on the understanding that scientific and operational data are core assets of the Antarctic community and are valuable in their own right. When these data are managed well, they provide vital resources for the future and facilitate the development of interdisciplinary research and international collaborations.

SCADM is the entity responsible for the Antarctic Data Management System (ADMS); it provides:

- A single portal for recording information about data holdings - the Antarctic Master Directory (AMD), and;
- A distributed system for storing and providing access to that data – the National Antarctic Data Centres (NADC).

SCADM's recently reviewed and re-drafted Terms of Reference (listed below) reflect these two responsibilities:

1. To promote long-term preservation and accessibility of data relating to Antarctica and the Southern Ocean in sustainable repositories,
2. To assist in establishing Antarctic data management policies, priorities and best practices,
3. To support the establishment and ongoing work of National Antarctic Data Centres, in accordance with ATCM XXII Resolution 4.1 (1998),

4. To encourage submission of metadata and data to the Antarctic Data Management System,
5. To further improve and populate the AMD and provide guidance to the AMD host,
6. To provide linkages to other relevant data management systems and thereby enhance the ADMS,
7. In partnership with SCAGI to work with SCAR SSGs and SRPs, COMNAP and the Antarctic Treaty Secretariat to identify and develop fundamental datasets of value to the Antarctic Community.

SCADM is essentially a coordinating body, comprising over 30 representatives from countries with an interest in SCAR data management. The Committee meets annually in various Member countries, on a rotational basis, although only a sub-set of member states (usually less than a 1/3) generally fund participation in these meetings. It is important to understand how SCADM currently operates so that there is a common perception of its capabilities, as these issues significantly influence the strategies detailed later in this document.

A 2007 email survey of SCADM members indicated that relatively few of the 30 nations involved in SCADM are actually operating a National Antarctic Data Centre (NADC) – see Table 1. Most countries have assigned a single individual person as an Antarctic data management coordinator and often this person has dual roles in many cases also acting as the national science coordinator or science liaison officer. In these cases it is highly unusual for the coordinator to (a) handle any Antarctic data and (b) know exactly where all data from their Antarctic programs reside, although it is common for these officers to be involved in metadata submission. There is also a highly variable level of data management proficiency between coordinators (and Data Centres) within SCADM, whether data are being handled or not. Many respondents to the survey readily acknowledge this latter point as a significant issue.

Table 1 Existing NADCs

Countries Reporting An Established NADC	NADC
Argentina	Argentinean Antarctic Institute (Instituto Antártico Argentino - Dirección Nacional del Antartico)
Australia	Australian Antarctic Data Centre
China	Chinese National Antarctic Data Center(CN-NADC) -Polar Research Institute of China (PRIC)
Italy	PNRA - SIRIA Project
Japan	Polar Data Center (PDC) in the National Institute of Polar Research(NIPR)
Netherlands	Royal Netherlands Institute for Sea Research (NIOZ)
Norway	NADC within the Norwegian Polar Institute (NPI)

Spain	Polar National Data Centre - located in the Spanish Geological Survey
United Kingdom	National Antarctic Data Centre (the AEDC) – within the Environment and Information Division of the British Antarctic Survey
United States of America	There is no "One" dedicated data center - a variety of Government funded institutions (National Science Foundation, National Oceanic and Atmospheric Administration, NASA) and Universities provide various levels of service.

Where there is a designated Data Centre, in all cases reported through the survey, the number of people directly involved with Antarctic data management is commonly 2 full-time staff or less. Australia and the UK are exceptions and have the largest Centres (Australia with 8 staff in one dedicated Centre and the UK with 8 distributed staff loosely coordinated through a dedicated Antarctic Data Manager). **This is a fundamental issue for SCAR because very few SCADM Members are operating NADC's and yet coordinating these Centres is one of SCADM's key goals**. Many of the countries with the larger programmes have NADCs and are actively involved in SCADM, although at the time of writing Brazil, France, India, Republic of Korea, Russia, Germany and South Africa had not indicated to SCADM or SCAR that they maintained an NADC. Due to resource implications, it is expected that nations with larger programs are able to contribute more. But even smaller programs could assign limited NADC responsibilities to existing national institutions.

The current Antarctic Data Management System in reality is founded on metadata management and activities surrounding the development and ongoing population of content to the Antarctic Master Directory, rather than on a system that is collaboratively organised to manage, publish and archive actual scientific data and information. The discrepancy between the stated goals of SCADM and the reality of what can be achieved with the current resources creates unrealistic expectations of SCADM within the SCAR science community and has in the past lead to dissatisfaction with performance.

This resourcing issue is well known to many in SCAR and it was highlighted in the 2005 review of JCADM (Rickards et al, 2005). Unfortunately, no suggestions have been forthcoming about how this situation should, or could be remedied, quite possibly reflecting SCAR's lack of a coherent and integrated data and information management strategy. The same 2005 review recognised several other key short-comings with the current system which include:

a. poor compliance by SCAR Members with Antarctic Treaty provisions relating to data access (i.e. a lack of support for making national data freely and publicly available),
b. the need for more outreach from SCADM to SCAR science projects and programmes,

c. the lack of a SCAR strategy for data and information management, particularly one which pays appropriate attention to the future use and management of data,
d. the political role that the SCAR Executive needs to play in urging SCAR member states to safeguard data for future use,
e. the need to involve science/national/international funding bodies in ensuring adequate data management within funded scientific programmes,
f. that there are insufficient resources within SCADM and the NADCs to expand from managing metadata into managing the data and associated generation of products,
g. the need to forge better links with existing global initiatives such as CCAMLR, IODE, OBIS, GBIF, SuperDARN and JCOMM,
h. lack of an explicit forward plan for the GCMD-based AMD software,
i. a view expressed by the Review Team that it is too difficult for SCADM to work with SCAR projects or programmes where the data remains with the scientists rather than within a professionally managed archive, and
j. a number of relatively minor issues related to the operation and functioning of the AMD.

To SCADM's credit, in the three years since the 2005 JCADM review was conducted, it has attempted to address all of the metadata related issues and a number of other short-comings raised by the reviews of 2005 and 2008, including the development of this Strategy. Two key things have thwarted further development of the data and information system. Lack of an over-arching strategy for data and information has impeded development in the most appropriate directions, and an over-reliance on simply bringing together NADC managers in the belief that some strategic self-organisation would emerge has restricted thinking, impeded the development of links to the users, and placed the burden too heavily on the shoulders of the few Members capable of providing extensive support for data and information management. In addition, **SCADM** (previously JCADM) **continues to be hampered by a fundamental lack of resources**. The only budget SCADM has had through the years has been the annual contribution to the GCMD for managing the Antarctic Master Directory. No assistance has been provided, for example, for the managers of less well-endowed NADCs to attend SCADM meetings. An inability to attract active members has in turn lowered morale and if left to continue unaddressed will eventually render the network unviable. Not surprisingly, the larger, better-resourced Data Centres have tended to take on the lion's share of SCADM responsibilities in anticipation that capabilities within the network would grow over time, and others would eventually rise to play similar roles within the community. This has not happened. The strategy has been flawed. Fundamental problems have been masked. Expectations have not been met. Both SCAR and SCADM now recognise the need to correct and strengthen the system.

Action has recently been taken to improve the operational circumstances of SC-AGI, which (as the Expert Group for Geographical Information) had been situated under the Standing Scientific Group for the Geosciences during the reconstruction following the SCAR review of 2000. That strategy did not work,

and to improve matters the former EGGI was given the status of a Standing Group in its own right in 2006. That change, the appointment of a new chairman, and the provision of much stronger support from the SCAR Executive Committee in the group's operational activities and in the promotion and marketing of SC-AGI products, has considerably improved SC-AGI's overall performance. Nevertheless, further improvements are needed. SC-AGI is working to reduce a long wish list of items down to a set of key high priority items. The Committee is attempting to discern how many of the nationally-centric tasks fit into a coherent plan for strategically building Antarctica's spatial data infrastructure. Committee funding is minimal, and business relies on the contributions of SCAR Members in getting their representatives to SCAGI meetings. Attendance is, as a result, uneven. Despite this lack of resourcing, an active cluster of nations (mostly those with the largest investment in Antarctic research) has in the past managed to produce a number of important products, which if better patronised and more strategically utilised would greatly assist SCAR members to put in place an AntSDI. As in the case of SCADM, it is somewhat surprising that some of the larger and more well-endowed programmes fail to participate; their involvement could be one key to success.

2.3.2 SCAR Science Lead Data Management Activities

Although the SCADM and SCAGI terms of reference indicate that these peak groups should be leading the development of the Antarctic Data and Information Management System (ADMS) through coordinating links to existing "systems" and by leveraging facilities at the NADCs**, it is usually the science programs that forge such links and/or develop their own niche systems, generally without reference to either SCADM or SCAGI**. This is not surprising because neither SCADM nor SCAGI have the resources to respond to calls for assistance from research programs, perhaps with the exception of having the capability to assist with metadata preparation.

Under these circumstances, within the SCAR portfolio of research, SCAR scientists are using a very wide variety of data management systems and data exchange networks. Some of these are large, global infrastructures with well-resourced, professionally managed archive centres (e.g. Argo Float data - http://www.argo.ucsd.edu/FrAbout_Argo.html; GCOS observing system network - http://www.wmo.ch/pages/prog/gcos/index.php?name=about), others are much smaller, Antarctic-themed or project specific. There is, however, no SCAR-wide description available of the various systems currently being patronised by SCAR science groups. **Although guidance is given explicitly and actively by SCADM to SCAR science groups about where and how to deposit metadata, no advice is given about where best to deposit actual datasets**.

While SCAR members may be individually involved in one, or a number, of existing data management initiatives or data networks, **there is little coordinated effort across NADCs to patronise any particular systems,** possibly with the exception of two biodiversity databases. SCAR-MarBIN (a Belgian-based, marine-themed data aggregator node of the Ocean Biogeographic Information System – OBIS), directly supports the SCAR Census of Antarctic

Marine Life Project and is being actively promoted by the SCAR Life Sciences Standing Group and SCADM. Its complement is the terrestrial-themed, Evolution and Biodiversity in Antarctic (EBA) Database, also promoted heavily by the SCAR Life Sciences Group.

2.3.3 SCAR Data Management and the ATS

As SCADM and SCAGI are subsidiary bodies of SCAR, their interactions with the bodies of the ATS must be coordinated with SCAR's Standing Committee on the Antarctic Treaty System (SCATS), which is tasked with managing these interactions and responding to requests. While SCAR recognizes the Treaty's agreements on reporting and sharing of data, these are rather generic. SCAR's interests must remain within its mission of science and scientific advice. It is not within the remit of SCAR, or within its ability, to serve as the repository of data or the primary avenue of data collection related to national obligations under the Antarctic Treaty. That is the responsibility of the Antarctic Treaty Secretariat. While SCAR might benefit from information and data exchanged by Parties under Treaty directives, SCAR operates solely as a third party, objective observer. Nevertheless, the existence of a comprehensive SCAR data and information management strategy and of an Antarctic Data and Information Management System will undoubtedly go some considerable way to fulfilling the requirements of Resolution 4 (1998) referred to under item 2.2.1 (above). Indeed, if the Treaty Parties can be persuaded to see the ADMS as a mechanism of value to them in meeting the requirements of Article III (1)(c), then they may urge their operators and scientists to use it to the full, making SCADM's task that much easier.

SC-AGI has developed several products (e.g. Antarctic Digital Database, Map Catalogue and Gazetteer) that may be of assistance to the Antarctic Treaty Secretariat (ATS). The Secretariat has developed a web site and associated databases related to Treaty obligations for information exchange. Closer co-operation between SCAR and the ATS is essential to avoid duplication of effort, but the intent and remit of the two organizations must remain independent.

SCADM and the ADMS might be of particular utility and value to the Committee on Environmental Protection (CEP), which is a primary point of contact between SCAR and the Antarctic Treaty System. What this might entail and how SCAR might accommodate or assist the needs of the CEP is a subject for further discussion between the two organizations.

Both SCAGI and SCADM products might also be of value to COMNAP. It would be beneficial if COMNAP and SCAR collaborated on issues that have common data requirements.

2.4 Strategic Recommendations

This report's fundamental strategic recommendation is to build an Antarctic Data Management System (ADMS), capable of supporting Antarctic science and SCAR activities within the Antarctic Treaty System. The ADMS should be viewed as a science enabler and a service that supports practical applications of various kinds. Strategic success would mean that the ADMS is a tangible operational entity, deemed an indispensable tool for doing scientific research in Antarctica and the Southern Ocean. The cost of supporting the system should be an integral part of budgets of national programmes. Funding bodies should be encouraged to view ADMS development and maintenance as a necessary cost of doing science.

Through a range of individual activities SCAR is already making progress towards achieving parts of the vision illustrated in the testimonials presented earlier in this report. But much more can be achieved. The likelihood of realising the desired goals would be far greater if appropriate strategic foundations were put in place now to enable better coordination of these individual and often disconnected efforts. These strategic foundations would necessarily include:

a. better articulated governance arrangements and strong leadership, suitable for driving the development of a distributed, but loosely federated, shared infrastructure;
b. fostering a culture willing to share and collaborate on data management related activities;
c. leveraging existing SCAR and non-SCAR systems, capabilities and resources and supplementing these where there are obvious deficiencies (the primary purpose of such leveraging would be to create a network of designated permanent data archives capable of the long-term management and publication of all types of SCAR related data);
d. agreement on and implementation of standards that support the interoperation of technology platforms and data transport protocols. In particular development or adoption of standards to describe and encode data objects, equipment, processing techniques and instruments that ultimately function to permit data integration and aggregation; and
e. education, outreach and guidance on all facets of the systems operation, protocols and functions.

Each of these foundations is discussed below in the form of strategic actions.

2.4.1 Policy, Leadership, Coordination and Governance

To clarify obligations with regard to data access, SCAR requires a clearly enunciated data management policy so that there is unambiguous guidance on what is expected of SCAR members in relation to data access and data management. Two suitable policies that can readily be adapted are the IPY Data Policy (http://ipydis.org/data/) and the IOC data exchange policy (http://www.iode.org). The creation of Data Management Plans for all major SCAR Scientific Research Programmes is now part of the function of SCAR

Science Planning Groups, and a Data Management Planning template should be an annex to the SCAR Data Policy as guidance for these groups on recommended Plan content (see Appendix A for a Draft SCAR Data Policy, encompassing a Data Management Plan Template and reference to a sample Plan).

Recommendation 1: Development of a SCAR Data Policy.

Admittance to the SCAR family, from a data management perspective, currently entails a Member nominating someone to represent data interests on SCADM. The SCADM "Rules of Procedure" have recently been drafted and state that: "Nominees should be professional data managers or scientists with expertise in data management, who are closely affiliated with either the Member's National Antarctic Data Centre (NADC), or in the absence of an NADC, another national scientific data repository". Implicitly SCADM has an NADC recruitment campaign, but there is no SCAR policy requiring its Members to nominate a functioning NADC to participate in the ADM network. There is also no guidance as to what constitutes an NADC, nor any consistently applied criteria about the function of such a facility. There is also no onus on any member nation to designate any type of permanent data archiving facilities.

Clear obligations on SCAR members would partially remedy some of the more significant barriers to advancement of the AMDS. Data centric obligations of SCAR membership might for instance include:

a. Nomination of a national Antarctic data management coordinator, **actively** engaged in national data management coordination, and appropriately resourced to do so, plus
b. Nomination of an **operational** national Antarctic data hosting repository willing, and resourced to participate as a designated National Antarctic Data Centre and become part of the AMDS, and
c. Agreement to adhere to the SCAR Data Policy.

As in the current situation, a national Antarctic data management coordinator nominated to serve on SCADM need not come from the same institution in which an NADC resides. The coordinator could be affiliated with the Centre or operate outside of it. The intent, however, where the NADC and coordinator are in separate institutions, is that a close working relationship between the NADC and the coordinator should be encouraged. In addition there should be explicit and clearly articulated expectations of an NADC. Some suggested core NADC functions might include:

- assistance to users in using the AMD and preparing metadata,
- collation of all data generated through national Antarctic science projects,
- provision of data archiving services that permit the long-term re-use of data,
- publication of data from national Antarctic science programs to one or more SCAR endorsed data access networks, and

- active participation as a node in the ADMS (involving input into the development of network infrastructure standards and conformance with community data standards and protocols)

The primary function of the NADCs is to act as long-term repositories for archiving and publishing Antarctic data or to coordinate this activity. To ensure that this is undertaken in a consistent manner across all NADCs, guidance should be provided as to what constitutes acceptable archiving and publishing procedures from a SCAR perspective. Archiving standards such as the OAIS Reference Model (CCSDS, 2002) could readily be adapted to suit the SCAR NADC operating context, but equally others may be appropriate. Providing NADC obligations can be met, a member State could organise for its Antarctic data to be hosted through a virtual facility formed from the collaboration of several national institutes. A lead agency, however, would need to be nominated to take responsibility for interactions with SCADM and SCAR on data management.

Recommendation 2: Ensure that a SCAR Data Policy includes any expectations regarding the appointment, functions and obligations of an NADC.

To ensure that smaller, less well resourced nations are not penalised by any new obligations, where there is a lack of suitable national facilities, they should have the option of establishing a partnership with a larger nation probably located in their geographic region. This Regional Centre could perform a data hosting service.

This model of geographically dispersed designated Centres is common in the IODE and ICSU WDC Systems (see IODE Strategic Plan - http://www.iode.org/index.php?option=com_oe&task=viewDocumentRecord&docID=1524). The concept of having a number of designated facilities dispersed globally that could act as hosting facilities for less well resourced nations should therefore be explored in conjunction with ICSU and the IODE, both of whom are currently revamping their institutional systems and facilities. It is highly likely that a formal partnership with both, or either of these entities, would significantly strengthen the ADMS.

Recommendation 3: SCAR take advantage of the current revamping of IODE and the ICSU World Data Centre System to explore the potential for formal partnerships.

SCADM has a part-time Chief Officer and two part-time Deputies. The majority of SCADM inter-sessional work has, in the past, remained largely unfulfilled from year to year, in part because the SCADM Executive lacks the time to drive the effort, but importantly because there are insufficient resources within SCADM to undertake the work.

While SCAR is relying on the goodwill of members and a largely voluntary work-force to establish the ADMS, it is important that there is a dedicated individual available to lead and coordinate the more important strategic actions. An

analogy can be drawn between the SCAR data management workforce and an open source community. In an open source community people coalesce to tackle tasks of mutual interest and then contribute their labour for free to derive a solution. What is perhaps less well known is that in successful open source communities there is generally a paid core of individuals who form the nucleus of the workforce and this core almost always includes an individual who dedicates themselves to leading, monitoring and continuously stimulating the activity. Open source projects with poorly articulated visions, ill-defined goals, poor coordination and communication and cultures that don't adequately provide any intrinsic or extrinsic rewards usually fail to attract volunteers and soon wither on the vine. People are attracted by success (Schweik & Semenov, 2002; Latterman & Stieglitz, 2005).

If the AMDS is to be effective it will have to expand, through the recruitment of operational NADCs, possibly involving partnerships with other entities. Under the circumstances the need for a full-time, or near full-time person to drive and coordinate these efforts is considered essential. Because SCAR relies on the contributions of members, and current contributions do not include the funds to pay for such a position, implementing this element of the strategy would require a commitment from individual nations willing to temporarily re-assign their staff to help take care of specific strategic actions. The SCAR Executive Committee could also seek to obtain sponsorship to support SCAR data activities via lobbying private companies, funding bodies or philanthropic foundations. A preferred initial solution is to obtain staff secondments from those operational agencies that can be persuaded to see the eventual benefits of developing the system; these are most likely to be major Antarctic operators.

Recommendation 4: SCAR Executive to explore the possibility of using secondments from members to help deliver on Strategy outcomes, and/or work with members to raise external funding to assist in implementing strategic actions.

If SCAR is successful in recruiting adequately resourced, professionally managed NADCS from SCAR Members to participate in the ADMS, and these Centres are represented through SCADM, by default the SCADM membership will become more capable of responding collectively to meet the needs of SCAR science groups. If, however, the ADMS mainly expands through alignment with other types of Centres affiliated with various global data exchange networks, not necessarily directly associated with SCAR or with membership of SCADM, the ongoing role and function of SCADM itself may need to be reviewed. Whilst it is in SCAR's interest to ensure that all of its members have a voice in how data management is executed and coordinated in support of SCAR science projects, if SCADM is not representative of those engaged in actual SCAR data management, it will not be capable of fulfilling its current ToRs (as listed in 2.3.1).

Recommendation 5: Review the role, membership and function of SCADM in light of development of a SCAR Data and Information Strategy Implementation Plan.

2.4.2 Cultural Change and Incentives

SCAR is a diverse mix of nations, people and projects. Fostering a data management culture is already being actively pursued within SCAR, particularly through current SCADM activities, action by the SCAR Executive Director and Secretariat, and SCAR's involvement in IPY. But moving towards a culture that highly values the sharing and wide accessibility of well-described data is difficult in the SCAR environment. That is partly because, as participants in an international collaborative venture, SCAR Members can encounter potential sovereignty issues that may act to impede data flow. Another potential defect is that the scientific reward system is heavily geared towards individualism and small cliques of collaborators, where being the first to publish on a subject is still the most often used and highest weighted measure of scientific performance. In such a competitive context restricting timely access to data that you have invested significantly in capturing is therefore very tempting to individuals (and some nations). Many commentators have recognised these issues as impediments to cultural change, and almost unanimously suggest that creating a dataset citation system would go a long way towards encouraging data sharing.

Even so, developing a data citation system is not a simple task. For instance, apart from establishing a publishing system and determining how to garner and harvest content, datasets, unlike articles or books, may not be considered to be final products. Datasets can continuously evolve, which means devising ways of referencing potentially dynamic objects that can change with time. In cases where a dataset is part of a larger database, there is debate about how a portion of a database might be referenced and how datasets should be refereed. Despite these complexities, other groups are currently investigating ways to tackle this problem (see - http://www.scor-int.org/Publications/wr207.pdf). The Scientific Committee on Oceanic Research (SCOR) has been working with the IODE on this particular issue. SCOR's intent is to run a series of pilot implementation projects. SCAR could form a partnership with SCOR for testing such citation systems, through SCAR's existing NADCs. Scientists will find it easier to patronise systems and processes that are tangible and where the benefits are clear. Cultural change can often be facilitated through these types of demonstration systems and pilot projects.

Recommendation 6: SCAR to seek a partnership with SCOR to pursue the concept of instituting a dataset citation system.

There are significant amounts of money available globally at present to support the development of data access networks, particularly those involving exploration of interoperability issues and those geared to support observing systems. The UK invested £120 million in the first phase of its e-science program, the European Union is investing heavily through programs such as INSPIRE and its ICT programmes, and the US continues to fund data management projects through the NSF and various philanthropic trusts. Given the value of applying Antarctic data to many significant environmental issues of global scale, freeing up access to Antarctic sourced data should be a global priority. Given the difficulty that SCADM and SC-AGI have with providing sufficient resources to

implement a fully operational ADMS (including an AntSDI), they should be strongly encouraged to seek external grants to implement key aspects of their work programme.

It is therefore recommended that SCADM use existing European or US-based funding sources to undertake collaborative pilot projects centred on the management and publication of SCAR sourced data. Ideally these funded projects should include collaborators in the Asia-Pacific and Southern-hemisphere and an equitable way of resourcing such activities should be creatively established.

Recommendation 7: SCADM and SC-AGI, led by the SCADM Chair, should develop collaborative funding proposals targeting key European and US funding bodies. All SCADM members should keep a watching brief on calls for proposals.

One way of freeing up access to funds for data management is to educate funding agencies about the value that good data management practices can add to outcomes achieved from their grants. A key output of any scientific research is the data captured, or generated during the course of the research. Without adequate data management this output can be lost to future generations of researchers, thus significantly de-valuing the overall scientific outcomes. Because a significant number of SCAR research projects are in fact funded through national agencies and national funding bodies, a generic education program aimed specifically at funding bodies about the value of good data management practices, could improve the quantum of funds allocated for data management in science grants. Funding bodies that encourage scientist to explicitly and adequately factor data management into their grant proposals, and who monitor the delivery of data management tasks in assessing payment milestones, would assist to change the current scientific culture in favour of improved data management practices.

Recommendation 8: SCAR, through SCADM, to develop a generic education package that can be used to educate funding agencies and national institutes about the value of good data management practices.

2.4.3 Leveraging Resources and Systems

SCAR data and information management will continue to be a largely voluntary, collaborative activity for the foreseeable future. As previously discussed, SCAR must therefore capitalise on the activities of existing operational data management networks, more effectively coordinate any data management efforts emanating from within SCAR, and continue to build capabilities across and between Members in order to distribute the workload.

The SC-AGI is currently undergoing a revitalisation in terms of its activities, having just been made a SCAR Standing Committee. There is a new Chair, a commitment to build on past achievements and a recognition that the Group needs to better articulate strategic directions. Fundamentally this Group's vision is to build an Antarctic Spatial Data Infrastructure (SDI). An SDI theoretically

comprises networked, spatially-enabled databases or datasets that are accessible for downloading or manipulation by using contemporary technologies, usually according to explicit institutional arrangements, and which are supported by policies, standards and human capital. Given this vision, it is clear that it is not achievable without significant involvement from SCADM and operational NADCs. Therefore, the strategic directions of SC-AGI and SCADM must be aligned if this SDI vision is to be realised. In recent years, in response to technological advancements, the distinctions between spatial data management and general data management have largely disappeared. With the few resources available to undertake work in the data management arena it makes no sense for these two groups to be pursuing independent agendas. Having cross-membership between the two groups as at present, while useful, is not sufficient to ensure such alignment.

The SCAR Executive, SC-AGI and SCADM Chief Officers must ensure that the strategic directions articulated by both SCADM and SC-AGI are complimentary and that the work-plans of the two groups are synergistic, with some tasks being jointly pursued. In due course there may be a case for merging the two groups, but there is much work to be done first to ensure appropriate alignment.

Recommendation 9: SCADM and SC-AGI Chairs collaborate to ensure synergies with respective work-plans and an alignment of strategic goals. SCADM and SC-AGI should jointly define the specific goals to be met in developing an Antarctic Spatial Data Infrastructure and then engineer work-plans to meet these goals directly. Consideration should also be given to combining SCADM and SC-AGI in the future.

The ADMS must leverage initiatives, systems and funding wherever they exist. Partnering with the IODE and ICSU WDC systems has already been mentioned. But there is also a range of other large-scale activities that receive substantive international support and an element of core funding, and which could play important roles in Antarctic data management. These initiatives bring together both expertise and infrastructure, and often are directly aligned with the specific type of science that they support (e.g. Genbank – distributed databases of nucleotide sequences @ http://www.ncbi.nlm.nih.gov/Genbank/; Climate Variability and Predictability data network - @ www.clivar.org; Global Biodiversity Information Facility @ www.gbif.org to name only a few). The success and longevity of these activities is nearly always dependent on the ability of these systems and their infrastructure proponents to acquire content and steadily grow their user-base. This makes them very receptive to expanding their services to appease broader audiences. It is much cheaper to be a content provider to such networks and to leverage existing infrastructure than to build infrastructure from scratch.

From a strategic perspective the ADMS could be consolidated and shaped by subscribing to an appropriate range of relevant global networks. Existing NADC's could then work towards joining these endorsed networks and contribute content and/or services. If SCAR actively endorsed networks that had the capacity to interoperate because of similarities in their operating standards and

protocols, it would be easier to access multi-themed Antarctic data through one, or a number of Antarctic-centric web portals designed to draw data from these systems. Alternatively, SCAR could influence one or more of the network partners to provide such a facility as an extension to its existing services.

Initially a smaller number of endorsements with good levels of collaboration would be preferable to a larger number of endorsements in a misguided effort to try to cover all SCAR disciplines. NADCs will be too stretched in terms of their capacity to contribute if required to distribute data to multiple networks using many different protocols. Data access systems are emerging entities, not maturing ones, and as such their standards and protocols are often quite dynamic. In the future, when systems are better embedded it will be much easier to distribute data to multiple networks without significant effort on the part of an NADC. But this is not the current reality. Only relatively robust networks should be considered for endorsement and should be chosen in liaison with the SCAR science community.

While SCAR members individually regularly leverage national and international funding sources to support SCAR science (e.g. SCAR's recent success in winning Total Foundation funding for MarBIN), SCADM and SC-AGI as entities have not applied for funding to support Antarctic data management. SCAR, through one if its members, has recently sought funding for data management through IPY activities and has been relatively successful in securing a US National Science Foundation (NSF) grant to assist in establishing an IPY Data and Information Service (IPY DIS). The character and scope of this service is still emerging; any expanded SCAR ADMS should necessarily include the IPY DIS. The proposed SCAR Donors Committee could be tasked with advising on funding for DIM initiatives to help implementation of this strategy.

Recommendation 10: SCAR science bodies should be consulted to choose a (small) number of existing and complementary data access networks with which SCAR should be officially affiliated and then endorse and promote NADC collaboration in these networks.

2.4.4 Standards and Interoperability

SCAR is fortunate in having a particularly strong partnership with the Global Change Master Directory (hosted by NASA) and a robust and user-friendly metadata tool in the Antarctic Master Directory (AMD) system. Other international networks that have been operating for significantly longer (e.g. IODE) still lack a common approach to this issue, and this lack of adherence to a common standard or toolset continues to hamper development of the IODE system. Although the AMD metadata standard is based on the Directory Interchange Format (DIF), the GCMD is ensuring that all AMD (DIF) metadata records comply with ISO 19115 (the international standard for metadata records with any spatial attributes). All ADMS NADCs are urged to submit metadata to the AMD, regardless of whether they have their own national systems.

The AMD is a fully-hosted system, so the metadata records are either posted directly to the US-based host, or harvested from existing national systems and copied into the AMD master database. The toolset of the AMD satisfies most basic needs for metadata management and now contains a substantial number of records. One advantage of the AMD is that it is a node in a much larger system, which permits interested users to search for other types of data using the same software. The tool's utility will grow as the GCMD widens its content acquisition programme to encompass harvesting of other metadata systems.

SCAR pays an annual fee to have this service provided and receives very good value for the money. As more SCAR members realise the value in contributing to, and using this system, SCAR will become substantially reliant on this key service, and so it is reasonable in such a partnership to expect some input into the development plans for the system. The GCMD should therefore be encouraged, as part of its partnership with SCAR to provide annual updates on proposed future strategic directions and to seek input from SCADM on those directions.

Recommendation 11: SCADM Chief Officer to request that the GCMD provide a forecast of future directions of the GCMD over a 3-year time horizon and ensure that SCAR members are able to provide feedback which is taken into account by the GCMD.

While the AMD is designed to manage metadata records as well as links to the datasets that they describe and any associated services, it is not currently a versatile data hosting and data publishing facility. These types of niche services are being provided, albeit in a fragmented way, by some of the NADCs but usually through SCAR science project partners.

The ADMS must ultimately be capable of operating within the contexts that SCAR science projects demand, which means:

(a) having a common set of reference data models, encodings and communication protocols;
(b) maintaining the flexibility to map from one system's formats, encodings and protocols to another; and
(c) developing appropriate and shared infrastructure to support (a) and (b) above.

Culturally and strategically these are the hardest issues to address in a voluntary, collaborative environment, because appropriate vision and leadership are often lacking, traditional business drivers are not readily apparent, progress can often appear slow and cumbersome, and the temptation to "just get on and do your own thing" in the absence of a well articulated framework is overwhelming. While creativity and novel approaches are important ingredients, building interoperable systems must be a planned and coordinated activity on a number of levels (i.e. at the communication, data transport, data and application levels).

The Internet is now the main communication tool for exchanging data and information. The Hypertext Transfer Protocol (HTTP) underpins this technology

and eXtended Mark-up Language (XML), developed by the World Wide Web Consortium (W3C), has rapidly proliferated as the language of choice for many web users wishing to exchange information. XML variants (e.g. RDF, GML, KML, SensorML) are in heavy use throughout scientific communities. There are also other encoding languages, or rather formats that have been in popular use within the scientific community for some time. For example, netCDF is widely used for the exchange of in-situ marine observations and measurements and its usage is not going to diminish.

Our strategic data management directions must be cognisant of the existence and importance of these data exchange encoding standards, particularly in terms of our ability to interoperate with other non-SCAR based scientific research and global data management programmes.

While HTTP has already been mentioned as the most ubiquitous communication protocol in use today, there is a range of other protocols, which usually operate at a level above HTTP, and which are, or will soon be, significant for the SCAR science community. Some of these deserve special mention.

Service-oriented-architectures (SOA), where software services are bundled with self-describing interfaces, which are then registered and advertised in service registries so that they can be found and used by human or computer based resources, are increasingly being deployed within the science community. Often this approach is coupled with architectures that support distributed processing, as in the development of scientific grids. The use of grids to run complex, data intensive models that sometimes require massive amounts of parallel processing is an increasing phenomenon, and one which is already heavily patronised by the astronomical, particle physics, meteorological and atmospheric sciences.

The use of OGC-based web services within the scientific community for both data visualisation and exchange is now significant, and their use is apparent in many disciplines. Some alternative protocols, however, are more firmly anchored in specific areas of science. For example, the Distributed Generic Information Retrieval (DiGIR) protocol (http://digir.sourceforge.net/), which is a set of tools for linking a community of independent databases into a single, searchable "virtual" collection, is used primarily by the biological sciences. SCAR-MarBIN, for example uses DiGIR as its preferred data exchange protocol as does GBIF and OBIS.

While the biological sciences use protocols such as DiGIR, the physical sciences, particularly the oceanographic and atmospheric communities have been using a protocol called, OPeNDAP (http://www.opendap.org/). OPeNDAP provides a way for ocean researchers to access oceanographic data anywhere on the Internet from a wide variety of new and existing programmes. Scientists that regularly use OPeNDAP will generally also subscribe to a "type" of SOA architecture built from a specific set of middleware called THREDDS (Thematic Real-time Environmental Distributed Data Services).

Understanding that different scientific groups use different data encodings, formats and protocols to exchange data is fundamental to developing an inclusive ADMS, one that can address the diverse needs of the SCAR scientific community. Most communities have already made a significant investment in one or more of these foundational technologies and SCAR should acknowledge this existing investment.

At the "data" level, communities should agree on how they attribute and structure datasets so that their content is widely and unambiguously understood by both machine and human interpreters. The tasks of efficiently integrating, filtering, transforming and analysing exchanged data, requires that we structure and codify information using agreed language syntax and known semantics.

Communities of interest are now working on standardising vocabularies for species registers, place name gazetteers, chemical compounds, datums and reference systems, geological processes, geological objects, and geographic features to name only a few.

Recommendation 12: Development of the ADMS must recognise the diversity of standards and protocols required to meet SCAR scientific needs. A SCAR Data and Information Strategy Implementation Plan must explicitly address how to build a loosely federated system encompassing these standards and protocols.

2.4.5 Outreach and Guidance

SCADM has a well-managed web site (www.jcadm.scar.org) but it could be used to greater effect in terms of its role in education and communication. SC-AGI also has a web site (www.antsdi.scar.org), but its long-term maintenance is currently under review. Both sites currently mix content that should be tailored for two different audience types (i.e. data users vs standing group members) and both focus predominantly on their own group members as the primary target of the web site. A clearer distinction should be made about content that is useful for data and service consumers, such as SCAR scientists, and other information that is generally of more interest to SCADM and SC-AGI group members. Content should then be communicated in a manner that best suits these different audiences.

The general SCAR web site provides more links to SCAR data products than either the SCADM or SC-AGI sites. The SCADM site has substantive information content but is relatively poor at providing access to SCADM member's data and services. This is a significant issue because SCAR scientists looking for assistance to access data, through SCADM, should be able to use the SCADM website as a portal into available SCAR data management services.

Other features which would improve the reach of both sites are the inclusion of regularly updated news items, an area for frequently asked questions, and prominent posting of templates, guidelines and work-plans (cognisant of the different audience types using the web sites).

To ensure that national delegates are fully informed of the activities of SC-AGI and SCADM when participating in SCAR decision-making forums, they should be regularly made aware of the work-plans of both groups. Similarly, all Chief Officers of SCAR Research Programmes should also be kept fully informed. The standing group web sites are good vehicles for delivering these updates. The SCADM and SCAGI Chief Officers should regularly post information to these sites and ensure that stakeholder groups are notified that new information is available.

Recommendation 13: SCADM and SC-AGI should review their use of the web, and better target content and its communication to different audience types. The web should also be used to ensure that services/products and work-programs are well-known to all SCAR national delegates and scientific research programmes through targeted information dissemination campaigns

Many SCADM members have indicated that their capabilities could be significantly improved if they had access to training, mentoring and professional advice on data management issues and on polar data management in particular.

SCADM generally runs one or two-day data management workshops in conjunction with SCADM meetings, but often the very nations that would benefit from these sessions do not attend. During these workshops it is often hard to pitch the material at a level suitable for all attendees, and language barriers also make the transfer of information difficult. Recognising these limitations, SCADM has already acknowledged the need to broaden its approach to encompass modes of training suitable for a range of situations.

An important role for existing NADCs is to help develop new NADCs. Four mechanisms could be tried to deliver the mentoring required:

- NADCs could indicate their availability to conduct short training courses on establishing and running a polar data centre;
- NADCs could host interns for periods of time with the view to training the intern in one, or all aspects of managing a polar data centre;
- NADCs could collaborate on the development of multimedia product(s), distributable via DVD, that include training material on operating a polar data centre; and
- The more developed NADCs could be allocated as mentors to help the less developed NADCs – through a 'buddy system'.
- Video-conferencing or telephone conferencing could be used to engage representatives of those SCAR Members financially unable to attend meetings of capacity building courses.

For existing NADCs to invest the effort required to host and train people at their own expense, there needs to be an understanding that the members receiving such training will then invest in establishing a visible and active NADC. The

whole network is then strengthened by this activity and the better-resourced NADCs gain their returns from the improved capabilities of a growing network.

SC-AGI also holds an annual meeting where members primarily provide updates on activities in their national programmes as well as reporting on progress with collaborative tasks. Considerably more communication on SC-AGI matters occurs in the few weeks leading up to these face-to-face SC-AGI meetings than at any other time in the year. The group would benefit from a more even and planned set of communications on key issues throughout the course of the year, so that issues can be aired and worked on prior to attendance at annual meetings.

Recommendation 14: SCADM and SC-AGI should review how they approach training and/or mentoring of new, or less capable NADCs.

3.0 Next Steps

This report makes a series of recommendations concerning the operation of SCADM and SC-AGI, the two standing groups appointed by SCAR to lead the SCAR data and information management effort. In making these recommendations it is appreciated that much of the data management that occurs within SCAR science projects, is conducted under circumstances outside of the influence of either SCADM or SC-AGI. The network of NADCs on which the SCAR ADMS should be founded needs to be expanded and become interdependent with other, successful thematic and global data networks, that are currently being patronised by SCAR research programs or which could add value to SCAR science.

Several opportunities exist to more closely align SCAR data management with large international data management facilities and networks (notably the ICSU WDCs, IODE, the WMO Information System and the IPY DIS, all of which also need to align themselves with existing niche networks already directly supporting scientific research).

To realise this vision SCAR needs to develop a roadmap to achieve the strategic recommendations in this report in the form of a Data and Information Strategy Implementation Plan. A recent (2009) successful proposal to ICSU, by CODATA (the Committee on Data in Science and Technology, an inter-disciplinary scientific committee of ICSU), has secured funding to support the development of an Implementation Plan for establishing a Polar Information Commons (PIC). This is intended to be a bi-polar approach for creating better access to polar research data, particularly data resulting from IPY. Ideally, development of the SCAR Data and Information Strategy Implementation Plan should be a key component of any emerging PIC Implementation Plan.

Recommendation 15: SCAR needs to develop a Data and Information Strategy Implementation Plan to action the recommendations in this Strategy. A collaboration with ICSU's PIC Steering Committee should be explored to enable joint development of the PIC and SCAR Data and Information Strategy Implementation Plans, so as to ensure complementarity between them.

References

Consultative Committee For Space Data Systems. (2002). Reference Model For An Open Archival System (OAIS). CCSDS 650.0-B-1. Retrieved January 2006 from the WWW: http://public.ccsds.org/publications/archive/650x0b1.pdf.

ICSU. (2007). A Framework for the International Polar Year 2007-2008. Accessed on June 2009. Available athttp://www.icsu.org/Gestion/img/ICSU_DOC_DOWNLOAD/562_DD_FILE_IPY_web_version.pdf

Lattermann, C. & Stieglitz S. (2005) Framework for Governance in Open Communities. *Proceedings of the 38th Hawaii International Conference on System Sciences*, Hawaii, USA.

Rickards L., Capra, A., Candidi M., Huiskes A., Lyons B., Summerhayes C. (2005). JCADM Review. NIOZ, Texel, Netherlands.

Schweik, C. M. & Semenov, A. (2002) The Institutional Design of Open Source Programming: Implications for Addressing Complex Public Policy and Management Problems. *First Monday* 8(1): http://firstmonday.org/issues/issue8_1/schweik/index.html.

Glossary of Terms

Antarctic Data Management System: A distributed, but networked data discovery, data access and data services system capable of supporting SCAR science programs. It would encompass an AntSDI.

AntSDI: An Antarctic themed spatial data infrastructure that provides access to spatial datasets, shared services and tools, that can be accessed from anywhere, supported by appropriate standards, protocols and institutional arrangements.

Data Citation: A method of appropriately referencing datasets used in the development of a product, or scientific research that appropriately acknowledges data originator(s).

Data Encodings: The various methods used to represent data types like integers, floating point numbers, strings, or complex types like objects, arrays and graphs.

Data Management Plan: A plan that outlines what data will be captured and/or generated as part of a project and the methods that will be used to manage, manipulate, publish and archive the data and the resources required.

Data Network: A linked set of data sources (or repositories) that transmit and exchange data.

Data Policy: A policy that outlines the rules governing data availability and access for a particular community or communities and may include guidance on how these data are to be managed and archived.

Datum: A datum is a reference from which measurements are made. In surveying and geodesy, a **datum** is a set of reference points on the earth's surface against which position measurements are made, and (often) an associated model of the shape of the earth (ellipsoid) to define a geographic coordinate system

Gazetteer: A dictionary of place names.

Interoperability: Is the ability of two or more systems, or components to exchange information and to use the information that has been exchanged.

National Antarctic Data Centre: A centre established by a sovereign nation for the purposes of managing Antarctic related data.

Protocols: A formal set of rules, conventions and data structures that govern how computers and other network devices exchange information.

Scientific Grids: Scientific Grids can be defined as an expandable, scalable set of resources applied to solve a single or a set of problems - usually a scientific or technical problem that requires a large number of computer processing cycles. It has an architecture that enables dynamic allocation of resources to varying workloads in accordance with scientific needs.

APPENDIX A

Draft SCAR Data Policy

Executive Summary

SCAR (see www.scar.org) is charged with initiating, developing and coordinating high quality international scientific research in the Antarctic region, and advising on the role of the Antarctic region in the Earth system. The scientific business of SCAR is conducted by its Standing Scientific Groups in the Physical-, Life- and Geo-Sciences which represent the scientific disciplines active in Antarctic research. These groups share information on disciplinary scientific research being conducted by national Antarctic programmes; identify research areas or fields where current research is lacking; coordinate proposals for future research by national Antarctic programmes to achieve maximum scientific and logistical effectiveness; identify research areas or fields that might be best investigated by a major SCAR Scientific Research Programme; and establish Action and Expert Groups to address specific research topics within the discipline.

SCAR related research data is highly multidisciplinary and disparate. This policy aims to provide a framework for these data to be handled in a consistent manner, and to strike a balance between the rights of investigators and the need for widespread access through the free and unrestricted sharing and exchange of both data and metadata. This policy is compatible with the data principles of SCAR's parent body, ICSU and other relevant international agencies (e.g. WMO), and with the goals of Article III 1 c of the Antarctic Treaty.

Since SCAR coordinates a distributed programme of research, generally implemented through a number of nationally self-managed projects, the principles enshrined in this Data Policy should be applied to data in each SCAR-endorsed Project. In order to be considered part of a SCAR Research Program, each Project should follow the SCAR Data Policy, submit metadata and linked datasets to the Antarctic Master Directory (AMD - **gcmd**.gsfc.nasa.gov/Data/portals/**amd**/) in a reasonable timeframe, and should have an appropriately funded data management plan in place before the Project begins.

Nations affiliated with SCAR are urged to establish a National Antarctic Data Centre (NADC) or assign NADC responsibilities to an existing national institution capable of carrying out NADC obligations. NADCs in collaboration with SCAR Research Projects and Programs will work towards developing a SCAR Antarctic Data and Information System (ADMS).

The SCAR Standing Committee on Antarctic Data Management (SCADM) is responsible for this Data Policy. Questions about the policy and its implementation should be directed to the SCADM Executive (see http://scadm.scar.org).

Data Definition

SCAR data are those data generated under the auspices of a SCAR-sponsored Research Project. This policy applies specifically to those data. It should be recognized, however, that SCAR researchers will use SCAR-relevant data from non-SCAR sources, such as from existing operational data networks and historical national sources. Wherever possible, data used in SCAR Projects that are not SCAR-generated, should be treated similarly to SCAR data, where copyright permits and it is practical to treat these data according to SCAR Data Policy norms. A small subset of data both generated and used by SCAR may require special policy and access considerations, because they need to be legitimately restricted in some way. Access to these data may be restricted because they are about human subjects, because there may be intellectual property issues, or because there is a situation where release of the data may cause harm to the public or environment (e.g., the location of nesting sites for an endangered species).

Figure 1 below conceptually illustrates the different classes of data typically used by SCAR Projects.

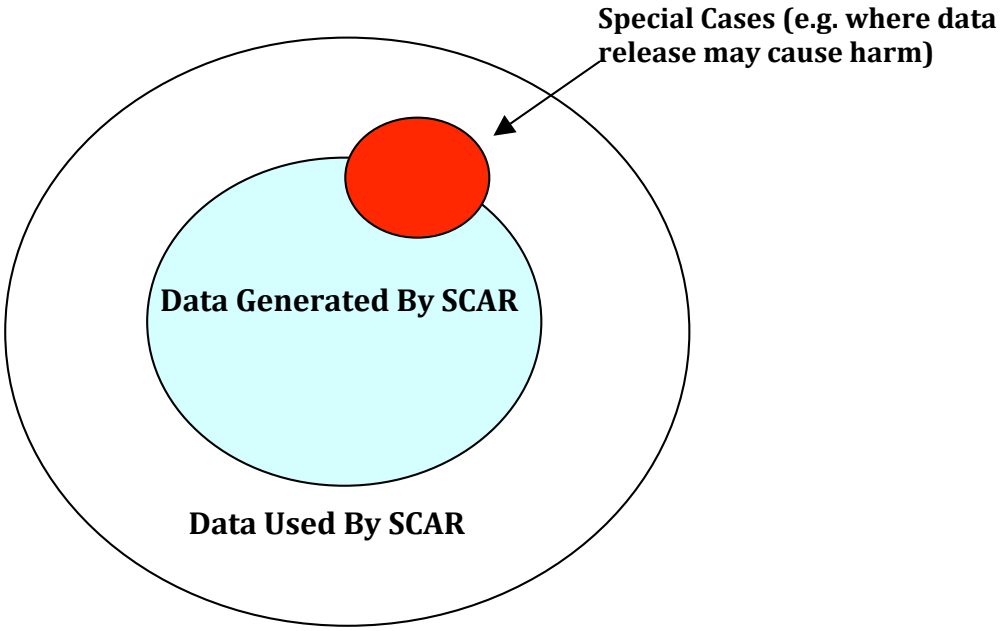

Figure 1. Graphical definition of "SCAR data" (inner blue circle), "SCAR-related data" (outer circle), and Special Cases (red circle).

Data Availability and Exchange

In accordance with

- the Twelfth WMO Congress, Resolution 40 (Cg-XII, 1995)

- the Thirteenth WMO Congress, Resolution 25 (Cg XIII, 1999)
- the ICSU 1996 General Assembly Resolution
- the ICSU Assessment on Scientific Data and Information (ICSU 2004b)
- Article III-1c from the Antarctic Treaty
- the Intergovernmental Oceanographic Commission Data Exchange Policy

and in order to maximize the benefit of data gathered under the auspices of SCAR Projects, the SCAR Executive Committee (EXCOM) requires that SCAR data, including operational data delivered in real time, are made available fully, freely, openly, and on the shortest feasible timescale.

The only exceptions to this policy of full, free, and open access are:

- where human subjects are involved, confidentiality must be protected
- where data release may cause harm, and where specific aspects of the data may need to be kept protected (for example, locations of nests of endangered birds).

ICSU (2004b) defines "Full and open access" as equitable, non-discriminatory access to all data preferably free of cost, but some reasonable cost-recovery is acceptable. WMO Resolution 40 uses the terms "Free and unrestricted" and defines them as non-discriminatory and without charge. "Without charge", in the context of this resolution means at no more than the cost of reproduction and delivery without charge for the data and products themselves.

Metadata are essential to the discovery, access, and effective use of data. All SCAR data should be accompanied by a full set of metadata that completely document and describe the data. In accordance with the ISO standard Reference Model for an Open Archival Information System (OAIS) (CCSDS 2002), complete metadata may be defined as all the information necessary for data to be independently understood by users and to ensure proper stewardship of the data. Regardless of any data access restrictions or delays in delivery of the data itself, all SCAR Projects should promptly provide basic descriptive metadata of collected data to the Antarctic Master Directory (AMD) system.

Data Preservation

Recognizing that the true value of scientific data is often realized long after these data have been collected, and to ensure the lasting legacy of SCAR Projects, it is essential to facilitate long-term preservation and sustained access to SCAR data. All SCAR data should be archived in their simplest, useful form and be accompanied by a complete metadata description. SCADM national contacts (http://scadm.scar.org/index.php?id=368) can help Projects identify appropriate long-term archives and data centers, but it is the responsibility of individual SCAR Projects to make arrangements with long-term archives to ensure the preservation of their data. It must be recognized that data preservation and access should not be an afterthought and needs to be considered when data collection plans are developed. SCADM should work with

the relevant national institutions, NADCs, and other organizations to ensure the preservation of SCAR-related data.

Nations affiliated with SCAR are urged to establish a National Antarctic Data Centre (NADC) or assign NADC responsibilities to an existing national institution capable of carrying out NADC obligations. Providing NADC obligations can be met, a SCAR member country could organise for its Antarctic data to be hosted through a virtual facility formed from the collaboration of several national institutes. Under these circumstances a lead institution must be nominated for the purposes of contact and coordination. The responsibilities of an NADC should include:

- assistance to users in using the AMD and preparing metadata,
- collation of all data generated through national Antarctic science projects,
- provision of data archiving services that permit the long-term re-use of data,
- publication of data from national Antarctic science programs to one or more SCAR-endorsed data access networks, and
- active participation as a node in the ADMS (involving input into the development of network infrastructure standards and conformance with community data standards and protocols)

For an NADC to participate fully as a node in the ADMS it should have, as a minimum:

- a Portal on the AMD, (http://gcmd.nasa.gov/KeywordSearch/amd/nadc_portals.html),
- a publicly accessible web site providing access to national Antarctic data,
- a published operational plan describing how data is managed and archived to permit re-use,
- a publicly accessible national Antarctic data policy complementary to the SCAR Data Policy, and
- a commitment and the capability to publish data to SCAR-endorsed data distribution networks.

An operationally active representative from the NADC should be considered for nomination to participate in SCADM.

Development and ongoing enhancement of the SCAR ADMS should be under-pinned by a Data Strategy and an implementation roadmap.

Data Management Planning

All SCAR-endorsed Projects should be required to prepare a Data Management Plan which outlines how any data captured, modelled or acquired will be managed both during the life of the project and beyond. All Data Management Plans should articulate the resources required to implement the plan and outline where data will be hosted for long-term curation. A Data Management Plan

template is attached which demonstrates the types of issues to be considered in such a Plan (see Appendix below).

Data Acknowledgment Norms

To recognize the valuable contributions of data providers (generally scientists who collect, synthesise, model or prepare analysed data) and to facilitate repeatability of research results, users of SCAR data should formally acknowledge data authors (contributors) and sources. Where possible, this acknowledgment should take the form of a citation, such as when citing a book or journal article. Some journals already require the formal citation of data used in articles that they publish. However, most current journals do not, but as a professional courtesy all data consumers operating under the auspices of SCAR Projects, should formally acknowledge the datasets that they use in their research.

References

The Antarctic Treaty. 1959. Available at http://www.ats.aq/

CCSDS (Consultative Committee for Space Data Systems). 2002. *Reference Model for an Open Archival Information System (OAIS).* CCDSD 650.0-B-1. Blue Book. Issue 1. Washington, DC: CCSDS Secretariat. [Equivalent to ISO 14721:2002]. Available at http://public.ccsds.org/publications/archive/650x0b1.pdf

Intergovernmental Oceanographic Commission Oceanographic Data Exchange Policy. 2003.Available at http://www.iode.org/contents.php?id=200

ICSU (International Council for Science). 2004b. *ICSU Report of the CSPR Assessment Panel on Scientific Data and Information.* Available at http://www.icsu.org/1_icsuinscience/DATA_Paa_1.html

Resolution of the International Council of Science (ICSU) General Assembly (24-27 September 1996). Available at http://globalchange.gov/policies/icsu-1996.html

World Meteorological Organization Congress, Resolution 40 (Cg-XII, 1995): *WMO Policy and Practice for the Exchange of Meteorological and Related Data and Products Including Guidelines on Relationships in Commercial Meteorological Activities* Available at http://www.nws.noaa.gov/im/wmocovr.htm

World Meteorological Organization Congress, Resolution 25 (Cg XIII, 1999). *Exchange of Hydrological Data and Products*. Available at http://www.wmo.ch/web/homs/documents/english/res25eng.html

Appendix to SCAR Data Policy

Data Management Plan Template

The purpose of a data management plan is to anticipate likely data capture activities and data flows within a project and cognisant of desired project outcomes and outputs, it should spell out a comprehensive strategy for managing data, metadata and products associated with the project. The data management plan should provide best practice guidance to project participants on mechanisms to efficiently manage project data entities while the project is being executed, as well as suggest practices that will ensure the longer-term re-usability of project data once the project is complete.

In such a document "efficiency" and "best practice" are characterised by management actions that:

(a) Track the provenance of data,
(b) Adequately annotate data and products so that their fitness for use can readily be judged by potential users (including those not connected with their generation),
(c) Provide for data/product discovery,
(d) Minimise the number of redundant copies of data within the system,
(e) Minimise the number of times data transformations are unnecessarily undertaken (e.g. format changes as data is passed between project participants),
(f) Protect the accidental release of any embargoed or confidential data,
(g) Provide intuitive data access mechanisms, and
(h) Preserve source data and its subsequently processed variants for future use.

A data management plan typically addresses six basic data management functions, examines the inter-relationships between these functional activities and makes recommendations about 'best practice" techniques which should be used within the context of a particular project. It also identifies the resources anticipated to implement the plan. The six data management functions addressed include:

- Data capture
- Post capture data processing
- Data storage
- Data analysis/product generation
- Data publishing
- Data archiving

A typical Table of Contents for a Data Management Plan might include the following topics and a sample Data Management Plan can be found at (http://SCADM.SCAR.ORG/XXX):

1.0 Data Management Plan Objectives & Project Overview

1.1 Data management plan objectives
1.2 Brief project background and rationale
1.3 Anticipated project output and outcomes
1.4 Anticipated project deliverables and timelines

2.0 Data Capture

2.1 Description of discrete data capture activities
2.2 Description of all observed/measured data types and anticipated data volumes.
2.3 Data capture/sampling protocol issues
2.4 Data flow from field/lab to main data stores

3.0 Post Collection Processing

3.1 Descriptions of required image, video (?) processing
3.2 Descriptions of required physical sample (?) processing
3.3 Descriptions of required observation/measurement (?) processing

4.0 Project Data Stores

4.1 Descriptions of managed data stores

4.1.1 Local project data-store(s)
4.1.2 Central or distributed data-store(s)
4.1.3 Use of sample tracking database/systems (?)
4.1.4 Metadata System

4.2 Descriptions of data-store data access methods for project personnel

4.2.1 Local project data-store(s)
4.2.2 Central or distributed data-store(s) & metadata
4.2.3 Sample Tracking Database/System (?)

5.0 Data Analysis

5.1 Analysis methods affecting data management
5.2 Provenance tracking for post-processed and analysed data

6.0 Data Publishing

6.1 Description of method(s) that will be used to make data accessible post project completion.
6.2 Data citation and usage guidelines

7.0 Data Archiving

7.1 Archiving practices
7.2 Archiving facilities

8.0 Resource Requirements