Scientific Committee On Antarctic Research (SCAR)

Data and Information Strategy

2008 - 2013

Draft SCAR Report 33

Prepared by

Joint Committee on Antarctic Data Management (JCADM)
&
Standing Committee on Antarctic Geographic Information (SC-AGI)

This document is a draft version -originally prepared in June 2007 with edits made up until May 2008.

Authored by
Kim Finney (Manager, Australian Antarctic Data Centre)
Australian Antarctic Division
kim.finney@aad.gov.au
2008

On behalf of JCADM and SC-AGI

Edited by
JCADM & SC-AGI Executive and members

Table of Contents

Page No.

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.

Data Quality: Data quality refers to a dataset's fitness for use for a particular purpose.

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

Feature Catalogue: A Feature Catalogue provides detailed descriptions of real world objects or features and may contain information about synonyms, the source of the definition and the relationships that might exist between different features.

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 governs how computers and other network devices exchange information.

Registers: A formal place for recording specific types of information. Web-based service registries often record information on where a particular service can be found, who owns it and how it can be accessed.

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.

Symbology Catalogue: A catalogue of symbols used to display or portray certain types of data, features or objects.

Topology: Topology stores the relationships of one spatial element with respect to another.

Virtual Observatories: A term originally connected with solar and astronomical observations where data repositories are made accessible through grids and web services. This term is now routinely applied to any type of observational data accessed in this manner.

Executive Summary

International, Antarctic-based research is loosely coordinated under the auspices of the Scientific Committee on Antarctic Research (SCAR), which is an inter-disciplinary committee of the International Council of Scientific Unions (ICSU). SCAR is governed by an Executive Committee, National Delegates and an Executive Director, and operates via Scientific Standing Groups. ICSU and the World Meteorological Organization (WMO) are responsible for coordinating the 2007-09 International Polar Year (IPY), a globally coordinated set of science projects with the capacity to significantly advance human understanding of earth systems. SCAR is contributing directly to the management of the IPY 2007-09, which is the fourth in a set of International Polar Years that began in 1882. Each International Polar Year left a global legacy. The third IPY became known as the International Geophysical Year (IGY) of 1957-58. One part of the legacy of the IGY was the creation of SCAR in 1958, to continue the coordination of Antarctic research begun during the IGY. Another part of the legacy of the IGY was the creation of the World Data Centre System. Post 2008 it is hoped that IPY will leave a series of long-lived, global observing systems capable of monitoring and detecting changes in the behaviour of the oceanic, atmospheric, cryospheric and biospheric elements of the overall Earth System. The need for such observatories has never been more important in the history of mankind given recent consensus on the existence of human-induced climate change and its potential impact on society and natural systems.

Sustained and effective global observing systems will be collaborative ventures shared by nations and under-pinned by professional data management, agreed data and information standards, and shared data and information infrastructure. This infrastructure will require planning, a coordinating framework, governance and appropriate resourcing. In addition to being able to draw upon national research programs, SCAR members are fortunate in having already put in place collectively some of the elements of the infrastructure required for supporting these sustainable observing systems.

In 1992 SCAR presented a proposal (No. WP5) at the Antarctic Treaty Consultative Meeting (ATCM XVII) for development of an Antarctic Data Management System which was agreed to by all Parties. By 1998 SCAR and the Council of Managers of National Antarctic Programs (COMNAP) reported to ATCM (XXII) that they had established a Joint Committee of Antarctic Data Managers (JCADM) to pursue development of the system. More recently a SCAR group of experts providing advice and services related to building an Antarctic Spatial Data Infrastructure (AntSDI) was elevated to become a SCAR Standing Group (the Standing Group on Antarctic Geographic Information – SC-AGI). JCADM has been in operation for over 10 years and has been successful in raising the profile of data management within the SCAR community, but its work to date has not been supported by a comprehensive Strategy that defines how an effective Antarctic Data Management System (ADMS) could be implemented. Similarly, SC-AGI has developed a range of quality products that could be used to build components of a spatial data infrastructure, but they are not in wide use within the Antarctic community and there is no roadmap detailing how an AntSDI could be developed. Recognising these deficiencies, the SCAR Executive called for JCADM to develop a SCAR Data and Information Strategy. In 2006 COMNAP, co-sponsor of JCADM, reported to the SCAR Executive Committee that it would be withdrawing financial support for JCADM activities in June 2008 stating that "*The group now showed maturity and seemed to have* *reached the point where it could sustain itself".* Accordingly, from July 2008 JCADM will seek to become a Standing Committee of SCAR.

Theses pages describe a SCAR Data and Information Strategy. The Strategy recommends that the number and alignment of the current, loosely federated system of independently operating National Antarctic Data Centres (NADCs), be upgraded and augmented to create a sustainable and network-based infrastructure capable of meeting the data and information management challenges that accompany the technological advances which we now enjoy at the beginning of 21st Century. It also provides the larger and currently missing context for existing data and information bodies like JCADM and SC-AGI, each of which addresses a part of the overall picture – but currently in a relatively uncoordinated way. An important constraint applied throughout the Strategy is that a SCAR Antarctic Data Management System should be interoperable with other existing global infrastructures and initiatives, particularly with the reform of the ICSU World Data Centre System, and should leverage existing and emerging global standards and protocols wherever possible.

The Strategy indicates that several obstacles must be overcome to build a better data management capability in support of SCAR science. There are significant implications for both JCADM and SC-AGI.

Inadequate resourcing of the JCADM Executive and JCADM/SC-AGI activities severely limits the development of data and information management services for SCAR's science. The Strategy recommends that:

- SCAR examines the feasibility of staffing the JCADM Chair role with someone able to devote the majority of their time to implementing the Strategy. Possible mechanisms for doing so are outlined in Section 5.2.3.
- it should be SCAR Policy that Membership entails the establishment of an NADC. If retrospectively applied, this would bolster national resources committed to participation in SCAR data management activities. Section 5.2.2 of the Strategy discusses this issue and its implications.
- a closer working relationship be fostered between SCAR science programs, NADCs and JCADM, through an obligation on endorsed SCAR science programs to submit a data management plan as part of the SCAR proposal and endorsement process. By getting adequate data management factored into science proposals there is a greater chance that data management aspects of the science projects will be better resourced and funded. JCADM should be called upon to assist in developing such Plans and to provide a template to ensure consistency in the matters that should be addressed. Section 5.2.1 elaborates on why data management plans are recommended and the benefits that they could bring.

To date SCAR data management has mainly been undertaken as the marginal cost of doing science. The task has been accomplished largely through voluntary labour provided by Members, particularly where shared infrastructure services have been developed. Many of the more mature national programs are now experiencing decreases in their program budgets making it difficult to continue substantive voluntary contributions that primarily support the common good, as opposed to national interests. A focus on national interests neglects the benefits that can accrue through taking a pan-Antarctic approach. Now is the time to start seeking external funding sources to support some of the key data management activities required to develop the pan-Antarctic Data Management System (ADMS) from which all Members should benefit. Section 5.4.2 makes this case and also posits that linking with a small number of well-resourced, emerging data networks would give SCAR a kick-start in terms of leveraging some existing core capability.

There is still a strong culture amongst scientists NOT to share data, despite Treaty Parties agreeing to make scientific data public under Article (III).(1).(c) of the Treaty. This culture is being reinforced by existing reward and recognition systems operating within national science programs. Section 5.3 discusses this issue and suggests that a recognised data citation system, that operates similarly to a publication citation system, could influence scientists to publish more data. The SCAR Executive, possibly through IPY activities, or in conjunction with an appropriate peak library or publication body, is in a position to explore the development of a SCAR data citation system that would encourage scientists to publish data by providing recognition.

The data products and services already developed by SC-AGI members at considerable cost and effort on the part of national contributors are not being effectively exploited. A closer partnership between SC-AGI members, who develop these products, and potential consumers, such as JCADM's NADCs, should be formed to better utilise these tools. Such a partnership would provide an excellent mechanism for fine-tuning these tools to meet actual user needs and for weeding out developments that are not operationally robust or useful. The need for closer cooperation between these two SCAR groups is outlined in section 5.4.1.

The restricted nature of available resources to undertake development work related to building data systems will require an incremental approach to building an ADMS encompassing a Spatial Data Infrastructure (SDI). As core capability grows over time, more users and developers should be attracted to assist in growing the capability further because of the intrinsic rewards that are more likely to be obtainable from participating in the development of a functioning system. To start this approach, the Strategy suggests setting up a series of well-bounded pilot projects that could begin to build the necessary components of an ADMS, ideally centred around developing virtual observatories to support observing networks. Section 5.5 discusses this concept and provides examples of the types of activities that might be pursued.

Building the foundation data layers of the proposed AntSDI could be greatly facilitated by closer cooperation amongst those nations with an active national Antarctic mapping program. As a first step, obvious gaps in data availability should be identified and agreed upon. Secondly, forecasts of likely mapping activity by national programs should be sought to assess how these gaps might be best filled. SC-AGI members should be able to readily assist in these exercises given information that they already have at hand and because of their links into national mapping programs. Once this information is available, SC-AGI should provide advice to SCAR, on possible mechanisms to fill key gaps. This issue is also discussed in more detail in Section 5.5.

Capacity-building will be crucial to develop the work force and to spread the work-load, currently predominantly carried by the more mature SCAR contributors. Section 5.4.4 outlines a range of possibilities for training and mentoring activities.

The Strategy concludes by anticipating that both JCADM and SC-AGI will develop comprehensive work-plans derived from guidance presented in this Strategy and will regularly work together to review progress and adapt this Strategy to fit emerging opportunities or changes that will inevitably occur in SCAR's operating contexts.

The overall vision pursued by this Strategy is provided in the form of two hypothetical testimonials which immediately follow this Executive Summary.

Strategic Vision

Consider the following imaginary 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. This is the vision pursued by 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 a 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 that has been promoted by JCADM, and is now widely accepted as a legitimate performance indicator by SCAR participants, 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. 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.

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."

Summary of Recommendations

Table 1 summarises all the recommendations made within the body of the strategy. The recommendations and tasks have been prioritised, and with the exception of the SCAR strategic tasks, have been grouped under the headings used within the JCADM and SCAR Terms of Reference (the terms of reference are highlighted below by underlined text). This is to facilitate the development of work-plans for each group focussed around the recommendations in this strategy, and the Terms of Reference of each group.

Table 1 –SCAR and JCADM strategic tasks

SCAR Strategic Tasks

High Priority	**Responsible Party**
1. *SCAR Executive to invite CCAMLR to join JCADM.*	SCAR Executive
2. *SCAR Executive to request that all SCAR endorsed projects develop a data management plan.*	SCAR Executive
3. *SCAR Executive ensures that new SCAR members are aware of data centric obligations and that existing SCAR members are made aware of these retrospectively.*	SCAR Executive
4. *SCAR Executive considers options for appointing a full-time JCADM Chair.*	SCAR Executive

Medium-Term Priority	**Responsible Party**
1. *SCAR seeks a partnership with an appropriate institution, industry body, or Library peek body to pursue the concept of instituting a dataset citation system.*	SCAR Executive

JCADM Prioritised Recommendations

<u>To assist in establishing Antarctic data management policies, priorities and best practices</u>

High Priority	**Responsible Party**
1. *JCADM tasked to develop a SCAR Data Policy.*	JCADM
2. *SCAR Executive to request that all SCAR endorsed projects develop a data management plan. JCADM should provide a plan template and advice to the SCAR Executive on the adequacy of any developed plans prior to projects receiving SCAR endorsement.*	JCADM
3. *JCADM develops a profile of a typical NADC, including its functions and obligations as part of the ADMS.*	JCADM

4. *JCADM should review the tasks it undertakes at face-to-face meetings and the frequency of these meetings. The JCADM Chair should encourage the use of new communication technologies to pursue inter-sessional work.*	JCADM Executive
5. *JCADM and SC-AGI review their use of the web and better target content and its communication to different audience types.*	JCADM
6. JCADM and SC-AGI should ensure that their services/products and work-programs are well-known to all SCAR national delegates and scientific research programs through targeted information dissemination campaigns	JCADM

<u>Promote long-term preservation and accessibility of data relating to Antarctica and the Southern Ocean in sustainable repositories.</u>

Medium Priority	**Responsible Party**
1. *JCADM Chair explores the potential for establishing regional data hosting facilities.*	JCADM Executive
2. *JCADM should review the types of data centric services it can supply and work towards the provision of a well-bounded set of services to increase its utility to SCAR science projects.*	JCADM

<u>To support the establishment and ongoing work of National Antarctic Data Centres, in accordance with ATCM XXII Resolution 4.1 (1998),</u>

High Priority	**Responsible Party**
1. *JCADM forms an NADC Technical Committee to focus on the technical aspects of building the ADMS.*	JCADM
2. *JCADM Chair ensures that members of the NADC Technical Committee are included in any Technical Committees of partner data access networks.*	JCADM Executive
3. JCADM to instigate a 'buddy system whereby more developed NADCs are allocated as mentors to less developed NADCs.	JCADM Executive

Medium Priority	**Responsible Party**
1. *JCADM, lead by the JCADM Chair should develop collaborative funding proposals targeting key European and US funding bodies. All JCADM members should keep a watching brief on calls for proposals.*	JCADM Executive
2. *JCADM should investigate the training requirements of the newer NADCs and develop a training program, harnessing the expertise in the longer-established NADCs.*	JCADM

To provide linkages to other relevant data management systems and thereby enhance the ADMS

High Priority	**Responsible Party**
2. *JCADM in liaison with SCAR science bodies chooses a (small) number of existing and complimentary data access networks with which to be affiliated and then endorses and promotes NADC collaboration in these networks.*	JCADM Executive

Medium-Longer Term Priority	**Responsible Party**
1. *A small number of pilot projects should be endorsed, scoped and implemented to demonstrate capability, develop skills and incrementally develop the ADMS.*	JCADM
2. *JCADM through its NADCs should participate in data access trials of existing larger networks, which are using new technologies such as web services. NADCs should also contribute to data modelling activities within these networks aimed at supporting semantic data interoperability.*	JCADM

To encourage submission of metadata and data to the Antarctic Data Management System

High Priority	**Responsible Party**
1. *JCADM and the GCMD explore better ways to link datasets directly to metadata records and improve user interfaces.*	JCADM/GCMD

To further improve and populate the AMD and provide guidance to the AMD host

High Priority	**Responsible Party**
1. *JCADM Chair to request that the GCMD provide a forecast of future directions of the GCMD over a 3-year time horizon and ensure that JCADM members are able to provide feedback which is taken into account by the GCMD.*	JCADM Executive

In partnership with SCAGI, to work with SCAR SSGs, COMNAP and the Antarctic Treaty Secretariat to identify and develop fundamental datasets of value to the Antarctic Community.

Medium Priority	**Responsible Party**
1. *JCADM to partner with SC-AGI in expanding the use of existing*	JCADM/SCAGI

SC-AGI tools into the broader SCAR science and global data management community. JCADM to provide user-feedback on the tools.	

SC-AGI Prioritised Recommendations

Provide Antarctic fundamental geographic information products and policies in support of science programs.

High Priority	**Responsible Party**
1. *SC-AGI to continue development, maintenance and management of products such as the SCAR Composite Gazetteer, Feature Catalogue, Map Catalogue, Antarctic Digital Database and Symbology Catalogue*	SC-AGI

Medium Priority	**Responsible Party**
1. *JCADM to partner with SC-AGI in expanding the use of existing SC-AGI tools into the broader SCAR science and global data management community. SC-AGI to use this experience to further refine community products.*	JCADM/SCAGI

Integrate and coordinate Antarctic mapping and GIS programs.

High Priority	**Responsible Party**
1. *SC-AGI to undertake an audit of the SCAR Antarctic Digital Database (ADD) to ensure that it is meeting "user" needs, integrates where necessary with other SC-AGI products, and provides an example of best practice in terms of governance and production methods.*	SC-AGI

Medium Priority	**Responsible Party**
1. *SC-AGI to provide an assessment of the current extent and scale of mapping across Antarctica and a forecast of likely future national or global mapping activities, with the aim of identifying gaps.*	SC-AGI

Long-Term Priority	**Responsible Party**
1. *SC-AGI to identify opportunities for collaborative capture and/or compilation of mapping data, particularly in regions of high SCAR*	SC-AGI

scientific activity or where there are gaps in desirable coverage to progressively create a seamless map coverage of Antarctica at appropriate scales.	

<u>Promote an open standards approach to support free and unrestricted data access.</u>

High Priority	**Responsible Party**
1. *SC-AGI to promote the development of standards, protocols and technologies that permit the integration of spatial data and the interoperation of national Geographic Information Systems.*	SC-AGI

<u>Promote capacity building within all SCAR nations.</u>

High Priority	**Responsible Party**
1. SC-AGI and JCADM should review the tasks it undertakes at face-to-face meetings and the frequency of these meetings. The SC-AGI Chair should encourage the use of new communication technologies to pursue inter-sessional work.	SC-AGI Executive
2. SC-AGI and JCADM should ensure that their services/products and work-programs are well-know known to all SCAR national delegates and scientific research programs through targeted information dissemination campaigns	SC-AGI

1.0 Scope and Purpose of This Strategy

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 are embarking on another International Polar Year (IPY 2007-2008). It could be argued that the timing yet again is highly opportune given the massive advances in technology that have been evident over the last decade and since IGY.

From the mid 1990s we have witnessed rapid increases 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. For example, it is now possible for us 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 activity. 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 lead to an explosion of information. For example, 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 1950, computers are now ubiquitous in their use. 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 indeed been a quantum shift in our scientific capability.

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. The current IPY also has plans to provide a legacy, and its focus is on leaving 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.

Given this context, the following Data and Information Strategy focuses on the framework required to enhance the current data management system supporting science in the Antarctic and sub-Antarctic polar regions, so that it can provide the basis for a suitable data and information management system for the future as a key legacy of the IPY.

A significant amount of Antarctic-based research is loosely coordinated by the Scientific Committee on Antarctic Research (SCAR) through collaborations amongst SCAR Members (comprising countries represented through their national academies, and international scientific unions with Polar interests) primarily via national science programs. In 1992 SCAR, in collaboration with the Council of Managers for National Antarctic Programs (COMNAP) agreed to appoint a Joint Committee on Antarctic Data Management (JCADM) to pursue the development of an Antarctic Data Directory System. JCADM has been in operation for over 10 years and has been successful in raising the profile of data and information management within the SCAR community, but its work to date has not been supported by a comprehensive Strategy that defines the goals of an Antarctic data management system, and there is no community-agreed roadmap which outlines how such goals are to be achieved. The need for a data and information strategy to fill this gap was clearly identified by the SCAR Strategic Plan 2004-2010, and confirmed by the independent external review of JCADM that took place in Texel, The Netherlands, 31 March – 1 April, 2005.

The Data and Information Management Strategy outlined below addresses this gap. The recommendations made in the Strategy are designed to guide an upgrade to the current data and information management system. Importantly it hopes to stimulate the creation of a more sustainable and 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. An important constraint applied throughout the Strategy is that an Antarctic Data Management System should be interoperable with other existing global infrastructures and initiatives, particularly with the reform of the ICSU World Data Centre System and should leverage existing and emerging global standards and protocols wherever possible.

This document first provides an overview of SCAR and the SCAR data management coordinating bodies and then summarises the generic data management issues that are being faced by SCAR science projects. This is followed by an analysis of how Antarctic data management currently functions, noting positive and negative features of the model. The final section of the document makes a series of strategic recommendations to strengthen SCAR data management, covering the following:

- Components necessary for an Antarctic Data Management System (ADMS)

- Governance, Partnerships & Coordination
- Data Management Culture
- Leveraging Resources and Systems
- Standards & Interoperability Issues
- Education, Outreach & Guidance

The report concludes with a prioritisation of the recommendations made in order to provide guidance for the development of annual or 2-yearly JCADM and SC-AGI work-plans. These plans should contain specific tasks and implementation detail sufficient to achieve the objectives articulated in this Strategy.

2 Overview of SCAR, JCADM and SC-AGI

2.1 SCAR

SCAR is an inter-disciplinary committee of ICSU. It 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 Earth systems. The scientific business of SCAR is conducted by its Standing Scientific Groups, which represent the scientific disciplines active in Antarctic research.

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. The Antarctic Treaty System (ATS) is the whole complex of arrangements made for the purpose of regulating relations among States (known as Parties) in the Antarctic. At its heart is the Antarctic Treaty itself. The original Parties to the Treaty were the 12 nations active in the Antarctic during the International Geophysical Year of 1957-58. The Treaty was signed in Washington on 1 December 1959 and entered into force on 23 June 1961. The Consultative Parties comprise the original Parties and a further fourteen States that have become Consultative Parties by acceding to the Treaty and demonstrating their interest in Antarctica by carrying out substantial scientific activity there.

The primary purpose of the Antarctic Treaty is to ensure "in the interests of all mankind that Antarctica shall continue forever to be used exclusively for peaceful purposes and shall not become the scene or object of international discord" (Antarctic Treaty, 1961). To this end it prohibits military activity, except in support of science; prohibits nuclear explosions and the disposal of nuclear waste; promotes scientific research and the exchange of data; and holds all territorial claims in abeyance. The Treaty applies to the area south of 60° South Latitude, including all ice shelves and islands.

SCAR supports Treaty activities through the work of its Members who comprise the appropriate bodies of those national scientific academies, or research councils, which are the adhering bodies to ICSU and which are, or plan to be, active in Antarctic research, together with the relevant scientific Unions of ICSU. There are over 42 nations and organisations represented through SCAR (see SCAR web site at www.scar.org).

The main programs of SCAR are carried out primarily through its Standing Committees, of which there are 7, comprising:

- Standing Scientific Group on GeoSciences (SSG-GS)
- Standing Scientific Group on Life Sciences (SSG-LS)
- Standing Scientific Group on Physical Sciences (SSG-PS)
- Joint Committee on Antarctic Data Management (JCADM)
- Standing Committee on the Antarctic Treaty System (SC-ATS)
- Standing Committee on Antarctic Geographic Information (SC-AGI)
- Standing Committee on Finance

The Standing Scientific Committees coordinate scientific research. They are responsible for:

- Sharing information on disciplinary scientific research being conducted by national Antarctic programmes;
- Identifying research areas or fields where current research is lacking;
- Coordinating proposals for future research by national Antarctic programmes to achieve maximum scientific and logistical effectiveness;
- Identifying research areas or fields that might be best investigated by a SCAR Scientific Research Programme and establishing Scientific Programme Planning Groups to develop formal proposals to the Executive Committee; and
- Establishing 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, but with the coordination and leadership provided by two Standing Committees, these are the:

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

2.2 JCADM

The SCAR–COMNAP sponsored group that predated JCADM was the ad-hoc Planning Group on Antarctic Data Management. The terms of reference for the Planning Group included the development of a plan for the coordination and management of Antarctic data, taking into account SCAR's programmes and requirements under the Antarctic Treaty System, especially with respect to the Protocol on Environmental Protection (See Antarctic Treaty Secretariat web site at www.ats.aq). The Planning Group's first report (October 1992) proposed a committee for Antarctic data management whose role would be 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 recommendations of this report were accepted by the Antarctic Treaty Meeting in November 1992, and by the SCAR and COMNAP Executives in April 1993. As recommended, JCADM replaced the Planning Group on May 23, 1997. In the following year the Antarctic Treaty Parties re-confirmed their commitment to Article III.1.c of the Treaty, which urged Parties to exchange scientific observations and results from Antarctica and make them freely available (ATCM XXII). ATCM Resolution XXII-4 (1998) was also passed urging members, who had not yet done so, to establish National Antarctic Data Centres and link these to the Antarctic Data Directory System managed by JCADM. Further, in the same resolution Consultative Parties and their National Antarctic Data Centres were asked to 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 National Antarctic Data Centres for distribution through the Antarctic Data Directory System.

Some ten years on from its formation, JCADM now comprises 30 national representatives reponsible for Antarctic data management within their respective jurisdictions. JCADM's purpose is to advise SCAR on 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). JCADM 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. JCADM, in conjunction with the NASA-based, Global Change Master Directory (GCMD) has developed the Antarctic Metadata Directory (AMD) System to manage these metadata records. JCADM must now compliment this system with other components, that can provide a more holistic framework for Antarctic data management.

The JCADM Chief Officer reports anually to the SCAR Executive Committee (EXCOM) . The Chief Offier is elected by EXCOM , 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. In 2006 JCADM also appointed 4 Liaison Officers whose roles are to forge links with the SCAR Science Standing Groups and the CEP, to facilitate communication between JCADM and SCAR scientists.

2.3 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 JCADM, the Chief Officer is 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 JCADM 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 in the future. In practical terms there is very little difference between the stated goals of the two groups, as both want to initiate a comprehensive system for managing and utilising data and information derived from Antarctic research and national Antarctic operations.

3.0 General Data Management Issues Facing SCAR Science Groups

Through the activities of its 3 Standing Scientific Groups, their associated expert and action groups, and SCAR's 5 coordinated research programs (http://www.scar.org/researchgroups/), SCAR members are involved in capturing an extremely diverse range of data types. The individual data management requirements of all SCAR related activities are as varied as the data types being captured, but there are a number of commonly recurring requirements, which can be exemplified by reference to specific SCAR research interests that should be generically addressed in a whole-of-SCAR Data and Information Strategy. These basic requirements are articulated and discussed below.

3.1 Data Discovery

The trend in funding science is towards support for multi-disciplinary research. Given this trend and the increased capacity for even poorly funded researchers to conduct sophisticated, temporally and spatially repeated measurements without belonging to a large SCAR endorsed project, it is no longer reasonable to assume that all applicable sources of data are well known to scientists. Some required data can often be sourced from outside of Antarctic national programmes and may be captured outside of the Antarctic region of interest. These situations, combined with two key objectives of SCAR - which are to (a) share information on disciplinary scientific research being conducted by national Antarctic programmes and (b) to identify research areas, or fields, where current research is lacking, indicates that SCAR science needs effective ways to advertise and consume information about all of the datasets available that can support Antarctic research.

JCADM invested the majority of its effort in its formative years in providing scientists with a flexible data discovery mechanism through the the vehicle of the AMD. This tool is capable of providing a window into data of significance to Antarctic science, but it is also part of a system with broader thematic and geographical coverage with the capability to link to other catalogues of interest. Adequate data discovery mechanisms are therefore already in place for SCAR scientists to leverage, but these tools are only as useful as the content they operate on. Although JCADM has expended a considerable amount of energy in promoting an Antarctic metadata system, the system is still relatively under-patronised given the scope of the data that is regularly being collected by national Antarctic programmes. Of significance to the success of any such system as the AMD are the cultural issues associated with the motivation of individuals to contribute content when the “pay-offs” for contribution might appear nebulous in comparison to the effort required to register requisite information.

Scientists need to be shown the benefits of publicising their data for broader use, and until there is sufficient traction (as evidenced by systems containing large volumes of readily searchable data) artificial incentives may need to be applied to motivate contributions. It is worth bearing in mind that this is equally true of attempts to develop national data and information management systems. The cultural problem is generic and has led, for example, to individual countries, and entities like the European Commission, developing carrot and stick data and information policies to encourage the population of data banks.

3.2 Data Access

Data streams produced through SCAR activities can be thought of as belonging to three groups: real-time data, delayed mode data and modelled data.

Real-time data is any data stream that is conveyed in “real”, or “near-real” time. This type of data is usually transported:

- using the Global Telecommunications System (GTS) (e.g. meteorological data exchanged regularly under the auspices of the World Meteorological Organisation);

- via global networks such as the ARGOS telemetry system (as used by many biological programs tracking animal movements, or Argo floats); or,
- using dedicated private networks established specifically to capture and transmit data from in-situ sensors to one or a number of collaborators (e.g. automatic weather station data captured by national programmes that is relayed back to a local Antarctic base or national headquarters).

Very little post-processing is performed initially on these data before they are received by users.

In contrast, delayed-mode data is generally more highly processed, may be qualified and is available some time after the collection event. Nearly all SCAR research programmes utilise this type of data.

Modelled data can be envisaged as synthetic data in that it is generated from algorithms (often embedded in software), rather than being captured from direct measurement and observation. This category of data is generated by most SCAR science programmes at some stage in their research (e.g. modelled habitat regions, predicted ocean circulation patterns, and upper atmosphere thermal tides).

There are many existing initiatives underway aimed at establishing on-line data access networks that cover these different types of data. Some of the larger and more relevant activities to SCAR members include:

(a) IODE Ocean Data Portal (http://www.iode.org): A system that will use web-oriented technologies to access non-homogeneous and geographically distributed marine data and information. It is intended to be a "system of systems" but will require conformance to a range of agreed standards.

(b) SeaDataNet (http://www.seadatanet.org/): A pan-European marine data and information infrastructure built to eventually link 35 data centres involved in data management by integrating databases of standardised quality on-line. The quality, compatibility and coherence of the data issuing from so many sources, will be ensured by adopting standardized methodologies for data checking, and by dedicating part of the activities to training and preparation of synthesised regional and global statistical gridded products from the most comprehensive in-situ and remote sensing data sets made available by the participants.

(c) Global Biodiversity Information Facility - GBIF (http://www.gbif.org/): GBIF is an infrastructure for sharing biodiversity related information from the molecular to the ecological level.

(d) Ocean Biogeographical Information Systems - OBIS (http://www.iobis.org/) and SCAR-MarBIN (http://www.scarmarbin.be/): These two related global systems are concerned with sharing marine biodiversity data. MarBIN is exclusively Antarctic, and provides the Southern Ocean node of OBIS.

(e) BioMoby (http://www.biomoby.org/): The MOBY-S system defines an ontology-based messaging standard through which a client is able to automatically discover and interact with task-appropriate biological data and analytical service providers, without requiring manual manipulation of data formats as data flows from one provider to the next. Used primarily by genomics and plant scientist.

(f) The Geosciences Network - GEON (http://www.geongrid.org/index.php): GEON is a collaboration among a dozen PI institutions and a number of other partner projects, institutions, and agencies to develop cyberinfrastructure in support of an environment for integrative geoscience research.

(g) SuperDarn (http://superdarn.jhuapl.edu/) is an international radar network for studying the upper atmosphere, so involves data sharing between national data providers.

While SCAR members may be individually involved in one or a number of such initiatives, there is no planned or coordinated effort amongst NADCs to patronise any particular systems, possibly with the exception of SCAR-MarBIN, which directly supports the SCAR Census of Antarctic Marine Life Project and is being actively promoted by JCADM.

It should also be mentioned that there are some "super" networks planned for the future, which intend to bring together existing and new hardware, software and newtorks to provide integrated and comprehensive data to support environmental decision-making. The Global Earth Observing System of Systems (GEOSS) (http://www.epa.gov/geoss/) is one such programme. These ambitious "system of systems" projects are yet to get any real traction, but they do have considerable political support globally and therefore need to be kept visible on the horizon. GEOSS comprises land, ice, ocean and atmosphere components (such as the Gloal Ocean Observing System for instance), each of which currently has its own way of managing the data relevant to its needs.

3.2.1 Duplicates, Versioning and Language Issues

All data access systems that attempt to aggregate data from disparate sources, particularly those with delayed mode and real-time data feeds, face similar data management obstacles. Effective and comprehensive solutions for many of these issues are still under development.

For example, real-time data streams may not contain the same degree of data resolution as their delayed mode counterparts, usually because filtering has been applied before publishing to reduce data transmission volumes. Because both types of dataset (real-time and delayed mode) can be transported over the same network, it is not always immediately apparent which dataset is the parent and which is the filtered derivative.

In addition, delayed-mode data may be processed or subjected to quality control measures several times, by any number of individuals, particularly in collaborative pursuits. After each set of processing another set of slightly different, but otherwise similar data are produced. Again, all of these variants may be launched over the same network but will often lack a reliable electronic method of identifying which datasets are original source files and which are derivatives.

Similarly, a model may be run several times each time with slightly different parameterisations or random starting points, and so many modelled datasets may be available for almost identical generating scenarios. It is important for both a researcher and a data manager to be able to identify that some data streams are simply different versions derived from the same source.

In addition to being able to identify different versions of a dataset, it is also important to be able to detect when two or more datasets are duplicates of each other. In an era where the Internet is the main data access medium, and it is now popular practise to establish aggregator portals delineated by theme, research programme, discipline, or geographic region, it is particularly important that we are able to detect duplicate data. The utility of search and discovery tools is quickly compromised if large volumes of data that are retrieved by users are found to consist of numbers of duplicate datasets.

If data is publicly available and on-line it is not difficult for that data to be harvested for republication elsewhere on the Internet. There is also pressure on science programmes and Data Centres to distribute data to several different aggregator networks simultaneously, which may then also harvest data from each other (GBIF and OBIS for example regularly exchange data). In some cases, re-published data can be re-packaged, or slightly modified so that detection of inexact duplicates is very difficult. Inexact duplicates are data that are essentially identical but for a few minor modifications.

Within the Life Sciences field, the concept of assigning unique Life Science Identifiers (LSIDs) to both biological objects and taxonomic concepts has recently gained ground not just to address the duplicates issue, but also to ensure that biologists can refer to objects of similar classification without ambiguity. It is anticipated that governance mechanisms will soon emerge, sufficient to manage the allocation and use of these globally unique labels. An LSID conforms to the Universal Resource Name (URN) name standard. Every LSID may consist of up to five parts: the Network Identifier (NID); the root Domain Name System (DNS) name of the issuing authority; the Namespace chosen by the issuing authority; the Object ID unique in that Namespace; and finally an optional Revision ID for storing versions of the information. Each part is separated by a colon to make LSIDs easy to parse. Each LSID names and refers to one unchanging data object. Unlike the familiar URLs of the World-Wide-Web, LSIDs are location independent. This means that a programme or a user can be certain that what they are dealing with is exactly the same data if the LSID of any object is the same as the LSID of another copy of the object obtained elsewhere. Digital Object Identifiers (DOIs) managed by the International DOI Foundation (see http://www.doi.org/) provide the same service, but for a fee.

The use of a persistent identifier for all data objects exchanged within SCAR, regardless of the discipline from which it derives, is something which has significant merit and requires further investigation.

In terms of data access, it should also be remembered that SCAR is a community of international collaborators, so we need to be mindful of how we provide on-line access to datasets in terms of the language options used in our web sites, when developing data products and any accompanying material that we distribute with SCAR endorsed datasets.

3.2.2Gated vs Public Access To Data

It could be argued that the Antarctic Metadata Directory (AMD) has not so far achieved the acceptance levels hoped for with SCAR scientific groups because it has focused on metadata content when what is really desired is a discovery mechanism that leads to direct access to those data that have been returned as a result of the user's search criteria. Being able to achieve such a goal, however, requires that there must be a genuine willingness on the part of investigators to make their data widely accessible, and that they will link these data to the metadata records. Not all scientists are willing to provide that link, despite their countries signing up to the Antarctic Treaty which states that "*to the greatest extent feasible and practicable" ... "scientific observations and results from Antarctica shall be exchanged and made freely available*". The level of conformance with this aspiration is patchy, and responsibility for meeting the aspiration must to a great extent rest at the local level with individual national Antarctic programmes. Where national scientists work in SCAR programmes, there is likely to be a greater degree of international data exchange than where that is not the case.

Investigators, when questioned privately, will often agree to "gated" access to their data but not "public" access. If "gated" access is required, so that only an investigator's closest collaborators can gain access to specific datasets, variable levels of authentication must be built into the systems we are using. Currently this is not the case. Implementation of secure authentication systems is not impossible, but is a technological challenge where the user domain is spread across geographical and organisational boundaries and the number of potential security groups that any single user can belong to is highly variable. Simple password protection, while not particularly secure, is a much easier technology to implement.

An issue open for debate is whether the AMD would be more heavily patronised and seeded with accessible datasets, if it provided for optional secure access to hyperlinked datasets and whether the introduction of such a facility is even palatable given the public access obligations implied by the Treaty. SCAR's data policy is for free and unrestricted exchange. National policies may differ.

3.3Data Exchange

The topic of data exchange can be considered to include the data access issues outlined above, but the issues articulated under this heading relate primarily to the various modes available for communicating and encoding data streams that are collated and used by SCAR members.

3.3.1 Data Encodings & Formats

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. Despite its propensity for creating what some may argue are unnecessarily verbose data files, its popularity is assured for the foreseeable future. XML is a tagging language and has a syntax that allows users to encapsulate text inside tags that are meaningful to the author. The semantics of the tagged text may not, however, be meaningful to someone not involved in authoring the mark-up. So using XML generally involves some form of prior user-community agreement on what the tags represent, so that both producers and consumers understand the semantics implied by the tags.

XML as a language has now been supplemented with several utilities: XMLQuery, a language primarily for filtering and restructuring XML data; XPath a language that permits a user to parse an XML file and address a specific point in the document; XLink which allows elements to be inserted into XML documents in order to create and describe links between resources and XSLT which is used for transforming XML documents into other XML documents. This group of utility languages are considered *de facto* standards when working on the Internet and have been very broadly adopted by the global business and commercial software companies. It is impossible to ignore their pervasiveness as accepted technologies in all of the environments in which SCAR member nations operate.

Less well adopted and perhaps more focussed and bounded in use are the Resources Description Framework Language (RDF), another W3C standard – which is used quite extensively in the Life Sciences; the Geography Mark-up Language (GML), which arose out of the geospatial community; Keyhole Mark-up Language (KML), used by Google Earth, and Sensor Model Language, for describing all types of sensors and their behaviours.

RDF is used regularly for modelling the relationships between biota, biotic communities, their environments and their interactions as well as for exploring and encoding the relationships between functional elements at the cellular level, particularly gene sequences. Recent advances in network-based visualisation tools, which can operate on RDF, have made this encoding language popular. GML, championed by the Open Geospatial Consortium (OGC), is gaining in popularity amongst scientists and data managers alike as the commercial software companies responsible for producing GIS, web-based mapping and spatial data analysis tools have adopted GML as the *de facto* industry standard for data exchange. GML is an XML-based language tuned for representing spatial objects and spatial object relationships. A less expressive, but simpler language as an alternative to GML is Google Earth's Keyhole Mark-up Language (KML). This language (similar to GML version 1) has gained recent popularity because of its use in Google Earth products, and isn't used widely outside of this environment. Sensor Model Language (Sensor ML), based on XML and also championed by the OGC, is a language for defining the geometric, dynamic, and observational characteristics of a sensor. Its popularity within the physical sciences is anticipated to escalate as more sensors are deployed and linked into sensor networks. The IPY goal of leaving a comprehensive observation system as a legacy for IPY should expedite the adoption of Sensor ML.

The languages discussed above can be considered relatively contemporary, at least in terms of their mainstream adoption within the scientific community, but there are other encoding languages, or rather formats, that have been in popular use within the scientific community for some time. netCDF is widely used for the exchange of in-situ marine observations and measurements. netCDF is often used for 'coverage' or gridded type data where a variable or variables have some regularity across either horizontal, vertical or temporal axes. Like XML it suffers from requiring prior community agreement on the semantics of the contained variables and attributes, but unlike XML it is compact in its encoding. netCDF uses a binary file format so its contents are not human readable, unlike XML, but there are utilities available for rendering the netCDF file into a human readable format. This format is simply for human consumption to enable viewing of the binary encoding. The use of netCDF is not going to diminish and there are already several initiatives underway to ensure that netCDF files can be used in conjunction with GML in order to exploit the wide variety of web-based spatial data analysis, visualisation and discovery tools available, whilst maintaining the efficiency of encoding.

The encodings specifically mentioned above are not the only formats in use within the SCAR scientific community, but they could be considered to be the most ubiquitous ones used, or are viewed as being emergent standards that cannot be ignored by SCAR members. 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.

3.3.2 Data Exchange Protocols

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. Several of these deserve special mention.

There has been a trend in last 5 years towards using a services approach to the development of software, particularly software that involves the Internet as a transport medium. 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 (see Figure 1). 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.

SOAs are supported by a range of protocols; the most commonly used at present are:

- Universal Discovery, Description and Integration (UDDI) protocol for composing registries (Clement, Hately, von Riegan & Rogers, 2004),
- Simple Object Access Protocol (SOAP) that defines how to format a message to communicate between applications (W3C, 2003),
- Web Services Description Language (WSDL), a language for describing services and how to access them (Christensen, Curbera, Meredith & Weerawarana, 2001),
- Web Service Business Process Execution Language (WS-BPEL) uses web services standards to describe business process activities as Web services and defines how they can be composed to accomplish specific tasks (http://www.oasis-open.org/news/oasis-news-2007-04-12.php).

In the past few years, the geospatial community, mainly led by the OGC and the ISO Technical Committee 211 for "Geographic Information/Geomatics, has embraced the SOA model but developed alternate standards which are designed specifically to deliver or discover geospatial data payloads (see http://www.opengeospatial.org/standards and http://www.isotc211.org/). For example, instead of UDDI and WSDL, the OGC has developed a registry interface standard, the OpenGIS Catalogue Service (CS-W) and three types of web services with their own messaging formats, the Web Maps Service (WMS - http://www.opengeospatial.org/standards/wms), Web Feature Service (WFS - http://www.opengeospatial.org/standards/wfs) and Web Coverage Service (WCS). These web services and other spatially-based standards are expressed in GML. While the OGC and its ISO counterpart are currently in the process of harmonising these standards with mainstream IT protocols (OGC, 2005), a degree of specialisation is currently still required until mainstream IT interface standards efficiently cater for the discovery, transfer and manipulation of data with a spatial component.

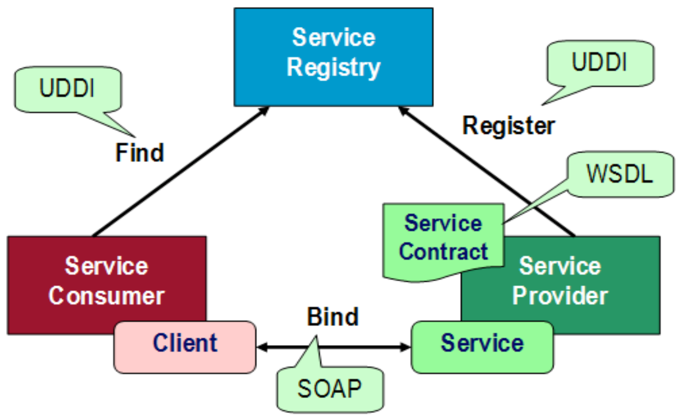

Figure 1. Service-oriented architecture publish, find, bind paradigm (source: Brauer & Kline, 2005).

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 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. The DiGIR protocol was developed by BRC Informatics in collaboration with the Museum of Vertebrate Zoology at UC Berkeley and the California Academy of Sciences. It provides a uniform interface for managing XML-based search requests to a community of dissimilar data sources. Each institution in the DiGIR community implements an interface application called a DiGIR provider. The DiGIR provider hides the details of the underlying database and presents a uniform "virtual" view of the data to the network. SCAR-MarBIN, for example uses DiGIR as its preferred data exchange protocol as does GBIF and OBIS.

DiGIR will eventually be replaced, although not in the immediate future, with a new protocol called the TDWG Access Protocol For Information Retrieval (TAPIR) – see http://wiki.tdwg.org/twiki/bin/view/TAPIR/. TAPIR specifies a standardised, stateless, HTTP transmittable, XML-based request and response protocol for accessing structured data that may be stored on any number of distributed databases of varied physical and logical structure. TAPIR extends the features of DiGIR to create a new and more generic means of communication between client applications and data providers.

While the biological sciences use protocols such as DiGIR and TAPIR, the physical sciences, particularly the oceanographic and atmospheric communities have been using a protocol called, OPeNDAP (http://www.opendap.org/). "OPeNDAP" stands for "Open-source Project for a Network Data Access Protocol". The OPeNDAP provides a way for ocean researchers to access oceanographic data anywhere on the Internet from a wide variety of new and existing programmes. By developing network versions of commonly used data access Application Program Interface (API) libraries, such as those for netCDF and others, the OPeNDAP project can capitalize on years of development of data analysis, and display packages that use these existing APIs, allowing users to continue to use programmes with which they are already familiar. Using flexible data types suitable for many uses, the OPeNDAP servers deliver data directly to the client programme in the format needed by that client. To expand the universe of data available to a user, OPeNDAP incorporates a data translation facility, so that data may be stored in data structures and formats defined by the data provider, but may be accessed by the user in a manner identical to the access of local data files on the user's own system.

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). The goal of this software is to simplify the discovery and use of scientific data and to allow scientific publications and educational materials to reference scientific data. At the heart of THREDDS (http://www.unidata.ucar.edu/projects/THREDDS/) is a cataloguing system built using XML documents that describe on-line datasets and contain (non-standardised) metadata. THREDDS users have now built bridges into systems such as the GCMD to leverage the metadata and services housed in such systems. The THREDDS software uses a common data model to read datasets in various formats, and serves them through OPeNDAP, OGC Web Coverage Service, netCDF subset, and bulk HTTP file transfer services.

Understanding that different scientific groups use different data encodings, formats and protocols to exchange data is fundamental to developing a data management strategy 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 the SCAR Data and Information Strategy must acknowledge this investment.

3.4Data Quality

Data quality is something that must be judged by a user and will necessarily relate to the uses to which the data will be put. By accompanying datasets with structured, high level descriptive metadata, a user can get some sense of a dataset’s fitness for use. The types of information generally provided in descriptive dataset level metadata, which help an assessment to be made, might include information about the scale of capture of the data, positional and temporal accuracy, geographic and temporal coverage, and broad descriptions of the instruments used in data capture etc. This type of metadata does not, however, generally present information about individual data points in the dataset (although the ISO 19115 metadata standard does cater for this), nor does it explain how the data may have been processed. A good dataset metadata record will be accompanied by processing notes if processing has taken place, otherwise it is extremely difficult for a user to make a well informed judgement about the quality of data for his or her purposes.

Many disciplines require a comprehensive understanding of a dataset's lineage before it can be used. Take for example climate studies, which in some cases rely on very small variations in the temperature of the ocean (of only tenths of a degree) over time, geography and depth to deduce changes in ocean circulation patterns. In these types of research, for data to be of any value, it is imperative that instrument types and calibration details are known, and that all quality control processes applied to the data are understood - including who did the processing and what degree of reliability has been placed on each data point in the dataset. All of those details could affect decisions made about the quality of the data.

This stringent requirement for understanding data quality has led many scientific disciplines to "flag" datasets and data points after processing has been applied. Unfortunately there are numerous "flagging" and "processing systems" in place, many of which are not well documented. Obviously, while it is difficult to standardise on tests, processes and flags across disciplines, it should be possible to develop some consistency within disciplines. The World Ocean Circulation Experiment (http://www.awi-bremerhaven.de/GEO/eWOCE/), one of the largest oceanographic collaborations in history, demonstrated that it is possible to develop common systems and conventions for processing and documenting data.

The lack of processing and flagging consistency or documentation conventions can result in the same tests being applied to data over and over again when data is passed from one user to another, simply because each user is unaware of what has already been applied to the data. The Joint Commission on Marine Meteorology (JCOMM) has rated this problem as a significant issue and one that it will be addressing with its partners (Keeley, 2007).

3.5 Data Integration

Thus far, most discussion has centred on issues related to the discovery, access and exchange of data, but only cursory mention has been made about attributing and structuring datasets so that their content is widely and unambiguously understood by both machine and human interpreters. This topic was touched on briefly in section 3.3.1 during discussion on XML, RDF and GML.

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, often supported by logical constructs. There are many initiatives underway across many different scientific communities of interest, to approach these tasks with more rigour than has perhaps been applied in the past. Scientific data users and generators alike now recognise the value to be derived from combining a myriad of differentially sourced datasets to answer questions that could not otherwise be answered. Many aspects of global climate modelling fall into this category. It is the sum total of all current and previous efforts that combine, often through the data, that makes complex interdisciplinary scientific problems tractable.

There are a range of activities, projects and communities working on standardisation issues to improve the ability to integrate data from different sources that are relevant to the SCAR domain. Three categories of activity worthy of special mention are expanded upon below.

3.5.1 Development of Species Registers

In biodiversity research and many other biological fields of study, species are the common currency with a well-established, standardised code of nomenclature. Unfortunately, despite the existence of the Linnean System for naming species, different names are often used for the same species and the same names for different species (Costello, 2000). This can cause considerable confusion when exchanging and integrating biological datasets, particularly when machine-mediated transactions take place in the absence of a human user.

To address this issue there are efforts by a wide range of communities to develop species lists or taxonomic registers. These lists usually commonly contain core information on taxonomy associated with particular specimens, the taxonomic authority, synonyms, common names and sometimes codes allocated to specific taxons by various naming authorities or institutions. Using such lists as mediators, equivalent taxon concepts can be mapped as synonyms and common names mapped to their taxonomic concept counter-parts, making integration of data that much easier and more accurate.

Recognising the need to combine all of these individual efforts into a coherent whole there are a few global-in-scope aggregation projects that through a federated approach aim to create a validated checklist of all the world's species (plants, animals, fungi and microbes). Species list projects that are already being contributed to, or leveraged by SCAR, include The Register of Antarctic Marine Species (http://www.scarmarbin.be/scarramsabout.php) and The Catalogue of Life (http://www.catalogueoflife.org/search.php) – and amalgamation of Species 2000 (http://www.sp2000.org/) and the Integrated Taxonomic Information Systems (http://www.itis.gov/).

3.5.2 Standardising Terminology

Without semantic interoperability, the type and value of the services that can be launched to operate within any deployed data access infrastructure that relies on multiple data feeds from a wide range of sources will be severely restricted. Computers are not innately able to interpret or infer the meanings of data elements, nor their relationship to related or similar elements. To assist this process we must be able to be explicit about the definition of exchanged objects and their properties, and state their relationship to other objects in the system, ideally using formal logic. The concept of the semantic web is based on this premise.

Communities are beginning to develop and agree upon terminologies to describe real-world objects and concepts that are routinely encountered in their domain of interest. For example, this might include parameter dictionaries, equipment and instrument descriptions, standardised units of measure, observation techniques, phenomena dictionaries and process descriptions. These types of terminologies are often called ontologies. Ontologies are referred to as either being “lightweight” (i.e. essentially little more than taxonomies of terms) or “heavyweight” (taxonomies with added axioms and constraints). Formal (heavyweight) ontologies are generally expressed in an ontology language, of which there are now many, but the most popular is the Web Ontology Language (OWL) – built on RDF. Using OWL and inference engines (now available through the open source community) it is possible to build "intelligent" web applications that are able, in a limited sense, to infer information about particular data elements based on their context. This is good from both a data discovery and a data integration perspective. But one of the most widely used applications of ontologies at present is simply to define the universe of discourse for a particular domain and then leverage this for mapping between domain vocabularies for data integration purposes.

Standardising terminology is a particularly onerous task and if it is possible to leverage and/or extend the work of an existing community, this should be encouraged. Several communities have taken a lead in this area including the GCMD with its GCMD keywords; the Marine Metadata Interoperability Project (http://marinemetadata.org/); SEEGRID (http://www.seegrid.csiro.au/); GEON – the Geosciences community (http://www.geongrid.org/); SEEK – the ecological community (http://seek.ecoinformatics.org/) and numerous niche medical communities under various e_Health initiatives.

3.5.3 Special Issue of Place Names (Gazetteers)

Like taxonomic terms, place names deserve special mention as a type of terminology requiring particular attention in Antarctic data management. The Antarctic continent is a jointly managed entity. Generally in most national contexts, various scales of mapping are usually ascribed to one or more jurisdictional authorities, which manage the naming of features and places, or cooperate with those national bodies that have that mandate. National mapping authorities then perform the task of ensuring that place names appear consistently and reliably on mapping products.

In Antarctica we have the situation where there are often several nations naming the same features and places, using their own national naming systems. Frequently, even similarly named features or places are ascribed different geographical coordinates. This poses difficulties for the production and compilation of common mapping products. For example, let's assume that there is an incident in Antarctica requiring several nations to cooperate in a search and rescue operation. Each nation begins the search using their own national maps of a particular region. Each nation has named similar places and features by a different name, and it becomes almost impossible for a rescue coordinator to communicate effectively with the different search parties about their location, particularly given that the original distress call may have only provided a place name rather than a precise geographical coordinate. Similarly, coordinating multi-national scientific campaigns would suffer identical problems without some way of harmonising these various national naming conventions.

Fortunately, development of a SCAR Composite Gazetteer has been underway for some time to help national operators and scientists circumvent this kind of problem. Up until recently the Gazetteer was hosted by Italy, but it has now been transferred to the Australian Antarctic Division. The Gazetteer maps the various place and feature names applied by nations to a common identifier for interoperability purposes. Italy continues to moderate the Gazetteer's content by harnessing the inputs of various contributing nations. Urging nations to continuously provide the requisite input requires a surprisingly large effort applied by a very small group of SC-AGI members.

3.5.4 Feature and Symbology Catalogues

Having agreed on common vocabularies and ontologies, we have to manage these community standards in web accessible repositories as part of a data access infrastructure so that they are available for community use. Ontology libraries and Feature Catalogues are two types of tools that provide the mechanics for managing domain vocabularies and data models. The ISO 19000 series of standards includes various specifications for defining feature types and for developing Feature Catalogues. A feature catalogue holds the various definitions and relationships that exist between commonly used features found in the real world (e.g. a "Mountain" is a type of feature and a "Peak" is "part-of" a "Mountain").

Feature catalogues and other semantic concept libraries usually rely on other types of applications to provide information about how data objects are portrayed when they are displayed in a map or some other type of visualisation medium. Symbology catalogues are tools that perform a role in data portrayal. For example, in a data access system using web services it should be possible to source data from any provider capable of deploying a standards-based data service, and integrate and render the supplied data in a web map for a client by using consistent symbology and styling.

Style Layer Description standards developed by the OGC (http://www.opengeospatial.org/standards/sld) have been developed to facilitate this process, but this type of visual integration of data from disparate sources requires prior agreement on a symbology set, which in turn must be linked to an existing ontology describing the concepts to be symbolised. The Australian Antarctic Data Centre (AADC) has developed a Symbology Catalogue service (http://aadc-maps.aad.gov.au/aadc/symbology/) which integrates with the SCAR Feature Catalogue (http://aadc-maps.aad.gov.au/aadc/ftc/) but now requires partners to help populate the catalogues and assist with trialling the deployment of the tools in operational situations. Other communities investigating the use of such tools include the IHO (http://www.iho.shom.fr/); the Intergovernmental Committee on Surveying and Mapping (ICSM) (http://www.icsm.gov.au/icsm/) and collaborators in the Infrastructure for Spatial Information in the European Community (INSPIRE) project (http://www.ec-gis.org/inspire/).

3.5.5 Datums and Reference Systems

Integrating data, particularly those which have a spatial component, necessitates a common understanding of the frames of reference in which these data have been captured or placed.

Prior to satellite mapping technology, the best approximation of the shape of the Earth was the mathematically calculated geoid, which evolved into slightly flattened spheroids or ellipsoids. Geographic coordinate systems use a spheroid to calculate positions on the Earth. A datum defines the position of the spheroid relative to the center of the Earth (Price, 2001). Knowing which datum a dataset belongs to matters significantly when you want to bring data from different datums together, because the same feature represented in one datum will be offset in location when it is represented in another datum. As there is a range of datums and geographic coordinate systems to choose from, particularly as datum models have improved over time, bringing multi-national datasets together can be a challenge when all the datums are different and, worse still, un-stated.

Other reference systems often used in capturing and representing scientific data can be termed engineering reference systems, and might describe the topology of a particular dataset. Again, understanding these systems and having them explicitly stated is the first step required before data integration can be undertaken. SC-AGI members in particular have considerable interest in the standards associated with datums and reference systems, and where practical they have appointed liaison officers to a limited number of standards bodies active in these areas.

3.6 Data Archiving

Ultimately, data collected from SCAR scientific endeavours must be appropriately archived if they are to be re-used many times over, often for purposes not envisaged at the time of capture. SCAR requires well enunciated archiving standards and guidelines and these should be available to National Antarctic Data Centres. Consideration should also be given to establishing a network of capable, designated facilities that are geared to archiving data from all Antarctic programmes and projects where archiving capacity does not already exist.

There is a range of ways in which a designated network of facilities could be reliably achieved, including but not limited to harnessing the existing ICSU World Data Centre System, by forging explicit partnerships between existing SCAR National Data Centres, by aligning with the IODE, or by exploring a partnership with the Global Change Master Directory to provide data hosting, as well as a metadata services. In all cases the goal required is to identify Centres willing and able to archive data according to SCAR archiving guidelines, and for those Centres to participate in a virtualised network for data discovery and data re-use purposes.

4.0 Antarctic Data Management – The Current System

In the preceding section an overview was provided of the types of issues that need to be taken into account in any modern scientific data management system. This section discusses the ability of the current institutional arrangements to address these requirements and the nature of the systems that have already been established.

JCADM 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.

JCADM is the entity ostensibly responsible for the Antarctic Data Management System (ADMS) and purportedly 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).

JCADM'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, COMNAP and the Antarctic Treaty Secretariat to identify and develop fundamental datasets of value to the Antarctic Community.

JCADM is essentially a coordinating body, comprising over 30 officers who represent their country's interests in SCAR data management. The Committee meets annually in various member's States, on a rotational basis, although only a sub-set of member states (usually less than a 1/3) generally funds participation in these meetings. It is important to understand how JCADM currently operates so that there is a common perception of its capabilities, as these issues, as well as the scientific data management needs mentioned earlier, have significantly influenced the character of the strategies detailed later in this document.

A 2007 email survey of JCADM members indicated that relatively few of the 30 nations involved in JCADM are actually operating a National Antarctic Data Centre (NADC). Most countries have assigned a single individual officer 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 JCADM, whether data are being handled or not. Many respondents to the survey readily acknowledge this latter point as a significant issue.

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 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 who are loosely coordinated through a dedicated Antarctic Data Manager). This is a fundamental issue for SCAR because very few JCADM member States are operating NADC's and yet coordinating these Centres is one of JCADM's key goals. 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. The discrepancy between the stated goals of JCADM and the reality of what can be achieved with the current resources, creates unrealistic expectations of JCADM within the SCAR science community and has in the past lead to tensions.

However, 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 were forthcoming about how this situation should, or could be remedied. That assessment also 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,
(b) the need for more outreach 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 JCADM 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 that it is too difficult for JCADM 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 JCADM's credit, in the three years since the 2005 review was conducted, it has attempted to address all of the metadata related issues and a number of other short-comings, including the development of this Strategy, but it continues to be hampered by a fundamental lack of resources. The inability to attract active members has lowered morale and if left to continue unaddressed will eventually render the network unviable. The larger, better-resourced Data Centres have in the past taken on the lion's share of 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, and perhaps that particular strategy has been flawed by allowing fundamental problems to be masked and therefore allowing a certain level of complacency to occur at senior levels within SCAR and JCADM in terms of recognising the need to correct and strengthen the system.

In contrast, action has recently been taken to improve the operational circumstances of the SC-AGI Group, where it was taken out from under the GeoScience Standing Group wing and given the stature of a Standing Group in its own right. There has been a change in chairmanship and strong support from the SCAR Executive Committee in the group's operational activities and in the promotion and marketing of SC-AGI products. Like JCADM, SC-AGI, like its predecessor, suffers from inadequate resourcing. The group has a large and ambitious work-plan, but many of the tasks have no definitive end date and have been on the group's "to-do" list for some time. More importantly, although each task on the work-plan is no doubt contributing value, it is difficult to discern how many of the nationally-centric tasks fit into a coherent plan for strategically building Antarctica's spatial data infrastructure. Despite a lack of resourcing, an active cluster of nations has in the past managed to produce a number of important products, which if better patronised and more strategically utilised will greatly assist SCAR members to put in place an AntSDI.

The Strategy that follows in the next section suggests a mechanism to bolster and expand the existing JCADM group and Antarctic Data Management System (encompassing an AntSDI), in addition to recommending a range of data centric activities designed to assist SCAR science groups with their data management needs, given the various contexts in which they are operating. This Strategy of necessity can only deal with common and generic issues faced by the SCAR science groups. It does not seek to address project-specific data management problems, but provides a framework which individual scientific projects should be able to leverage to their advantage. While many of the SC-AGI and JCADM data management issues overlap there are some cases where recommendations are made that address matters specifically related to building the AntSDI.

5.0 SCAR Data & Information Strategy

The SCAR Data and Information Strategy is directed towards building an Antarctic Data Management System, capable of supporting Antarctic science and aligned activities of nations participating in the Antarctic Treaty System. The ADMS should be viewed as a science and operations enabler. We will know when this Strategy has succeeded because the ADMS will be a tangible operational entity, it will be deemed an indispensable tool for doing scientific research and logistical operations, much like the Internet works now for business operations, and the cost of supporting the system will be routinely factored into national programmes.

5.1 Necessary Components of A Globally Distributed Antarctic Data & Information Network

SCAR, through a range of individual activities, is already making some progress towards a system such as the one illustrated in the testimonials presented earlier on pages 9 & 10. But much more could be achieved, and the likelihood of getting a successful outcome would be far greater, if appropriate strategic foundations were put in place now. These foundations would necessarily include:

- Better articulated governance arrangements and strong leadership, suitable for driving development of a distributed and shared infrastructure,
- Fostering a culture willing to share and collaborate on data management related activities,
- Introducing a data citation system, possibly in conjunction with the publishing industry or other partner,

- Leveraging existing SCAR and non-SCAR systems, capabilities and resources and supplementing these where there are obvious deficiencies,
- Creating a network of designated permanent data archives capable of the long-term management of all types of SCAR related data,
- Agreement on, and implementation of, standards that support interoperation of technology platforms and data transport protocols,
- Harmonisation of standards for facilitating data discovery and data evaluation,
- 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
- Education, outreach and guidance on all facets of the systems operation, protocols and functions.

5.2 Policy, Governance, Partnerships & Coordination – Strengthening The Existing System

To help clarify obligations with regard to data access, SCAR requires a clearly enunciated data management policy that elaborates upon Article III 1-c of the Antarctic Treaty 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 adopted or 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 endorsed science projects should be a mandatory component of the SCAR policy.

Governance aspects of SCAR data management need to be reviewed and the use of partnerships with existing networks should be aggressively pursued to bolster the resources and infrastructure available to SCAR science. In addition, to ensure that SCAR is well informed on the needs and activities of the wider Antarctic community, such as those involved in environmental management, JCADM should seek to involve groups requiring access to, or management of Antarctic science data. As a first step, the Commission for the Conservation for Antarctic Marine Living Resources (CCAMLR), Data Manager should be invited to join JCADM and the CCAMLR Data Centre should be asked to consider becoming an NADC node in the ADMS.

Recommendation: JCADM tasked to develop a SCAR Data Policy.

Recommendation: SCAR Executive to invite CCAMLR to join JCADM.

5.2.1 Data Management Plans – Why Are They Required?

In complex, multi-disciplinary science, particularly research that uses a combination of observed and modelled data, with collaborators widely geographically dispersed, effective data management needs to begin at the time the research is planned. Deferring data management considerations to later phases of the research, when observing programmes are already well underway, or have been completed can significantly compromise what can later be achieved in terms of research outcomes. The two hypothetical situations expressed below can be used to exemplify this point.

(i) *A coordinated set of voyages is planned under a large science programme that amongst other activities will collect plankton samples. One scientific objective of the collaboration is to better understand the distribution and abundance of planktonic biota and how it varies according to a range of environmental parameters. All collaborators are clear on the objective, but each collaborator has a different tacit understanding of how the objective will be achieved. Half of the collaborators involved analyse the samples that they have collected and record the resultant taxonomic data as presence - absence data, to the best taxonomic resolution possible and these results are grouped according to voyage transect segments 5km in length. The other collaborators collect and analyse the samples using 10's of km transects, record taxa to "family" level only, and don't record absences, just presence data. After the voyages the collaborators start to share their data and discover that their divergent data capture and analysis methods will have an appreciable impact on their collective ability to integrate the data for broader analysis and will almost certainly affect achievement of their original goals.*

(ii) *On the same voyages other collaborators have been funded to work together to produce a "habitat" map of the Antarctic region, which will be pieced together from the data captured across all of the voyages. Again each collaborator has been able to describe his/her scientific objectives to colleagues, there is consensus on the broad aims of the research, but each collaborator is left to design and conduct his/her own observation programme. While there is loose agreement on what variables will be observed or measured there is no detailed pre voyage data management planning or sampling programme design. Subsequently each collaborator uses different types of instrumentation, with varying degrees of accuracy and resolution to collect their data and the post voyage processing of data is performed according to a range of differing habitat classification systems. Each collaborator's timetable for working up the data is also very different and not previously agreed. When the collaborators start to exchange their data they discover that their habitat datasets do not readily integrate, and because many collaborators have already managed to publish their specific outcomes, they have moved on to other projects and are reluctant to re-process their data to achieve an integrated dataset.*

These types of situations could be readily avoided if adequate data management planning was instituted for all SCAR sponsored research programs. Data management plans can also help to ensure that research projects are allocated sufficient resources to undertake data management related activities. These plans also make very explicit who will be responsible for specific data management tasks, they outline collaborator expectations and the protocols required to achieve effective data sharing. JCADM should develop a template Plan to make it easier for science projects to pitch their planning at the appropriate level of detail and should assist if requested in developing such plans.

A cursory examination of documentation available for the 5 current SCAR research programmes, from the SCAR web site (http://www.scar.org/researchgroups/) demonstrates that the need to consider data management up-front in research planning has been recognised, but there is considerable variability in the degree to which each programme has articulated data management aspects of the research. Some of the more mature programmes have established databases for the collation of data but there is very little other overt guidance for collaborators on how data management operates within the research programmes. Some of the newer programmes mention the development of multi-parameter database products to support the research but it is unclear how these products will be developed, what relationship or overlap they have with existing thematic and global databases and there is no indication of how, or if, these products will be maintained into the future.

Recommendation: JCADM tasked to develop a basic data management plan template.

Recommendation: SCAR Executive to request that all SCAR endorsed projects develop a data management plan in close consultation with their JCADM representative. JCADM should provide advice to the SCAR Executive on the adequacy of such plans prior to their endorsement.

5.2.2 NADCs

The Antarctic Data Management System is currently a loose, primarily administrative collaboration between individuals who are usually based in national institutions whose responsibility is for carrying out Antarctic science, or the coordination of Antarctic science. For a variety of reasons, these particular individuals or institutions may not always be best placed to manage the data sourced from Antarctic research, and an alliance is necessary with other national institutes in the same country that do have such capacity, even if their focus may not be on Antarctic science. Some JCADM members have such arrangements but in other cases there are no suitable institutes with which to partner, or suitable partnerships are yet to be formed. While JCADM may have 30 members, there are less than 8 designated NADCs.

Admittance to the SCAR family from a data management perspective currently entails nominating someone to represent data interests on JCADM. Implicitly JCADM has an NADC recruitment campaign, but there is no SCAR policy requirement 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.

Some small, but potentially far-reaching changes to the obligations placed on SCAR membership could serve to remedy some of the more significant barriers to advancement of the AMDS. To have impact these obligations would need to be retrospectively applied, perhaps with a suitable transition period, to enable all existing members to comply with new obligations.

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 an Antarctic data management coordinator need not come from the same institution in which an NADC resides. The coordinator could be affiliated with this Centre or operate outside of it. The intent, however, where the NADC and coordinator are in separate institutions, would be that there is a close working relationship between the NADC and the coordinator. In addition there would be explicit and clearly articulated expectations of an NADC, which would need to be met by the member and monitored by the Coordinator and the JCADM Chair. Some suggested core NADC functions include:

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

The primary function of the NADCs should be to act as long-term repositories for archiving and publishing Antarctic data, or 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. 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 JCADM and the JCADM Chair on Antarctic data management.

To ensure that the 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. It is anticipated that this would involve some negotiation and mutual obligations, with an expectation that the smaller, less well resourced member would eventually develop the capacity internally to fully participate in the system in their own right.

This model of geographically dispersed designated Centres is common in the IODE and ICSU WDC Systems. 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: SCAR Executive to ensure that new SCAR members are aware of data centric obligations and that existing SCAR members are made aware of these retrospectively.

Recommendation: JCADM to develop a profile of a typical NADC, including its functions and obligations as part of the ADMS.

Recommendation: JCADM Chair to explore the potential for establishing regional data hosting facilities.

5.2.3 Governance Of The ADMS

JCADM has a part-time Chair and two part-time Deputies. The Chair's role at present focuses primarily on outreach, communication, liaison and reporting, with little time remaining for any actual leadership, coordination and monitoring of the JCADM Group and work-plan. The majority of JCADM inter-sessional work remains largely unfulfilled from year to year, in part because the Chair lacks the time to drive the effort, but importantly because there are insufficient resources within JCADM 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, full-time individual available to lead and coordinate these efforts. 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. Importantly people are attracted by success (Schweik & Semenov, 2002; Latterman & Stieglitz, 2005).

If the AMDS is to expand through the recruitment of operational NADCs, possibly involving partnerships with other entities, the need for a full-time, or near full-time JCADM Chair is considered mandatory. Because SCAR relies on the contributions of members, this would require a large commitment from an individual nation willing to temporarily re-assign their staff to full-fill this Executive role. The current 4-year term for the JCADM Chief Executive position would tend to be a disincentive for nations to make such a commitment, so perhaps a much shorter term (e.g. 1 to 2 years) should be considered. Alternatively (or in addition), the SCAR Executive Committee, could seek to obtain sponsorship to support a full-time Chair position by lobbying private companies, funding bodies or philanthropic foundations.

While JCADM itself should continue to be open to all SCAR members, there should also be a smaller sub-group of individuals responsible for providing advice to the JCADM Chair on establishing and growing the ADMS. This technically oriented sub-group, probably comprising of no more than 5 individuals with expertise in data management and in developing data access systems, would have its membership regularly rotated on a staggered basis to enable wide international participation over time. Ideally these individuals would be affiliated with the SCAR NADCs and also with members from SC-AGI with expertise in OGC-type standards. This specialist group would provide a forum for the discussion of issues on standards, protocols and interoperability, while the broader JCADM group could concern itself with policy, communication, outreach and user-centric issues. Members of this sub-group could also be nominated by SCAR to participate in any technical bodies formed by potential partners such as the IODE and WDC Systems to promote fruitful cross-fertilisation between the partners. The technical group should work closely with the JCADM representatives for the SCAR Science Groups (SSG), to ensure that technical development remains closely aligned with their requirements.

In addition to the full-group, face-to-face meeting each year, JCADM members should undertake inter-sessional work using remote communication technologies. A one-day session, specifically for the core technical group and the JCADM representatives for the SSGs, should also be set aside at the beginning of each year's annual meeting. Any JCADM members should also be welcome to attend such meetings if they wish given that the meeting will precede the main JCADM meeting.

Recommendation: SCAR Executive to consider the feasibility of, and options for, appointing a full-time JCADM Chair.

Recommendation: JCADM to form an NADC Technical Committee to focus on the technical aspects of building the ADMS and convene yearly meetings with the JCADM SSG representatives in conjunction with the annual JCADM member's meeting.

Recommendation: JCADM Chair to ensure that members of the NADC Technical Committee are included in any Technical Committees of partner data access networks.

5.3 Data Management Culture – Data Citation

SCAR is a diverse and eclectic mix of nations, people and projects. Fostering a data management culture is already being actively pursued within SCAR, particularly through current JCADM activities 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. As an international collaboration, SCAR members can encounter potential sovereignty issues, which often act to impede data flow. The scientific reward system is also heavily geared towards individualism and small cliques of collaboration, where being the first to publish on a subject is still the most often used and highest weighted measure of scientific performance. Restricting timely access to data that you have invested significantly in capturing is therefore very tempting to individuals (and some nations) in such a competitive context. 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.

Rather than wait for a citation system to emerge, SCAR could take the lead in implementing such a system. The issue is not, however, straightforward. 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, how might only a portion of that database be referenced and how would, or should, datasets be refereed? Despite these issues this Strategy recommends that SCAR explores the possibility of creating a dataset citation system for SCAR related science. A path forward might involve the development of a partnership with the GCMD or the Polar Library community during IPY that is additionally supported by a research scholarship for a Masters or PhD student who could pursue much of the necessary work.

Scientists will find it difficult to patronise systems and processes that are intangible and where the benefits are unclear. Cultural change can often be effectively facilitated through the use of demonstration systems and pilot projects. Pilot projects, very publicly sponsored or driven by JCADM, which are well-bounded, have wide utility and which can be delivered in short-timeframes are good vehicles with which to gain credibility and pique people's interest. Several such pilots are suggested later in Section 5 of this document.

Recommendation: SCAR to seek a partnership with an appropriate institution, industry or Library peek body to pursue the concept of instituting a dataset citation system.

5.4 Leveraging Resources & 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, efficiently coordinate any data management efforts directed from within SCAR, and build capabilities across members in order to distribute the workload.

5.4.1 JCADM & SC-AGI Interaction

The SC-AGI is currently undergoing a revitalisation in terms of its activities, having just been made a SCAR Standing Group in its own right. There is a new Chair, a commitment to consolidate 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. It is clear that this vision of an SDI is not achievable without the involvement of JCADM and operational NADCs. So the strategic directions of SC-AGI and JCADM must be in alignment if this SDI vision is to be realised. In recent years, given 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 JCADM Chairs must ensure that the strategic directions articulated by both JCADM and SC-AGI are complimentary and that the work-plans of the two groups are synergistic, with some tasks being jointly pursued.

Recommendation: JCADM and SC-AGI Chairs collaborate to ensure synergies with respective work-plans and an alignment of strategic goals. JCADM 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 be given to combining JCADM and SC-AGI in the future.

5.4.2 International Initiatives & Funding

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 serve to play important roles in Antarctic data management. These initiatives bring together both expertise and infrastructure. 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. NADC's could then work towards joining these endorsed networks and contribute content and/or services. If SCAR endorsed networks that had the capacity to interoperate because of similarities in their operating standards and protocols, it would not be a difficult task 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, through JCADM, could influence one or more of the network partners to provide such a facility as an extension to its existing services.

A smaller number of endorsements with good levels of collaboration would be preferable to a large number of endorsements in a misguided effort to try and 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 slightly 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 to multiple networks without significant effort on the part of an NADC. But this is not the current reality. Relatively robust networks that could be considered by SCAR include SCAR-MarBIN

(OBIS), GBIF, SeaDataNet (IODE), but these 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, JCADM 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 not yet fully detailed and only a portion of the money sought for the service was provided. However, any expanded SCAR ADMS must necessarily include the IPY DIS.

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, global-in-scale environmental issues, freeing up access to Antarctic sourced data should be a global priority. Given the difficulty that JCADM 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 JCADM 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: JCADM and SC-AGI in liaison with SCAR science bodies should choose a (small) number of existing and complimentary data access networks with which to be affiliated and then endorse and promote NADC collaboration in these networks.

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

5.4.3 ADMS Products & Services

As previously discussed, the NADCs should work together to create an ADMS that is more than just the sum of the individual AMDS institutions. This requires that the AMDS as a network should meet a number of goals, some of which should be achievable in the short-term and others in the longer-term. Achieving such goals will depend on the ability of JCADM and the NADCs to garner sufficient resources to develop collective capability and interest from SCAR scientists in using AMDS products and services. Some typical products and services that a useful AMDS might offer (but not necessarily from every node in the network) include:

- Metadata management services (which already currently exist through the GCMD),
- Data archiving services (with a variety of data retrieval options),
- Data publishing services (offering publication using a number of protocols and formats),
- Data translation services (converting data from one format or encoding standard to another),
- Data mapping services (semantically mapping data from one source to another to permit data integration),
- Data aggregation services (where different nodes may act as the point of truth for continuously updated fundamental datasets that are composites of data derived from a variety of sources – e.g. digital topographic database; bathymetric database, habitat classifications and regionalisations; sea surface temperature database; sea-ice thickness database etc),
- Infrastructure hosting services (e.g. manage and serve: taxonomic names lists; feature catalogues; services registries; gazetteers; symbology libraries or manage mirror sites), and
- Specialised data visualisations (where different nodes might specialise in the analysis and portrayal of different types of data)

Pilot projects are a useful mechanism to help incrementally develop capability. In response to the JCADM member survey, several members responded that they do run operational data centres (e.g. Italy, Japan, Australia, UK, Argentina, China, Norway, Spain), even if Antarctic data management occupies only a small proportion of their total data centre effort. These Centres in particular could begin to provide the core of the ADMS network and collaborate to provide one, or a number, of the types of services mentioned above through a series of pilot projects. Ideally these pilots could be partially funded through grants and involve SCAR science projects.

The most strategic types of projects, given current emphases on providing observing systems as a major legacy of the IPY, would be those that help build the basic components of virtual observatories. However, typical projects could include:

- Nomination of an ongoing SCAR host for compiling a widely used Antarctic fundamental dataset. Development of a web site for visualising and publishing this dataset. Development of content contribution guidelines, standards and data exchange mechanisms for this particular type of data between partners.
- Working with the Southern Ocean READER project being undertaken by the SCAR Antarctica in the Global Climate System (AGCS) Scientific Research Programme, to develop a system that actually extracts data from all of the sources independently listed on the web site and which then permits the user to visualise (map and graph) different parameters.
- Joint development of a range of data manipulation tools, executed as services which could be registered with the GCMD and used on certain datasets made accessible via the GCMD metadata system,

- Agreement to publish data via an existing global scale data access network, and, independently or in conjunction with the existing network partners, develop mechanisms to visualise and extract Antarctic-themed data,
- Development of 'derived' or filtered dataset products that meet the needs of a broad cross-section of SCAR science projects,
- Collaborative development of a range of web map or web feature services that depict environmental and biological data that could be accessed via sites like SCAR-MarBIN or the IODE Oceans Data Portal, or the Encyclopaedia of Life.
- Development of rules and mechanisms for using Life Science Identifiers for SCAR biological data.
- Work within an existing data exchange community to develop data processing flags and explore how these can be exploited and promulgated by SCAR NADCs and science users.

Recommendation: JCADM should review the types of data centric services it can supply and work towards the provision of a well-bounded set of services to increase its utility to SCAR science projects.

Recommendation: A small number of pilot projects should be endorsed, scoped and implemented to demonstrate capability, develop skills and incrementally develop the ADMS.

5.4.4 Capacity-Building

Many JCADM 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.

JCADM generally runs one or two day data management workshops in conjunction with JCADM 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, JCADM 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 indicate their availability to conduct short training courses on establishing and running a polar data centre and JCADM manages the nomination process and brokers attendance at such courses. Members undertaking the training would have to bear travel and accommodation costs which could perhaps be supplemented by SCAR,
- NADCs indicate their willingness to host interns for periods of time with the view to training the intern in one, or all aspects of practically managing a polar data centre,
- NADCs collaborate on the development of multimedia product(s), distributable via DVD, that includes training material on operating a polar data centre,
- More developed NADCs to be allocated as mentors to less developed NADCs – in terms of a 'Buddy system'.

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.

The workshop sessions currently held in conjunction with JCADM meetings might be better utilised by targeting specific issues of relevance to all attending NADCs, rather than catering for newer members. This face-to-face time could also be used for developing funding proposals, fleshing out non-funded collaborative projects, knowledge transfer in a specific technical domain, or project reviews for collaborative projects already underway.

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 occurs 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: JCADM should investigate the training requirements of the newer NADCs and develop a training program, harnessing the expertise in the longer-established NADCs.

Recommendation: JCADM to instigate in the first instance a 'buddy system whereby more developed NADCs are allocated as mentors to less developed NADCs.

Recommendation: JCADM/SC-AGI should review the tasks undertaken at face-to-face meetings and the frequency of these meetings. The JCADM/SC-AGI Chairs should encourage the use of new communication technologies to pursue inter-sessional work.

5.5 Standards & Interoperability

The most ubiquitous standards in use by JCADM and SC-AGI member are the ISO TC211 series of standards, which cover geographic data, metadata and services. These standards are therefore a good starting point on which to base the creation of interoperable systems. It is acknowledged, however, that they are not the only standards that are in use within the SCAR community (refer back to Section 3), and that they will need to be supplemented to provide for the diversity and complexity demanded of a credible, scientific data network.

5.5.1 Metadata Systems

SCAR is fortunate in having a particularly strong partnership with the GCMD and an extremely robust and user-friendly metadata tool in the Antarctic Metadata Directory (AMD) system. Other 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 are ISO 19115 compliant. All ADMS NADCs are required 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 using the OAI-OMH Protocol. 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 only 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 it outlays. As more JCADM 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 the Memorandum of Understanding with SCAR, to provide annual updates on proposed future strategic directions and to seek input from JCADM on those directions.

JCADM, in collaboration with the GCMD should also investigate whether the AMD would be more highly patronised if it provided more consistent and direct methods of linking to available scientific data. Continuous improvements to the GCMD user interfaces to permit a more tailored and efficient user-centric search experience may also yield a higher degree of system usage.

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

Recommendation: JCADM and the GCMD explore better ways to link datasets directly to metadata records and improve user interfaces.

5.5.2 Beyond Metadata Systems To Interoperating Data Networks

While the AMD is designed to manage metadata records as well as links to the datasets 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 the NADCs and SCAR science project partners. It is useful to be able to provide links to such services via the metadata records, so that the AMD can function as a one-stop-shop for metadata discovery purposes, but a network of interoperable NADCs requires that a number of elements is in place before these services can be made available to be discovered and potentially integrated.

Sections 3.4 and 3.6 have already outlined the types of data exchange and data integration issues that are faced by SCAR science projects. 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,
(c) Developing appropriate and shared infrastructure to support (a) and (b) above.

Culturally and strategically these are the hardest issues to work on 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 planned and coordinated on a number of levels.

If SCAR is to pursue the vision outlined in the previous hypothetical testimonials, several activities need to be better coordinated and pursued by the community. These activities could be initiated by a core set of NADCs or preferably be conducted in partnership with some of the larger data access initiatives mentioned earlier in this Strategy. In summary JCADM and SC-AGI need to encourage:

(i) Development of shared terminology and representation for objects that we exchange.

To date SC-AGI and its predecessor have focussed on the development of a number of on-line tools, primarily centred on cartographic and mapping requirements that were designed to assist with developing an SDI. These tools are extremely useful and accessible but have very limited penetration at present within the SCAR community. For example, SC-AGI has developed a SCAR Composite Gazetteer that lists place names in Antarctica and this tool is linked to a catalogue that describes the types of places or real-world features that these names represent. These tools, if all members contribute content, provide important baseline information that many other existing and future SCAR-based science applications should be able to draw upon when requiring access to internationally standardised nomenclature. Although the Feature Catalogue content is at present dominated by terminology focussed towards topographic mapping, it should be used much more broadly to house descriptions and definitions of any type of data object encountered while undertaking Antarctic science.

There is significant international activity at present surrounding the development of Feature Catalogues to support interoperability of data systems, particularly those systems that are OGC compliant and which use OGC web services to deliver data. Catalogue use is not necessarily tied to these systems or protocols. SCAR has had an operational catalogue for at least 2 years, but is failing to capitalise on the benefits of such a system because its advantages and uses are not widely understood within the community. This situation could be remedied if the catalogue was used within one of the pilot projects mentioned earlier to showcase its function in managing community terminology and vocabularies.

The SCAR community should also leverage work already taking place in other communities that are actively defining common terminologies, data models and vocabularies (e.g. MMI, INSPIRE, TDWG).

(ii) The exchange of these objects using some standard protocols and encodings and the registration and advertising of these data services with appropriate networks.

There are many OGC-based web service interoperability projects currently in progress that the SCAR community could join. The SCAR community could develop a well-bounded pilot project involving a number of NADCs to trial publication of some services using common datasets and encodings. NADCs could focus on how to map between OGC standard services and other data exchange protocols mentioned previously, such as netCDF and TAPIR. There are already data models and encodings that can support both netCDF and GML in the one service e.g. see CSML- (Woolf, et al, 2005). This blending of encodings provides for harmonisation between tools traditionally used by climate scientists and OGC adherents.

(iii)Development of re-usable client software capable of rendering and visualising the objects that we are exchanging

The concept of re-usable software that can be deployed in a range of environments is not a new concept, but one that has merit within the SCAR environment. The US Federal Geographic Data Comittee (FGDC) has been promoting the concept of re-usable portlets since 2006. If we can standardise the way in which we encode and represent data, or at least agree on a common interchange model, then there is scope for building software that can operate on these standardised representations. If we develop software in a modular fashion with transparent interfaces and few system dependencies, these reusable components can be readily exchanged or even chained together in suitable contexts to rapidly build complex user interfaces and functionality.

The open source community already has a large range of foundation libraries capable of interacting with OGC and OpenDAP protocols that could be leveraged in some bounded pilot projects.

(iv)Development of a seamless map-base at desirable scales.

A seamless foundation layer of topographic-type data to act as a backdrop, or a spur, for the science conducted in Antarctica and the Southern Ocean is an eminently achievable goal if we had a greater level of coordination, given the availability of new remote sensing technologies and the number and variety of nations operational in Antarctica.

Systematic compilation of updated bathymetric data for the Southern Ocean is currently being undertaken by an expert Sub-group of the SCAR Standing Scientific Group on Geosciences to create a new International Bathymetric Chart of the Southern Ocean

(IBCSO). This charting activity is unlikely to focus on near-shore coastal regions, which are becoming a focus for international interest. SC-AGI has also been compiling a digital database of Antarctica (the SCAR ADD), and already has a map catalogue that points to nationally produced paper and digital maps. A prototype GIS of King George Island has been developed to demonstrate the utility of producing seamless spatial data for a jointly occupied region of Antarctica.

The SCAR ADD is an important flagship product for SC-AGI and it should set the standard for SCAR digital data products in terms of the governance processes used for its compilation, the quality processes used for its generation, as well as the technical standards used in its production. It should be a demonstration product for showing the benefits of using shared datums, coordinate reference systems, metadata, data quality indicators, common terminologies and symbol sets. At present this product is generated and managed largely using the resources of only one SCAR member (i.e.UK). The product has undergone a number of important revisions and updates over time. However, it may now be timely to ensure that the product continues to meet user needs, given the pervasiveness of new web GIS technologies and the ever-increasing demand for additional digital data thematic layers. There is also a need to ensure that the SCAR ADD appropriately integrates with other SC-AGI products such as the Feature and Symbology Catalogues.

It is highly likely that SCAR could play an even greater role than it currently does in fast-tracking the compilation of foundation spatial datasets, which are drawn from national scientific campaigns and territorial management activities, if SCAR better understood what gaps existed in coverage at different spatial scales, and sought a consensus on which gaps to fill first. SCAR members could then actively consider how these gaps might be filled when planning their own national mapping programmes and, closer cooperation may prove achievable between nations occupying adjacent or overlapping territories. The UK and USA have already collaborated to produce a Landsat Image Mosaic of Antarctica (LIMA, http://landsat.usgs.gov/index.php) with a spatial accuracy of 30m, which will benefit a wide range of users.

Recommendation: JCADM to partner with SC-AGI in expanding the use of existing SC-AGI tools into the broader SCAR science and global data management community.

Recommendation: JCADM through its NADCs and SC-AGI should participate in data access trials of existing larger networks, which are using new technologies such as web services. NADCs should also contribute to data modelling activities within these networks, aimed at supporting semantic data interoperability.

Recommendation: SC-AGI to undertake an audit of the SCAR ADD to ensure that it is meeting "user" needs, integrates where necessary with other SC-AGI products, and provides an example of best practice in terms of governance and production methods.

Recommendation: SC-AGI to provide an assessment of the current extent and scale of mapping across Antarctica and a forecast of likely future national or global mapping activities, with the aim of identifying gaps.

Recommendation: SC-AGI to identify opportunities for collaborative capture and/or compilation of mapping data, particularly in regions of high SCAR scientific activity, or where there are gaps in desirable coverage, to progressively create a seamless map coverage of Antarctica at appropriate scales.

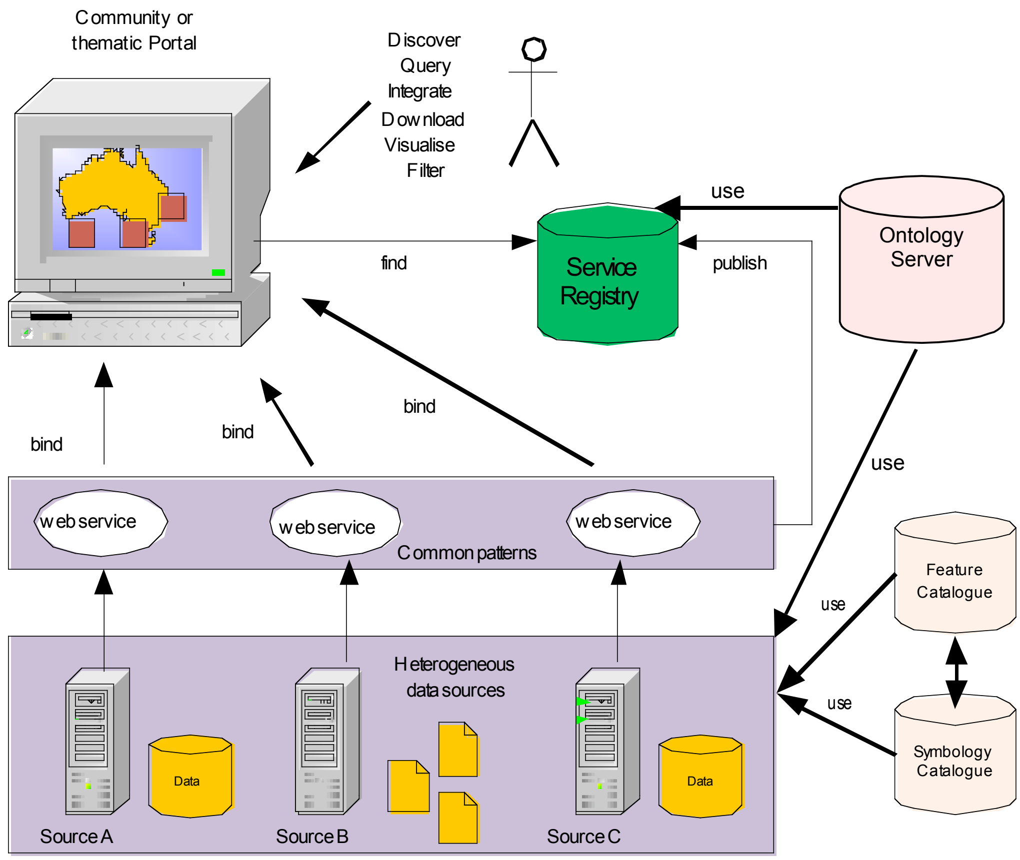

Figure 2 – Components of a distributed services-oriented data access system

5.6 Education, Outreach & Guidance

JCADM and SC-AGI already have well managed web sites (www.jcadm.scar.org & www.antsdi.scar.org) but they could be used to greater effect in terms of their role in education and communication. Both sites currently mix content that should be tailored for two different audience types (i.e. data users and 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 JCADM and SC-AGI group members. Content should then be communicated in a manner that best suits these different audience types.

The JCADM web site has a lot of information content but is relatively poor at providing access to JCADM member's data and services. This is a significant issue because SCAR scientists looking for assistance to access data, through JCADM, should be able to use the JCADM 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 JCADM when participating in SCAR decision-making forums, they should regularly be made aware of the work-plans of both groups. Similarly, all Chairs of SCAR Research Programmes should also be kept fully informed.

Recommendation: JCADM and SC-AGI review their use of the web, and better target content and its communication to different audience types.

Recommendation: JCADM and SC-AGI should ensure that their services/products and work-programs are well-known to all SCAR national delegates and scientific research programmes through targeted information dissemination campaigns

6.0 JCADM & SC-AGI Work-Plans

To create focus for JCADM and SC-AGI, particularly given the limited resources that each group can mobilise, it is suggested that the work of both groups be bounded by an annual or two-yearly work-plan, informed primarily by the strategic recommendations made in this document and the terms of reference for each group. It is anticipated that these work-plans will be relatively detailed, with actions prioritised around those bodies of work that require attention in the short (within year 1), medium (within years 2-3) and long-term (within 3+ years).

The execution of these plans will rely to a large degree on the SCAR Executive also taking a lead role in several of the initiatives raised within this document. The prioritised list of recommendations can be found in Table 1 (at the beingining of this document).

References

Antarctic Treaty. (1961) Retrieved June, 2007 from the WWW: http//www.scar.org/treaty/at_text.html

Brauer, B & Kline S. (2005). SOA Governance: A key ingredient of the Adaptive Enterprise. Retrieved October 7, 2006 from the WWW: http//devresource.hp.com/drc/resources/soa_gov/index.jsp.

Clement, L., Hately, A., von Riegen, C. & Rogers, T. (2004) OASIS UDDI Version 3.0.2 - UDDI Spec Technical Committee Draft.. Retrieved December, 2006 from the WWW: http://uddi.org/pubs/uddi-v3.0.2-20041019.htm

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.

Costello, M. (2000). Developing Species Information Systems. The European Register of Marine Species (ERMS). Oceanography Vol 13 (3).

Christensen, E., Curbera, F., Meredith, G. & Weerawarana, S. (2001) W3C Web Services Description Language (WSDL) 1.1. Retrieved October 2006 from the WWW: http://www.w3.org/TR/2001/NOTE-wsdl-20010315

Keeley R. (2007). Intergovernmental Oceanographic Commission (of UNESCO) JCOMM Data Management Strategy. IOC/IODE-XIX/48.

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

Price, M. (2001) ARC USER Newsletter – April-June 2001. Retrieved December 2007 from the WWW: http://www.esri.com/news/arcuser/0401/datum.html.

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

Rickards L. (2007). Intergovernmental Oceanographic Commission (of UNESCO) Strategic Plan for Oceanographic Data and Information Management – Draft 3 March 2007. IOC/IODE-XIX/45.

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.

W3C (2003). SOAP Version 1.2. Retrieved November, 2006 from the WWW: http://www.w3.org/TR/soap/

Woolf A., Lawrence B., Lowry R., Kleese Van Dam K., Cramer R., Gutierrez M., Kondapalli S., Latham S., O'Neill K., Stephens A. (2005). "Climate Science Modelling Language: Standards-based markup for metocean data", 85th meeting of American Meteorological Society, San Diego, Jan 2005.