

Australia's Role In Creating a Polar Information Commons

Kim Finney

Australian Antarctic Data Centre, Kingston, Tasmania, kim.finney@aad.gov.au

INTRODUCTION

Australia is a signatory to the Antarctic Treaty which is a unique instrument covering the region of the earth below 60° South – a region encompassing the Antarctic continent, its surrounding ocean and atmosphere. The Treaty obliges Australia and other signatory nations to conduct only those activities that are compatible with peaceful endeavours and with the pursuit of science. Under the Antarctic Treaty System information and data are deemed freely exchangeable goods. There is no corresponding Treaty for the Arctic and as yet no internationally agreed Arctic policy regarding data transfer. But given the importance that both the Antarctic and Arctic are known to play in the global climate system there is a pressing need for the communities involved in bipolar research to develop an effective, internationally supported data exchange infrastructure. This infrastructure must be capable of freeing up data to fuel more predictive science and strategically enable the type of environmental change detection that is useful for formulating potential management action in response to undesirable ecosystem shifts. CODATA, the International Council For Science's (ICSU) Data and Technology Committee is championing the development of a bipolar data network to meet these goals called the Polar Information Commons (PIC). Key partners include the Scientific Committee on Antarctic Science (SCAR), International Arctic Science Council (IASC), International Polar Year Program Office (IPO), World Meteorological Organisation (WMO) and the ICSU's Strategic Coordinating Committee on Information and Data (SCCID).

PIC will not be built from scratch but will instead rely on bringing together various existing community efforts in the data management space. It will require strengthening and extension of both the institutional and technological frameworks that currently underpin polar data management and will use Science Commons [1] approaches to foster 'scientific norms' that have the capacity to culturally motivate on-line publication of data. Because the Antarctic has a formalised community of Antarctic data managers, organised under the auspices of the SCAR Standing Committee on Antarctic Data Management (SCADM), it is likely that this group will play a significant role in laying the foundations of PIC.

This paper outlines Australia's contribution to establishing an Antarctic data infrastructure, identifies lessons learned from that experience and postulates what will be required to develop a robust and sustainable PIC.

THE ANTARCTIC DATA INFRASTRUCTURE

Australia currently chairs SCADM, a 25 nation, international collaborative that takes the lead in managing and publishing Antarctic data and information. The community subscribes to one metadata system and uses the NASA-based Global Change Master Directory's (GCMD) Directory Interchange Format (DIF), recently harmonised with the ISO TC 211 19115 metadata standard. There are over 5,000 published metadata records in the Antarctic Metadata Directory (AMD) [2] of which Australia has contributed about 40%. The AMD functions as both a metadata repository and a data access mechanism. Contributing nations are encouraged to hyperlink on-line accessible datasets to the metadata records and there are facilities available to publish datasets as services (e.g. Open GIS Consortium or OPENDAP compliant web services). Although this facility exists, most nations in the SCAR Antarctic community are not publishing data as services. Australia, the UK and the USA are exceptions.

Several community-sponsored products, in addition to the AMD and individual national portals, currently form the core of the Antarctic data infrastructure. These products include a common gazetteer of Antarctic place names; a pan-Antarctic digital topographic database; a map and image repository; a feature catalogue containing representations of real-world objects commonly found in Antarctica, a digital sea-ice extents repository and an Antarctic terrestrial biodiversity database. All of these products, with the exception of the topographic database, are managed and maintained on-line by the Australian Antarctic Data Centre (AADC) on behalf of the SCAR community, albeit with community input. The governance arrangements surrounding each component in the infrastructure differ but all are coordinated under the SCAR institutional framework and importantly all components are publicly accessible.

A recent addition to the Antarctic cyber-infrastructure is the SCAR Marine Biodiversity Information Network (SCAR-MarBIN), hosted and managed by Belgium. SCAR-MarBIN is a marine biodiversity data aggregator service which republishes composed and contextualised data via the SCAR-MarBIN portal and also offers an Antarctic marine species taxonomic names register service (RAMS). It is an example of a facility that both exploits the current infrastructure and adds to it. Australian sourced data was used to boot-strap the service.

REQUIREMENTS FOR A SUSTAINABLE PIC

Australia's experience in working internationally to help develop the Antarctic data commons gives us a unique insight into what some of the key hurdles will be for establishing and maintaining a PIC. The Antarctic Treaty System obliges the free exchange of data, but in reality this has relatively little influence on national compliance. Far greater influence is exerted by the data policies of those individual institutions actively participating in Antarctic science, irrespective of the public stance taken at the national level. Even then the cultural practices of institutional scientists will have a strong bearing on the availability of data from individual countries. Motivating individual scientist participation is therefore crucial to the success of establishing the PIC, particularly from a content provision perspective. It is clear from experience with Antarctic scientists that this means infrastructure tooling must be simple for a user to operate, robust, always available, accessible from almost anywhere and geared to publicise a data provider's contribution. Mainstreaming dataset citation as part of the scientific performance and review process is an avenue now being increasingly explored by those building data networks [3, 4]. This however, is a relatively complex issue to tackle because it must necessarily involve co-operation between data management repositories, journal publishers and the institutions that structure the reward and recognition systems for scientists.

Unlike the building of some types of infrastructure (e.g. telecommunications), building an international scientific data exploitation network often relies heavily on the contributions of unfunded individuals and community volunteers. In this regard the PIC is much more like an open source project than a fully bank-rolled venture with people in paid positions working to a regulated and orchestrated set of plans and budget. However, to succeed both types of development model require a vision, a core group of full-time individuals to drive the process and a roadmap that explains what must be done to get there. What is often poorly appreciated by science community leaders, when it comes to data management activities, is that even predominantly volunteer-based developments must be seed funded and resourced sufficiently to form a core of activity that can be built upon by volunteer labour [5]. To this end the PIC Steering Committee will need to co-opt national institutions with the capability and resources to form the nucleus of the PIC infrastructure and/or identify funding opportunities that can be leveraged to incubate the PIC. There is also a fine balance to be achieved in the infrastructure development cycle between delivering tangible results early and frequently while avoiding delivery of poor quality components that fail to attract interest from users or those who might subsequently create enhancements, and gestating the project for too long and in so doing losing your potential support-base to another promising activity. Being aware of this last issue the PIC Steering Committee is suggesting developing the PIC as both a loosely organised, accessible collection of very lightly documented datasets located somewhere in the internet cloud and as a set of structured (perhaps semantically enriched) data that is hosted by institutional repositories and which is discoverable via more sophisticated search mechanisms through a common gateway. This concept involves migrating data from the PIC cloud into repositories through pre-arranged agreements. The hope is that by exposing deposited data almost immediately in a simple, publicly accessible archive, a multitude of creative internet developers could start to develop tools for mining and exploiting the data without constraint, leading to innovations that wouldn't emerge as quickly in the repository-centric environment.

SUMMARY

Whilst many individual components necessary to underpin a useful data infrastructure already exist, particularly covering scientific data captured in the Antarctic, more must be done to increase community participation and service uptake and to motivate content providers to publish their data if we are to create a sustainable PIC. Australia currently plays a lead role in polar data infrastructure development and is represented on the ICSU's SCCID [6]. This Committee (established in May 2009) is charged with reforming and modernising the ICSU World Data Centre System (WDCS). Australia considers PIC to be an ideal use-case for driving WDCS reform.

REFERENCES

1. Science Commons. Available from <http://sciencecommons.org/>. Accessed on 1st June 2009.
2. Antarctic Master Directory (AMD). Available from <http://gcmd.nasa.gov/KeywordSearch/Home.do?Portal=amd&MetadataType=0> accessed on 1st June 2009.
3. UNESCO - IOC Project Office for IODE. (2008). SCOR/IODE Workshop on Data Publishing. Workshop Report No 207. Available from <http://www.scor-int.org/Publications/wr207.pdf> accessed on 1st June 2009.
4. Costello, M.J. (2009). Motivating Online Publication of Data. *BioScience* 59(5): 418-427
5. Finney, K.T. (2007). A "Bottom Up" Governance Framework For Developing Australia's Marine Spatial Data Infrastructure (SDI). *Data Science Journal, Volume 6*: 64-90
6. ICSU. (2008). Ad hoc Strategic Committee on Information and Data Final Report to the ICSU Committee on Scientific Planning and Review. Available at http://www.icsu.org/Gestion/img/ICSU_DOC_DOWNLOAD/2123_DD_FILE_SCID_Report.pdf. Accessed on 1st June 2009.