Polar Data Activities in Global Data Systems Communiqué

Recommendations & Observations Arising From the 'International Polar Data Forum'

15–16 October 2013, Tokyo (Japan)

Participants in the International Polar Data Forum (comprising of data managers, scientists, and research coordinators) share their observations about the current state of polar data activities and their recommendations for enhancing and sustaining core data services into the future.

General Remarks

Despite the focus generated by the International Polar Year 2007–2008 (http://www.ipy.org/) there are still unresolved deficiencies in the way polar data are managed. These shortcomings continue to hamper our ability to discover and reuse existing and new data assets. The significant public investment in polar science made through national and international research programs is accompanied by an expectation that data should be preserved and be openly available for reuse and verification purposes. This is the responsibility of both scientists and science funders supported by data managers. Achieving these goals requires implementation of open data policies, development of long-term funding strategies to support data repositories, and a change in scientific practices to require the sharing and citation of data. In addition, the ideas and interpretations that have traditionally underpinned research publication need to be updated accordingly. Scientists must factor the costs of managing and publishing data in their research/monitoring funding proposals, and explicitly address these tasks in research/observing system plans.

The Scientific Committee on Antarctic Research (SCAR) and the International Arctic Science Committee (IASC), the two lead non-governmental organizations coordinating international polar research, are now embarking on long-term science planning activities (i.e., the SCAR Science Horizon Scan and the 3rd International Conference on Arctic Research Planning). Both organizations are committed to recommending—through their strategic-direction setting initiatives—not only that the development of robust polar data networks is seen as a science-funded activity but also that this activity has visibility in IASC- and SCAR-sponsored science plans and strategies. Through their policies and actions, IASC and SCAR member countries are encouraged to promote the funding of data management as an integral part of science implementation and to contribute to the design and development of shared global data infrastructure. Participation in the ICSU World Data System (ICSU-WDS; a global system of accredited scientific repositories and data service providers) is one mechanism for member countries to engage in such collaborative infrastructure and capability development.

Forum Observations & Recommendations

- Improving polar data discovery, data preservation, and reusability relies in part on building more pervasive systems interoperability. This interoperability is now a commonly stated goal for polar research organizations, but it is recognized that interoperability needs to be addressed at a number of different levels and covers both social and technical aspects; the combination of which is difficult to address. Considerable investment

is required to develop robust solutions implementable across disciplines that are concurrently sustainable and cost-effective. Much of the long-term activity invested to date has been through sporadic, often-fragmented spurts of voluntary labour; making progress slow. New initiatives such as the Research Data Alliance (https://rd-alliance.org/) potentially offer coordinated mechanisms for addressing interoperability problems in a more efficient, interdisciplinary manner.

- 'Brokering' technology (http://www.eurogeoss.eu/broker) is often a good solution to unify disparate systems whilst preserving domain-specific requirements. However, the heterogeneity and inconsistency of metadata (descriptive information about data) that typically results from disciplinary differences, and which underpins such approaches, can reduce the utility of the unified system. Brokering approaches now being taken by the Global Earth Observing System of Systems and EarthCube (http://www.nsf.gov/geo/earthcube/) are seeking to address these brokering deficiencies, and disciplinary communities are encouraged to engage with such programs to help deliver enhanced solutions.

- Trusted systems (social and technical) must exist to build interoperability, transparency, and data reuse. This will necessitate the development of networks by the polar community. Scientific repository and service provider accreditation systems can help engender trust and can generally lift global capacity for interoperability and data reuse. Development, global harmonization and advertising of such trusted systems should be actively pursued. ICSU-WDS is taking a lead role in this area. Polar repositories and service providers can benefit by engaging in, and affiliating with, accredited networks such as ICSU-WDS.

- The scientific publication and peer assessment process, which is already in transformation and is adapting to accommodate new social forms of communication and practice, must highly prioritize the inclusion of dataset citation as a new norm. This necessarily involves developing new methods to assess data quality and validity, and these should then be encompassed in peer-based acknowledgement and reward systems.

- Being able to uniquely identify, describe, and access resources (e.g., publications, datasets, dataset components, people, organizations, places, projects) using Hypertext Transfer Protocol offers exciting possibilities to interlink currently disparate silos of information and data. This 'Linked Data' approach then has the capacity to enhance the data discovery and integration process.

- Convergence around existing standards for common and domain-based activities is a desirable goal if such standards already cover a significant percentage of requirements. We recommend that disciplinary communities stop developing standards *de novo* when collaborative efforts could be used to enhance existing standards to satisfactorily accommodate requirements.

- Despite the pervasiveness of 'open' data policies, behaviours that constrain access to scientific data are still seen by many as conferring personal, institutional, or national 'competitive advantage'. The polar community must continue efforts to change these attitudes.

- The long tail of smaller, heterogeneous, and often unstructured datasets (those without metadata, mark-up, and not in databases) usually receive minimal data management consideration by both the scientists who produce them and the repositories that manage them in the long-term. However, utilizing the inherent structure of any digital resources provides an objective framework to discover their relationship in a manner that complements existing content and context management solutions. More attention should be paid to making such data web-accessible.

- There are numerous exemplars of data management best practice both inside and outside of the polar community. Generally these 'exemplars' are willing to share their experiences and often their technologies and methods. Frequently, what are lacking are appropriate communication channels and the necessary social connections to capitalize on this best practice. Members of the Arctic community have taken the lead in setting up social media avenues (e.g., https://arctichub.net/groups/adcn and the Twitter feed:@ArcticDCN) to foster better communication, and are now inviting participation from across the polar community.