

THE STATE OF IPY DATA MANAGEMENT: THE JAPANESE CONTRIBUTION AND LEGACY

M Kanao^{1*}, *A Kadokura*¹, *M Okada*¹, *T Yamnouchi*¹, *K Shiraishi*¹, *N Sato*¹, and *M A Parsons*²

^{*1}*National Institute of Polar Research, Research Organization of Information and Systems, 10-3, Midori-cho, Tachikawa-shi, Tokyo 190-8518, Japan*

*Email: * kanao@nipr.ac.jp, kadokura@nipr.ac.jp, okada.masaki@nipr.ac.jp, yamanou@nipr.ac.jp, kshiraishi@nipr.ac.jp, nsato@nipr.ac.jp*

²*National Snow and Ice Data Center, University of Colorado, 449 UCB, Boulder, Colorado 80309-0449, USA*
Email: parsonsm@nsidc.org

ABSTRACT

Diverse data accumulated by many science projects make up the most significant legacy of the International Polar Year (IPY2007-2008). The Polar Data Center (PDC) of the National Institute of Polar Research (NIPR) has a responsibility to manage these data for Japan as a National Antarctic Data Center (NADC) and as the World Data Center (WDC) for Aurora. During IPY, a significant number of multidisciplinary metadata records have been compiled from IPY- endorsed projects with Japanese activity. A tight collaboration has been established between the Global Change Master Directory (GCMD), the Polar Information Commons (PIC), and the newly established World Data System (WDS).

Keywords: International Polar Year, National Antarctic Data Center, Data Management, Metadata Portals, Polar Information Commons, World Data System

1 INTRODUCTION

The International Polar Year (IPY 2007-2008) was the world's most diverse international science program. It was conducted during the 50th anniversary of the International Geophysical Year (IGY 1957-1958). The IPY greatly enhanced the exchange of ideas across nations and scientific disciplines to unveil the status and changes of planet Earth as viewed from the polar regions (Rapley, et al., 2004). This sort of interdisciplinary exchange helps us understand and address grand challenges such as rapid environmental change and its impact on society.

The IPY 2007-2008 was jointly led by the International Council for Science (ICSU) and the World Meteorological Organization (WMO). A Joint Committee of WMO and ICSU (IPY-JC) was established in 2004 to arrange the whole IPY program. The same year in Japan, the IPY national committee was initiated under the Science Council of Japan (SCJ). Eventually, Japanese researchers participated in a total of 63 projects endorsed by the IPY-JC (Sato, et al., 2011).

The scientific results from IPY now begin to emerge, but it is clear that deep understanding will require creative use of myriad data from many disciplines. Many of these projects provided well-coordinated observation platforms, and many continue in the post-IPY era. The huge amount of data accumulating during and after IPY should be the most important legacy for IPY if it is well preserved and utilized (Parsons, et al., 2011a; 2011b).

The Polar Data Center (PDC; <http://www.nipr.ac.jp/english/polar-information01.html>) of the National Institute of Polar Research (NIPR) has served as the Japanese National Antarctic Data Center (NADC) with a strong relationship with the Scientific Committee on Antarctic Research (SCAR) under ICSU. During the IPY, we compiled much of the polar data from the endorsed projects involving Japanese activities. In this paper, the state of IPY data management involving Japan, particularly the tasks of the PDC, are demonstrated. A tight linkage is conducted with other science bodies of ICSU, such as the Committee on Data for Science and Technology (CODATA) and the new World Data System (WDS) (Figure. 1).

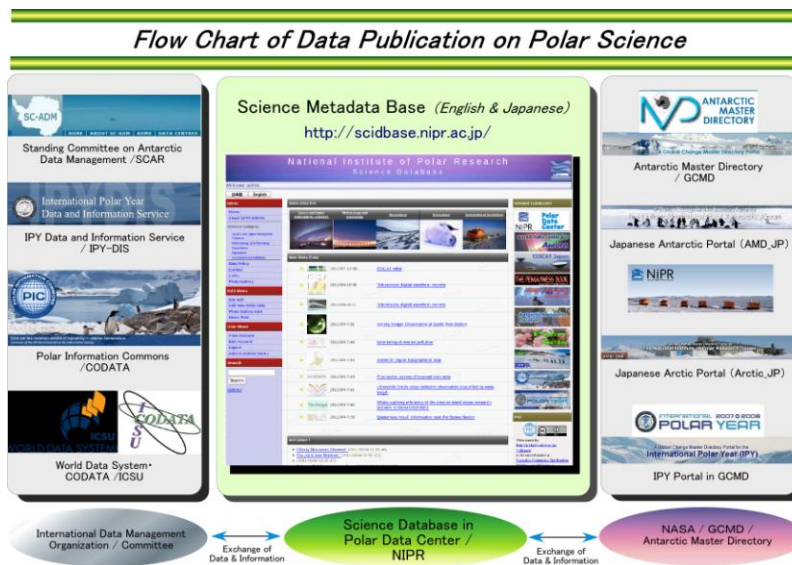


Figure 1. Flow chart of data publication on polar science involving PDC, including IPY data.

2 POLAR DATA CENTER

Recent, rapid, technological improvement and development of Earth observation by satellites and ground observation networks both in the Arctic and the Antarctic has led to a large quantity of polar observation data being collected every day. The processing and utilization of these data is an important issue to promote polar science. Our mission is twofold: scientific data management and management of information infrastructure.

At the 22nd Antarctic Treaty Consultative Meeting (ATCM) in 1998, affiliate countries were obliged to ensure that scientific data collected from Antarctic programs could be freely exchanged and used. Following Article No.III.1.c of the 1998 Antarctic Treaty, each country is required to establish a National Antarctic Data Centre (NADC) and to properly provide the data collected from involved scientists. The PDC at NIPR has performed the function of a NADC for Japan. The PDC established a data policy in February 2007, based of the requirements of the Standing Committee on Antarctic Data Management (SCADM) of SCAR. This contributed to the subsequent SCAR Data and Information Management Strategy (SCAR-DIMS; Finney, 2009; de Bruin, T., & Finney, K., 2011).

Regarding aurora data, in particular, we have administered the World Data Centre (WDC) for Aurora since 1981. The WDC for Aurora (<http://polaris.nipr.ac.jp/~aurora/>) is responsible for data archiving and dissemination of all-sky camera observations, visual observations, other optical observations, auroral image and particle observations from satellites, geomagnetic observations, and observations of the upper atmosphere phenomena associated with aurora such as Ultra Low Frequency (ULF), Very Low Frequency (VLF) and Cosmic Noise Absorption (CNA) activities (Kanao, Kadokura, Yamanouchi, & Shiraishi, 2008).

Outside these obligations, the PDC is responsible for archiving and analysis of Earth observing satellite data (Polar Operational Environmental Satellite: POES of NOAA), seismological data (short-period and broadband seismometers) and crustal movement data (Global Positioning System;GPS, Very Long Baseline Interferometry;VLBI) around the Syowa Station (SYO, 69S, 30E), East Antarctica. Finally, the PDC manages various information infrastructures such as: (1) a mainframe and a workstation system, (2) network systems of domestic and related facilities such as Syowa Station, and (3) Earth observing satellite data reception facilities.

3 METADATA MANAGEMENT

The PDC has the significant task to archive and deliver the digital data obtained from the polar regions. Summary information of all the archived data (metadata) is available to the polar science community as well as more general interests. The compiled metadata describe all kinds of observed/collected science disciplines (space and upper atmospheric sciences, meteorology and glaciology, geoscience and bioscience) from both long- and short-term projects in the Arctic and Antarctic, particularly data collected by the Japanese Antarctic Research

Expedition (JARE) (Kanao, Kadokura, Yamanouchi, & Shiraishi, 2008). In the science meta-database provided by PDC, a total of 150 metadata records had been accumulated as of October 2011 including metadata from IPY endorsed projects (<http://scibase.nipr.ac.jp/>). A new content management system for providing the metadata has been in place since April 2011.

The science database provided by PDC has a tight connection with the Antarctic and Arctic Master Directories (AMDs) in the Global Change Master Directory (GCMD) of the National Aeronautics and Space Administration (NASA) (Fig. 1). In addition to the IPY-related data, data from Japanese national and other international projects have been compiled. Moreover, 210 metadata records have been compiled in the Japanese Antarctic portal (URL; http://gcmd.nasa.gov/KeywordSearch/Home.do?Portal=amd_jp&MetadataType=0) in GCMD.

PDC stores its metadata in our own original format, but this includes the main items listed in the GCMD Directory Interchange Format (DIF). There are tight cross-linkages in corresponding metadata held in the AMD and PDC. Metadata collected by IPY projects for Japan have also been compiled in an IPY portal within the GCMD (<http://gcmd.gsfc.nasa.gov/KeywordSearch/Home.do?Portal=ipy&MetadataType=0>). More than 140 metadata records contributed from Japan are in the IPY portal as of October 2011. This constitutes a significant proportion of all IPY metadata contributed to the GCMD.

It is also noted that there is an Arctic metadata portal in GCMD, describing data about Japanese activities in the Arctic (http://gcmd.gsfc.nasa.gov/KeywordSearch/Home.do?Portal=arctic_jp&MetadataType=0), but the portal server includes only 15 records at the moment.

4 POLAR INFORMATION COMMONS

The Standing Committee on Antarctic Data Management (SCADM) under SCAR has been strongly connected with the activities of the IPY data-management community (IPY Data and Information Service; IPY-DIS). The IPY data policy (http://classic.ipy.org/Subcommittees/final_ipy_data_policy.pdf) emphasizes the need to make data available on the “shortest feasible timescale.” Rapid changes in the polar regions, particularly in the Arctic, make this need to share data more acute because alone, no single investigator or nation can understand these changes. In accordance with the IPY data policy, the data community (IPY-DIS) explicitly recommends that data be formally cited when used, and the IPY Data Committee developed initial guidelines for how data should be cited (Parsons, Duerr, & Minster, 2010). These guidelines harmonize different approaches and they have been adopted by many data centers around the world.

After the end of IPY, a new initiative, the Polar Information Commons (PIC), began as a framework for open and long-term stewardship of polar data and information (Parsons, et al., 2011a). The PIC would serve as an open, virtual repository for vital scientific data and information, and would provide a shared, community-based cyber-infrastructure fostering innovation, improved scientific understanding, and encourage participation in research, education, planning, and management in the polar regions. The PIC builds on the legacy of the IPY and also seeks active participation and ideas from national governments, international organizations, and the scientific and data management communities at large to build this common resource. The PIC was initiated by the International Council of Science (ICSU) Committee on Data for Science and Technology (CODATA) in collaboration with several multidisciplinary science bodies including the World Meteorological Organization (WMO), the International Arctic Science Committee (IASC), the International Union of Geodesy and Geophysics (IUGG), SCAR, Creative Commons, and others. The PIC was officially launched during the IPY Conference in Oslo, June 2010.

The PIC has developed specialized tools that produce a small, machine-readable “badge” that is attached to the metadata or data. This badge asserts that the data are openly available and allows generic search engines or customized portals to automatically identify and locate relevant data, but the badge also requests data users to adhere to basic ethical norms of data use including proper data citation. This service is coupled with a cloud-based data repository for data that may not have a suitable archive elsewhere (<http://www.polarcommons.org/>). NIPR and other Japanese organizations have made significant contributions to the PIC, both by attaching the data/metadata badges and by registration in the cloud-based repository. As of October 2011, Japan was a leading PIC participant and had contributed more than 50 data sets to the PIC.

Polar data can have great relevance for modern, global environmental research well beyond the polar regions. It is critical to explore new approaches like the PIC to develop an effective framework for open, and long-term stewardship of polar data. Data coming from the poles and elsewhere will continue to grow in size and

complexity. The experience of handling IPY data can serve as a valuable case study to examine data management approaches seeking to address issues around complex interdisciplinary science (Parsons, et al., 2011b).

5 WORLD DATA SYSTEM

Through a decision of the 29th General Assembly of ICSU in 2008, a new World Data System (WDS) was established based on the 50-year legacy of two ICSU science bodies – the World Data Centers and the Federation of Astronomical and Geophysical Data Analysis Services. The new WDS aims at a transition from existing standalone WDCs and individual services to a common, globally interoperable, distributed data system that incorporates emerging technologies and new scientific data activities, including polar data as a legacy of the IPY. The new system will build on the potential offered by advanced interconnections between data-management components for disciplinary and multidisciplinary applications.

More than 100 data centers expressed interest to join the new WDS (<http://www.icsu-wds.org/>). The WDC for Aurora, in PDC of NIPR, also expressed interest in joining the new WDS. In October 2010, the ICSU Executive Board accepted the offer from the Japanese National Institute of Information and Communications Technology (NICT) to host and financially support the International Program Office (IPO) for WDS. The office manages and coordinates the establishment and operations of WDS and takes responsibility for outreach and promotional activities.

The first ICSU WDS Conference – Global Data for Global Science – was successfully held at Kyoto University in September 2011, under collaboration with CODATA and the Integrated Risk and Disaster Research (IRDR) of ICSU. It was the first international WDS meeting, and it sought to construct a smooth human network with advanced interconnections between data-management components for disciplinary and multidisciplinary applications across the globe.

The WDS policy of full and open access to data will benefit the international scientific community and ultimately society at large. Many concepts of data publication and data citation should be adopted and promoted by the WDS to facilitate timely release of data. The WDS has agreed to take the necessary steps to archive IPY data and to work with the PIC to preserve, curate, and add value to data in the PIC cloud in order to preserve the legacy of data of the IPY (WDS-SC, 2009).

6 CONCLUSION

The status of IPY data-management in Japan has been summarized in this short paper. Many dedicated data service tasks have been conducted by the staff of PDC in NIPR as a member of NADC under SCAR. Several different aspects of scientific data collected in the polar region have great significance for global environmental research in this century. To construct an effective framework for long-term strategy of the polar data, data must be made available promptly and new Internet technologies such a repository network service like the PIC must be employed.

In addition to the activities in polar science communities of SCAR and the International Arctic Science Committee (IASC), tighter linkages must be established with other cross-cutting science bodies under ICSU, such as CODATA, and WDS. Linkages among these data-management bodies need to be strengthened in the post IPY era.

7 ACKNOWLEDGEMENTS

The authors would like to express their special appreciation to a significant number of collaborators associated with the IPY activities both in national and international projects. They also acknowledge the members of SCADM of SCAR, as well as the IPY Data sub-committee under the IPY-JC for their great efforts to adhere to the data-management issues during the IPY. The authors appreciate the committed individuals of WDS and CODATA for their fruitful discussion and arrangement to initiate the PIC, as well as to create the new data strategy of ICSU. The authors would like to express appreciation to Prof. T. Watanabe and other members of WDS-SC for their arrangement in publishing a special issue on CODATA Data Science Journal, as a proceeding

of the 1st ICSU WDS Conference in Kyoto, 2011.

8 REFERENCES

de Bruin, T., & Finney, K., (2011) The SCAR Data Policy. *SCAR newsletter*, 27, p3.

Finney, K., (2009) SCAR Data and Information Management Strategy (DIMS) 2009 – 2013. In Summerhayes, C., & Kennicutt, C., (Eds.), *SCAR Ad-hoc Group on Data Management*, 34, Cambridge, Scott Polar Research Institute.

Kanao, M., Kadokura, A., Yamanouchi, T., & Shiraishi, K., (2008) The Japanese National Antarctic Data Centre and the Japanese Science Database. *JCADM newsletter*, 1, p10.

Parsons, M.A., de Bruin, T., Tomlinson, S., Campbell, H., Godoy, Ø., LeClert, J., & the IPY Data Policy and Management Subcommittee, (2011a) The State of Polar Data—the IPY Experience. In Krupnik, I., Allison, I., Bell, R., Cutler, P., Hik, D., Lopez-Martinez, J., Rachold, V., Sarukhanian, E., & Summerhayes, C., (Eds.), *Understanding Earth's Polar Challenges: International Polar Year 2007-2008 –Summary by the IPY Joint Committee-*, 3.11, 457-476, Edmonton, Alberta, Art Design Printing Inc..

Parsons, M.A., Godoy, Ø., LeDrew, E., de Bruin, T., Danis, B., Tomlinson, S., & Carlson, D. (2011b) A Conceptual Framework for Managing Very Diverse Data for Complex, Interdisciplinary Science. *J. Information Sci.*, 1-21, DOI: 10.1177/0165551500000000.

Parsons, M.A., Duerr, R., & Minster, J.-B., (2010) Data Citation and Peer Review. *EOS Transactions*, American Geophysical Union, 91(34), 297-298.

Rapley, C., Bell, R., Allison, I., Bindschadler, P., Casassa, G., Chown, S., Duhaim, G., Kotlyakov, V., Kuhn, M., Orheim, O., Pandey, P. C., Petersen, H., Schalke, H., Janoschek, W., Sarukhanian, E., & Zhang, Z., (2004) A Framework for the International Polar Year 2007–2008. *ICSU IPY 2007–2008 Planning Group*. ICSU, Paris, pp 57.

Sato, N., Ito, H., Kanao, M., Kanda, H., Naganuma, T., Ohata, T., Watanabe, K., & Yamanouchi, Y., (2011) Engaging Asian Nations in IPY:Asian Forum for Polar Sciences (AFoPS) (Japanese Section). In Krupnik, I., Allison, I., Bell, R., Cutler, P., Hik, D., Lopez-Martinez, J., Rachold, V., Sarukhanian, E., & Summerhayes, C., (Eds.), *Understanding Earth's Polar Challenges: International Polar Year 2007-2008 –Summary by the IPY Joint Committee-*, 5.3, 555-574, Edmonton, Alberta, Art Design Printing Inc..

World Data System - Scientific Committee (WDS-SC) (2009) IPY & WDS -Executive Summary-. 2/7.1 Ver. 1.1