

PRACTICE PAPER

Science Metadata Management, Interoperability and Data Citations of the National Institute of Polar Research, Japan

M Kanao¹, M Okada¹, J Friddell² and A Kadokura¹¹ Polar Data Centre, National Institute of Polar Research, Research Organization of Information and Systems, 10-3 Midori-cho, Tachikawa-shi, Tokyo 190-8518, JP² Polar Data Catalogue and Canadian Cryospheric Information Network, University of Waterloo, 200 University Avenue West, Waterloo, Ontario N2L 3G1, CACorresponding author: M. Kanao (kanao@nipr.ac.jp)

The Polar Data Centre (PDC) of the National Institute of Polar Research (NIPR) has a responsibility to manage polar science data as part of the National Antarctic Data Centre and the Science Committee on Antarctic Research. During the International Polar Year (IPY 2007–2008), a remarkable number of data/metadata involving multi-disciplinary science activities were compiled. Although the long-term stewardship of the accumulation of metadata falls to the data center of NIPR, the work has been in collaboration with the Global Change Master Directory, the Polar Information Commons, the World Data System and other data science bodies/communities under the International Council for Science. In addition, links with other data centers, such as the Data Integration and Analysis System Program of the Global Earth Observation System of Systems and the Polar Data Catalogue of Canada were initiated in 2014 using the Open Archives Initiative Protocol for Metadata Harvesting. The metadata compiled by the PDC were recently modified using an automatic attributing system and DataCite through the Japan Link Center.

Keywords: Polar Data Centre; Metadata management; Interoperability; Data citation; DOI

1 Introduction

The Polar Data Centre (PDC) of the National Institute of Polar Research (NIPR) forms part of the National Antarctic Data Center (NADC) for the Scientific Committee on Antarctic Research (SCAR) of the International Council for Science (ICSU). Close cooperation has been achieved with several global data bodies under the ICSU including the Standing Committee on Antarctic Data management (SC-ADM) of the SCAR, the Arctic Data Committee (ADC) of the International Arctic Science Committee (IASC), involved Working Groups/Export Groups within the Data for Science and Technology (CODATA) and World Data System (WDS), together with other groups and committees related to global data issues.

During the International Polar Year program (IPY 2007–2008) and beyond, the PDC/NIPR mined a significant amount of scientific data involving polar activities mainly based on Japanese endorsed projects (Parsons et al., 2011a; 2011b; Kanao et al., 2013). In this paper, the present status of metadata/data management employed by the PDC/NIPR after the IPY era are investigated by focusing on several new trials on data interoperability with other data centers, as well as initiatives for metadata/data citation/publication by attributing the Digital Object Identifier (DOI) for the purpose of the utilization of data by the polar/global communities (e.g., Duerr et al., 2011; Ball & Duke, 2012; ESIP, 2012). Interoperable metadata linkage and promoting data citations could provide an efficient model in a framework for long-term preservation and publication of polar data by the global system.

2 Metadata Management

The PDC/NIPR is charged with archiving and delivering data obtained from the polar regions. Summary information (metadata) of all the archived data is made available to both the polar communities and the public domain. The metadata compiled by the PDC/NIPR describe various scientific research disciplines (space and upper atmospheric science, meteorology, glaciology, geoscience, and biosciences) from both long- and short-term projects in the Arctic and Antarctic, particularly with respect to data collected by the Japanese Antarctic Research Expedition (JARE) (Kanao et al., 2014). These categories cover almost all of the studies on environmental change and earth evolution viewed from the polar regions. A total of 365 records including the data obtained from IPY projects have been compiled in the portal server for scientific metadata (<http://scidbase.nipr.ac.jp/>) as of October 2016.

The NIPR metadata portal server is connected with the Antarctic and Arctic Master Directories (AMD; <http://gcmd.gsfc.nasa.gov/KeywordSearch/Home.do?Portal=amd&MetadataType=0>) in the Global Change Master Directory (GCMD) of the National Aeronautics and Space Administration (NASA). In addition to the IPY data, the metadata from other national and international projects have been compiled in the GCMD, and almost 300 metadata records have been amalgamated (as of October 2016) in the Japanese Antarctic portal (http://gcmd.nasa.gov/KeywordSearch/Home.do?Portal=amd_jp&MetadataType=0; **Figure 1**). Although the PDC/NIPR portal server stores all the metadata in their original form in an initial stage after registration into the portal, the majority of the field items in the GCMD Directory Interchange Format (DIF) are also included, and metadata in both the AMD and the PDC/NIPR metadata portal are closely linked to each other.

The Polar Information Commons (PIC; <http://www.polarcommons.org/>) was envisioned after the IPY in 2010 as a shared virtual resource mirroring the geographic commons (Parsons et al., 2011). The PIC could 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 metadata portal of PDC/NIPR has also been providing their data to the PIC (**Figure 1**). Metadata from a total of 20 projects have been compiled so far inside the PIC cloud server.

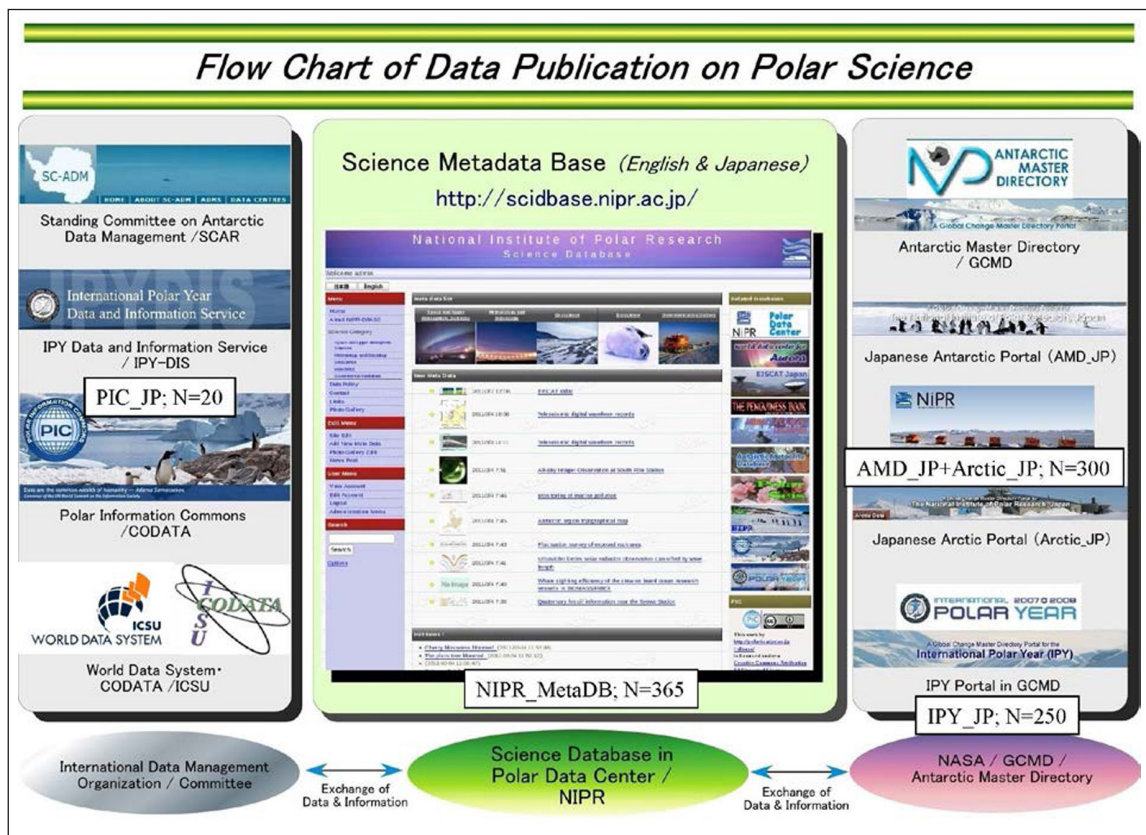


Figure 1: Flowchart of data publication on polar science involving the NIPR metadata portal server (<http://scidbase.nipr.ac.jp/>). Metadata of numerous projects have been registered at each portal site (AMD/GCMD, PIC, etc.). N = number of projects.

3 Interoperability

Providing the metadata of the PDC/NIPR portal server to “A Search and Discovery System for the Datasets” project of “Data Integration and Analysis System Program (DIAS; <http://www.editoria.u-tokyo.ac.jp/projects/dias/?locale=ja>, <http://www.diasjp.net/en/>)” started in March 2014, by cooperating with the University of Tokyo, Kyoto University, the National Institute of Informatics (NII), and other institutions (Figure 2).

The goals of DIAS are to collect and store Earth observation data; to analyze such data in combination with socio-economic data and convert it into information useful for crisis management with respect to global-scale environmental issues, natural disasters, and other threats; and to make this information available both within Japan and overseas. The DIAS program also aims to help resolve global issues through policy-making assistance; the development of applications and tools through cooperative planning and production with the industrial world; and the creation and social implementation of new public benefits. On the international stage, DIAS is also connected to data centers across the world participating in the Global Earth Observation System of Systems (GEOSS), positioning it as an international contribution to the project.

All the metadata produced by the PDC/NIPR portal server (<http://scibase.nipr.ac.jp/>) are available from DIAS by updating them to reflect the latest datasets every day, although the metadata are provided only for English datasets. The metadata records follow the DIF standard format that has been offered for many years.

The link between the PDC/NIPR and the Polar Data Catalogue of Canada (<https://www.polardata.ca/>) was initiated in May 2014. The Polar Data Catalogue is a database of metadata and data that describes, indexes, and provides access to diverse datasets generated by Arctic and Antarctic researchers (Friddell et al., 2014). The records in the Polar Data Catalogue follow the ISO 19115 standard format in order to enable metadata exchange with other data centers. These records cover a wide range of disciplines from natural sciences and policy, to health and social sciences. The metadata in the portal server of PDC/NIPR has been provided to the Polar Data Catalogue using the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH; <http://www.openarchives.org/OAI/2.0/openarchivesprotocol.htm>).

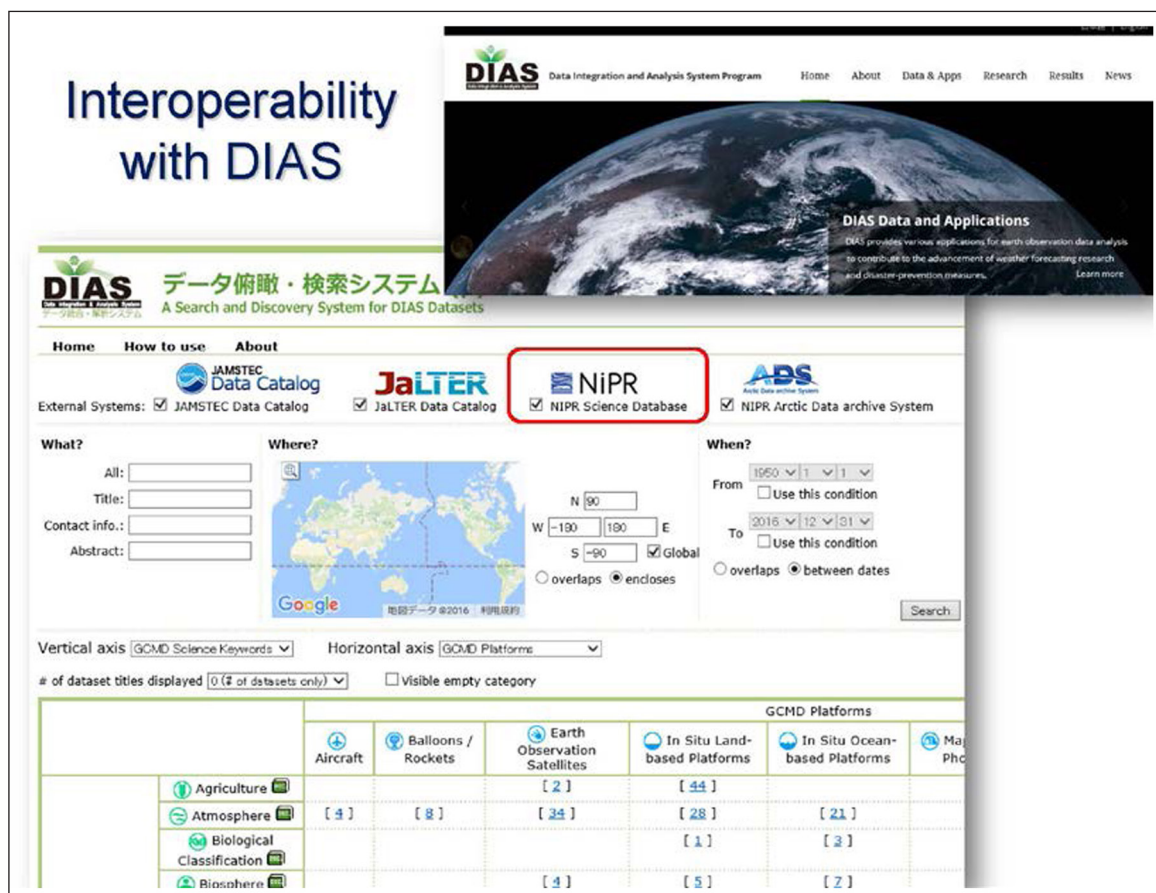


Figure 2: Screenshot of the DIAS website (<http://www.diasjp.net/en/>). Metadata from NIPR are available on the site.

The OAI-PMH is a low-barrier mechanism for repository interoperability using a set of six verbs or services that are invoked within HTTP. Data providers handle the repositories that expose structured metadata via OAI-PMH. Service providers then make OAI-PMH service requests to harvest that metadata. Great efforts have been made to match the corresponding fields of each metadata record by both the PDC/NIPR and the Polar Data Catalogue. The compiled data can cover both polar regions: the Arctic and Antarctic.

4 Data Citations

A software system that can automatically attribute the DOI for the compiled metadata from the PDC/NIPR was recently installed on the portal server. The DOIs can be requested using DataCite (<https://www.datacite.org/>) through the gateway interface provided by the Japan Link Center (JaLC; <https://japanlinkcenter.org/>). The JaLC is the only Japanese organization authorized as a registration agency which can provide the DOIs.

Using various evaluation procedures, the metadata and their associated data are attributed DOIs with a prefix of 10.17592. Under the NIPR DOI auto-numbering rule, the suffix part of the DOI (i.e., the character string ordering) is generated arbitrarily in a manner defined by the metadata portal of PDC/NIPR (**Figure 3**). The landing page of the data attributed by its DOI is initially oriented to the English version of the corresponding metadata in the PDC/NIPR portal server (<http://scidbase.nipr.ac.jp/>) (**Figure 4**).

After receiving offers to obtain DOIs from the data providers/managers, the quality of each data record will be strictly evaluated by the NIPR data management committee, followed by attributing the DOIs with a sufficient quality for opening/publishing in the public domain. There are several evaluation terms before assignment of the DOIs regarding data quality, publishing methodology, long-term maintenance strategy, and data policy, and these should be overcome in both the description of the metadata record itself and the quality of the corresponding actual dataset.

Significant effort has been expended on the systems mentioned in this paper by the staff of the NIPR. Multi-disciplinary scientific data collected in the polar regions have great value for researchers of global

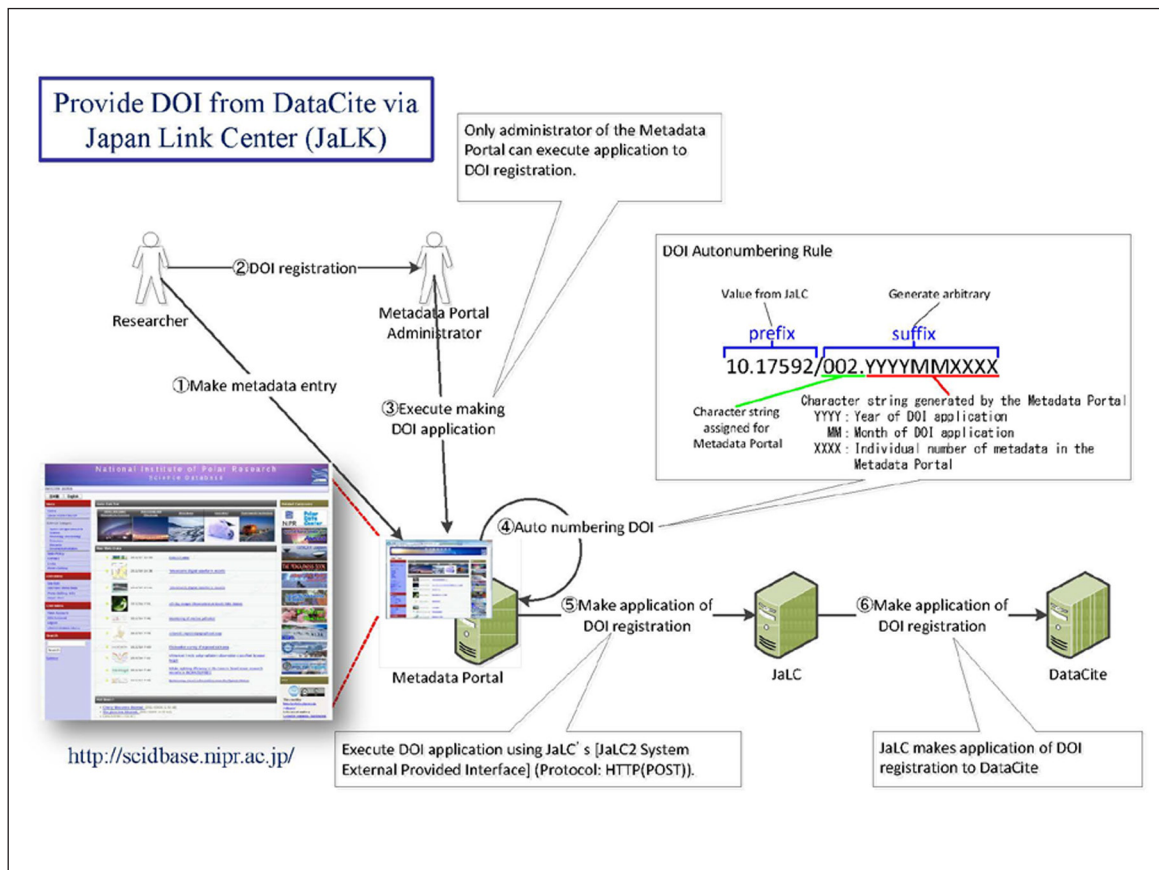


Figure 3: Flowchart to obtain the DOI prefix from DataCite to the PDC/NIPR metadata portal server (<http://scidbase.nipr.ac.jp/>) via the JaLC gateway interface.

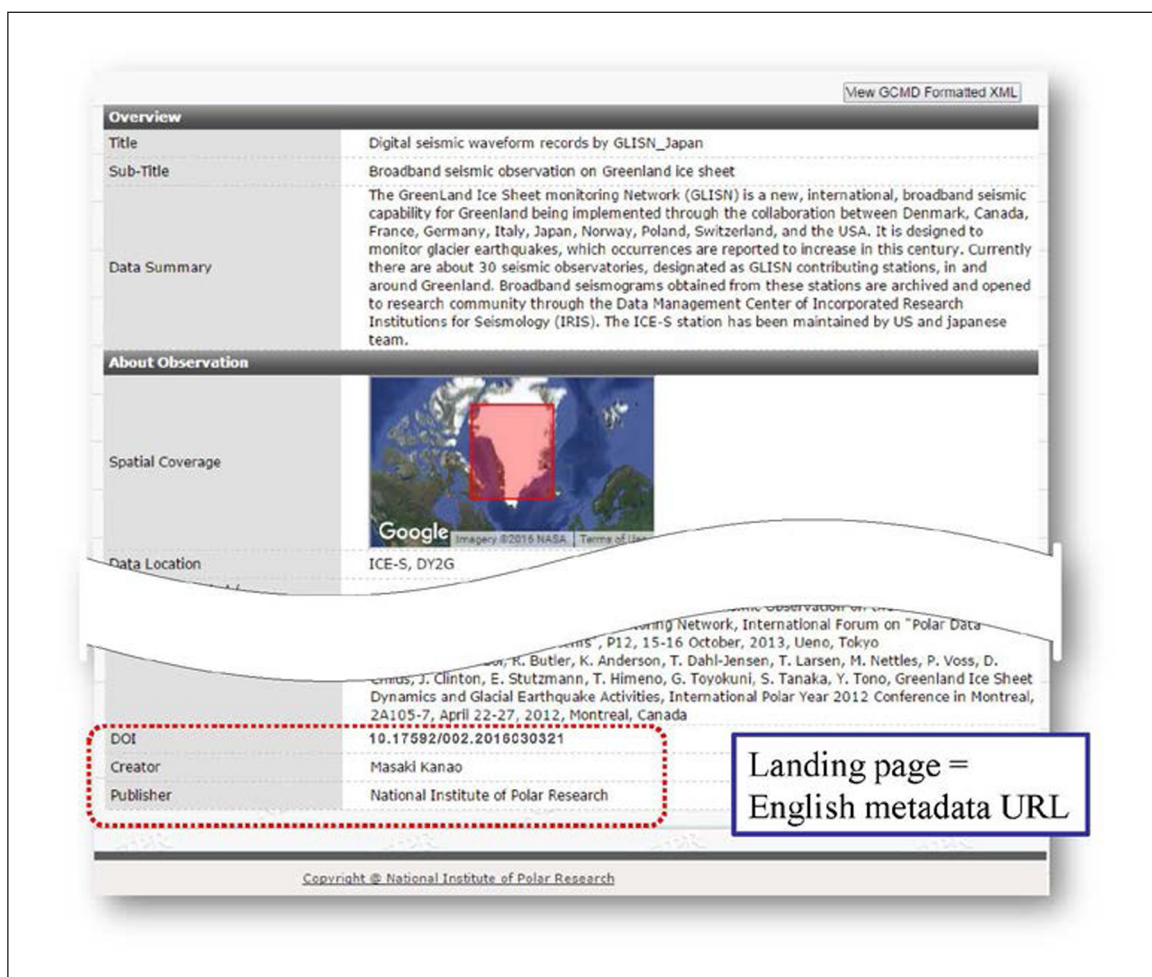


Figure 4: Sample image of the landing page for the DOI attributed metadata in PDC/NIPR portal server (<http://scibase.nipr.ac.jp/>).

environmental change (Parsons et al., 2011b). Interoperable metadata linkage and promoting the data citation system introduced in this paper could demonstrate a model case with an effective framework for a long-term strategy for the publication and preservation of polar data. Moreover, the approach of interoperability and data citation/publication conducted by this study could be relevant to other applications (Parsons & Fox, 2013; Lawrence et al., 2011).

The next generation of NIPR database will be designed to compile all datasets by using integrated applications for the public domain. The future integrated database will be composed of Arctic and Antarctic data, as well as earth/environmental/bioscience and social/human science information. The new database will also provide information to related data centers and libraries.

5 Acknowledgements

The authors would like to express their appreciation to the many collaborators involved in polar data management and in particular the members of PDC/NIPR, SC-ADM of SCAR, ADC of IASC, WDS, CODATA, and IPY. They also appreciate for comments from the two anonymous reviewers and acknowledge the organizing committee of "SciDataCon 2016" for arranging this special issue of the *CODATA Data Science Journal*, following SciDataCon, held in Denver, USA, in September 2016.

Competing Interests

The authors have no competing interests to declare.


References

- Ball, A and Duke, M** 2012 How to Cite Datasets and Link to Publications. *DCC How-to Guides*. Digital Curation Centre. <http://www.dcc.ac.uk/resources/how-guides/cite-datasets>.
- Duerr, R, Downs, R, Tilmes, C, Barkstrom, B, Lenhardt, W, Glassy, J, Bermudez, L and Slaughter, P** 2011 On the utility of identification schemes for digital earth science data: an assessment and recommendations. *Earth Science Informatics*, 4: 139–160. DOI: <https://doi.org/10.1007/s12145-011-0083-6>
- ESIP (Federation of Earth Science Information Partners)** 2012 Data Citation Guidelines for Data Providers and Archives. In: Parsons, M A, Barkstrom, B, Downs, R R, Duerr, R, Tilmes, C, and ESIP Preservation and Stewardship Committee (Eds.) *ESIP Commons*. <http://commons.esipfed.org/node/308>.
- Friddell, J E, Michaud, J, Vincent, W F and LeDrew, E F** 2014 The Polar Data Catalogue: Best Practices for Sharing and Archiving Canada's Polar Data. *CODATA Data Science Journal*, 13: PDA1–PDA7. DOI: <https://doi.org/10.2481/dsj.IFPDA-01>
- Kanao, M, Kadokura, A, Okada, M, Yamnouchi, T, Shiraishi, K, Sato, N and Parsons, M A** 2013 The State of IPY Data Management: The Japanese Contribution and Legacy. *CODATA Data Science Journal*, 12: WDS124–WDS128. DOI: <https://doi.org/10.2481/dsj.WDS-0211>
- Kanao, M, Okada, M and Kadokura, A** 2014 Metadata Management at the Polar Data Center of the National Institute of Polar Research, Japan. *CODATA Data Science Journal*, 13: PDA27–PDA31. DOI: <https://doi.org/10.2481/dsj.IFPDA-05>
- Lawrence, B, Jones, C, Matthews, B, Pepler, S and Callaghan, S** 2011 Citation and peer review of data: Moving towards formal data publication. *International Journal of Digital Curation*. 6(2). DOI: <https://doi.org/10.2218/ijdc.v6i2.205>
- Parsons, M A, de Bruin, T, Tomlinson, S, Campbell, H, Godoy, Ø, LeClert, J, et al.** 2011a The State of Polar Data—The IPY Experience. In: Krupnik, I, Allison, I, Bell, R, Cutler, P, Hik, D, Lopez-Martinez, J et al. (Eds.), *Understanding Earth's Polar Challenges: International Polar Year 2007–2008—Summary by the IPY Joint Committee 3.11*, Edmonton, Alberta: Art Design Printing Inc, 457–476.
- Parsons, M A, Godoy, Ø, LeDrew, E, de Bruin, T, Danis, B, Tomlinson, S and Carlson, D** 2011b A Conceptual Framework for Managing Very Diverse Data for Complex, Interdisciplinary Science. *Journal of Information Science*, 1–21. DOI: <https://doi.org/10.1177/016555150000000>
- Parsons, M A and Fox, P A** 2013 Is Data Publication the Right Metaphor? *CODATA Data Science Journal* 12, WDS32–WDS46.

How to cite this article: Kanao, M, Okada, M, Friddell, J and Kadokura, A 2018 Science Metadata Management, Interoperability and Data Citations of the National Institute of Polar Research, Japan. *Data Science Journal*, 17: 1, pp. 1–6, DOI: <https://doi.org/10.5334/dsj-2018-001>

Submitted: 20 October 2016 **Accepted:** 01 December 2017 **Published:** 24 January 2018

Copyright: © 2018 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

 *Data Science Journal* is a peer-reviewed open access journal published by Ubiquity Press.

OPEN ACCESS 