



Activities of the Polar Environment Data Science Center of ROIS-DS, Japan

COLLECTION:
WDS ASIA-OCEANIA
CONFERENCES 2019

PRACTICE PAPER

AKIRA KADOKURA

MASAKI KANAO

HIRONORI YABUKI

YOSHIMASA TANAKA

KOJI NISHIMURA

**Author affiliations can be found in the back matter of this article*

]u[ubiquity press

ABSTRACT

The Polar Environment Data Science Center (PEDSC) is one of the centers of the Joint Support-Center for Data Science Research (DS) of the Research Organization of Information and Systems (ROIS), which was established in 2017. The purpose of the PEDSC is to promote the opening and sharing of the scientific data obtained by research activities in the polar region led by the National Institute of Polar Research (NIPR). Activities of the PEDSC have been carried out along a five year plan with the following seven specific tasks since 2017: (1) construction of an integrated database; (2) upgrade and interoperable use of the three existing database systems (NIPR Science Database, Arctic Data archive System (ADS), and Inter-university Upper atmosphere Global Observation NETWORK system (IUGONET)); (3) processing of the time-series digital data; (4) processing of the sample data; (5) data publication in the Polar Data Journal; (6) collaboration with external communities; and (7) promoting data science using the database and database system.

CORRESPONDING AUTHOR:

Akira Kadokura

Polar Environment Data
Science Center, Joint
Support-Center for Data
Science Research, Research
Organization of Information
and Systems, Tokyo, JP

kadokura@nipr.ac.jp

KEYWORDS:

Data activity; polar science;
open data; data science; polar
region

TO CITE THIS ARTICLE:

Kadokura, A, Kanao, M, Yabuki,
H, Tanaka, Y and Nishimura,
K. 2022. Activities of the Polar
Environment Data Science
Center of ROIS-DS, Japan.
Data Science Journal, 21:
12, pp. 1–8. DOI: [https://doi.
org/10.5334/dsj-2022-012](https://doi.org/10.5334/dsj-2022-012)

1. INTRODUCTION

The Polar Environment Data Science Center (PEDSC) was launched in 2017 as one of the centers of the Joint Support-Center for Data Science Research (DS) of the Research Organization of Information and Systems (ROIS). In the following sections, brief histories of the ROIS, DS, and PEDSC are introduced in the sections 2 and 3, the data treated by the PEDSC and previous data activities and current status are described in the sections 4 and 5, the tasks and activity plan of the PEDSC are described in section 6, and the activities of the PEDSC during 2017-2019 are described in section 7.

2. HISTORY OF ROIS AND DS

The ROIS was established in 2004 as an Inter-university Research Institute Corporation, consisting of four national institutes: the National Institute of Polar Research (NIPR), the National Institute of Informatics (NII), the Institute of Statistical Mathematics (ISM) and the National Institute of Genetics (NIG). The third term of the ROIS started in 2016. The DS was established in 2016 at the start of the third term as a flagship platform of the ROIS. The main purpose of the DS is to support activities for data sharing, data analysis, and data scientist human resource development in universities and other external communities, responding to the needs of the times for the open science and data centric science. The history and current organization structure of the ROIS can be seen in the Web sites (<https://www.rois.ac.jp/en/about/rois.html> and <https://www.rois.ac.jp/en/about/organization.html>, respectively), and the structure of the DS can be seen in the Web site (https://ds.rois.ac.jp/en/aboutus/en_outline/).

3. HISTORY OF PEDSC

The PEDSC is a center of the DS designed to support the data activities associated with polar science. The planning phase for the PEDSC started in 2015, and a preparatory office was made in the NIPR in 2016. After detailed planning for the schedule and structure of the PEDSC in the preparatory office, the PEDSC has been formally launched in 2017. The PEDSC has a close relationship with the scientific observation and research activities of the NIPR. The main purpose of the PEDSC is to strengthen and promote the opening and sharing of the polar science data for further broader communities, closely collaborating with the polar science research community led by the NIPR. Internationally, the PEDSC intends to play a key role of the National Data Center (NDC) for polar science in Japan, corresponding to international bodies for the Antarctic and Arctic sciences. The goals of the PEDSC are to play a central role of the data activities in the polar science research in Japan, to create a new data centric polar science, and to contribute to the advances of the global environmental studies. The current staff of the PEDSC consists of five research members, three assistants, and six concurrent advisory members from the NIPR. Details of the PEDSC can be seen in its Web site (<http://pedsc.rois.ac.jp/en/>).

4. DATA TREATED BY THE PEDSC

Data treated by the PEDSC are obtained by scientific research observations and activities in the Antarctic and Arctic regions. Various observations and research activities have been carried out in both polar regions by research scientists collaborating with the NIPR in various research fields, including space and upper atmospheric sciences, meteorology, marine science, glaciology, solid earth science, and bioscience. The data are obtained in various forms, e.g., digital data in various recording media, or samples stored in various containers, etc. Those data are processed and analyzed afterward and converted to the secondary physically meaningful data, and then used for obtaining scientific results. In order to obtain accurate scientific results, reliability of the data should be ensured. For that purpose, it is required that the data should be securely stored, not be lost, not be degraded, not be falsified, and be reusable for everyone to obtain the same results. On the other hand, it is also required for doing multidisciplinary or cross-disciplinary studies, such as the study of global environmental change, to use data in various research fields together to invent new findings. In such a case, information on the location and attribute of those various data (metadata) should be treated in a unified manner.

And also, it could be possible that some data in some research field are applied to the other unrelated research field to get unpredictable new findings and new values. For such a case, public property and accessibility of the data are important. The task of the PEDSC is to support such activities for processing, analyzing, archiving, sharing, and opening polar science data to promote further collaboration with external communities, as shown in [Figure 1](#).

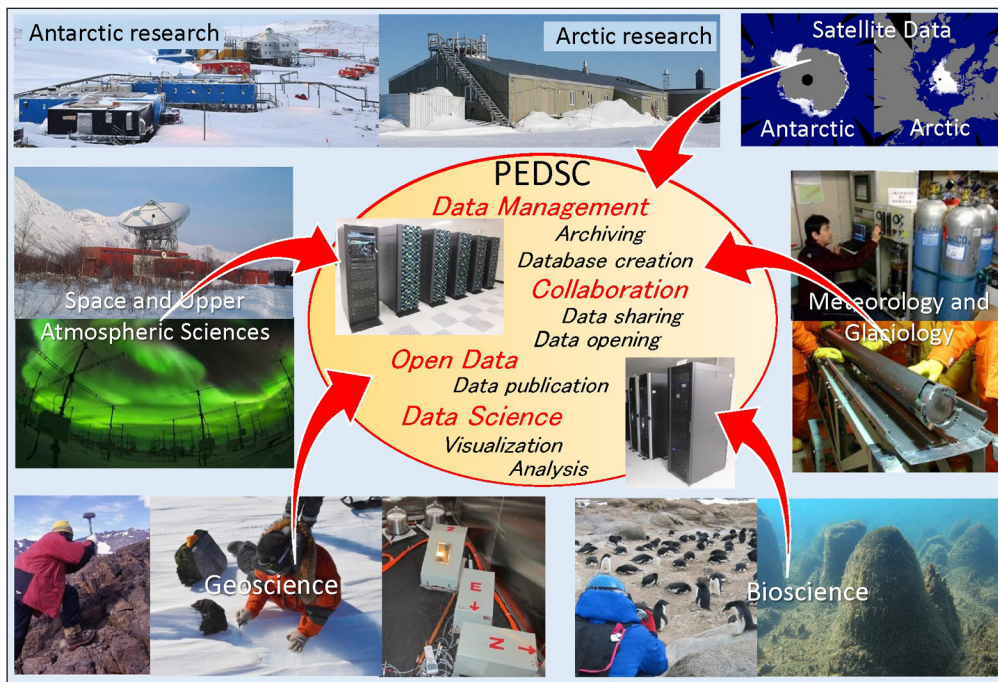


Figure 1 Conceptual figure showing the tasks of the PEDSC to support polar science using the data obtained in both polar regions.

5. PREVIOUS DATA ACTIVITIES AND CURRENT STATUS

The Japanese Antarctic Research Expedition (JARE) was started in 1956, and Japanese Antarctic station Syowa was established in 1957. In the Arctic region, the NIPR Arctic research station was established in 1991 at Ny-Ålesund on Spitsbergen in the Svalbard Archipelago. In the history of the scientific activities in both polar regions over the past 60 years, databases have been created in a variety of research fields. Those data activities were done by data managers in each research field, and depend on the human, hardware, and software resources unique to each team, resulting in various current status of the data processing, database creation, and opening data to the public for each observation dataset. It is also a current problem that an integrated database covering all the research fields of polar science has not yet been constructed, and it is difficult to do a cross-disciplinary search, use, and access. On the other hand, some database systems to treat various data in various research fields have been developed, such as the ‘NIPR Science Database (NIPR-SDB)’ (<https://scidbase.nipr.ac.jp/>), ‘Arctic Data archive System (ADS)’ (<https://ads.nipr.ac.jp/>), and ‘Inter-university Upper atmosphere Global Observation NETwork system (IUGONET)’ (<http://www.iugonet.org>).

The NIPR-SDB has been constructed and operated since 2007 as a metadata database covering almost all the data in both polar regions. The NIPR-SDB has a close relationship with international data activities, especially with the Standing Committee on Antarctic Data management (SCADM) of the Scientific Committee on Antarctic Research (SCAR) and the Global Change Master Directory (GCMD) of the National Aeronautics and Space Administration (NASA) (Kanao, et al., 2013; 2014; 2018). The NIPR-SDB is some kind of a complete data catalog. The ADS has been constructed and operated since 2012 mainly for the specific Arctic projects in Japan (GRENE (Green Network of Excellence) (<https://www.nipr.ac.jp/grene/e/index.html>) and ArCS (Arctic Challenge for Sustainability) (<https://www.arcs-pro.jp/en/index.html>)). The ADS is not only a metadata database system but also has a function for the online visualization and analysis of actual data. The ADS has a close relationship with international Arctic research community such as the International Arctic Science Committee (IASC). The IUGONET has been developed and operated as a Japanese inter-university research project among four Japanese universities and the NIPR (and 6 collaborative organizations), as shown in <http://www.iugonet>.

org/about/organization.jsp?lang=en, since 2009 to build the metadata database for ground-based observations of the Earth's upper atmosphere. Currently, ROIS-DS-PEDSC plays a central role in the IUGONET project in the sense that one research member of the PEDSC is the leader of the IUGONET system development team and the PEDSC is doing a financial support to employ the chief engineering research staff in the development team. The PEDSC is also responsible for supplying the related polar science data of the NIPR to the IUGONET system. The IUGONET also develops the software to analyze the observation data provided by various universities/institutes, closely collaborating with the Space Physics Environment Data Analysis System (SPEDAS) community in the United States (Hayashi, et al., 2013; Tanaka, et al., 2013; Yatagai, et al., 2014; Angelopoulos, et al., 2019). Data treated by the IUGONET are obtained not only in the polar region but also in the low and middle latitudes. As described above, each purpose and background of each existing database system is different from each other, and an integrated database system to search, visualize, and analyze all the data in polar region in all the research fields has not yet been constructed.

As for the publication of the polar science data, in 1968 the NIPR launched the 'JARE Data Reports' (<http://polaris.nipr.ac.jp/~library/publication/pub/pube.html#anchor200820>) for the data obtained in the JARE activities, and the 'NIPR Arctic Data Reports' (https://nipr.repo.nii.ac.jp/?action=repository_opensearch&index_id=784) in 1996 for the data obtained in the Arctic region. In January, 2017, a new data journal 'Polar Data Journal' (PDJ) (<https://pdr.repo.nii.ac.jp/>) was launched by the NIPR, collaborating with the Japanese Institutional Repositories Online Cloud (JAIRO Cloud) system of the NII. The PDJ is a free-access, peer-reviewed online journal, and publishes not only the data paper but also the related data itself with assigning its own DOI (Digital Object Identifier). The NIPR Arctic Data Reports were merged into the PDJ in 2017, and the JARE Data Reports will also be merged into the PDJ in the future.

6. TASKS AND ACTIVITY PLAN OF THE PEDSC

Based on the above-mentioned current status and scientific needs, the PEDSC defines the following 7 tasks to undertake between 2017 and 2021.

1. To construct an integrated database system to cover all the data in all the research fields of polar science.
2. To make the existing database systems (NIPR-SDB, ADS, IUGONET) interoperable with each other for the polar science data.
3. To promote archiving, opening, and sharing of the time-series digital data in each research field.
4. To promote archiving, opening, and sharing of the sample data in each research field.
5. To promote publication of the scientific data, collaborating with the Polar Data Journal of the NIPR.
6. To promote collaboration with universities and other institutions in Japan and international communities.
7. To promote data science using the database and database system.

Figure 2 shows the relationship among the NIPR, PEDSC, and external community associated with the above seven tasks. The PEDSC plays a role of a bridge between the NIPR and the external community to support the data activity in polar science. **Figure 3** shows the annual schedule of the PEDSC activity during the five years associated with the seven tasks, which was planned at the start of the PEDSC in 2017. It should be noted that the research staff in the PEDSC consists of the persons who are in charge of the NIPR-SDB, ADS, and IUGONET systems, and the editorial board members of the PDJ.

7. ACTIVITIES OF THE PEDSC DURING 2017–2019

1. Construction of the integrated database

The integrated database to be constructed should have the following functions: (1) to accept some different metadata formats (e.g., ISO19139, SPASE, etc.) used in the polar science; (2) to read some domain standard format files (CDF, NetCDF, etc.) and convert

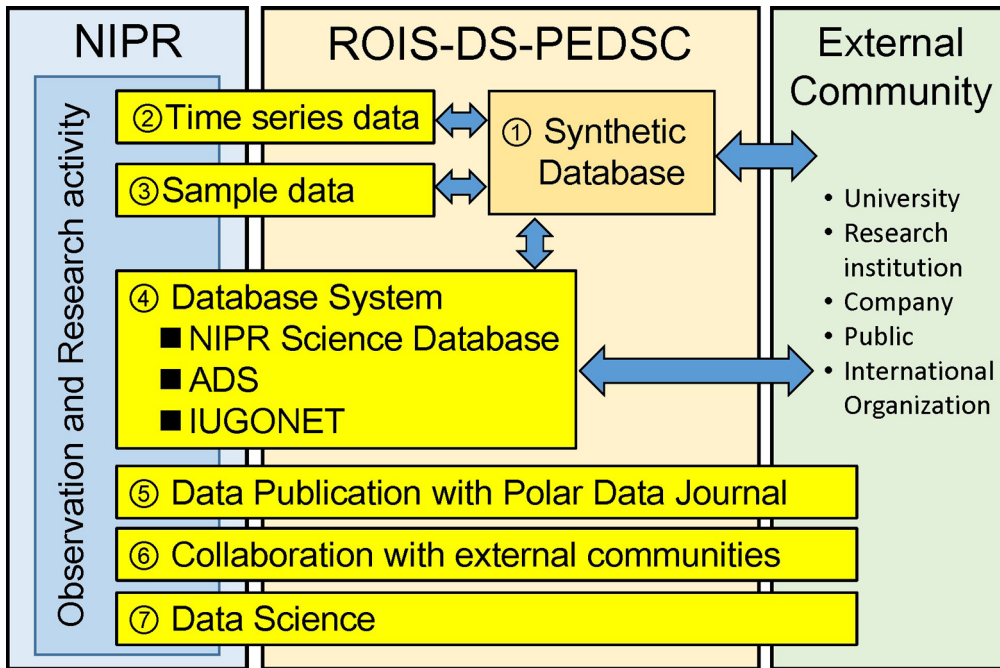


Figure 2 Diagram of relationship among the NIPR, PEDSC, and external community associated with the seven tasks of the PEDSC.

	Item	H29 (2017)	H30 (2018)	H31 (2019)	H32 (2020)	H33 (2021)
1	Synthetic Database		Development	Operation		
2	Database system upgrade and interoperate					
2.1	ADS	Continuous operation				
	Upgrade to AADS (Arctic & Antarctic Data archive System)		Development	Operation		
2.2	IUGONET	Continuous operation				
	System upgrade	Occasionally				
	Apply to other research fields	Occasionally				
2.3	NIPR Science Database	Continuous operation				
	Upgrade to Unified Database		Development	Operation		
	Interoperate with AADS		Development	Operation		
	Interoperate with IUGONET		Development	Operation		
3	Database creation of time series data in each research field					
	PANSY	Continuous				
	EISCAT	Continuous				
	Space and Upper Atmospheric Sciences data	Continuous				
	Geoscience data	Continuous				
	Meteorology and Glaciology data		Development	Operation		
4	Database creation of sample data in each research field					
	Rock sample		Development	Operation		
	Biological sample			Development	Operation	
	Meteorite				Development	Operation
	Snow and Ice				Development	Operation
5	Data Publication with the Polar Data Journal	Continuous				
6	Collaboration with external communities	Continuous				
	International meeting					
7	Data science and research activities with external scientists	Continuous				

Figure 3 Annual schedule of the PEDSC activity during the five years associated with the seven tasks planned at the start of the PEDSC in 2017.

them to the ASCII files; (3) to search and display both the metadata and actual data in the easiest way; (4) to analyze a correlation and relationship among actual data in various research fields for multidisciplinary or cross-disciplinary studies.

System design of the integrated database has been done since 2018, and a system construction had been finished in 2019. System test has been carried out in 2019 by registering some metadata in various research fields, such as upper atmosphere data, meteorites, bioscience samples, seismological data, infrasound data, etc.

2. Existing database systems (NIPR-SDB, ADS, IUGONET)

Interoperability among the metadata registered in the NIPR-SDB, ADS, and IUGONET has been considered for the polar data. The NIPR-SDB has started to assign DOI to the already registered data after some approval process since 2018 (https://scidbase.nipr.ac.jp/modules/site/index.php?content_id=19&ml_lang=en). The data archive system of the ADS has treated not only Arctic data but also Antarctic data more explicitly than before, attempting to become a larger database for both polar regions in future (<https://ads.nipr.ac.jp/antarctic>). As a collaboration with the JARE operation, the PEDSC has become

a responsible organization to handle the data obtained in the NIPR associated JARE programs since 2018. A data management plan for the JARE data has been constructed, and the PEDSC started to collect the metadata and actual data obtained by the JARE-60 summer party in 2018-2019 season by using the ADS system (<https://ads.nipr.ac.jp/antarctic/sheet-download>).

3. Processing of time-series digital data

For seismological data, a real time monitor system had been constructed in 2017. For auroral data, a Web-based system to look at the data obtained at multiple ground-based observatories in both polar regions has been constructed (<http://133.57.20.115/www/AQVN/>). For the data of the PANSY radar (Program of the Antarctic Syowa MST/IS Radar) at Syowa Station (<http://pansy.eps.s.u-tokyo.ac.jp/en/>) (cf. Sato, et al., 2014), new data processing algorithms have been developed to extract and analyze new physical values from original radar signals. In the collaboration program of the ROIS-DS, a database system and a Web site for the cosmic-ray data obtained at Syowa Station has been constructed since 2018 (<http://polaris.nipr.ac.jp/~cosmicrays/en.html>). For the satellite data received at Syowa Station, a data publishing platform has been constructed in the ADS system (<https://ads.nipr.ac.jp/satelliteGallery/#/>). For the satellite imagery data around Syowa Station and the live camera data at Syowa Station, data viewing Web sites have been constructed also in the ADS system to support the JARE activity (https://ads.nipr.ac.jp/shirase_monitor/gallery?image and <https://ads.nipr.ac.jp/vishop/#/monitor/type=SYOWA>, respectively).

4. Processing of sample data

For rock sample data, a database system ('NIPR Rock Repository' (<https://ads.nipr.ac.jp/nrr/>)) has been constructed in 2018. For Ice core sample data, metadata of the already published data has been created and registered in the ADS system since 2018. For meteorites and biological sample data, their metadata and CDF (Common Data Format) data have been created and registered in the integrated database in 2019. In the collaboration program of the ROIS-DS, a database system for the historical archive data of the NIPR ('NIPR Digital Archive' (<https://ads.nipr.ac.jp/image/>)) has been constructed in 2019.

5. Data publication with the Polar Data Journal

In total, twelve papers were submitted and eight ones were published so far from the launching of the PDJ in January, 2017. Number of the DOI assignment to the actual data by the ADS system is 12, so far.

6. Collaboration with external communities

The number of accepted projects in the collaboration program of the ROIS-DS (ROIS-DS-JOINT) (https://ds.rois.ac.jp/en_crp/en_calling/) related with the PEDSC in 2017, 2018 and 2019 are 3, 8 and 9, respectively. Their titles are listed in [Table 1](#). As for the international collaboration, two international workshops were held by the PEDSC in 2017 and 2018 as shown in [Table 2](#).

NO.	TITLE
2017-01	Building "digital archive" system for utilizing historical pictures about Antarctic and Arctic regions
2017-02	Data publishing framework for multi-station airglow imaging with IUGONET system
2017-03	Development of Antarctic rock samples repository
2018-01	Database of optical flow using data from Longyearbyen and Iceland
2018-02	Development of Antarctic rock samples repository (cont.)
2018-03	Development of database and data analysis system of ground-based atmospheric electric field observation data for global electrical circuit study
2018-04	Development of a real time database system for Cosmic Ray data observed at Syowa Station
2018-05	Construction of archiving/information disclosure system for infrasound observation in Antarctica
2018-06	Development of integrated database using various kinds of optical data obtained in South pole and McMurdo stations
2018-07	Redundant data server system for the service of the real-time information of geomagnetic field data and indices

Table 1 Titles of the projects accepted in the collaboration program of the ROIS-DS related to the PEDSC.

NO.	TITLE
2018-08	“Science and Data” -- In search of future directions of Open Science and Data-driven Science (meeting)
2019-01	Development of a real time database system for Cosmic Ray data observed at Syowa Station (cont.)
2019-02	Sharing and Visualization of 1-100Hz ELF Waveform Data Obtained at Syowa Station
2019-03	Method for detection of the snowy cloud and estimation of snow falling amount based on NOAA/AVHRR images
2019-04	Development of Antarctic rock samples repository (cont.)
2019-05	Construction of archiving/information disclosure system for infrasound observation in Antarctica (cont.)
2019-06	Application of machine learning to hypocenter determination for Seismic observation network in Showa base
2019-07	Development of a signal processing method for improving quality of wind and turbulence products obtained by atmospheric radars
2019-08	Research for participatory consortium for accelerating data-driven science
2019-09	Problems on mutual use and integrated analysis of solid earth science data (meeting)

TITLE	DATE	VENUE	ATTEND	OVERSEA	PRESENTATIONS
International Workshop on Sharing, Citation and Publication of Scientific Data across Disciplines http://polaris.nipr.ac.jp/~pseis/data.ws-2017/main.dwt	5-7 Dec. 2017	Tachikawa, Tokyo	73	8	43
International Workshop on Data Science – Present & Future of Open Data & Open Science – https://ds.rois.ac.jp/article/dsws_2018/	12-15 Nov. 2018	Mishima, Shizuoka	120	10	58

7. Data science using the database and database system

The data stored in the ADS have been widely used in the Arctic research community, e.g., to analyze the temporal variation of the sea ice concentration in the Arctic ocean, and to find an appropriate shipping route in the Arctic Ocean. The IUGONET system has been used for the data analysis of the upper atmospheric science data in the polar region. One of the staffs of the PEDSC is a specialist in radar signal processing, who is in charge of the maintenance and improvement of the PANSY radar system, and has developed a cutting-edge technique to derive new scientific results from the PANSY radar data.

Table 2 International workshop held by the PEDSC in Japan.

COMPETING INTERESTS

The authors have no competing interests to declare.

AUTHOR CONTRIBUTIONS

Akira Kadokura, Masaki Kanao, Hironori Yabuki, Yoshimasa Tanaka, and Koji Nishimura are all members of the Polar Environment Data Science Center (PEDSC), and are involved with all of the activities of the PEDSC described in this manuscript. A. K. wrote this manuscript. M. K. is primarily responsible for the description on the NIPR Science Database; H. Y. is primarily responsible for the description on the ADS system; Y. T. is primarily responsible for the description on the IUGONET system and integrated database; and K. N. is primarily responsible for the description of the PANSY radar data.

AUTHOR AFFILIATIONS

Akira Kadokura  orcid.org/0000-0002-6105-9562

Polar Environment Data Science Center, Joint Support-Center for Data Science Research, Research Organization of Information and Systems, Tokyo, JP

Masaki Kanao  orcid.org/0000-0001-6330-4691

Polar Environment Data Science Center, Joint Support-Center for Data Science Research, Research Organization of Information and Systems, Tokyo, JP

Hironori Yabuki  orcid.org/0000-0001-6597-1615

Polar Environment Data Science Center, Joint Support-Center for Data Science Research, Research Organization of Information and Systems, Tokyo, JP;
National Institute of Polar Research, Research Organization of Information and Systems, Tokyo, JP

Yoshimasa Tanaka  orcid.org/0000-0001-7276-4868

Polar Environment Data Science Center, Joint Support-Center for Data Science Research, Research Organization of Information and Systems, Tokyo, JP

Koji Nishimura  orcid.org/0000-0001-8824-9844

Polar Environment Data Science Center, Joint Support-Center for Data Science Research, Research Organization of Information and Systems, Tokyo, JP;
Research Institute for Sustainable Humanosphere, Kyoto University, Kyoto, JP

Kadokura et al.
Data Science Journal
DOI: <https://doi.org/10.5334/dsj-2022-011>

REFERENCES

- Angelopoulos, V, Cruce, P, Drozdov, A**, et al. 2019. The Space Physics Environment Data Analysis System (SPEDAS). *Space Sci Rev*, 215: 9. DOI: <https://doi.org/10.1007/s11214-018-0576-4>
- Hayashi, H, Koyama, Y, Hori, T, Tanaka, Y, Abe, S, Shinbori, A, Kagitani, M, Kouno, T, Yoshida, D, UeNo, S, Kaneda, N, Yoneda, M, Umemura, N, Tadokoro, H** and **Motoba, T**. 2013. IUGONET project team, Inter-university Upper Atmosphere Global Observation NETWORK (IUGONET). *Data Sci. J.*, 12: WDS179–WDS184. DOI: <https://doi.org/10.2481/dsj.WDS-030>
- Kanao, M, Kadokura, A, Okada, M, Yamnouchi, T, Shiraishi, K, Sato, N** and **Parsons, MA**. 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-021>
- 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. *CODATA Data Science Journal*, 17(1): 1–6. DOI: <https://doi.org/10.5334/dsj-2018-001>
- 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>
- Sato, K, Tsutsumi, M, Sato, T, Nakamura, T, Saito, A, Tomikawa, Y, Nishimura, K, Kohma, M, Yamagishi, H** and **Yamanouchi, T**. 2014. Program of the Antarctic Syowa MST/IS Radar (PANSY). *J. Atmos. Solar-Terr. Phys.*, 118A: 2–15. DOI: <https://doi.org/10.1016/j.jastp.2013.08.022>
- Tanaka, Y, Shinbori, A, Hori, T, Koyama, Y, Abe, S, Umemura, N, Sato, Y, Yagi, M, Ueno, S, Yatagai, A, Ogawa, Y** and **Miyoshi, Y**. 2013. Analysis software for upper atmospheric data developed by the IUGONET project and its application to polar science. *Adv. Polar Sci.*, 24: 231–240. DOI: <https://doi.org/10.3724/SP.J.1085.2013.00231>
- Yatagai, A, Tanaka, Y, Abe, S, Shinbori, A, Yagi, M, Ueno, S, Koyama, Y, Umemura, N, Nosé, M, Hori, T, Sato, Y, Hashiguchi, NO** and **Kaneda, N**. 2014. IUGONET project team, Interuniversity upper atmosphere global observation network (IUGONET) metadata database and analysis software. *Data Sci. J.*, 13. DOI: <https://doi.org/10.2481/dsj.IFPDA-07>

TO CITE THIS ARTICLE:

Kadokura, A, Kanao, M, Yabuki, H, Tanaka, Y and Nishimura, K. 2022. Activities of the Polar Environment Data Science Center of ROIS-DS, Japan. *Data Science Journal*, 21: 12, pp. 1–8. DOI: <https://doi.org/10.5334/dsj-2022-012>

Submitted: 03 January 2020

Accepted: 11 April 2022

Published: 20 April 2022

COPYRIGHT:

© 2022 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.