International Workshop on Sharing, Citation and Publication of Scientific Data across Disciplines

PROGRAMME and ABSTRACTS

Joint Support-Center for Data Science Research (DS),
Tachikawa, Tokyo, Japan

5–7 December 2017

Table of Contents :

International Workshop on Sharing, Citation and Publication of Scientific Data across Disciplines

Joint Support-Center for Data Science Research (DS), Tokyo

5–7 December 2017

Scope of Workshop :

The Workshop will focus on recent topics of interest in the field of scientific data, which are attributed to play a crucial role in global trend on accelerating "Open Science" and "Open Data". Contributions from all scientific disciplines are welcome, including life and bio science, social and human science, as well as polar science. Inter-disciplinary orientated topics on data management are especially encouraged.

Presentations will be given that span the entire range of topics of effective scientific data management by individual data center; from management planning and policy, to submission of metadata and actual data, to share the data so as to facilitate new inter-disciplinary science, to long-term preservation and stewardship with global and social perspectives.

Topics on data citation, data publication and data journal are strongly encouraged in terms of "Open Science" for the retrieved and archived data. Contributors will report on successes and challenges recently encountered, best practices learned, and what must yet be done to ensure the past data legacy. Fruitful discussions on data publication issues between multi-branch of science are expected to give a new horizon on data management and to achieve inter-disciplinary linkages.

By this Workshop, it is expected that mutual understanding on data activities in different fields of science could be obtained, and future activities are progressed, such as a promotion of new inter-disciplinary sciences and new research collaboration using multi-disciplinary data, as well as a contribution to global data activities based on facilities provided by the Joint Support-Center for Data Science Research of ROIS.

Session Themes :

- **Data Center - Best Practise & Activity:** Current Accreditation Schemes and their Benefits, Positives and Negatives of Current Approaches of each Data Center, Data Management Planning, Data Policy, etc.
- **Database, Data System - Network & Administration:** Database, Data System, Metadata, Vocabularies, Ontology, Cloud Computing and Storage, Repository Practises & Standards, etc.
- **Data Sharing & Inter-Operability:** Data Sharing, Virtual Observatories, Information and Communications Technology Infrastructure Protocols and Architectures, Sustainability and

Governance Models, Real-Time Data Handling, etc.

- **Data Citation & Publication, Data Journal:** Best Practise of Data Citation, Data Publication, Data Journal, Scientific Reward and Recognition Systems, etc.
- **Future of Data Sharing, Citation & Publication across Discipline:** Any Theme as long as the Data Sharing, Citation and Publication is strategically forward looking across Discipline

Schedule Summary :

Tuesday 5 December 2017 @ Lecture Room (4F, NIPR)
Wednesday 6 December 2017 @ Multi-Purpose Meeting Room (2F, NINJAL)
16:30–18:00 "Polar Data Journal" Editorial Board & Advisory Board meeting (closed)
@ Multi-Purpose Meeting Room (1F, DS)
Thursday 7 December 2017 @ Multi-Purpose Meeting Room (2F, NINJAL)

Organization Committee : (* Chair)

Susumu Goto (Database Center for Life Science, DS, ROIS)
Akira Kadokura (Polar Environment Data Science Center, DS, ROIS)
*Masaki Kanao (Polar Environment Data Science Center, DS, ROIS)
Asanobu Kitamoto (Center for Open Data in the Humanities, DS, ROIS)
Yasuhiro Murayama (National Institute of Information and Communications Technology)
Shinya Nakano (The Institute of Statistical Mathematics, ROIS)
Koji Nishimura (Polar Environment Data Science Center, DS, ROIS)
Hideaki Takeda (National Institute of Informatics, ROIS)
Yoshimasa Tanaka (Polar Environment Data Science Center, DS, ROIS)
Seiji Tsuboi (Japan Agency for Marine-Earth Science and Technology)
Hironori Yabuki (Polar Environment Data Science Center, DS, ROIS)
Ryozo Yoshino (Social Data Structuring Center, DS, ROIS)

Abbreviation :

DS : Joint Support-Center for Data Science Research
NIPR: National Institute of Polar Research
NINJAL: National Institute for Japanese Language and Linguistics
ROIS: Research Organization of Information and Systems

Website :

http://polaris.nipr.ac.jp/~pseis/data.ws-2017/main.dwt

Contact Address :

data.ws.oc-2017@nipr.ac.jp

Programme Summary :

Monday 4 December 2017 @ "Southern Cross"

17:00–18:30 Registration

18:30–20:30 Icebreaker (@ "Southern Cross" in front of Polar Science Museum)

Tuesday 5 December 2017 @ Lecture Room (4F, NIPR)

09:00–10:00 Registration

10:00–10:30 Opening Remarks

10:30–12:30 Session 1: Data Center - Best Practise & Activity

12:30–14:00 Group Photo & Lunch

14:00–15:50 Session 2: Database, Data System - Network & Administration

15:50–16:20 Coffee Break

16:20–18:10 Session 3: Data Sharing & Inter-Operability

18:30–20:30 Workshop Reception (@1F floor of DS building)

Wednesday 6 December 2017 @ Multi-Purpose Meeting Room (2F, NINJAL)

09:30–11:00 Session 4: Data Citation & Data Publication

11:00–11:30 Coffee Break

11:30–12:30 Session 5: Data Journal - Best Practise

12:30–14:00 Lunch & Poster Session

14:00–16:10 Session 6: Future of Data Sharing, Citation & Publication across Disciplines

16:10–17:00 Poster session

16:30–18:00 "Polar Data Journal" Editorial Board + Advisory Board Members' meeting (closed) @ Multi-Purpose Meeting Room (1F, DS)

18:30–20:30 Workshop Banquet (@Tachikawa Grand Hotel)

Thursday 7 December 2017 @ Multi-Purpose Meeting Room (2F, NINJAL)

09:30–11:00 Session 7a: Toward Inter-Disciplinary Data Sharing & Publication

11:00–11:30 Coffee Break

11:30–12:30 Session 7b: Toward Inter-Disciplinary Data Sharing & Publication

12:30–14:00 Lunch

afternoon Free Time (Participate to 8th NIPR Annual Symposium, etc.)

18:00–20:30 DS Town Hall Meeting (@ Restaurant nearby Palace Hotel Tachikawa)

For Participants :

Language :

The official conference language is English. There is no simultaneous interpretation service will be provided.

Name Badge :

The participant name badge will be provided at the registration desk of the workshop. All participants are required to wear the badge throughout the conference. Marks for each side events shall be pasted to the participant badge after payment.

Group Photo:

After the morning session on 5 December (1st day, Session 1), all participants are invited to take a "group photo" of the conference. The place is planned in front of the entrance of "Data Science (DS)" building.

WiFi service:

At the venue of NIPR/NINJAL, WiFi service is available. Detail information about SSID & Password will be given at the registration desk.

For Presenters :

Oral Presentation :

General presenters are allocated for 20 min. including questions and discussion time.
Keynote speakers are allocated for 30 min. including questions and discussion time.

A Window PC is available for presentations, but presenters can use their own PC as well.
Please bring presentation file (ppt, pdf) by USB when using the PC in conference room.

Poster Presentation :

Please prepare a poster with the maximum size of A0 x 2, 841 mm (width) x 1189 mm (height) x 2, although the posting board has a size of 1200 mm (width) x 1800 mm (height).

All posters can be posted during two days on 6-7 December 2017, in front of the Multi-Purpose Meeting Room (2F, NINJAL)

Core Times for Poster presentations are allocated for;
Wednesday 6 December 13:10–14:00, & 16:10-17:00

Registration :

Registration Desk :

Registration desk will open at the following date & time.
On-site registration can be acceptable for each day. There is no registration fee required.

- Monday 4 December 2017
17:00-20:00 at "Southern Cross", where in front of Polar Science Museum.
- Tuesday 5 December 2017
09:00-18:00 in front of Lecture Room @ NIPR 4F
- Wednesday 6 December 2017
09:00-17:00 in front of Multi-Purpose Room @ NINJAL 2F
- Thursday 7 December 2017
09:00-12:00 in front of Multi-Purpose Room @ NINJAL 2F

Pre-Conference Registration :

Those who are intending to attend the workshop prior to the conference date, please send the following information to the email address of Organizing Committee.
(data.ws.oc-2017@nipr.ac.jp).

Family Name; (Middle Name); First(Given) Name;
Affiliation 1 (Institution, University, Company, etc.);
Affiliation 2 (Department, Section, Position, etc.); Country; Email address

The Organizing Committee add to the participant list & prepare name badges before the conference.

Side Events :

- Monday 4 December 2017

13:00–18:00 ***Trustworthy Data Repositories*** @ Lecture Room (4F, NIPR)
- Forum for Sharing Practical Information about CoreTrustSeal Certification -
(only for Japanese participants)

Fifth CODH Seminar / Second DIAS Open Science Seminar / First Meeting of RDUF SIG "Construction of Network among Domain Repositories in Japan":

http://codh.rois.ac.jp/seminar/coretrustseal-20171204/

18:30–20:30 ***Workshop Icebreaker*** (@"Southern Cross" open space nearby DS building, just in front of "Polar Science Museum")

Registration desk will open from 17:00- at the "Southern Cross". Participants to the "Icebreaker" need to pay 2,000 JPY at the desk.

- Tuesday 5 December 2017

18:30–20:30 ***Workshop Reception*** (@1F open floor of DS building)

Participants to the "Reception" need to pay 2,000 JPY at the registration desk (NIPR 4F during day time; or the desk in DS 1F just before the party)

- Wednesday 6 December 2017

16:30–18:00 ***"Polar Data Journal" Editorial Board + Advisory Board Members' meeting***
(closed only) @ Multi-Purpose Meeting Room (1F, DS)

18:30–20:30 ***Workshop Banquet*** (@ Tachikawa Grand Hotel, joining with Banquet of the 8th NIPR symposium)
http://tachikawa.khgrp.co.jp/en/

Participants to the "Banquet" need to pay 3,000 JPY at the desk inside the "Grand Hotel".

* Shuttle buses will be served between NIPR and the Grand Hotel before the Banquet. Buses will leave the south exit of NIPR main building (polar science museum side) between 17:30 and 18:10 on Dec. 6th. Departure times are planning at 17:30, 17:40, and 18:00, 18:10.

- Thursday 7 December 2017

18:00–20:30 ***DS Town Hall Meeting*** (@ Japanese style Restaurant near Palace Hotel Tachikawa; "Namiki-An")

Participants to the "DS Town Hall Meeting" need to pay 3,000 JPY at the "Namiki-An" restaurant.

The Restaurant is famous for its classical Soba noodle menu, where is 200 m eastward from Place Hotel Tachikawa.

Accommodation :

Lunch Service:

Lunch boxes are for selling at the1st floor of NIPR building during the workshop dates. Participants are able to order their lunch boxes (500 JPY) at the reception desk of data workshop on 5 & 6 December.

Moreover, there are several restaurants for taking lunch near the NIPR/DS/NINJAL area. Please refer to the lunch map by another sheet which will be distributed at the registration desk. To take your lunch outside conference buildings, the map may satisfy your purpose.

Hotel Accommodation :

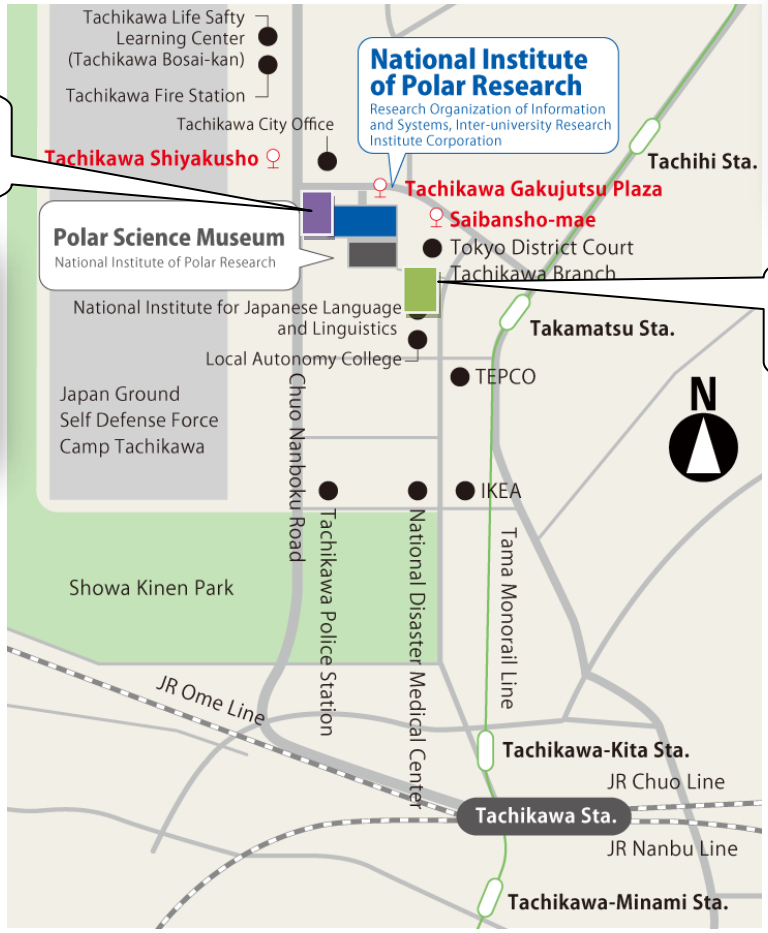
The Organization Committee keep several rooms for discount price to foreign participants in;
"The Palace Hotel Tachikawa"
http://www.palace-t.co.jp/english/

where is 5 min. northward walk from JR Tachikawa station, & 20 min. southward walk from DS, NIPR and NINJAL.

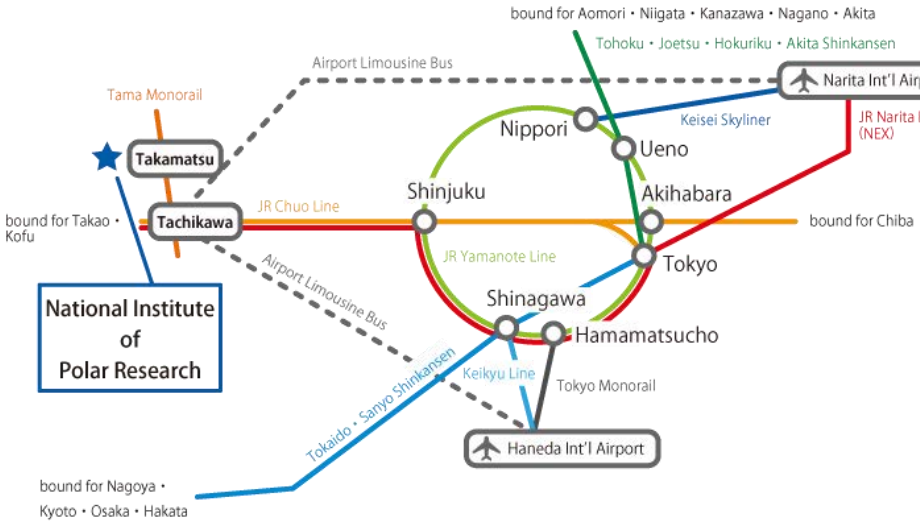
Access information to the Hotel is as follows; there is a direct connection from Narita/Haneda airports by using Limousine Bus Transportation.
http://www.palace-t.co.jp/english/access.html

Venue & Access :

Data Science Building
Joint Support-Center for Data Science Research



National Institute for Japanese Language and Linguistics



International Workshop on Sharing, Citation and Publication of Scientific Data across Disciplines

PROGRAMME

Joint Support-Center for Data Science Research (DS),
Tachikawa, Tokyo, Japan

5–7 December 2017

大学共同利用機関法人
情報・システム研究機構
Research Organization of Information and Systems

大学共同利用機関法人 情報・システム研究機構
データサイエンス共同利用基盤施設
Joint Support-Center for Data Science Research (DS)

International Workshop on Sharing, Citation and Publication of Scientific Data across Disciplines

Joint Support-Center for Data Science Research (DS), Tokyo

5–7 December 2017

PROGRAMME

Monday 4 December @ "Southern Cross"
17:00–18:30 **Registration**
18:30–20:30 **Icebreaker (@ "Southern Cross" nearby Data Science Building & Polar Science Museum)**

Tuesday 5 December @ Lecture Room (4F, NIPR)
09:00–10:00 **Registration**
10:00–10:30 **Opening Remarks** • **Ryoichi Fujii** (President, Research Organization of Information and Systems (ROIS)) • **Asao Fujiyama** (Director-General, Joint Support-Center for Data Science Research (DS), ROIS) • **Takuji Nakamura** (Director-General, National Institute of Polar Research, ROIS) • **Akira Kadokura** (Director, Polar Environment Data Science Center, DS, ROIS) • **Masaki Kanao** (LOC Chair), agenda of workshop & practical information
10:30–12:30 **Session 1: Data Center - Best Practise & Activity** (Session chair: Akira Kadokura) **Keynote 1**: *The Evolution of Data Publication and the Role of Persistent Identifiers and Linked Open Data in Dynamic Data Mobilization* **Peter Pulsifer** (National Snow and Ice Data Center, University of Colorado) (30') **Keynote 2**: *The Antarctic Master Directory, sharing Antarctic (meta)data from multiple disciplines* **Anton Van de Putte** (Royal Belgian Institute for Natural Science) (30') • *Magnetic Data at WDC Kyoto—Services under International Collaborations* **Toshihiko Iyemori** (Data Analysis Center for Geomagnetism, Kyoto University) (20') • *Management of Marine-Earth Science Data and Samples in JAMSTEC* **Seiji Tsuboi** (Japan Agency for Marine-Earth Science and Technology) (20') • *Sharing Real Business Purpose Datasets for Academic Research* **Keizo Oyama** (National Institute of Informatics, ROIS) (20')
12:30–14:00 **Group Photo @ Entrance of DS Building -> Lunch**
14:00–15:50 **Session 2: Database, Data System - Network & Administration** (Session chair: Seiji Tsuboi) **Keynote**: *FAIRsharing - Describing and Connecting Standards, Databases and Policies Across Disciplines* **Peter McQuilton** (Oxford e-Research Centre, University of Oxford) (30')

- *Life Science Database Integration Based on Semantic Web Technology*
  Susumu Goto (Database Center for Life Science, DS, ROIS) (20')
- *Outline of the Arctic Data Archive System (ADS)*
  Hironori Yabuki (Polar Environment Data Science Center, DS, ROIS) (20')
- *Development of Research Data Management Service for Open Science in Japan*
  Yusuke Komiyama (National Institute of Informatics, ROIS) (20')
- *Data Processing and Archive System for the Antarctic PANSY Rader*
  Koji Nishimura (Polar Environment Data Science Center, DS, ROIS) (20')

15:50–16:20 **Coffee Break**

16:20–18:10 **Session 3: Data Sharing & Inter-Operability** (Session chair: Susumu Goto)

Keynote: *SeaDataNet, a Network of Distributed Oceanographic Data Centres Now Going to the Cloud*
Serge Scory (Royal Belgian Institute for Natural Science) (30')

- *Inter-University Upper Atmosphere Global Observation Network (IUGONET) Metadata Database*
  Yoshimasa Tanaka (Polar Environment Data Science Center, DS, ROIS) (20')
- *Data Sharing at the National Research Institute for Earth Science and Disaster Resilience*
  Katsuhiko Shiomi (National Research Institute for Earth Science and Disaster Resilience) (20')
- *Development of Data Sharing and Archiving on International Relations*
  Kiyohisa Shibai (Social Data Structuring Center, DS, ROIS) (20')

Discussion (20')

- **Facilitator: Rorie Edmunds** (International Program Office, ICSU World Data System)

18:30–20:30 **Workshop Reception (@1F floor of DS building)**

- **Welcome Remark: Takeshi Watanabe** (International Program Office, ICSU World Data System)

Wednesday 6 December @ Multi-Purpose Meeting Room (2F, NINJAL)

09:30–11:00 **Session 4: Data Citation & Data Publication** (Session chair: Hironori Yabuki)

Keynote: *Data Publishing and Data Citation - Are We There Yet?*
Jens Klump (Mineral Resources, CSIRO in Australia) (30')

- *Data Citation at World Data Center for Geomagnetism, Kyoto*
  Masahito Nose (Data Analysis Center for Geomagnetism, Kyoto University) (20')
- *Data Citation Procedure on Science Database of Polar Research*
  Masaki Kanao (Polar Environment Data Science Center, DS, ROIS) (20')
- *The Polar Data Catalogue: Best Practices for Sharing and Archiving Canada's Polar Data*
  Julie Friddell (Polar Data Catalogue in Canada) (20')

11:00–11:30 **Coffee Break**

11:30–12:30 **Session 5: Data Journal - Best Practise** (Session chair: Shinya Nakano)

- *Data paper of JAMSTEC Report of Research and Development*
  Daisuke Suetsugu (Japan Agency for Marine-Earth Science and Technology) (20')
- *"Polar Data Journal" ; A new data publishing platform for polar science*
  Akira Kadokura (Polar Environment Data Science Center, DS, ROIS) (20')
- *Polar Science and Data Journal by "Elsevier Publisher"*
  Takeshi Yamanouchi (National Institute of Polar Research, ROIS) (20')

12:30–14:00 **Lunch (+ Poster session)**

14:00–16:10 **Session 6: Future of Data Sharing, Citation & Publication across Disciplines**

(Session chair: Mustapha Mokrane, Yoshimasa Tanaka)

Keynote: *The Canadian Consortium for Arctic Data Interoperability: An Emerging Initiative for Sharing Data Across Disciplines*
Shannon Vossepoel (University of Calgary) (30')

- *Emerging domain agnostic functionalities on the handle-centered networks*
  Kei Kurakawa (National Institute of Informatics, ROIS) (20')
- *Support Project for Data Fusion Computation: Current status and future prospects*
  Shinya Nakano (The Institute of Statistical Mathematics, ROIS) (20')
- *The Present Situation of Open Data Usage in the Social Sciences and Related Problems*
  Yusuke Inagaki (Social Data Structuring Center, DS, ROIS) (20')
- *Open Data in the Humanities: Data Sharing and Publication for Triadic Co-Creation*
  Asanobu Kitamoto (Center for Open Data in the Humanities, DS, ROIS) (20')

Discussion (20')

- **Facilitator: Yasuhiro Murayama** (National Institute of Information and Communications Technology)

16:10–17:00 **Poster session**

16:30–18:00 **"Polar Data Journal" Editorial Board + Advisory Board Members' meeting (closed only) @ Multi-Purpose Meeting Room (1F, DS)**

18:30–20:30 **Workshop Banquet (@Tachikawa Grand Hotel)**

Thursday 7 December @ Multi-Purpose Meeting Room (2F, NINJAL)

09:30–11:00 **Session 7a: Toward Inter-Disciplinary Data Sharing & Publication** (Session chair: Asanobu Kitamoto)

Keynote: *Strengthening International Data Sharing Networks*

Mustapha Mokrane (International Program Office, ICSU World Data System) (30')

- *Vocabulary Broker Application Connecting Data, Information and Literature Across Various Scientific Domains*
  Bernd Ritschel (Kyoto University) (20')
- *Recent Trends of Open Publication and Policy Development for Open Science toward Inter-Disciplinary Data Sharing & Publication*
  Kazuhiro Hayashi (National Institute of Science and Technology Policy) (20')
- *Data and Metadata Management at DIAS: Toward More Open Earth Environmental Information Platform*
  Toshiyuki Shimizu (Kyoto University) (20')

11:00–11:30 **Coffee Break**

11:30–12:30 **Session 7b: Toward Inter-Disciplinary Data Sharing & Publication** (Session chair: Masaki Kanao)

- *Building Trust in Scientific Data: Certification & the CoreTrustSeal*
  Rorie Edmunds (International Program Office, ICSU World Data System) (20')
- *Interdisciplinary Online Data Sharing Service on ADS*
  Takeshi Sugimura (International Arctic Environment Research Center, NIPR, ROIS) (20')
- *Toward An Open Science Ecosystem Including Sharing, Citing and Publishing Research Data*
  Yasuhiro Murayama (National Institute of Information and Communications Technology) (20')

Closing of Workshop

12:30–14:00 **Lunch**

afternoon **Free Time** (Participate to 8th NIPR Annual Symposium, etc.)

18:00–20:30 **DS Town Hall Meeting (@ Restaurant nearby Palace Hotel Tachikawa)**

Poster Presentation @ outside of Multi-Purpose Meeting Room (2F, NINJAL)

Core Time; Wednesday 6 December 13:10–14:00, 16:10-17:00

- *P-1: "Standing Committee on Antarctic Data Management (SCADM)"*
  Anton Van de Putte (Royal Belgian Institute for Natural Science)
- *P-2: " Antarctic Biodiversity Portal"*
  Anton Van de Putte (Royal Belgian Institute for Natural Science)
- *P-3: History of polar data management in Japan; before and after the IPY2007-2008*
  Masaki Kanao (Polar Environment Data Science Center, DS, ROIS)
- *P-4: Antarctic rock samples database: current status and future perspectives*
  Tomokazu Hokada (Polar Science Resources Center, NIPR, ROIS)
- *P-5: Activities of Polar Environment Data Science Center*
  ***A*kira Kadokura (**Polar Environment Data Science Center, DS, ROIS)
- *P-6: "Polar Data Journal" ; A new data publishing platform for polar science*
  Yasuyuki Minamiyama (Library, National Institute of Polar Research, ROIS)

- *P-7: Possibility and prevention of data tampering in the referee process of data journal*
  Takeshi Terui (International Arctic Environment Research Center, NIPR, ROIS)
- *P-8: Inter-University Upper Atmosphere Global Observation Network (IUGONET) Metadata Database*
  Yoshimasa Tanaka (Polar Environment Data Science Center, DS, ROIS)
- *P-9: Web service for reproducible multidisciplinary data visualization*
  Koji Imai (National Institute of Information and Communications Technology)
- *P-10: Current activities towards data sharing at Center for Global Environmental Research, National Institute for Environmental Studies (CGER/NIES)*
  Yoko Fukuda (National Institute for Environmental Studies)

Friday 8 December @ Seminar Room (3F, NIPR)
10:00–12:00 **Inter-Disciplinary session on "Polar Data Science" of the 8th NIPR Annual Symposium**
13:00–16:00 **Inter-Disciplinary session on "Polar Data Science" of the 8th NIPR Annual Symposium**

Abbreviation & Location-Access URL;

DS : Joint Support-Center for Data Science Research

NIPR: National Institute of Polar Research

NINJAL: National Institute for Japanese Language and Linguistics

International Workshop on Sharing, Citation and Publication of Scientific Data across Disciplines

ABSTRACTS

Joint Support-Center for Data Science Research (DS),
Tachikawa, Tokyo, Japan

5–7 December 2017

Building Trust in Scientific Data: Certification & the CoreTrustSeal

***Rorie Edmunds*[1]*, *Mustapha Mokrane*[1], *Ingrid Dillo*[2]**

[1]* *ICSU World Data System International Programme Office, Tokyo 184-8795, Japan*
[2] *Data Archiving and Networked Services, 2593 HW Den Haag, The Netherlands*
Email: rorie.edmunds@icsu-wds.org

Summary. Data created and used by scientists should be managed, curated, and archived in trustworthy data repositories to enable their reuse and ensure the integrity of science. Trustworthiness and sustainability of a data repository raise important organizational, technical, financial, and legal challenges, and depend on the quality and transparency of their data management processes, the use of established standards, their efforts for long-term preservation, and how suitable their services are for their designated community. Repository certification—such as the CoreTrustSeal Certification—gives an independent and objective evaluation of a repository's reliability and durability, and helps researchers, funders, librarians, and publishers ascertain which repositories to use. This talk will focus on the core certification of data repositories by the CoreTrustSeal and of networks by the ICSU World Data System. It will also touch on the future of core certification of other entities within the scientific research process.

Keywords. Trustworthiness, Certification, Data Repository, Scientific Data.

1. Introduction

Data repositories are increasing valued as a key element of global research infrastructure, playing a central role in the long-term preservation of research data that continue to escalate in volume and diversity.

Scientific integrity and norms dictate that data created and used by researchers should be managed, curated, and archived in a data repository to ensure that science is verifiable and reproducible while preserving initial investment in data collection. However, to guarantee that generated datasets remain available, useful, and meaningful into the future, research stakeholders—scientists, funders, librarians, and publishers—must be able to establish the trustworthiness of a data repository.

The need for trustworthiness and (in particular) trustworthy data repositories is therefore recognized as a prerequisite for efficient scientific research and data sharing.

2. Certification of Data Repositories

Data repository certification is the process whereby data repositories supply evidence to, and are assessed by, an independent authority for their trustworthiness and sustainability against defined criteria through a transparent and objective procedure. Certification helps data communities—producers, repositories, and consumers—to improve the quality and clarity of their practices, and to increase awareness of, and compliance with, established standards.

Nowadays, certification standards for data repositories are available at three different levels:

- Core – CoreTrustSeal Certification [1], which replaces the Data Seal of Approval (DSA) [2] and the ICSU World Data System (ICSU-WDS) Certification of Regular Members [3].
- Extended – The nestor-Seal/DIN 31644 [4].
- Formal – ISO16363 Audit and Certification of Trustworthy Digital Repositories [5].

Even at the core level—in which a repository first conducts a self-assessment that is then reviewed by community peers—certification offers many benefits to a repository and its stakeholders. In fact, completing a self-assessment is very useful

to a repository whether or not it wishes to apply for core certification, since it enables the repository to appraise its internal procedures with respect to the criteria and to update them where necessary. The current status of the repository is thus made apparent, in addition to serving for prospective certification.

3. Certification at the Core Level: The CoreTrustSeal

DSA and ICSU-WDS historically offered separate core certification standards. Drawing from their respective criteria, and within the framework of the Research Data Alliance, the two communities have now created and adopted a new set of harmonized *Core Trustworthy Data Repositories (TDR) Requirements* [6] for certification of data repositories at the core level.

The Core TDR Requirements consist of 16 universal guidelines intended to reflect the characteristics of a trustworthy data repository. An ad hoc Standards and Certification Board of DSA and WDS representatives has initially been responsible for implementation of the new standard, in which certification is granted by the Board after peer-review of self-assessments based on the Core TDR Requirements. It was recently announced that the standard will be further developed under the branding *CoreTrustSeal*.

CoreTrustSeal Certification is a minimally intensive process that accounts for the specific aims and context of a repository. By submitting its self-assessment for review, a repository's procedures and documentation are evaluated by external professionals who give the repository independent insights as to how it may mature and further increase its trustworthiness. This core certification moreover offers a solid foundation if the repository hopes to attain higher-level certification in the future.

4. WDS Certification of Networks

ICSU-WDS is striving to build worldwide 'communities of excellence' for scientific data services by certifying holders and providers of data or data products from wide-ranging scientific domains. Individual data centres and data analysis services must now become CoreTrustSeal Certified Repositories before they may be (re)accredited as WDS Regular Members.

Networks (umbrella bodies) are outside the scope of the CoreTrustSeal at present, and so the accreditation of WDS Network Members [7] for now remains a WDS-only focus. Since networks vary greatly in their makeup and remits, the WDS Scientific Committee has developed criteria and a procedure to ensure the trustworthiness of WDS Network Members; in particular, how they take responsibility for the competence and ongoing performance of their component nodes.

5. Conclusions

Whilst the verifiable level of trust in the scientific process and the feasibility of reproduction are greatly enhanced by having reliable access to supporting datasets through certified data repositories, also needed are the used protocols and methods, standards, digital/physical samples, software, vocabularies and ontology services, and so on. Hence, there are many other elements of research activity for which some form of trusted service or infrastructure component are required. The CoreTrustSeal is already exploring the development and provision of core-level certifications to ensure the trustworthiness of such services and components.

References

1. CoreTrustSeal, www.coretrustseal.org/ [accessed on: October 2017]
2. DSA, www.datasealofapproval.org/ [accessed on: October 2017]
3. ICSU-WDS, www.icsu-wds.org/ [accessed on: October 2017]
4. nestor-Seal, goo.gl/NvFDsD [accessed on: October 2017]
5. ISO16363, goo.gl/78yatU [accessed on: October 2017]
6. Core TDR Requirements, doi.org/10.5281/zenodo.168411 [accessed on: October 2017]
7. Accreditation of WDS Network Members, goo.gl/1kCZDT [accessed on: October 2017]

The Polar Data Catalogue: Best Practices for Sharing and Archiving Canada's Polar Data

Julie E. Friddell[1*], Gabrielle Alix[1], Yunwei Dong[1], David Friddell[1], Frank Lauritzen[1], and Ellsworth LeDrew[1]

[1*] *Canadian Cryospheric Information Network/Polar Data Catalogue, University of Waterloo, 200 University Avenue West, Waterloo, ON, N2V 3G1, Canada*
Email: julie.friddell@uwaterloo.ca

Summary. Since its online launch in 2007, the Polar Data Catalogue has become Canada's primary online source for Arctic and Antarctic research data and information. Through stable partnerships, the PDC has developed infrastructure and policy for reliable security, discoverability, and access to Canada's polar data and metadata. Recent activities include increased user and community engagement and collaboration, online release of a new bilingual data entry application, registration of DOIs for hundreds of PDC datasets, and conversion of the PDC database to an open format. The PDC is a member of the World Data System and is Canada's National Antarctic Data Centre.

Keywords. Data management, Data access, User engagement, Best practices.

1. Introduction

Scientific research in the polar regions has increased tremendously over the past few decades, with programs such as the ArcticNet Network of Centres of Excellence, Canada's International Polar Year (IPY) programme, the Canadian High Arctic Research Station (CHARS) of Polar Knowledge Canada (POLAR), other government and academic programs, and organizations and communities in northern Canada generating massive amounts of new scientific data and information on Arctic Canada and Antarctica. With these data comes the need to build systems to manage the burgeoning new data resources, to ensure their proper preservation, stewardship, and access. An appropriate data management system must not only respect confidentiality requirements and researchers' rights to publication but also, even more fundamentally, be able to accommodate the vast amounts of data and the huge diversity of topics and fields represented.

2. The Polar Data Catalogue

To help address the data management challenge, ArcticNet, the Canadian IPY programme, Canada's Department of Fisheries and Oceans, Noetix Research Inc., and the Canadian Cryospheric Information Network (CCIN) at the University of Waterloo joined together to develop the Polar Data Catalogue (PDC; https://www.polardata.ca). With additional collaborators Environment and Climate Change Canada; the Beaufort Regional Environmental Assessment, the Northern Contaminants Program, and the Nunavut General Monitoring Plan of Indigenous and Northern Affairs Canada; the international Circumpolar Biodiversity Monitoring Program (CBMP); and numerous others, the PDC has been developed into one of the largest repositories of polar data in Canada. Current holdings include over 2,500 descriptive metadata records of datasets and other polar data resources, almost 2.9 million datafiles, and more than 28,000 satellite images of northern Canada and Antarctic, all of which are available for free download.

3. User Engagement & Collaboration

To enhance management and accessibility of the data in the PDC archive, CCIN staff have actively sought input from partners and users whom we seek to serve, namely researchers and students, northern Canadian community members, and policy and decision makers at all levels of

government and in northern organizations. Recent activities include review of the PDC Lite low-bandwidth data search application (https://www.polardata.ca/pdclite/) by Inuit partners which resulted in a number of interface improvements to better meet user needs. We have also conducted a targeted survey of user requirements for snow and ice data in Canada. In addition to receiving large numbers of specific requests for particular types of snow and ice data and products, the survey demonstrated that people seek raw data as well as interactive visualizations, graphs, and map products which help them to understand the data more quickly and easily.

A significant effort over the last few years has involved hosting three meetings to advance management of polar data within Canada and at the international level. In 2015 and in 2017, CCIN hosted two Canadian Polar Data Workshops (https://secondcanadianpolardataworkshop.wordpress.com) to facilitate collaboration within the growing polar data community in Canada, and, in 2015, co-hosted the Polar Data Forum II (http://www.polar-data-forum.org) to bring the international polar data community together for making progress on priority themes and challenges.

A third engagement activity with three of our partners has resulted in publication of *Data Management Principles and Guidelines for Polar Research and Monitoring in Canada* (https://www.canada.ca/en/polar-knowledge/publications/data-management-principles-and-guidelines-2017-may.html). This document is intended to guide and unify data management requirements and expectations for polar researchers in Canada, streamlining and simplifying work for scientists and for data managers.

4. Website and App Development

To better serve our contributing researchers, some of whom have requested improvements and updates to the functions and usability of the PDC metadata and data Input application, we undertook a major project in 2016 to completely redesign and rebuild our data ingest app. The new application (https://www.polardata.ca/pdcinput/) has a completely modern and improved user interface, uses updated web technologies, is mobile-enabled, and is bilingual in English and French to accommodate the two official languages of Canada. Our next large projects are to redevelop and modernize the CCIN website (https://ccin.ca/) and the PDC Search application, both of which are currently in advanced prototype stages and are expected to be released in 2018.

5. DOIs to Improve Access to Data

Digital Object Identifiers (DOIs) are standard online identifiers that provide long-term links to data, improving the discoverability, accessibility, and citability of the data to which they are assigned. DOIs are viewed by researchers as an incentive to provide their data and thus make depositing their data attractive for researchers. In partnership with the National Research Council, Canada's member of the international DataCite initiative, CCIN has developed a process to assign DOIs and has registered DOIs to over 300 PDC datasets. We can provide DOIs early in the data management process so that the DOI can be recorded in a researcher's journal publications which report on the archived dataset.

6. Conclusions

The collaborative and technical activities over the last couple years have helped CCIN and the PDC grow into a well-respected repository of Arctic and Antarctic data. In recognition of the professionalism of our operation, the PDC has become the 100th member of the World Data System and has been selected as Canada's National Antarctic Data Centre. To continue the progress, we plan to strengthen the dialogue initiated via the Canadian and international Polar Data Workshops and Forum and receive and address additional user needs for improving our operation, so that we can optimally serve our users and contributors in the best possible ways.

Acknowledgments. We would like to thank our partners and supporters who have funded development of the advances reported here as well as numerous additional improvements over the last several years. We also thank the hundreds of researchers who willingly provide their time and data for archiving and sharing.

Current activities towards data sharing at Center for Global Environmental Research, National Institute for Environmental Studies (CGER/NIES)

Yoko Fukuda[1*], ***Tomoko Shirai***[1], ***Nobuko Saigusa***[1]

[1*] *Center for Global Environmental Research, National Institute for Environmental Studies, 16-2 Onogawa, Tsukuba 305-8506, Japan*
Email: fukuda.yoko@nies.go.jp

Summary. The National Institute for Environmental Studies (NIES) undertakes a broad range of environmental research in an interdisciplinary and comprehensive manner. To promote data sharing, NIES established a basic data policy in April 2017 and is planning implementation of an institutional repository for research data management. As one of the 9 research centers and branches of NIES, the Center for Global Environmental Research (CGER) is conducting various activities toward data sharing on global environmental research, such as a multifunctional database and an informational website for observation sites.

Keywords. Global Environment, Monitoring, Database, Data Sharing.

1. Introduction

Global environmental problems such as climate change are an important matter, which deeply concern a basis of the survival of human beings. For elucidation and future measures of the problem, international and multidisciplinary studies are essential. To investigate the global environmental system, CGER/NIES performs strategic global environmental monitoring. For long-term preservation and provision of quality-controlled observational data, sustainable data services are key components of scientific infrastructure. Here we introduce some activities towards data sharing conducted by CGER/NIES.

2. Global Environmental Database (GED)

GED (http://db.cger.nies.go.jp/portal/) is constructed to provide data and research results collected and compiled from natural and social sciences. The GED serves as a fundamental database related to global environmental problems with an emphasis on global warming and climate change.

The GED provides the following five menus.

- Database: Quality-controlled dataset is released using original format or NASA AMES format. DOI has been minted for the datasets since September 2016. The data can be easily displayed by Quick plot tool.
- Real time data: The latest greenhouse gases (GHGs) concentrations and their long-term trends over 20 years are hourly updated.
- Analysis support: Tools are provided for extraction of trend components, calculation of air trajectories and the residence time of particles released from a location.
- Associated Data: The data and results of CGER research projects in a wide range of fields are grouped into different research themes. Researchers conducted in collaboration with other centers at NIES are also introduced.
- Data Search: Data can be looked up in multi-discipline and multi-project combinations from the wide range of data provided by GED. Search can be made also for metadata or from the observation map.

3. Informational website for observation sites of GHGs in the Asia-Pacific region

To improve observational data coverage and estimate global or regional emissions of GHGs with high accuracy, CGER conducts global environmental monitoring by aircrafts, ships, terrestrial monitoring stations, and forest observation sites worldwide. For effective utilizations of existing observation datasets, CGER developed a website to provide information of observation sites for GHGs and related parameters conducted by NIES or other collaborators in the Asia-Pacific region. They are categorized into atmospheric/ecosystem monitoring or aircraft/ship monitoring. Users can easily search observation sites of their interest by selecting observation categories and parameters or by choosing geographical coordinates. Basic information of the selected observation sites, such as the category, the site name, the responsible institution, the location, and the observation parameter, is listed on a station table as shown in Figure 1. More information of each observation site is available from the link to the website of each observation group.

4. Conclusions

CGER/NIES is promoting data sharing by offering GED, the informational website, and other activities.

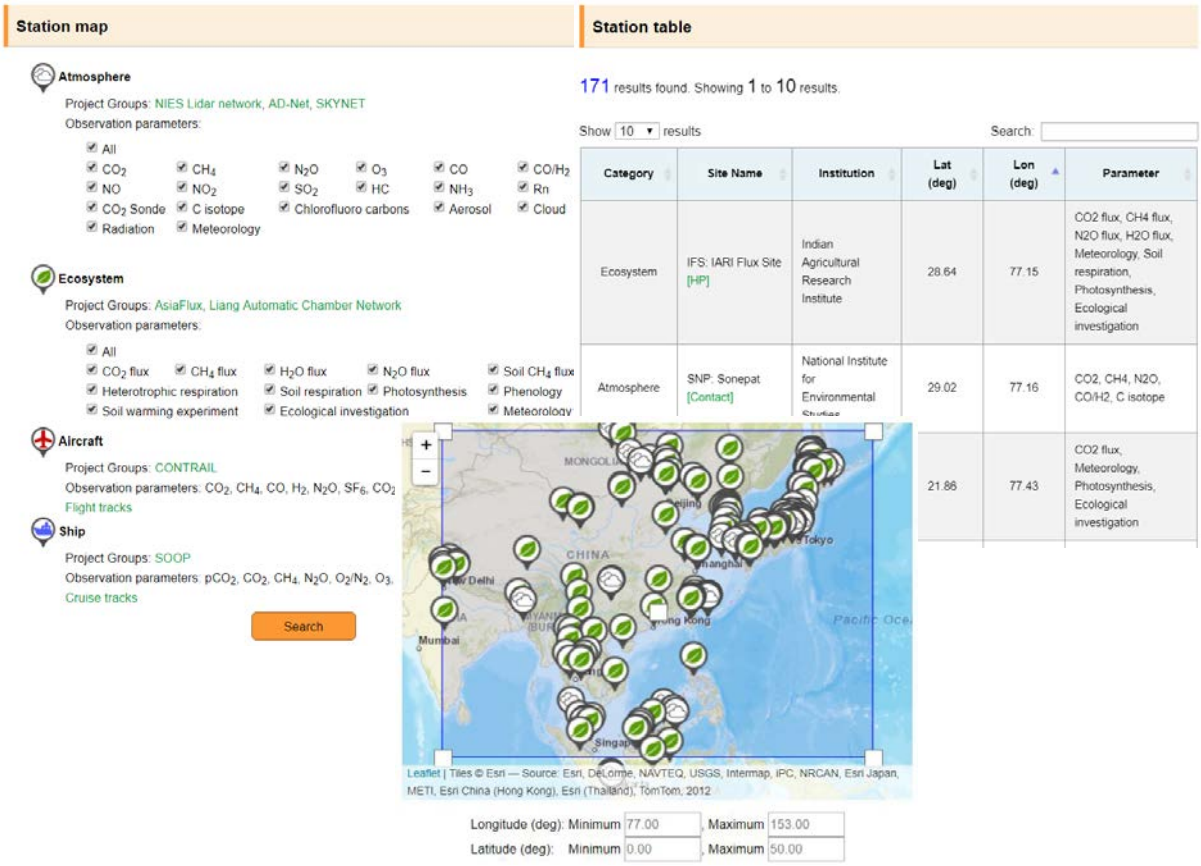


Figure 1. Station map and station table of informational website.

Life Science Database Integration Based on Semantic Web Technology

Susumu Goto[1*]

[1*] *Database Center for Life Science, Joint Support-Center for Data Science Research, Research Organization of Information and Systems, 178-4-4 Wakashiba, Kashiwa, Chiba, 277-0871, Japan*
Email: goto@dbcls.rois.ac.jp

Summary. More than 1,500 databases have been developed in the life science research field and most of them are available on the web. Searching and combining data across the databases for data sharing, reuse and analysis are still very difficult tasks for both data scientists and experimental biologists/clinicians, and even for bioinformaticians. Database Center for Life Science (DBCLS) has been developing tools and systems for the life science database integration using the semantic web technology, specifically, resource description framework (RDF). Here, we introduce activities of DBCLS regarding data sharing in the life science research field.

Keywords. Resource Description Framework, SPARQL, hackathon, natural language processing, multi-omics.

1. Introduction

More than 1,500 databases have been developed in the life science research field and most of them are available on the web [1]. They cover data from various research field including molecules, cells, tissues, individuals, phenotypes and environments. The computational representation for those data also varies depending on the research interests and the targets of data analyses. This situation has hindered efficient and effective sharing, inter-operability and reuse of life science data.

There have been efforts to solve this problem such as providing database catalogues and building portal sites. However, it is still difficult for data scientists, experimental biologists, clinicians and bioinformaticians to select a right database for their objectives and to combine several databases for their advanced data analyses. Even if they know the right databases, there are cases where different databases use different naming systems for the same data that further hiders integrated data analyses.

Database Center for Life Science (DBCLS) [2] has been established in 2007 to contribute to life science research by developing computational technologies for utilizing various life science databases. After the establishment of JST National Bioscience Database Center (NBDC) in 2011, DBCLS has conducted collaborative research with NBDC.

2. Database Center for Life Science

DBCLS focuses on the semantic web technology, specifically resource description framework (RDF), to establish bioinformatics environment that enables a unified way for handling data from different data sources. RDF represents all data in a simple form of triples, subject (S) – predicate (P) – object (O), for the relationship (semantics) between subject and object (data).

If all databases use well-defined common terminology, which we call ontology, for describing S, P and O, they can be easily integrated just by mixing the triples from each database that create a huge network of data, and there are several database management systems called triple stores for handling large network of data. However, the continuing efforts are necessary to create a real use case that uses RDF databases. DBCLS is trying to achieve this by the following research developments and activities.

- **Ontology and guideline development:** It is important to use common ontologies to

determine whether the objects or entities from different databases are same or not. Many ontologies have been developed so far and DBCLS has contributed to some of them, such as those for next generation sequencing data and microbial environmental data. To promote the use of common ontology, DBCLS provides a guideline for the database developers to indicate which ontology and identifiers should be used [3].

- **Tools and services development:** DBCLS develops tools and services to support various levels of users for accessing data without knowing detailed data structure, because the query language SPARQL for accessing triple stores are difficult for initial users. For example, LODQA provides a natural language query interface. DBCLS also supports database developers to convert their data into RDF formatted data, most of which are accessible via NBDC RDF portal site [4].
- **BioHackathon and related events:** DBCLS organizes both international and domestic biohackathons every year. About one week concentrated developments have accelerated database integration. The international BioHackathon contributed to the FAIR principle concept development [5]. DBCLS also organizes several related events, such as SPARQLthon, RDF summit and Biomedical Linked Annotation Hackathon, for promoting database integration and its application. Another important activity is lecture series, all of which are broadcasted via TogoTV. These events are quite useful to broaden the users in both development and application sides including those from developers of bioscience databases, data scientists and bioinformaticians to experimental biologists and clinicians, and to promote collaborative research projects.

3. Japan Alliance for Bioscience Information

There is no big one-stop centre in Japan for bioscience databases and bioinformatics research such as European Bioinformatics Institute in EU and National Center for Biotechnology Information in USA, Instead, DBCLS and other three institutes, NBDC for database catalogue and cross database search, DDBJ for nucleic acids sequence database and PDBj for protein 3D structure database, have built an alliance for bioscience information. Its portal site has been launched in July 2017 [6].

4. Conclusions

As one of the promising data sharing among independently developed databases all over the world, semantic web technology has been used in the database integration in DBCLS. Based on the RDF formulation and DBCLS RDFizing DB guidelines, 20 databases including different types of omics data are currently available from the NBDC RDF portal and several others are on the web. They will facilitate the reusability and inter-operability of the data in the life science research field and its easy integration will accelerate data sharing among the disciplines.

Acknowledgments. The author acknowledges all the members of DBCLS and NBDC for their contribution in developing tools, services, ontologies and databases, and arranging hackathon and lecture series. The activity of DBCLS is funded, in part, by JST NBDC and ROIS International Network Formation Project.

References

1. Rigden, D. J. et al., The 2016 database issue of Nucleic Acids Research and an updated molecular biology database collection. *Nucleic Acids Res.*, 44, D1–D6, 2016
2. DBCLS homepage, http://dbcls.rois.ac.jp/ [accessed on: Oct 2017]
3. DBCLS RDFizing DB guidelines, https://github.com/dbcls/rdfizing-db-guidelines [accessed on: Oct 2017]
4. NBDC RDF Portal, http://integbio.jp/rdf [accessed on: Oct 2017]
5. Wilkinson, M. et al. The FAIR guiding principles for scientific data management and stewardship. *Scientific Data,* 3, 160018, 2016
6. JBI Portal, http://jbioinfo.jp/ [access on: Oct 2017]

Recent Trends of Open Publication and Policy Development for Open Science toward Inter-Disciplinary Data Sharing & Publication

Kazuhiro Hayashi[1]*

[1]* *National Institute of Science and Technology Policy, 3-2-2, Kasumigaseki, Chiyoda-ku, Tokyo 100-0013, Japan*
Email: khayashi@nistep.go.jp

Summary. Open Science movement has been leveraged by Openness and Connectivity with ICT, aiming significant change of scholarship itself. Under this background, Open Science Policy has been developed including Japan, and current mission of Open Science is how to implement FAIR Data Science with data management policy, approved data repositories and research data infrastructures. On the other hand, this movement have developed some new styles of publication of articles such as using preprint server or publication platform by funding agencies. The key issue across the movement is reliability with standardization among researchers at first. As standardized process of publication with incentive models have organized robust and inter-disciplinal system "journal publishing" for researches for long years, standardized format for data must enhance data-sharing, interoperability, and sustainability, which already have been observed partially such as CIF and IIIF in natural science and humanities.

Keywords. Open Science, Open Publication, FAIR principles, standardized format, reliability.

1. Introduction

Open Science is one of important emerging issues around Science, Technology and Innovation Policy. There is still no definition of Open Science in general so far, and there have been many ways to implement it such as Open Access, Open (Research) Data, Citizen Science and so on.

In any ways, Open Science movement has been leveraged by Openness and Connectivity with ICT (Information and Communication Technology), aiming significant change of scholarship itself. [1] Especially EC sets Open Science in Horizon 2020 as one of important issues in Digital Single Market, and G7 Science and Technology Minister Meeting has had continuous discussion to make Open Science reality from 2013. The latest G7 SCIENCE MINISTERS' COMMUNIQUÉ says that "We recognize that ICT developments, the digitisation and the vast availability of data, efforts to push the science frontiers, and the need to address complex economic and societal challenges, are transforming the way in which science is performed towards Open Science paradigms." [2]

2. Open Science and FAIR Data Science

Under this background, Open Science Policy has been developed including Japan. The 5th Science and Technology Basic Plan sets an agenda of "Promotion of Open Science" in 2016. According to the report of the Cabinet Office for Open Science in 2015 [3], the outcomes of publicly funded research, such as published results and underlying data, should be accessible, unless they interfere with personal privacy, national security or direct commercial interests. All other countries have similar policies and they all are keen to research data-sharing.

Current mission of Open Science is how to implement FAIR Data Science. FAIR Data means Findable, Accessible, Interoperable, and Reusable Data.[4] In order to make research data FAIR, they focus on data management plan (DMP),

approved data repositories and research data infrastructures. Especially DMP is a catalyst to make stakeholders recognize the value of research data with their practical management of data. It is expected that accumulation of DMP would be a trigger to lead Open Science and innovation.

However, no one could have a confident vision of extreme data sharing world, which is supposed to develop with fostering a new culture of research for a long time.

3. New styles of publication of articles enhanced by potential of Openness

On the other hand, this movement with openness and connectivity have developed some new styles of publication of articles. [5]

Combination of posting draft articles on preprint server and publishing them on peer review Journal afterwards like arXive and Journals in Physics has been recognized as a good practice of exploiting openness. This combination has been challenged in other research fields recently. Furthermore, some domains such as Information Technology like deep learning only use preprint server because the time cycle of research is too short to conduct normal peer review.

Open Publication Platform by non governmental funding agency such as Wellcome Trust might change the way of publishing drastically because once funding agencies have their open publishing platform, publishers and libraries are no longer on the dissemination process in publication. These examples are so called incremental methods to make research outputs open focusing on articles, but it is reliable along with established publishing process.

4. Reliability and Standardization for Research

The key issue across the movement is reliability with standardization among researchers at first. Standardized process of publication with incentive models have organized robust and inter-disciplinal system "journal publishing" for researches for a long time. In addition, journal article is definitely one of interoperable standardized format of research data and it has played a significant role in research activity.

Also, standardized format for data must enhance data-sharing, interoperability, and sustainability, which already have been observed partially such as CIF (Crystallographic Information File) format in Chemistry Data and IIIF (International Image Interoperability Framework) format for image files in humanities. Accumulating those standardized data format would help develop a new inter-disciplinary framework of interoperable data sharing world.

5. Conclusions

All activities related Open Science should have been on the reliability among stakeholders and we could use properly both of established reliable publication process or emerging standardization of data format.

References

1. Open Digital Science - Final study report, 2016 https://ec.europa.eu/digital-single-market/en/news/open-digital-science-final-study-report [accessed on: Oct 2017]
2. G7 SCIENCE MINISTERS' COMMUNIQUÉ, 2017 http://www.g7italy.it/sites/default/files/documents/G7%20Science%20Communiqu%C3%A9.pdf [accessed on: Oct 2017]
3. Promoting Open Science in Japan "Opening up a new era for the advancement of science" Executive Summary Report by the Expert Panel on Open Science, based on Global Perspectives Cabinet Office, Government of Japan (2015) http://www8.cao.go.jp/cstp/sonota/openscience/150330_openscience_summary_en.pdf [accessed on: Oct 2017]
4. FORCE11, "Guiding Principles For Findable, Accessible, Interoperable And Re-Usable Data Publishing Version B1.0" : https://www.force11.org/fairprinciples [accessed on: Oct 2017]
5. Hayashi, K., Revolution of process on publishing and sharing towards Open Science enhanced by openness of scholarly communication. *STI Horizon,* 3, 35-39, 2017 (in Japanese) http://doi.org/10.15108/stih.00092 [accessed on: Oct 2017]

Antarctic rock samples database: current status and future perspectives

Tomokazu Hokada[1,2,*], Kazuyuki Shiraishi[1,2], Yoichi Motoyoshi[1,2], Yoshikuni Hiroi[1], Kenji Horie[1,2], Masaki Kanao[1,2,3], Hironori Yabuki[1,3]

[1*] *National Institute of Polar Research (NIPR), 10-3 Midori-cho, Tachikawa, Tokyo, 190-8518, Japan*
[2] *Department of Polar Sciences, The Graduate University for Advanced Studies (SOKENDAI), 10-3 Midori-cho, Tachikawa, Tokyo, 190-8518, Japan*
[3] *Joint Support-Center for Data Science Research (DS), 10-3 Midori-cho, Tachikawa, Tokyo, 190-8518, Japan*
Email: hokada@nipr.ac.jp

Summary. In Polar Science Resources Center of National Institute of Polar Research, the Rock Specimen Archive has collected and preserved some 20,000 rock and mineral specimens since the first Japanese Antarctic Research Expedition (JARE). The archive stores rocks and minerals not only from Antarctica but also Sri Lanka, India and Africa as international scientific research. Its collection is important for geological correlations and studies of earth's crust and mantle materials constituting Gondwana supercontinent. Specimens are organized according to year and region collection and are updated database. We will present the current status and future perspectives of the Antarctic rock samples database.

Keywords. Antarctic rock samples, Japanese Antarctic Research Expedition (JARE), Polar Science Resource Center.

1. Introduction

Antarctic rock samples belong to Polar Science Resource Center of National Institute of Polar Research. Rock samples and the related materials such as sample list, locality map and other information are also submitted and stored as archive. In order to arrange these to scientific purposes and museum exhibition, it is now on-going the construction of the catalogue.

2. Current status

Around 20,000 rock samples are stored in the Rock Storage Room of National Institute of Polar Research, and many other rock samples are still kept in university researchers at their own institutes. All those rock samples will be finally catalogued and stored in National Institute of Polar Research except a few rock samples kept in university museums and storages permanently. Because of the limitation of enough budget and man power, the cataloguing process is still underway.

3. Future perspectives

The establishment of Join Support-Center for Data Science Research in Research Organization of Information and Systems this year is good opportunity for organize the Antarctic rock samples database, and we have just started the arrangements. We will present the status of the Antarctic rock samples database.

Acknowledgments. The Antarctic rock samples database is supported by ROIS-DS-JOINT (007RP2017) to T. Hokada.

Web service for reproducible multidisciplinary data visualization

***Koji Imai*[1]*, *Yasuhiro Murayama*[1], *Ken Ebisawa*[2], *Daisuke Ikeda*[3], *Daisuke Kitao*[3]**

[1] National Institute of Information and Communications Technology, 4-2-1, Nukui-Kitamachi, Koganei, Tokyo, 184-8795, Japan*
[2] 3-1-1 Yoshinodai, Chuo-ku, Sagamihara City, Kanagawa Prefecture, 252-5210, Japan
[3] Kyushu University, 744, Motooka, Nishi-ku, Fukuoka, 819-0395, Japan
Email: koji.imai@nict.go.jp

Summary. We propose a new method for reproducible data visualization on a web browser. A web service, Cross-Cutting Comparisons (C3) has a query string (QS)-controllable system to make various interactive charts of earth, planetary and space sciences. By including information of data handling procedures in the QS in an orderly manner, the chart is easy to understand, remake and share via text-based communication tools.

Keywords. reproducibility, data visualization, geoscience, web service, cross-cutting research.

1. Introduction

Reproducibility is one of the foundations of the scientific method. Reproducible evidence is imperative to build new scientific knowledge. Since most natural scientists use figures, graphs, plots or diagrams (hereinafter collectively called charts) to understand a phenomenon in detail, reproducible charts are needed to promote scientific development. However, it is not easy especially for studies dealing with scientific data of different fields.

For this reason, we have been building a web service, Cross-Cutting Comparisons (C3) (https://darts.isas.jaxa.jp/C3/; [1]), to provide reproducible charts of earth, planetary and space sciences.

2. Data-flow design

C3 is an application of Data ARchives and Transmission System (DARTS) [2-3]. It consists of data and web servers. The data server currently stores opened data sets of the fields in earth, planetary and space sciences. The web server mediates between a client and the data server.

To handle a large number of requests from users, C3 checks the inputted setting and dynamically creates a query string (QS) on the client side. Then, the web server executes data handling procedures according to the client's requests.

3. Multidisciplinary data visualization

To capture the feature of various phenomena such as long-term variations or sudden events of earth, planetary and space sciences, it is necessary to examine scientific data by flexibly changing the time and space scales. C3 uses HTML and JavaScript libraries to visualize multidisciplinary data on a web browser.

Figure 1 shows examples of the charts: (a) is time series of total solar irradiance from the Solar Radiation and Climate Experiment (SORCE) [4], (b) is an altitude profile of temperature from the Atmospheric Chemistry Experiment-Fourier Transform Spectrometer (ACE-FTS) [5], (c) is a global map of ozone from the Superconducting Submillimeter-Wave Limb-Emission Sounder (SMILES) [6], and (d) is a focal depth from the United States Geological Survey (USGS) (https://www.usgs.gov/).

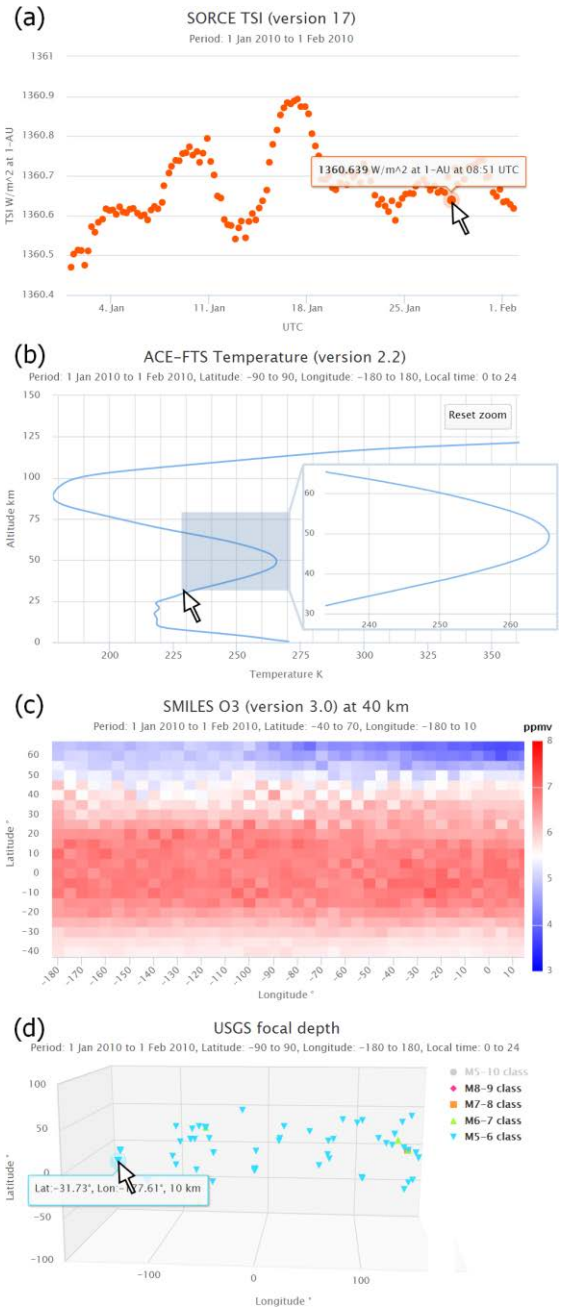


Figure 1 Examples of the interactive charts
(a): total solar irradiance (TSI) (b): altitude profile of temperature, (c): global map of ozone, (d): focal depth

4. Approach of reproducibility

All the charts showed in Figure 1 have a QS in an address bar. The structure of the QS of C3 is as follow:

header+selection+extraction+(analytic method).

The QS consists of three parts separated by plusses (+): header, selection and extraction. The header part includes information of language of the chart and the system version of C3 (default is the latest version), the selection part has information of metadata and figure type. The information of data extraction (e.g., time period, location and altitude) is described in the extraction part. Another part of data analysis (e.g., averaging, least square, correlation coefficient, size of mesh grid or a combination of these analyses) will be added to the QS in the next version of C3.

5. Conclusions

C3 has a QS-controllable system to make various interactive charts. By explicitly showing the inputted setting in an orderly manner in the QS, it is easy to understand how the chart is made. User can easily review the previous work, quickly access to the charts, and also reduce an amount of data. Our new approach making charts by a QS has the possibility to promote scientific development in the forthcoming epoch of Open Data.

Acknowledgments. C3 is maintained by the Center for Science-satellite Operation and Data Archive (C-SODA). This work was supported by JSPS KAKENHI Grant Number 15H02787.

References

1. Imai, K. et al., Quick look service for geoscience, Journal of Space Science Informatics Japan, Vol. 5, 93-109, 2016
2. Miura, A. et al., ISAS Data Archive and Transmission System (DARTS), ASPC, 216, 180, 2000
3. Tamura, T. et al., Data Archive and Transfer System (DARTS) of ISAS, ASPC, 314, 22, 2004.
4. Lawrence, G. M., G. Rottman, J. Harder, and T. Woods, Solar Total Irradiance Monitor (TIM), Metrologia, 37, 407, 2000
5. Bernath, P. F., et al., Atmospheric Chemistry Experiment (ACE): Mission overview, Geophys. Res. Lett., 32, L15S01, 2005
6. Kikuchi, K. et al., Overview and Early Results of the Superconducting Submillimeter-Wave Limb-Emission Sounder (SMILES), JGR, 115, D23306, 2010

The Present Situation of Open Data Usage in the Social Sciences and Related Problems

Yusuke Inagaki[1]*

[1]* *Center for Social Data Structuring / the Institute of Statistical Mathematics , 10-3 Midori-cho, Tachikawa, Tokyo 190-8562, Japan*
Email: yinagaki@ism.ac.jp

Summary. The main purpose of this presentation is to give some exeplanation on the circumstances in the open data usage in social sciences. In Japan, we have some data archives and archive centers of social survery data, however, their actual status is far behind compared to overseas distinguished archive centers such as GESIS in Germany and Roper Center in USA. I will present a brief description of our activities in the Center for Social Data Structuring to improve the present situation.

Keywords. Data Archive Center, Social Science, Social Survey, Privacy Protection, Response Rate.

1. Introduction

The enforcement of the Personal Information Protection Law and rising privacy protection consciousness of people caused thereby have brought a significant decrease in the response rates in many social surveys in Japan. The Surveys of the Japanese National Character, one of the oldest social surveys, conducted by the Institute of Statistical Mathematics is no exception [1] (see also Fig 1).

In response to this situation, web social surveys came to be actively carried out these days to get a large number of data in an inexpensive manner, and to make questionnaire result obtainable immediately. However, there are also some disadvantages in web surveys. For instance, some kind of sampling bias: Coverage error has been pointed out for a long time. Therefore, the Science Council of Japan has requested researchers to indicate the problems and the limits of web survey on announcing the results to prevent people's misunderstanding [2].

Under these circumstances, many researchers and organizations have been voicing their opinions requesting secondary use of past survey data sets, and to meet such requests some service systems for secondary/joint use of individual data research have been developed by certain research institutes.

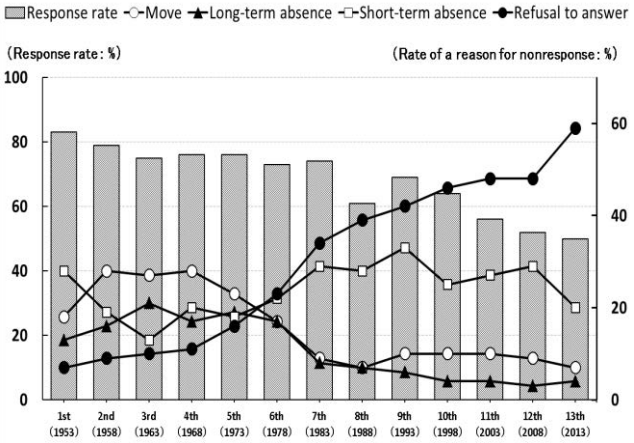


Fig 1. Transition of response rates the Surveys of the Japanese National Character

2. Situation in the world

USA and some European countries (especially Germany) have worked on that subject diligently, and they established archive centers almost a half-century ahead of Japan. Among them, Roper Center, ICPSR (Inter-university Consortium for Political and Social Research) at the University of Michigan and GESIS (German Social Science Infrastructure) have been played a central role for promoting archive service in all over the world.

With the construction of data archive centers in each country, they gradually started to

be recognized the importance of international cooperation and collaboration. Thus, CESSDA (Committee of European Social Science Data Archives) was organized among European countries in 1970, after that, IFDO (International Federation of Data Organizations) was formed including countries on the North American continent as literally an international institution.

3. Situation in Japan

Some data archive centers were established since the latter half of the 1990s. Familiar examples are SRDQ (Social Research Database on Questionnaires) at Osaka University, RUDA (Rikkyo University Data Archive) at Rikkyo University, SORD (Social and Opinion Research Database) at Sapporo Gakuin University and CSRDA (Center for Social Research and Data Archives) at the University of Tokyo. CSRDA plays a key role among Japanese social science data archives at this time, and they manage SSJDA (Social Science Japan Data Archive).

SSJDA provides well-known survey data sets like JLPS (the Japanese Life Course Panel Survey), NFRJ (National Family Research of Japan), and JGSS (the Japanese General Social Surveys). Therefore, many researchers and students use this archive ,and SSJDA has been chosen as practically the main depository institution by depositors who had conducted a social survey [3] (see also Fig 2).

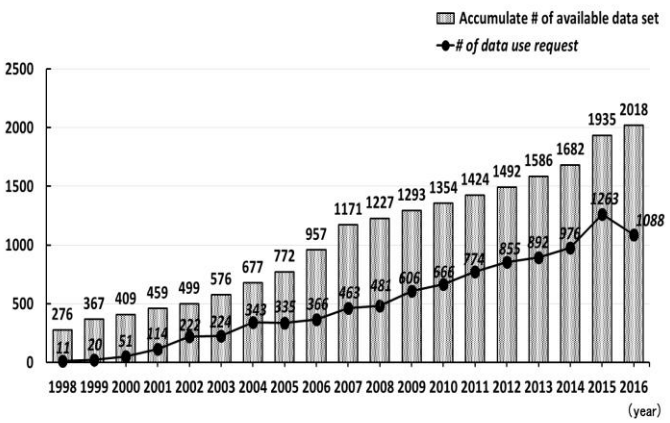


Fig 2. Transition of numbers of available data set and use request in SSJDA

4. Present problems in Japan

However, many Japanese archive centers or archives, except CSRDA and SSJDA, are no longer active or practicing. Moreover, if seen globally, even CSRDA and SSJDA are not really active because of the absolute number of the deposit data sets is small. Possible reasons for this present situation are as follows:

1. Procedures for data deposit or data use are complicated and confusing, or depositors and users think that procedures are difficult (but that is not true).

2. There is no incentive system for data deposit. Therefore, active researchers unwilling to confide their survey data with great difficulty.

3. Most social surveys contain confidential information which allows specification of an individual. Potential depositors have become cautious in depositing data from the viewpoint of privacy protection. It leads to the lack of improvement in survey data deposition.

Problems which described above should be solved for the utilization of open-data or joint usage. The Center for Social Data Structuring is working on developing a new system "Human-social data compliance management platform".

References

1. Nakamura, T., Yoshino, R., Maeda, T., Inagaki, Y., Shibai, K., (Eds.), A Study of the Japanese National Character: The Thitteenth Nationwide Survey (2013), *ISM Survey Research Report No.119,* The Institute of Statistical Mathematics, Tokyo, 2017
2. The Science Council of Japan, http://www.scj.go.jp/ja/info/kohyo/pdf/kohyo-23-t248-7.pdf [accessed on: October 2017] (in Japanese)
3. Center for Social Research and Data Archives, http://csrda.iss.u-tokyo.ac.jp/pdf/Brochure.pdf [accessed on: October 2017]

Magnetic Data at WDC Kyoto—Services under International Collaborations

***Toshihiko Iyemori*[1]*, *Masahiko Takeda*[1], *Masahito Nose*[1], *Hiroaki Toh*[1], *Yoko Odagi*[1] , *Noriko Takeuchi*[1]**

[1]* *Graduate School of Science, Kyoto University, Kyoto 606-8502, Japan*
Email: iyemori@kugi.kyoto-u.ac.jp

Summary. The World Data Center (WDC) for Geomagnetism, Kyoto was established in 1957 under ICSU/World Data Centre Panel and has been cooperating with the IAGA, International Association of Geomagnetism and Aeronomy, and also with the INTERMAGNET. Among regular data services, derivation of the AE and the Dst indices need most strong and stable international collaborations with geomagnetic observatories to keep the quality and constant dissemination. The users tend to request; (1) real-time and higher-time resolution data, (2) global coverage of observation sites, (3) quality assured data and (4) data analysis system with other type of data such as satellite data or the data in other discipline. To satisfy or to harmonize these requests at the same time is not easy for a data centre within limited resources. We have been trying to develop real-time data service of the AE and the Dst indices, new geomagnetic indices such as ASY/SYM, and WP, new distributed data system among universities (IUGONET). To continue and develop the data service, more social (or governmental) recognition on the importance of data centres and the resources necessary for operation. The collaboration with (or through) international data organization such as the World Data System (ICSU/WDS) and /or CODATA may become important also from this point of view.

Keywords. Geomagnetic data, international collaborations, World Data System, user requests.

1. Introduction

The World Data Center (WDC) for Geomagnetism, Kyoto, hereafter referred to as "WDC Kyoto", a regular member of the ICSU/WDS, belongs to the Graduate School of Science, Kyoto University. The official name as the latter is the Data Analysis Center for Geomagnetism and Space Magnetism (DACGSM). The DACGSM is also in charge of student education as well as scientific research on Geomagnetism, Space Magnetism and related Informatics.

The WDC Kyoto was established in 1957 under the ICSU/World Data Centre Panel and, in 1978, the DACGSM was established in the Graduate School of Science, Kyoto University to operate the WDC Kyoto.

The main task is to serve geomagnetic data collected from the world and to derive the Dst and the AE indices as well as to provide professional information/knowledge related to geomagnetic field such as geomagnetic main field model.

In this report, we present recent activity with international collaborations.

2. Derivation of geomagnetic indices

A geomagnetic storm index, Dst, and auroral electrojet index, AE, are very widely used geomagnetic indices. However, their stable derivation, in particular the derivation of the AE index has been difficult because the 12 geomagnetic observatories for the AE (AE stations) locate in the polar region surrounding the Arctic Ocean. The stable operation in Siberia has been especially difficult. To solve this problem, we had a project called "RapidMag" which was an international collaboration among AARI, NICT, APL, ROSHYDROMET and DACGSM. With this collaboration, we can now derive in near real time, the AE indices and serve from our web

site. However, we still often have problems, and we need to continue the collaboration among those institutions.

On the Dst index, we have been collaborating with Kakioka Magnetic Observatory, USGS (Honolulu and San Juan observatories), SANSA (Hermanus observatory) and IIG (Alibag observatory), and we could start near real-time derivation of the Dst index in 1996.

It should be noted that the resolution to support these activity by IAGA or ICSU has been effective to get domestic (or governmental) support to some extent.

3. Interdisciplinary data service

One of the main objectives of the World Data System is to promote interdisciplinary science and open science. As a regular member of the WDS, we are trying to promote open science and open data. The geomagnetism or space science is, in general, interdisciplinary by nature because various phenomena treated in this field are caused by the interaction among various parameters in wide range of space, i.e., from the Sun to the Earth via interplanetary space, magnetosphere and ionosphere.

The IUGONET, Inter university Upper atmosphere Global Observation NET work, was constructed for helping the researchers to use various data sets distributed in many institutions. The basic idea of the IUGONET is to have a common database of "metadata". With the database, we can search and connect distributed databases. Because we, i.e., IUGONET member institutions, adopted NASA's "SPASE" data model for IUGONET metadata database, we can also connect the system to other dataset in USA. However, because EU countries use different data model, we need to collaborate with European group to connect different systems, and we developed a prototype system called "vocabulary broker", which will be shown in another talk in this workshop, i.e., the paper by B. Ritschel et al..

Another open data activity is to promote using DOI for data sets. We already use a DOI for the Dst and Wp indices. This will be also presented in this workshop, i.e., in the paper by M. Nose et al..

4. How to harmonize contradictory user requests

Making quality controlled data set normally takes time to be prepared. On the other hand, we often have strong requests of (near) real-time data without quality check. Taking into account that the main purpose of our data centre under ICSU is to promote science and education, both of quality and timeliness are equally important. Therefore we should accept both requests. In our case, we categorize the data to "final data", "provisional data" and "real-time data" and attach some notes on the quality although people often miss or ignore the notes. In general, we should prepare enough metadata for each dataset including the description of data quality. Education on information literacy should include the importance of our attitude to check the data quality in handling given data sets.

5. Conclusions

Because of the nature of geomagnetism and space science, we need geomagnetic data observed worldwide and we also need interdisciplinary data. Therefore international and interdisciplinary collaboration is essential in our field, and the role of international organizations under the ICSU such as WDS, CODATA, IUGG, IAGA, etc. and collaboration with these organizations are important for us to make better data service.

Activities of Polar Environment Data Science Center

Akira Kadokura[1*]

[1*] *Polar Environment Data Science Center, DS, ROIS, 10-3, Midoricho, Tachikawa, Tokyo 190-8518, Japan*
Email: kadokura@nipr.ac.jp

Summary. Activities of the Polar Environment Data Science Center (PEDSC) are introduced. PEDSC has been established in the Joint Support-Center for Data Science Research (DS) of the Research Organization of Information and Systems (ROIS) in 2017. Purpose of the PEDSC is to promote collaboration with the data obtained by the research activities in the polar region, and to play a key role of the data activity in polar science to contribute to the global environment research.

Keywords. polar science, data science, open data, data publishing, data journal.

1. Introduction

Activity of the PEDSC is closely related with the research and observation activities of the National Institute of Polar Research (NIPR). Main purpose of the PEDSC is to promote the opening and utilization of the various data stored in NIPR in collaboration with universities and other institutions in Japan and foreign countries.

2. Data

Data to be handled by the PEDSC are obtained in both Antarctic and Arctic regions mainly by the four research groups in NIPR; Space and upper atmospheric sciences group, Meteorology and glaciology group, Geoscience group, and Bioscience group. Various data in various research fields have been obtained so far, e.g. aurora and upper atmosphere, meteorology and marine science, snow and ice, geology, geomorphology, seismology, gravity, and biology. Those data are stored and archived in various forms, and basically classified in two categories, time series data and sample data. The former data are sampled at a fixed interval and recorded continuously during some observation period. The latter data are obtained at some specific date at some specific locations as a sampled material, e.g. rock, meteorite, ice core, sea water, air, etc. For such sample data, both their catalogues and analysis data are created and stored.

3. Database system

Each data mentioned in the Section 2 are stored and archived in each database in each research field by each research group in NIPR in each form and style. There are also following general database systems for the data in NIPR, which can handle various data in various research fields.

3.1 Science database

Science database (https://scidbase.nipr.ac.jp/) is a metadata database for all the data in all the research fields of polar science, and has a close relationship with international data activities, e.g. NASA Global Change Master Directory (GCMD), Standing Committee on Antarctic Data Management (SCADM) under the Scientific Committee on Antarctic Research (SCAR), etc.

3.2 Arctic Data archive System

Arctic Data archive System (ADS) (https://ads.nipr.ac.jp/portal/index.action) is a metadata and actual data database system mainly for Arctic projects such as GRENE (Green Network of Excellence) and ArCS (Arctic Challenge for Sustainability). ADS is also equipped with online visualization and analysis tools, and is used for collaboration with Japanese and international communities for Arctic research

3.3 IUGONET system

IUGONET (Inter-university Upper atmosphere Global Observation NETwork) system (http://www.iugonet.org/) is a metadata

database system which is developed in an inter-university project among NIPR and 4 universities for upper atmospheric science research. IUGONET is also equipped with display and analysis software tools (UDAS: iUgonet Data Analysis Software) for actual data, and is close relationship with international data activities such as SPEDAS (Space Physics Environment Data Analysis Software) and SPASE (Space Physics Archive Search and Extract) groups. Workshop and school for the IUGONET system are regularly held for students and researchers in Japan and abroad.

4. Current problems

Current needs and problems on the data and database systems in NIPR are; (1) A synthetic database system to cover, search and utilize all the data and database in all the research fields is not constructed and needed to understand the polar science activities as a whole. (2) Status of the processing, archiving and opening of the data is in wide variety for each data and database, depending on status of the resources of manpower, hardware and software.

5. Activity plan of PEDSC

Current targets of the PEDSC are; (1) To construct a synthetic database for all the research fields of polar science. (2) To promote the processing, archiving and opening of each data in each research field both for the time series data and sample data. (3) To promote the data publication through the "Polar Data Journal ", data journal of NIPR. (4) To promote collaboration in data science with Japanese and international communities.

Staff of the PEDSC in 2017 JFY consists of a manager with one associate professor, three specially-appointed associate professors, and two office assistants.

Data Citation Procedure on Science Database of Polar Research

Masaki Kanao[1]*, ***Akira Kadokura***[1]

[1]* *Joint Support-Center for Data Science Research, Research Organization of Information and Systems, 10-3, Midori-cho, Tachikawa-shi, Tokyo 190-8518, Japan*
Email: kanao@nipr.ac.jp

Summary. The Polar Environmental Data Science Center (PEDSC) of the Joint Support-Center for Data Science Research (DS), the Research Organization of Information and Systems (ROIS) has a responsibility to manage the data for Japan as a National Antarctic Data Center (NADC) during the last few decades. At the International Polar Year (IPY2007-2008), a significant number of multi-disciplinary metadata/data have been compiled mainly from IPY- endorsed projects. These collected metadata have a tight collaboration with the Global Change Master Directory (GCMD), the Polar Information Commons (PIC), as well as several centers belonging to the World Data System (WDS). The compiled metadata by PEDSC, moreover, are recently equipped an automatically attributing system of the Digital Object Identifier (DOI) by requesting to the DataCite through the Japan Link Center.

Keywords. Polar Environmental Data Science Center, metadata management , DOI, data citation, polar science.

1. Introduction

The Polar Environmental Data Science Center (PEDSC) has a significant task to archive and deliver the data obtained from polar regions especially by Japanese related scientific activities. Summary information (metadata) of all the archived data are available to the usage of involved polar communities, together with more general interests by public domain. The compiled metadata describe various scientific disciplines (space and upper atmospheric sciences, meteorology and glaciology, geosciences and biosciences) from both long- and short-term projects in the Arctic and Antarctic, in which the majorities are obtained from Japanese Antarctic Research Expedition (JARE) [1]. These science branches cover almost all recent studies on environmental changes as well as earth evolution viewed from polar region. Inside the portal server for the scientific metadata (http://scidbase.nipr.ac.jp/), 380 records have been compiled as of August 2017.

In this report, present status on metadata management by PEDSC will be demonstrated, in particular focused on several new trials regarding data citation procedure by attributing the Digital Object Identifier (DOI) for the purpose of utilization to data users among polar/global communities [2]. Interoperable metadata linkage and promoting data citation as demonstrated here could provide a efficient model in a framework of long-term preservation and data publication regarding polar region within global system.

2. Data Citation

As for the compiled metadata for all science branches, a sophisticated system that can automatically attribute DOIs are recently equipped inside the portal server. The DOIs can be requested to the "DataCite (https://www.datacite.org/)" through a gateway interface provided by "Japan Link Center (JaLC; https://japanlinkcenter.org/)". The JaLC is Japanese organization authorized as one of the Registration Agency (RA) which can provide the DOIs.

Under the adequate evaluation procedures, the metadata and their associated actual dataset with enough quality of publication could be

attributed by their DOIs with a "prefix" of "10.17592". Under the DOI auto-numbering rule, the "suffix" part of the DOIs (i.e., the character string ordering) will be generated arbitrary in a manner defined by the metadata portal. After receiving offers to obtain DOIs from the data providers/managers, quality of individual data can be strictly evaluated by involved "data management committee", followed by attributing their DOIs for those with sufficient revel of quality for opening/publishing the data into a public domain. There are several evaluation terms before assignment of the DOIs; regarding data quality, publishing methodology, long-term maintenance strategy, and their data policy, etc.; however, these evaluation items should be overcome in both the description of the metadata itself and the quality of corresponding actual dataset.

Significant and dedicated works for serving the data/metadata as mentioned in this paper have been conducted by the staff of PEDSC for long years before and after the IPY. Multi-disciplinary scientific data collected in bi-polar regions have great merits for researches of global environmental change currently progressing [3-4]. Promoting data citation procedure introduced here could be a model case with an effective framework for long-term strategy of publication and preservation of polar data among global system. Moreover, the approach of data citation conducted by this study could have a potential as socially relevant applications to the public domain, in addition to polar community.

3. Conclusions

The status of metadata management in PEDSC of DS, ROIS is summarized in this short report. Many dedicated data service tasks have been conducted by the center staffs as well as the related members of other institutions / organizations. Several different aspects of scientific data collected in polar region have great importance for global environmental research in this century. In order to construct an effective framework for long-term strategy of the polar data, the data should be made available promptly and new Internet technologies such a repository network service must be employed. The next generation of database for polar science aims to compile all the kinds of datasets by use of integrated applications so as to offer to a public domain. The future integrated database will be composed by both the data from Arctic and Antarctic; beyond the existing science disciplines of earth/environmental/biosciences, as well as including social/human science branches in polar region. The new orienting database also aims to provide information to relating data centers, together with libraries where hold a plenty number of publications as their repositories.

Acknowledgments. The authors would like to express their appreciation to many collaborators involving polar data management, in particular to the member of PEDSC, data committee on SCAR, IASC, WDS, CODATA and IPY.

References

1. Kanao, M., Okada, M., Kadokura, A., Metadata Management at the Polar Data Center of the National Institute of Polar Research, Japan. *the CODATA Data Science Journal,* 13, PDA27-PDA31, 2014
2. Duerr, R., Downs, R., Tilmes, C., Barkstrom, B., Lenhardt, W., Glassy, J., Bermudez, L., Slaughter, P., On the utility of identification schemes for digital earth science data: an assessment and recommendations. *Earth Science Informatics,* 4, 139-160, 2011
3. Parsons, M. A., Godoy, Ø., LeDrew, E., de Bruin, T., Danis, B., Tomlinson, S., Carlson, D., A Conceptual Framework for Managing Very Diverse Data for Complex, Interdisciplinary Science. *Journal of Information Science*, 1-21, 2011
4. Parsons M. A., Fox, P. A., Is Data Publication the Right Metaphor? the *CODATA Data Science Journal* , 12, WDS32-WDS46, 2013

History of polar data management in Japan; before and after the IPY2007-2008

Masaki Kanao[1*], ***Akira Kadokura***[1]

[1*] *Joint Support-Center for Data Science Research, Research Organization of Information and Systems, 10-3, Midori-cho, Tachikawa-shi, Tokyo 190-8518, Japan*
Email: kanao@nipr.ac.jp

Summary. Diverse data accumulated by many science disciplines 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), followed by the Polar Environment Data Science Center (PEDSC) of the Joint Support-Center for Data Science Research (DS) have responsibility to manage these polar data in Japan as a National Antarctic Data Center (NADC). During the IPY, a significant number of multi-disciplinary metadata records were compiled from IPY- endorsed projects. 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). In this short report, a long-term history of data management for polar science is summarized, focusing on the era before and after the IPY.

Keywords. International Polar Year, Polar Information Common, data management, Polar Environmental Data Science Center, National Antarctic Data Center.

1. Introduction

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 Antarctic Treaty, each country is required to establish NADC and to properly disclose the data collected from involved scientists. The PDC/PEDSC has performed the function of a NADC for Japan and established a data policy in February 2007, based on the requirements of the Standing Committee on Antarctic Data Management (SCADM) of the Scientific Committee on Antarctic Research (SCAR). This contributed to the subsequent SCAR Data and Information Management Strategy [1].

Dedicated data services have been conducted by PDC/PEDSC as a member of NADC under SCAR. Several different aspects of scientific data collected in polar region have great significance for global environmental research. To construct an effective framework for long-term strategy of the polar data, data must be made available promptly by Internet technologies such a repository network service. In addition to activities in polar science communities of SCAR and the International Arctic Science Committee (IASC), tighter linkages expected to 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.

2. IPY Data Management

The International Polar Year (IPY 2007-2008) was the world's most diverse 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 polar region. The interdisciplinary exchange helped us understand and addressed grand challenges such as rapid environmental change and its impact on society. Eventually, Japanese

researchers participated to 63 projects endorsed by the IPY Joint Committee. The huge amount of data accumulating during IPY should be the most important legacy if it is well preserved and utilized [2].

The science database provided by PDC/PEDSC has a tight connection with AMD in GCMD. In addition to the IPY-related data, data from Japanese national and other international projects had been compiled. In totally, 300 metadata were compiled in Japanese Antarctic portal in GCMD. PDC/PEDSC stores its metadata in their original format, but this includes the main items listed in GCMD Directory Interchange Format (DIF). There are tight cross-linkages in corresponding metadata in AMD. Metadata collected by IPY projects for Japan have also been accumulated in an IPY portal of GCMD. More than 250 metadata from Japan were stored in the IPY portal [3]. This constitutes a significant proportion of all IPY metadata to GCMD.

3. Data Legacy of IPY

SCADM has been strongly connected with IPY data-management activities (IPY Data and Information Service: IPY-DIS). IPY data policy emphasized a need to make data available on the "shortest feasible timescale." In accordance with IPY data policy, IPY-DIS recommended that data be formally cited when used, and the IPY Data Committee has developed initial guidelines for how data should be cited [4]. The guidelines harmonized different approaches, and they adopted by many data centers relating polar area. 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 [2]. The PIC serves as an open, virtual repository for vital scientific data and information and provides a shared, community-based cyber-infrastructure fostering innovation and improved scientific understanding while encouraging participation in research, education, planning, and management in polar region. PIC developed specialized tools that produce a small, machine-readable "badge" that is attached to the data. However, the badge requested data users to adhere to basic ethical norms of data use including proper citation. This service was coupled with a cloud-based repository that may not have a suitable archive elsewhere. NIPR/DS have made contributions to PIC, both by attaching badges and registration in the repository. As of October 2017, Japan had contributed more than 50 data sets to the PIC.

Polar data have great relevance for modern, global environmental research well beyond the polar region. It is critical to explore the cloud approaches such as the PIC to develop an effective framework for open and long-term stewardship of polar data. The status of data-management before and after the IPY in Japan was introduced in this report. Several different aspects of the scientific data collected in the polar region have great advantage for global environmental research as well as in future.

Acknowledgments. The authors express their appreciation to all collaborators of the IPY activities. They also acknowledge the members of SCADM and IPY Data committee for their efforts to adhere to data-management issues.

References

1. Finney, K., SCAR Data and Information Management Strategy 2009-2013. *SCAR Ad-hoc Group on Data Management*, 34, 2009
2. Parsons, M. A., Godoy, Ø., LeDrew, E., de Bruin, T., Danis, B., Tomlinson, S., Carlson, D., A Conceptual Framework for Managing Very Diverse Data for Complex, Interdisciplinary Science. *Journal of Information Science*, 1-21, 2011
3. Kanao, M., Kadokura, A., Okada, M., Yamnouchi, T., Shiraishi, K., Sato, N., Parsons, M. A., THE STATE OF IPY DATA MANAGEMENT: THE JAPANESE CONTRIBUTION AND LEGACY, *Data Science Journal*, 12, WDS124-WDS128, 2013
4. Parsons, M. A., Duerr, R., Minster, J.-B., Data Citation and Peer Review. *EOS Transactions, AGU*, 91, 297-298, 2010

Open Data in the Humanities: Data Sharing and Publication for Triadic Co-Creation

Asanobu Kitamoto[1,2]*

[1] *Center for Open Data in the Humanities (CODH), Joint Support-Center for Data Science Research, Research Organization of Information and Systems, Japan*
[2] *National Institute of Informatics, 2-1-2, Hitotsubashi, Chiyoda-ku, Tokyo, 101-8430, Japan*
Email: kitamoto@nii.ac.jp

Summary. Center for Open Data in the Humanities (CODH) was established in April 2017 with the aim of developing data science research in the humanities. CODH has the concept of "triadic co-creation" in which three types of stakeholders, namely humanities scholars, machines (computer scientists), and citizens, collaborate together to advance data creation, analysis and utilization. In particular, we focus on the redefinition of the division of roles between humans and machines by providing large-scale open training data for machine learning (artificial intelligence). We already released several open datasets, such as books, characters, and recipes that are collaboratively created with National Institute of Japanese Literature under NIJL-NW project, and other datasets such as photographs, maps, geographic information, and magazines with other collaborators. We also work on data publication in the humanities, using IIIF (International Image Interoperability Framework) as one of platforms for sharing data as materials for humanities research.

Keywords. Humanities, Open data, Triadic co-creation, data publication, IIIF (International Image Interoperability Framework).

1. Introduction

Center for Open Data in the Humanities (CODH) is promoting research and support activities aimed at opening data and promoting data-driven collaboration in the humanities discipline. In the humanities research community, however, data science approach based on large-scale open data is still premature and we cannot expect that the usage of open data increases without effort. CODH will therefore develop and release databases and tools by adopting the methodology of Digital Humanities (Computers and Humanities), which shows a rapid growth in the global research community, and also hold seminars and tutorials to promote the utilization of research resources.

2. Triadic Co-Creation

CODH has the concept of "triadic co-creation" in which three types of stakeholders, namely humanities scholars, machines (computer scientists), and citizens, collaborate together to advance data creation, analysis and utilization. In particular, we focus on the redefinition of the division of roles between humans and machines by providing large-scale open training data for machine learning (artificial intelligence).

Thanks to the development of machine learning technology in recent years, some work can be transferred from humans to machines, but in order to ask machines to work, we need to provide training data in the first place. The lack of such open training data leads to the delay of introducing machines to humanities research, so CODH will work on constructing infrastructure for open training data on which researchers and citizens can learn about and co-create open training data.

3. Open Data

We already released several open datasets, such as books, characters, and recipes that are collaboratively created under "Project to Build an

International Collaborative Research Network for Pre-modern Japanese Texts" (NIJL-NW Project), promoted mainly by National Institute of Japanese Literature [1].

First, "Dataset of Pre-Modern Japanese Text (PMJT)" provides image data of 701 pre-modern books in a downloadable format. By assigning DOI (Digital Object Identifier) to each book, image data can be uniquely identified even when there are multiple books with the same title.

Second, "Dataset of PMJT Character Shapes" contains 3,999 character types and 403,242 characters of old Japanese characters (called Kuzushi-ji). The dataset can be used not only as learning material for humans but also as training data for machines to develop optical character recognition (OCR) software. These datasets are also used for our "Kuzushi-ji Challenge!" campaign to involve experts and citizens to develop best artificial intelligence software that recognizes old Japanese characters.

Third, "Dataset of Edo Cooking Recipes" is the collection of recipes from a cooking book "Tamago Hyakuchin," which contains more than 100 recipes about eggs. We translated the original recipes to modern Japanese and structured for citizen cookers. Moreover, those recipe data was released in Japan's largest recipe service "Cookpad," which triggered unexpected excitement among citizens [2].

4. Data Sharing and Publication

Sharing data as open data is one form of scholarly publication in the age of open science. To find the best practice of data publication in the humanities, we are now focusing on IIIF (International Image Interoperability Framework) as infrastructure for image-related projects.

IIIF has recently seen a rapid growth of adoption as interoperable, high-resolution image delivery from museums and libraries around the world, and CODH is one of them to use IIIF for Pre-modern Japanese Text, and Digital Silk Road Digital Archive of Toyo Bunko Rare Books. While being widely adopted, IIIF is still a premature specification without important functions for humanities research.

CODH focuses on a use case of collecting interesting images from IIIF contents all over the world. We proposed a new specification called Curation API and developed a reference implementation called IIIF Curation Viewer, applied to PMJT curation and IIIF global curation.

Cropping images and collecting them under a theme is the basic task in the humanities research, and sharing them is one form of data publication having value as the material of subsequent research. We are now studying the potential of this approach in the field of art history.

5. Conclusion

CODH has other types of open data such as photographs, maps, geographic information and magazines. To organize and utilize those datasets, CODH also develops web and mobile applications for researchers and citizens. To publicize and disseminate open data in the humanities, various programs such as seminars, tutorials and courses should be helpful. Moreover, wider involvement of researchers and citizens into data creation and analysis activities, or citizen science, is another relevant challenge. CODH's activities are naturally across disciplines, and beyond disciplines in the sense of trans-disciplinary science.

References

1. Kitamoto, A., Yamamoto, K., "High-throughput Collation Workflow for the Digital Critique of Old Japanese Books Using Computer Vision Techniques". *Sixth Annual Conference of the Japanese Association for Digital Humanities (JADH2016),* 2016
2. Kitamoto, A., "FAIRness for Citizens: Workflow and Platform for Open Data with a Case Study on Edo Cooking Recipes". *Seventh Annual Conference of the Japanese Association for Digital Humanities (JADH2017),* 14-16, 2017

Data Publishing and Data Citation - Are We There Yet?

Jens Klump[1]*

[1]* *CSIRO Mineral Resources, 26 Dick Perry Avenue, Kensington, Western Australia, 6151, Australia*
Email: jens.klump@csiro.au

Summary. The idea of treating data as a form of publication has been around for nearly twenty years. Technical barriers to the access to knowledge were minimised through the potential of online access over the internet but reality fell behind expectations. Looking at the comparatively low numbers of data publications we have to ask the question: are we offering the right incentives to researchers to share their data with others more freely? To bring about the changes in social norms around data we need to understand the balance between reputation gain and collaboration gain that motivates researchers to make data accessible - or not - and identify suitable points to intervene.

Keywords. Data publication, science policy, open access to data, social studies.

1. Introduction

The idea of treating data as a form of publication has been around for nearly twenty years [1]. In the past data were published as part of the original publication, primarily in the form of data tables. Over time the size of the data sets used in a scientific publication grew, and journal publishers started to cite page limits as a reason to exclude data tables from publications. As a result, data used as the basis of a publication are rarely published anymore, creating a structural barrier to the publication of data [2].

The technical barriers to the access to knowledge were minimised through the potential of online access over the internet but reality fell behind expectations. This discrepancy between expectations for broader access to knowledge and the barriers still encountered led to the formation of Open Access initiatives. The policy papers on the importance of making research data available appeal to the common good but make few suggestions how the desired cultural change is going to be achieved.

Discussions about creating broader access to data in science often revolve around 'finding the right incentives' (e.g. [3-4]). Given the important role of publication and citation in scholarship it is widely assumed that formal data publication would be an incentive for researchers to make data available [5-8]. Data publication follows the forms developed for the publication of research papers. But is it the data or the intellectual work that we are interested in as peers? Does the recognition gained by data publication merit the additional effort?

2. Does Data Publication Work?

In the years 2005 to 2016 roughly 30 million STM papers were published [9]. The volume of formal data publications through DataCite for the time period 2005 to 2016 is approximately 2.6 million data publications [10]. Comparing 30 million STM publications to three million data publications shows us that data sharing through research data repositories is still not the norm (e.g. [11-12]). Looking at the comparatively low numbers of data publications we have to ask the question: are we offering the right incentives to researchers to share their data with others more freely?

3. The Social Drivers

Data publication means sharing data with an anonymous data user. Surveys among researchers show that citation as a form of recognition is very important but it might not sufficient. Ethnographic observation among groups of researchers reported by Wallis et al. [13] are interpreted by the authors to support Hagstrom's [14] hypothesis that scholarship is characterised by a gift culture in which members

of the community make each other precious gifts. Putting data on the internet for anonymous users without being able to expect a gift in return is not an incentive in this model of scholarly culture as this violates the principle of reciprocity that is fundamental to the gift culture.

The interpretation of science as a gift culture has been disputed by Latour and Woolgar [15] who argue that science actually has a currency they identify as credibility. This currency can be transformed into other forms of capital which they identify as money, data, prestige, credentials, problem areas, argument, papers, and so on. The system of credibility and reward in the scholarly community has been described by Fecher et al. [6] as a reputation economy.

4. Conclusions

To bring about the changes in social norms around data we need to take into account the social norms of the reputation economy, and the balance between reputation gain and collaboration gain. Understanding these dynamics will allow us to identify suitable points of intervention that will influence behaviours. Research data policies need to recognise that from a researcher's perspective data are a form of social capital he or she will strategically invest in the reputation economy that characterises the scholarly community.

Acknowledgments. The author would like to thank Agi Gedeon, Peter Fox, Heidi Laine, Fiona Murphy, Simon Cox, Benedikt Fecher and Jillian Wallis for our discussions on social norms around data in the scholarly community.

References

1. Mundt, M., Der DOI (digital object identifier) ein verlagsorientiertes Indexierungswerkzeug auch anwendbar auf Datensätze?. *Fachhochschule Potsdam, Potsdam, Germany,* 1998 (in German) Available: http://doi.org/10.2312/GFZ.misc.370184
2. Klump, J. *et al.*, Data publication in the Open Access Initiative. *Data Sci. J.*, 5, 79–83, 2006
3. Nelson, B., Data sharing: Empty archives. *Nature*, 461, 7261, 160–163, 2009
4. Borgman, C. L., The Conundrum of Sharing Research Data. *J. Am. Soc. Inf. Sci. Technol.*, 63, 6, 1059–1078, 2012
5. Costello, M. J., Motivating Online Publication of Data. *BioScience*, 59, 5, 418–427, 2009
6. Fecher, B., Friesike, S., Hebing, M., What Drives Academic Data Sharing? *PLoS ONE*, 10, 2, e0118053, 2015
7. Kratz, J. E., Strasser, C., Researcher Perspectives on Publication and Peer Review of Data. *PLoS ONE*, 10, 2, e0117619, 2015
8. Van den Eynden, V., Bishop, L., Sowing the seed: Incentives and Motivations for Sharing Research Data, a researcher's perspective. *Knowledge Exchange, Colchester, Essex, UK,* 2014
9. Ware, M., Mabe, M., The STM Report - Celebrating the 350th anniversary of journal publishing. *International Association of Scientific, Technical and Medical Publishers, The Hague, The Netherlands,* 2015
10. THOR Project, THOR Dashboard. *THOR Dashboard*, 2016 [Online]. Available: http://dashboard.project-thor.eu/dashboard/. [Accessed: 04-Nov-2016]
11. Baronchelli, A., Felici, M., Loreto, V., Catiglioti, E., Steels, L., Sharp transition towards shared vocabularies in multi-agent systems. *J. Stat. Mech. Theory Exp.*, 2006, 6, P06014, 2006
12. Nature editorial, Share alike, *Nature*, 507, 7491, 140–140, 2014
13. Wallis, J. C., Rolando, E., Borgman, C. L., If We Share Data, Will Anyone Use Them? Data Sharing and Reuse in the Long Tail of Science and Technology. *PLoS ONE*, 8, 7, e67332, 2013
14. Hagstrom, W. O., Gift giving as an organising principle in science. in *Science in Context: Readings in the Sociology of Science*, Barnes, B., Edge, D., (Eds.) Milton Keynes, United Kingdom: The Open University Press, 21–34, 1982
15. Latour, B., Woolgar, S., The cycle of credibility. in *Science in Context: Readings in the Sociology of Science*, Barnes, B., Edge, D., (Eds.) Milton Keynes, United Kingdom: The Open University Press, 35–43, 1982

Development of Research Data Management Service for Open Science in Japan

Yusuke Komiyama[1*]

[1*] *National Institute of Informatics, Research Organization of Information and Systems, 2-1-2 Hitotsubashi, Chiyoda Ward, Tokyo 101-8430, Japan*
Email: komiyama@nii.ac.jp

Summary. National Institute of Informatics (NII) will provide three research infrastructures for publication, discovery, and management to promote open science in Japan. In fact, to introduce to universities and research institutions there are barriers such as classical computer systems, research conventions, and laws, organizational governance. We are developing a data infrastructure for open science that combines organizational issues with the researchers' usability. In this research, we report on the system development and trial results of GakuNin RDM which is the research data management service.

Keywords. Research Data Management Service, RDM, Open Science, research integrity, data platform.

1. Introduction

National Institute of Informatics in Research Organization of Information and Systems (NII, ROIS) develops three types of research data infrastructure for open science in Japanese academia. They are research data publication, discovery, and management services. This report will describe the research and development of the data management infrastructure.

We propose the function of research data management for governance management to chief information officer (CIO) to extend the open science infrastructure to Japanese academia. For example, it is that RDM has technology such as research trail and data encryption. In addition to improving researcher's workflow in RDM development, we also need to consider the merits of organization managers at the same time.

So, we will provide a new research data management (RDM) service, and named it GakuNin RDM.

2. System Development

2.1 Core System

We adopted Open Science Framework (OSF) as base system, the open source software developed by Center for Open Science (COS) in the US as the core system of RDM [1]. However, since the original OSF is not compliant with classical research practices of Japan, the Japanese version needs to be customized. It also extends management functions for IT administrators in academic institutions.

2.2 High-speed network and Cloud connection

We are preparing to connect service and cloud provider with L2VPN with SINET which is 100 Gbps high-speed network for science in Japan [2]. SINET has more than 800 usage agencies and can connect systems with more than 20 cloud providers. GakuNin RDM is built on this network.

2.3 Academic Access Management Federation

GakuNin RDM is compatible with Academic Access Management Federation of Japan; it is called GakuNin [3]. Users of universities and research institutions will be able to log-in to the system with the ID of their institutions. Also, the user can log-in to the service provider (SP) corresponding to GakuNin federation with single sign-on.

2.4 Cloud Storage Add-on

We have developed two cloud object storage plugins used in Japanese universities. One is the public Cloud Azure Blob Storage, and the private Cloud is Open Stack Swift (API v 2). These were

not implemented in the original OSF, so we feedback to COS on a pull request.

2.5 Research Data Repository Add-on

On the other hand, we also developed a plugin for the data repository for university libraries and research libraries. The system can send data from GakuNin RDM to WEKO2 which is a famous Japanese repository software.

2.6 Data Analysis Add-on

Also, we have implemented the ability to send data directly from the RDM to the data analysis platform to make the research workflow comfortable. We have adopted JupyterHub as a data analysis platform and currently only Python, it can be extended to various programming languages [4].

3. Trial and Use-case Reports

3.1 Trial for university IT centers

We conducted the first closed alpha test of GakuNin RDM in February 2016. Especially it was an experiment for faculty of six universities IT center and one IT department of a research institute. The number of participating faculties was about ten people. This experiment was conducted specifically to ask usability evaluation and request. As a review of GakuNin RDM, faculties at IT center pointed out that add-on control for each organization, management of users, a usability of a user interface, management of research trails and logs are insufficient.

To respond to these challenges, we are currently developing functions for institutional IT administrators.

3.2 Trial for laboratory in wide disciplines

Furthermore, we conducted the second closed alpha test of GakuNin RDM in October 2017. It is under experiment at the stage when the authors are writing this manuscript. The purpose of this experiment was to know the use cases when researchers use them in the laboratory and to receive reviews from them.

4. Conclusions

We have developed a prototype of research data management service GakuNin RDM for open science in Japanese academia. It is a nationwide RDM SaaS to promote super interdisciplinary research and experimental reproducibility by data-driven science, it promotes open science. The new service will be launched in 2020 as a business operation of NII.

In the first closed alpha test in fiscal 2016, faculty members of IT center of six universities and one research institution evaluated the function of the service. IT experts requested the management function to the service for administrative staffs in the university, and make the guideline of Cloud utilization. In the second Closed alpha test in fiscal 2017, researcher of five universities and one institution reviewed the service. At the second Closed alpha test, we investigated collaborative research method in various disciplines and cooperation with external systems such as research instruments or analytical software.

We have expanded the functions for institutional IT administrators to GakuNin RDM so far. In the future, we plan to enrich modules for RDM services for diverse academic fields in response to requests from researchers.

References

1. Foster, E. D., et al., Open Science Framework (OSF). *J. Med. Libr. Assoc.*, 105(2), 38, 2017
2. Kurimoto, T., et al., A fully meshed backbone network for data-intensive sciences and SDN services. *in 2016 Eighth International Conference on Ubiquitous and Future Networks (ICUFN)*, 909–911, 2016
3. Yamaji, K., et al., Attribute Aggregating System for Shibboleth Based Access Management Federation. *in 2010 10th IEEE/IPSJ International Symposium on Applications and the Internet*, 281–284, 2010
4. Grüning, B. A., et al., Jupyter and Galaxy: Easing entry barriers into complex data analyses for biomedical researchers. *PLoS Comput. Biol.*, 13(5), e1005425, 2017

Emerging domain agnostic functionalities on the handle-centered networks

Kei Kurakawa[1*], ***Takayuki Sekiya***[2], ***Yasumasa Baba***[3]

[1*] *National Institute of Informatics, 2-1-2 Hitotsubashi, Chiyoda, Tokyo, 101-8430, Japan*
[2] *The University of Tokyo, 3-8-1 Komaba, Meguro, Tokyo, 153-8902, Japan*
[3] *The Institute of Statistical Mathematics, 10-3 Midoricho, Tachikawa, Tokyo, 190-8562, Japan*
Email: kurakawa@nii.ac.jp

Summary. Core part of RDA WGs and IGs aims at establishing discovery and automatic processing way of data. This work illustrates automatic data processing framework in detail and implies an emerging domain independent functionalities on the framework.

Keywords. Digital Object, Handle, Automatic data processing, Kernel information, Domain independent functionalities.

1. Introduction

The infrastructure for scientists to share the scientific resources such as journal articles, conference proceedings, books, software, data, and any other available online resources has been developed over several decades. Recent trend of Open Access movement since 2000s reformulate the scientific community's mind so as to make the scientific resources open as default to be freely accessible by everyone. The scientific community's mind of Open Access has influenced on the divergence range of academic resources to share, as a result the movement reached at the slogan "research data sharing without barriers" of RDA (Research Data Alliance) among all disciplines.

Our standard procedures, which may be peculiar to each discipline, to aggregate and process the scientific data are a process of craftsmanship and a too much time consuming task compared to academic rewards, in that to deal with scientific data the data consumer needs to understand the semantics of data structure in domain dependent schemes and choose ordinarily a community standard of tools on a specific computational environment to process the data, which seems to be difficult to do the same things by outsiders of the expertise. The whole kind of the things must be easy and automatic even for the experts, needless to say for all who need the data.

Core part of RDA working and interest groups (WGs and IGs) concentrate on the issues and tackle with the PID (persistent identifiers) centric information model to invest the domain-independent automatic processing environment for very large and heterogeneous collections of distributed scientific data. This paper introduces the relevant information model produced by the WGs and IGs, and considers the emerging domain independent functionalities, which can be derived from the information model.

2. Data discovery and automatic data processing

To think of a next generation of research data infrastructure we need to analyze research data processing workflows of researchers. The major requirements are two kinds as follows.

One is to enhance discoverability of scientific research data that are explicitly separated from general resources on the Web. RDA Data Discovery IG focuses on the discoverable search interfaces and technical specific metadata for the scientific research data. For geophysical science a typical user interface has a browsing feature with longitude and latitude settings on geographical coordinate, which requires a metadata

enhancement from a common Dublin Core metadata scheme. Domain specific metadata schemes for a certain kind of disciplines may deductively produce available interfaces for search, but they are difficult to unify among a variety of scientific domains.

The other is to build automatic processing environment for scientific research data. Several groups of RDA (Data Fabric, Data Type Registries, PID Information Types, and PID Kernel Information) focus on this vision, especially in domain independent way, and discuss what the necessary information framework is for the future research environment.

3. Automatic data processing framework

Automatic data processing is an ideal and urgent way of function to deal with distributed large amount of data. The followings are the technological views to implement the framework.

3.1 Digital object and data types

Persistent identifiers (PIDs) as known as handles is a key to point out permanently a specific dataset on the network. The Handle Server is the implementation based on Digital Object Architecture [1] also known as Kahn-Wilensky Framework issued in 1995, which have influenced several digital library architecture development for a number of decades [2]. A Digital Object (DO) consists of data and key-metadata. The key-metadata includes a globally unique PID namely a handle and other metadata. In the architecture, the data of each digital object is associated with a type stored in a global data type registry.

3.2 Versioning and provenance information

To state the quality and context of data in general, versioning and provenance information used to be embedded in the metadata. Provenance information gives rich context of data such as quality, audit trail, replication, attribution, etc [3].

3.3 Domain dependent and independent metadata layers

Metadata consists of a variety of attributes pertaining to the data. Attributes of data are defined and cross-related in different granularity of concept, in different levels of domain-specificity, and in different aspects. Domain independent attributes are collected in Kernel Information profile as structural metadata [4].

3.4 Handle-centered networks

All kinds of things in this framework, e.g., data, types, metadata are embedded with a handle as a pointer. Metadata describes relationships among them by attributes to a handle. The collection of metadata produces a handle-centered network.

4. Domain independent functionalities on the framework

The handle-centered networks can be source for analytics to produce valuable knowledge, i.e.,

- Analysis of data
- Classification of data
- Recommendation of data
- Prediction of data.

5. Conclusions

This work illustrated the automatic data processing framework discussed in RDA community, and implied emerging domain independent functionalities on the framework.

Acknowledgments. This work is supported by the open collaborative research at National Institute of Informatics (NII) Japan (FY2017). The authors are thankful to all RDA Kernel Information WG members for their great discussions on remotely and in-person meetings.

References

1. Kahn, R., Wilensky, R., A Framework for Distributed Digital Object Services, doi: cnri.dlib/tn95-01, 1995
2. Nelson, M. L., Sompel, H. Van de, IJDL special issue on complex digital objects: Guest editors' introduction. *Int. J. Digit. Libr.,* 6,113-114, doi: 10.1007/s00799-005-0127-y, 2006
3. Simmhan, Y. L., Plale, B., Gannon, D., A survey of data provenance in e-science. *ACM SIGMOD Record,* 34(3), 31, doi:10.1145/1084805.1084812, 2005
4. RDA KI WG, Strawman PID Kernel Information Profile 17.04.05., http://bit.ly/2oH53XC, [accessed on October, 2017]

FAIRsharing – Describing and Connecting Standards, Databases and Policies Across Disciplines

***Peter McQuilton*[1]*, *Philippe Rocca-Serra*[1], *Alejandra Gonzalez-Beltran*[1], *Milo Thurston*[1], *Massimiliano Izzo*[1], *Allyson Lister*[1], *Delphine Dauga*[2], *Melanie Adekale*[3] *and Susanna-Assunta Sansone*[1]**

[1]* *Oxford e-Research Centre, Department of Engineering, University of Oxford, 7 Keble Road, Oxford, OX1 3QG, United Kingdom*
[2] *Bioself communication, 28 rue de la bibliothèque 13001, Marseille, France*
3 MD Health Consulting Ltd, 95 Bilton Road, Rugby, CV22 7AS, United Kingdom
Email: peter.mcquilton@oerc.ox.ac.uk

Summary. FAIRsharing (https://www.fairsharing.org) is a cross-discipline, manually curated, searchable portal of three linked registries, covering standards, databases and data policies. FAIRsharing maps the landscape and serves it up in a structured, human and machine-readable fashion, through a web front-end, API and embeddable visualisation tools. FAIRsharing covers 4 standard types at present – Reporting Guidelines/Checklists, Terminology artefacts/ontologies, Models/Formats, and identification schema, and links these to the databases that implement them, and the policies that endorse their use. FAIRsharing houses information on databases and data repositories captured from the literature, internet and through the community. The policy registry contains data policies curated from journal publishers, funders, organisations and societies. FAIRsharing is part of ELIXIR, the pan-European infrastructure programme, and works with a number of funders, journal publishers, data managers and researchers to ensure databases, standards and policies are Findable, Accessible, Interoperable and Re-usable, through the promotion of the FAIR principles.

Keywords. Metadata, standard, database, data policy, FAIR.

1. Introduction

FAIRsharing maps the ever-growing landscape of data repositories, databases, metadata standards and policies across disciplines. The growth of data science, across the life sciences, physics, chemistry, humanities and social science, along with a growing movement for reproducible and interoperable research, has led to a proliferation of data resources, and with that, the standards that they use. As the cost of data production and storage has fallen, the number of databases that can host the data has increased, leading to confusion as to which database to consult or where one should deposit their data. We have developed FAIRsharing (https://fairsharing.org), a curated, informative and educational resource with over 1700 records, to map this ever-changing landscape. Launched in 2011 as BioSharing [1], to reflect the initial focus on the life sciences, FAIRsharing is maintained as a community resource closely embedded in and co-sponsored by several infrastructure programmes, such as the ELIXIR EXCELERATE and Bioschema projects, or the NIH FAIR-metrics initiative. To encourage community interaction, and to ensure FAIRsharing truly reflects a need within the community, we have worked closely with a variety of stakeholders via a joint Research Data Alliance/Force11 working group which has recently created a set of recommendations for linking and describing standards, databases and policies [2].

2. Mapping the landscape

In the last 20 years, the amount and range of experimental data has increased almost exponentially. Concomitant with this, there are now a plethora of resources available, where a researcher can either deposit, obtain, or simply search data. This wealth of data resources has led

to an equally bewildering array of standards. FAIRsharing captures information on 4 types of standard – models/formats, terminology artefacts, reporting guidelines and identifier schema. Models/Formats captures information on schema or exchange formats, such as the FASTA sequence format [3] used for biological sequences. Terminology artefacts, such as thesauri or ontologies, allow data to be structured in a hierarchical fashion, such that users in multiple languages and disciplines can unambiguously refer to the same 'thing' at the same time. An example is the AGROVOC agricultural vocabulary [4] from the Food and Agriculture Organization (FAO) from the United Nations, which has over 32,000 terms and enables structured annotation over the whole of agricultural science. FAIRsharing currently houses 117 Reporting Guidelines or checklists, such as the ARRIVE guidelines [5] for in-vivo animal testing. As FAIRsharing has evolved from BioSharing, which focused on the life sciences, our database registry is heavily biased towards the life and biomedical sciences.

3. Claim your resource

FAIRsharing is a community-driven collaborative project. Every record can be claimed by the maintainer of the resource concerned. Once a record is claimed, it can be edited and updated at will. All records on FAIRsharing are also manually curated by a FAIRsharing knowledge engineer. If a record is updated by a maintainer, the FAIRsharing engineers are sent a notification, and vice versa. This cross-validation ensures the data therein is as accurate and timely as possible. If a resource is missing, a user can create an account and add the resource themselves. The record for that resource will be visible but not approved until a FAIRsharing engineer has confirmed that the record adheres to our curation guidelines.

4. Collating and connecting the dots

Resources can be grouped on FAIRsharing by project, domain, organisation or if they are recommended for use by a data policy. Each collection or recommendation is claimed by an individual or group related to the organisation. An example is the PLOS recommendation (https://fairsharing.org/recommendation/PLOS), based on the PLOS data policy. Each collection or recommendation can be viewed as a table or interactive graph network. Like every search on FAIRsharing, these can be refined further using our domains, disciplines and other filter matrices. Collections and recommendations are also available as embeddable widgets, enabling them to be accessed from 3rd party websites, such as those from a journal publisher.

5. Conclusions

FAIRsharing reflects a community need for a resource that collates and links databases, data standards and data policies, so mapping the landscape of interconnecting resources. FAIRsharing is also interlinked further with other resources in the ecosystem, such as the ELIXIR TeSS training events portal [6] and the Japanese National Bioscience Database Center's Integbio portal [7].

References

1. McQuilton, P., Gonzalez-Beltran, A., Rocca-Serra, P., Thurston, M., Lister, A., Maguire, E., Sansone, S-A., BioSharing: curated and crowd-sourced metadata standards, databases and data policies in the life sciences. *Database* , baw075, 2016
2. RDA/Force11 BioSharing Recommendation, http://dx.doi.org/10.15497/RDA00017
3. Lipman, D. K., Pearson, W. R., Rapid and sensitive protein similarity searches. *Science,* 227(4693),1435-41, 1985
4. Caracciolo, C., Stellato, A., Morshed, A., Johannsen, G., Rajbhandari, S., Jagues, Y., Keizer, J., The AGROVOC Linked Dataset. *Semantic Web Journal,* 4(3), 341-348, 2013
5. Kilkenny, C., Browne, W. J., Cuthill, I. C., Emerson, M., Altman, D. G., Improving bioscience research reporting: the ARRIVE guidelines for reporting animal research. *PLOS Biology,* 8(6), e1000412, 2010
6. TeSS – ELIXIR Life Science Training and Events portal, http://tess.elixir-europe.org/
7. Integbio life science database portal, http://integbio.jp

"Polar Data Journal"; A new data publishing platform for polar science

Yasuyuki Minamiyama[1]*, Akira Kadokura[1,2], Masaki Kanao[1,2], Takeshi Terui[1], Hironori Yabuki[1,2] , Kazutsuna Yamaji[3]

[1] Research Organization of Information and Systems, National Institute of Polar Research, 10-3 Midori-cho, Tachikawa, Tokyo, 190-8518, Japan*
[2] Research Organization of Information and systems, Joint Support-Center for Data Science Research, Polar Environment Data Science Center, 10-3 Midori-cho, Tachikawa, Tokyo, 190-8518 , Japan
[2] Research Organization of Information and systems, National Institute of Informatics, 12-1-2 Hitotsubashi, Chiyoda-ku, Tokyo, 101-8430, Japan
Email: minamiyama@nipr.ac.jp

Summary. Polar Data Journal (ISSN 2432-6771) is a free-access, peer-reviewed and online journal. It is dedicated for publishing original research data/dataset, furthering the reuse of high-quality data and the benefit to polar sciences. Polar Data Journal aims to cover broad range of research disciplines involving Arctic, Antarctic, or other polar regions, especially earth and life sciences. The Journal primarily publishes data papers, that provides detailed descriptions of research data/dataset (e.g. Methods, Data Records, Technical validation). The Journal does not require any new scientific findings, so the Journal also welcomes submissions describing past valuable data/dataset which has not published yet. In order to ensure data quality, Polar Data Journal requires to be passed our peer-review process. Before submitting your manuscript, authors should deposit their data/dataset to trustworthy data repository. Data authenticity is also guaranteed by publishing report of all review process, which will be published with author's manuscript at the same time.

Keywords. data journal, data publication, open science, JAIRO Cloud, polar science.

The National Institute of Polar Research (NIPR) has launched Polar Data Journal, a new data journal in January, 2017. Polar Data Journal is a free-access and peer-reviewed online journal. It is dedicated to publishing original research data/datasets, furthering the reuse of high-quality data for the benefit to polar sciences.

Polar Data Journal aims to cover a broad range of research disciplines involving polar regions, especially the earth sciences and life sciences domain. The journal primarily publishes data papers, which provide detailed descriptions of research data/datasets (e.g., Methods, Data Records, and Technical Validation). It is not required that the data papers published in this journal depict any new scientific findings; hence, the journal also welcomes submissions describing valuable existing data/datasets that have not been published to date. Some key features of the new journal are as follows:

- Polar Data Journal is a peer-reviewed journal that aims to provide high-quality data to researchers.
- It is a free-access journal.
- Polar Data Journal is thoroughly edited using an online editing system for quick publishing.
- The journal content is reviewed by an editing committee, which will disclose the reviewer's reports in each article of a volume.

The platform of Polar Data Journal is powered by WEKO (JAIRO Cloud), which is developed and operated by the National Institute of Informatics (NII), Japan.

For more information, please visit https://pdr.repo.nii.ac.jp.

Strengthening International Data Sharing Networks

Mustapha Mokrane[1]*, ***Sanna Sorvari***[2], ***Andrew Treloar***[3], ***Mark Parsons***[4], ***Carthage Smith***[5]

[1]* *ICSU World Data System International Programme Office, Tokyo 184-8795 Japan*
[2] *Finnish Meteorological Institute, 00101 Helsinki, Finland*
[3] *Australian National Data Service, Monash University, Melbourne, VIC 3145, Australia*
[4] *Rensselaer Polytechnic Institute, Troy, United States*
[5] *OECD Global Science Forum, 75775 Paris, France*
Email: Mustapha.mokrane@icsu-wds.org

Summary. International data sharing networks enable the dissemination of data within and between scientific disciplines and countries and thus provide the foundation for Open Science. Developing effective and sustainable international research data networks is critical for progress in many areas of research and for science to address complex global societal challenges. However, the development and maintenance of effective networks is not always easy, particularly in a context where public resources for science are limited and international cooperation is not a priority for many countries.

Keywords. Data Sharing, International Coordination, Data Preservation.

1. Introduction

The global landscape for data sharing is complex and includes many international data networks with highly variable structures. Some are linked to large intergovernmental research infrastructures with centralized services covering single disciplines. Some are distributed with less rigid governance structures and cover different domains. Most are somewhere in-between and cover different geographic regions. All provide a mix of data and associated data services.

What makes a network function effectively or not is unclear which means that there is no simple answer to what can usefully be done at the policy level to promote the development of effective and sustainable data networks. The present study analysed a variety of successful networks, explored the challenges they are facing and the lessons learned, and, where applicable, to translate this into potential policy actions.

2. Methodology

Detailed descriptive, operational and reflective information was collected on a total of 32 international data networks including several in the geosciences domain. The information gathering was done with a general survey and structured in-depth interviews. The overall findings were considered in a 2-day international workshop.

3. Challenges and Solutions

Several common challenges and potential solutions were identified. Many are related to policy mandates and incentives (including funding) that are beyond the control of network participants. These mandates and incentives will be affected by a greater policy-makers' understanding of the critical role international data networks play as a foundation for open science. This includes appreciation of what makes for a successful network and conversely what is likely to lead to failure. There is no simple one-size-fits-all answer, but effective actions to enable the successful functioning of international data networks can be taken at several scales and by a variety of actors.

3.1 The importance of policy

The main barrier to open sharing of research data across borders and scientific domains is the lack of policy coherence and trust between different communities. This is manifest in different interpretations of openness, different legal regimes for data sharing, and different ethical perspectives. Such differences need to be respected and understood but should not prevent a mutual understanding and workable international agreements being reached around the sharing of public research data. There are many successful examples of international sharing of data, including 'sensitive data', and the lessons from these cases are there to be built on.

3.2 Data networks as critical infrastructure

Several international data networks have been in existence for many decades and have become an invisible part of the infrastructure of science. At the same time, there is a massive increase in supply and demand for research and data networks need to be up-dated or redeveloped and maintained. These networks are the business-critical point of weakness in many research areas and for open science as whole.

3.3 Building successful networks

Developing and maintaining a successful international data network is dependent on several factors, both technical and social. Individual networks need to be tailored to the needs of specific data providers and users and they also need to evolve over time. They require an appropriate mix of long-term commitment, consistency and flexibility.

3.4 Funding and long-term sustainability

The current funding for international data networks is inadequate with increasing demand for data curation and data repositories struggling for support in many countries. There are additional funding and sustainability challenges for international data networks, many of which have no obvious sponsor for their critical central coordination functions. Often there is a considerable amount of un-costed 'in kind' support involved in networking and the amount of extra funding required for coordination is small but this should not mean that it is neglected since it can be critical to the network's success. There can be considerable overall cost benefits from federation of activities. At the same time, many networks do not currently have clearly articulated value propositions that can be used to justify additional investments.

4. Conclusion and recommendations

1. Responsible national authorities should be identified and work toward common definitions of, and agreements on, open data.
2. Governments need to work toward commonly agreed and enforced legal and ethical frameworks for the sharing of diverse types of public research data.
3. All stakeholders need to recognize international research data networks as a critical part of the generic infrastructure for open science.
4. Responsible national and international authorities must include data networks in long-term strategic planning and support processes for research infrastructure.
5. In establishing, developing, operating, and supporting international data networks critical 'organisational' aspects should be considered: user and data provider community, global landscape, governance models, roles and responsibilities across the network, level of standardization, assessment of the network, regional differences.
6. Funders and host institutions should view internationally coordinated data networks as long-term strategic investment and support and engage with them accordingly.
7. Networks should have clear business models, including value propositions and monitored measures of success that are relevant to their different stakeholders.
8. Funders should actively participate in relevant international discussions and forums to improve long-term functioning, support and coordination of data networks.

Toward An Open Science Ecosystem Including Sharing, Citing and Publishing Research Data

Yasuhiro Murayama [1*]

[1*] *National Institute of Information and Communications Technology, 4-2-1 Nukui-Kita, Koganei, Tokyo, 184-8795, Japan*
Email: Murayama@nict.go.jp

Summary. Open Science has increasingly become an important focus of international discussions for its possible big impacts on not only the way how scientific research is to be conducted, but also future development of the general society and national/regional economies. This international trend more and more includes a wider variety of stakeholders, researchers/research institutions, librarians, data managers, funders, publishers, governmental decision makers and so on. The intended change in the way of how to conduct scientific research is expected to be a big change of culture, norm, and legislative aspects of many of stakeholders involved in this Open Science development. This paper will attempt an overview of such possible change, and involved stakeholders, and what the society and scientists need to look at for inclusive welfare of our future.

Keywords. Open Science, Research Data Management, open research data, science policy.

1. Introduction

Open Science has increasingly become an important focus of international discussions for its possible impacts on not only the way how scientific research is to be conducted, but also future development of the general society and national/regional economies through digitization of the society. This international trend more and more includes a wider variety of stakeholders, researchers/research institutions, librarians, data managers, funders, publishers, governmental decision makers and so on. The intended change in scientific research ecosystem requires change of culture, norm, incentives/reward of research activities and legislative aspects, of many of stakeholders involved in this Open Science development. This paper will attempt an overview of such possible change, and involved stakeholders, and what the society and scientists need to look at for inclusive welfare of the future scientific research ecosystem.

2. Print and Electronic Technologies

In addition to the knowledge basis accumulated on printed records (books, articles) which has lasted for more than 300 years, dramatic advancement is expected to occur on the basis of digital technology and electronic information and communication infrastructure. Although the digital ICT technology has only an approx. 70-year history since the birth of its technical basis, our society has received its big impacts in past by its epoch-making technology innovations.

On the other hand, while it is becoming difficult to make bigger societal impacts by new engineering inventions of technology, a coupled social-technical system is increasing its importance and effectives in the society. Recent development of new concepts like the Internet of Things (IoT) is foreseeing a data rich and more useful and convenient daily life, industry, and society, based on electronically digitization of necessary information.

3. Information Asset

"Information asset"—or data files/bit streams—on Internet have a wide variety in size, contents,

velocity, and are extremely heterogeneous. At present a data and software system often requires specific design and specific coding for specific objective and application.

Electronic data in a context of Open Science is a relatively "static" information asset, and in particular "research data", which is not limited to academic research data but also includes any text and/or non-text highly-valued information asset which is produced by expert and/or professional works. Accessibility, sharing, interoperability of data and infrastructure, citeability, and reuse are important prerequisites. Regardless of the open or closed policy, the capacity to assess trustworthiness of datasets, to preserve and manage them in an organized way, and to enable professional and non-professional reuse to create new knowledge are expected to play a critical and essential role in Open Science research ecosystem and digitized society and economy.

4. Long Term Preservation

Long-term preservation of data raises questions such as the size of data we should preserve and storage capacity, the preservation time (50-100 years similar to academic articles?), the increasing costs of bigger size of data, and so on. In the scientific research data area, an international enterprise, the World Data Centres, was established in 1957-58 by the International Council for Science (ICSU) to exchange and store data that is important to the scientific community as books and microfilms in a machine readable form. With the unprecedented technical infrastructure available today to exchange electronic data over the world and the need for multidisciplinary data integration to solve the most pressing challenges facing humanity, ICSU decided to form the new ICSU-World Data System (WDS) in 2008 based on the strong legacy of ICSU's two data organizations (WDC and FAGS). The International Programme Office of WDS is now hosted by NICT in Japan, Tokyo. WDS works with its member organizations—many of them are data holders and data providers—to secure trustworthy, sustainable and findable data archives.

5. Coupled Technical & Social System

Libraries have been selecting books, improving their preservation, building the international network of information flows (such as International Inter-Library Loan program). We are now on the starting line to construct a similar infrastructure for electronic data resources where the optimized mixture and coupling of engineering and social technologies (when the word "technology" is defined more widely as "practical application of any scientific and scholarly knowledge") is indispensable to achieve our goal. These efforts to support and promote best practices, will lead to building a new layer on top of the current software/computer platform layers of the Internet and data processing systems. This new layer will be an essential component to enable wider and flexible reuse of datasets with greater interoperability, with much reduced cost burden (monetary, human, temporal costs etc.) on the society to achieve the same functions/services. Experiences and best practices of WDC and WDS of building and operating the real data-holders' community over the past 60 years will be invaluable to help this advancement and design tasks of future research ecosystem and Open Science-based scholarly community models.

Support Project for Data Fusion Computation: Current status and future prospects

Shin'ya Nakano[1,2]*

[1]* *The Institute of Statistical Mathematics, ROIS, Midoricho 10-3, Tachikawa, Tokyo, 190-8562, Japan*
[2] *School of Multidisciplinary Science, SOKENDAI, Hayama, Kanagawa, 240-0193, Japan*
Email: shiny@ism.ac.jp

Summary. Support Project for Data Fusion Computation (SPDFC) is a project aiming at providing our knowledge of novel statistical techniques for simulation researchers. Some statistical approaches are useful for enhancing the effectiveness of numerical simulation. One example is data assimilation which estimates a scenario of temporal evolution by incorporating a sequence of the observational data into a numerical simulation model. Another example is a statistical emulator which imitates a simulation model by a statistical model. This paper briefly explains these statistical techniques useful for simulation researches.

Keywords. data assimilation, statistical emulator, numerical simulation.

1. Introduction

Numerical simulation is widely used in variety of fields such as weather prediction, fishery prediction, aeroplane designing, building designing, and so on. Recent simulation models can compute the evolution of a system with high temporal and spatial resolution. Some of the simulation models achieves so high accuracy that a simulation result is hard to discriminate from a real phenomenon.

Numerical simulation just reproduces a scenario of temporal evolution of a system under given conditions. Thus, when conducting numerical simulation, it is crucial to give appropriate conditions such as initial conditions and boundary conditions. Statistical approaches are useful for appropriately setting the conditions to be given for simulation. However, simulation researchers are not necessarily familiar with statistical approaches which would be useful for enhancing the effectiveness of their simulation models. In order to promote application researches of statistical techniques for enhancing effectiveness of numerical simulation, Research Institute of Information and Systems (ROIS) has launched Support Project for Data Fusion Computation (SPDFC) as one of working groups under Joint Support-Center for Data Science Research in ROIS. Under this project, we provide the knowledge of novel statistical techniques for simulation researchers at request. We also give seminars and hands-on for a wide range of simulation researchers. In addition, we are developing statistical software for implementing the statistical techniques to various simulation programmes.

This paper explains key statistical techniques for enhancing simulation researches, which are promoted under this project. The activities for promoting such statistical techniques are also introduced.

2. Statistics for simulation

There are two important statistical techniques for enhancing numerical simulation: data assimilation and statistical emulation. In this section, we explain each of the two techniques.

2.1 Data assimilation

In order to accurately reproduce temporal evolution of a real phenomenon by a simulation, it is essential to give an initial condition and boundary condition which well correspond to the real situation. In most cases, the observational data are limited and we can observe only a small fraction of the entire initial state and boundary state. However, a time series of the observational

data enables us to constrain the initial and boundary states more effectively. In addition, a simulation model can also be used as a effective constraints about the temporal evolution of the system because the simulation model is composed on the basis of our knowledge on the physical laws governing the system.

Data assimilation [1] estimates a scenario of temporal evolution by incorporating a sequence of the observational data into a numerical simulation model. Data assimilation not only utilise the information of the observational data taken at various times but also exploit the simulation model as a constraint for the estimation. Data assimilation has been developed as a fundamental technique for numerical weather prediction. Numerical weather prediction uses an atmospheric simulation model which describes physical processes in the atmosphere. The atmospheric system can be regarded as a deterministic but chaotic system. In order to successfully predict a future weather, it is crucial to accurately estimate the initial condition. Even in other various fields, it would be essential to obtain a good initial condition to predict evolution of a system. The applications of data assimilation techniques are widely expanding.

2.2 Statistical emulator

In engineering applications, numerical simulation is conducted for finding an optimal value of tuneable design parameters. When designing a complex system, it is difficult to guess the effect of the change of such tuneable parameters. In such cases, numerical simulation is helpful for evaluating the effect of tuneable parameters. However, in order to find an optimal value of tuneable parameters for a complex system, simulation must be run many times until the optimal value is found. It tends to take much time to obtain one scenario by running a simulation model of a complex system. Thus, it is sometimes practically impossible to run such a simulation model repeatedly until the optimal value is found.

In order to avoid to repeatedly run a computationally expensive simulation model, a statistical model imitating a simulation model is considered [2-3]. Such a statistical model is called a 'statistical emulator' or an 'emulator'. A statistical emulator is derived by analysing the relationship between the input and the output using a statistical technique. The statistical emulator is a potentially useful tool for designing a complex system.

3. Concluding remarks

We are now conducting several collaborative researches of applications of data assimilation techniques and statistical emulation techniques. These collaborative researches cover a variety of fields including material science and social science. In addition, we are developing software in order to facilitate simulation researchers in various fields who want to conduct data assimilation. A data assimilation problem can be formulated by a common framework called a state space model. Thus, the software is designed to be applicable to problems in various fields as far as interfaces for a simulation model and data access are prepared in a pre-defined manner. Data assimilation techniques tends to requires high computational cost. We are thus developing parallelised software which can be run on a supercomputer.

References

1. Kalnay, E., Atmospheric modeling, data assimilation and predictability. *Cambridge University Press, Cambridge, UK,* 2003
2. Kennedy, M. C., O'Hagan, A., Bayesian calibration of computer models. *J. Roy. Statist. Soc. Ser. B,* 63, 425-464, 2001
3. Rougier, J., Efficient emulators for multivariate deterministic functions. *J. Comp. Graph. Statist.,* 17, 827-843, 2008

Data Processing and Archive System for the Antarctic PANSY Radar

***Koji Nishimura*[1,2]*, *Masaki Tsutsumi*[2], *Yoshihiro Tomikawa*[2], *Toru Sato*[3], *Taishi Hashimoto*[3], *Masashi Kohma*[4], *Kaoru Sato*[4]**

[1] *Polar Environment Data Science Center, Tachikawa 190-8518, Japan*
[2] *National Institute of Polar Research, Tachikawa 190-8518, Japan*
[3] *Dept. Communications and Computer Engineering, Kyoto Univ., Kyoto 606-8501, Japan*
[4] *Dept. Earth and Planetary Science, The Univ. of Tokyo, Tokyo 113-0033, Japan*
Email: knish@nipr.ac.jp

Summary. Program of Antarctic Syowa MST/IS (PANSY) Radar is the first and the only mesosphere-stratosphere-troposphere (MST) and incoherent-scatter (IS) radar in the Antarctic region. It coninuously observes the atmosphere from about 1,500 m through 500 km above the Syowa Station (69S, 40E) as of 2012. Since the station is remote both in physical access and data communications, special treatments are needed for archiving and transporting the large amount of data that the radar yields. *The PANSY data archiving system* (PANDA) is the integrated data management system devoted to PANSY radar, equipped with realtime data processing, archive, transfer, quick-look display, and search functions.

Keywords. PANSY radar, MST/IS radar, Syowa Station.

1. PANSY Radar

PANSY Radar, which has been installed and set up in the Japanese Syowa Station through 2012 to 2016, is the first and currently the only MST/IS radar in the Antarctic [1]. This radar has a capability of measuring the neutral wind as from the altitude of 1.5 up to about 100 km, and the ionized atmosphere above that. Syowa Station is isolated not only in physical access that is almost solely maintained by the Japanese icebreaker ship, but also in communications that is mainly supported by Intelsat. PANSY radar is continuously producing a large amount of data but, reflecting the limited bandwidth, the data is limited we can transfer realtime to Japan.

In this presentation, we present the scheme and the system for data transfer, archive and distribution for PANSY radar.

2. *PANDA*, Data Management System

PANSY radar is in operation 24 hours and yielding data of roughly 3 MB/min (~4 GB/day) as time series. In order to reduce the occupancy in the communication bandwidth, however, we temporally integrate the power spectra into ones with a data rate of about 400 kB/min and transfer them realtime via Intelsat to the hub in National Institute of Polar Research, Japan, and then to other spokes.

The original time series data are hand-carried once a year by the ship. Until the transport, for a risk of accidental losses, the data are multiply stored in physically separate buildings in the station. As the storage system is so distributed with a satellite channel inbetween, coherent data handling is needed.

Thus we have developed the integrated data management system PANDA, with some peripheral functions such as a quick-look display and a health monitoring of the system. In this talk, we present our data management and operation based on the PANDA system.

References

1. Sato, K., Tsutsumi, M., Sato, T., Nakamura, T., Saito, A., Tomikawa, Y., Nishimura, K., Kohma, M., Yamagishi, H., Yamanouchi, T., Program of the Antarctic Syowa MST/IS radar (PANSY). *J. Atmospheric Sol.-Terr. Phys.*, 118, 2-15, 2014

Data Citation at World Data Center for Geomagnetism, Kyoto

Masahito Nosé[1]*

[1]* *Graduate School of Science, Kyoto University,*
Oiwake-cho, Kitashirakawa, Sakyo-ku, Kyoto, 606-8502, Japan
Email: nose@kugi.kyoto-u.ac.jp

Summary. World Data Center for Geomagnetism, Kyoto (WDC-Kyoto) is a data center that collects and archives geomagnetic field data obtained at a few hundreds of observatories worldwide as well as provides the data to users. Recognizing the importance of publication and citation of scientific data that provide a lot of benefits to users, data providers, and data centers, WDC-Kyoto has been working with other solar terrestrial physics (STP) data centers to mint DOI to their database since August 2013. Four DOIs are introduced for geomagnetic field data that are archived in WDC-Kyoto (the AE, Dst, Wp indices, and magnetotelluric data) in addition to 14 data DOIs for STP database in Japan. This makes it possible to cite the magnetotelluric data in an article that are recently published in Journal of Geophysical Research. It is expected that DOI is minted to more dataset and data citation becomes more common in near future.

Keywords. Digital object identifier, data publication, data citation, open science.

1. Introduction

The Japanese government has found an importance of "Open Science" and is now going to promote its associated activities in Japan. In the end of March 2015, a report entitled "Promoting Open Science in Japan" [1] was published by the expert panel on Open Science, based on Global Perspectives, Cabinet Office. According to the report, research data should be made openly available, although they are subject to constraints that ensure ethical, legal, and commercial protections. To accelerate data availability, it is needed to prepare data identifiers, such as digital object identifiers (DOIs), and to foster a practice of citation for research data. This is because the citation for research data provides the following benefits: (1) Readers can more easily locate the data used in the paper, obtain necessary information of the data (i.e., metadata), and validate the findings of the paper; (2) readers can also easily discover datasets which are relevant to their interests but have not been noticed; (3) data providers/data centers can put necessary information about data (i.e., metadata) on their landing pages, and reduce labor to respond to user's inquiries; and (4) data providers/data centers can gain professional recognition, proper credit, and rewards for their labors to publish and manage data set in the same way as for traditional publications.

2. DOI-minting to Solar-Terrestrial Physics Data

Recognizing the importance of data citation, in August 2013, World Data Center (WDC) for Geomagnetism, Kyoto started a discussion about how to mint DOI to their own database with WDC for Ionosphere and Space Weather (National Institute of Information and Communications Technology (NICT)), and Integrated Science Data System Research Laboratory (NICT). The discussion finds that Japan Link Center (JaLC) is a proper agency to register DOI-URL mapping, because JaLC aims at public information services to promote science and technology in Japan and it handles scientific and academic metadata and content from holders nationwide, including national institutes and universities. We developed a web-based system to register metadata with JaLC and to create landing pages of data, to which DOIs are mapped. The system can handle

version of the landing pages when the data are updated.

JaLC started a 1-year pilot program to mint DOI to the database from October 2014, which is followed by a regular operation. We have been working closely with JaLC, resulting in 18 DOIs for solar-terrestrial physics (STP) data that include the mesospheric wind velocity data observed with MF radar (doi:10.17591/55838dbd6c0ad), the geomagnetic Dst index (doi:10.17593/14515-74000), the geomagnetic AE index (doi: 10.17593/15031-54800), and the ionograms (e.g., doi:10.17594/567ce8e9d3a52), as of October 2017. Table 1 compiles the 18 DOIs that we created. As far as we know, this is the first practice of the DOI-minting to scientific data in Japan.

3. Data Citation in Journal Article

One of these DOIs, the magnetotelluric data at Muroto (doi:10.17593/13882-05900), was cited in a recently published article [2], providing the first example of data citation in Japan. This article appeared in Journal of Geophysical Research: Space Physics, which is an internationally recognized journal in the STP discipline. It is expected that data citation in journal articles becomes common in near future.

4. Conclusions

DOI-minting and data citation are beneficial to data providers and data centers. STP data centers in Japan have been working to mint DOI to their database since August 2013. There are now 18 data DOIs for STP database in Japan, including 4 DOIs for geomagnetic field data that are archived in WDC for Geomagnetism, Kyoto (the AE, Dst, Wp indices, and magnetotelluric data). Data citation of the magnetotelluric data was practiced in the internationally recognized journal. Similar data citation for the Dst, AE, and Wp indices will be expected in future journal articles.

Acknowledgments. The author would like to thank Y. Murayama, T. Kinoshita, Y. Koyama, M. Nishioka, M. Ishii, M. Kunitake, K. Imai, T. Iyemori, T. Watanabe, and T. Sagara for their help in the DOI-minting practices.

References

1. Cabinet Office, Government of Japan, http://www8.cao.go.jp/cstp/sonota/openscience/ [accessed on October 2017]
2. Nosé, M., Uyeshima, M., Kawai, J., Hase, H., Ionospheric Alfvén resonator observed at low-latitude ground station, Muroto. *J. Geophys. Res.*, 122, doi:10.1002/2017JA024204, 2017

Table 1. Created 18 DOIs for solar-terrestrial physics as of October 2017

Name of Database	DOI	Date of Minting
Profiles of neutral atmosphere winds 30min average with MF radar at Poker Flat, Alaska	10.17591/55838dbd6c0ad	2015/06/19
Dst Index	10.17593/14515-74000	2015/12/30
Ionogram at Kokubunji, Japan	10.17594/567ce8e9d3a52	2016/04/01
Manually scaled parameters of Ionogram at Kokugunji, Japan	10.17594/567ced454d15b	2016/04/04
Automatically scaled parameters of Ionogram at Kokugunji, Japan	10.17594/567ced0bbccf9	2016/04/04
Ionogram at Wakkanai, Japan	10.17594/5704b5259137a	2016/04/06
Manually scaled parameters of Ionogram at Wakkanai, Japan	10.17594/5704641f8b11d	2016/04/06
Automatically scaled parameters of Ionogram at Wakkanai, Japan	10.17594/5704b5444c661	2016/04/06
Ionogram at Yamagawa, Japan	10.17594/5704b78099ac0	2016/04/06
Manually scaled parameters of Ionogram at Yamagawa, Japan	10.17594/5704b7b16d387	2016/04/06
Automatically scaled parameters of Ionogram at Yamagawa, Japan	10.17594/5704b79d253fd	2016/04/06
Ionogram at Okinawa, Japan	10.17594/5704b8b1d8dbc	2016/04/06
Manually scaled parameters of Ionogram at Okinawa, Japan	10.17594/5704b8e3a7ffa	2016/04/06
Automatically scaled parameters of Ionogram at Okinawa, Japan	10.17594/5704b8ce63d3b	2016/04/06
Wp index	10.17593/13437-46800	2016/08/10
Wind Profiler at NICT Tokyo (1993-2003)	10.17591/14791-10297	2017/01/25
Magnetotelluric Data at Muroto, Japan	10.17593/13882-05900	2017/02/14
AE index	10.17593/15031-54800	2017/08/20

Sharing Real Business Purpose Datasets for Academic Research

Keizo Oyama[1]*, ***Tomoko Ohsuga***[2]

[1] *National Institute of Informatics / SOKENDAI, 2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo, 101-8430, Japan*
[2] *National Institute of Informatics*
*Email: oyama@nii.ac.jp

Summary. Large scale real data is indispensable for informatics research. IDR service operated by National Institute of Informatics is mediating private companies that are willing to provide data obtained in their business and researchers that are requiring real data for their research. The authors briefly introduce the IDR's activities, including their significance, negotiations and consultations with companies, service to the researchers, and some statistics on the usage. In addition, discussions on the risks and countermeasures regarding personal information, copyright, privacy and so on which may reside in the datasets is presented.

Keywords. Large scale datasets, Real business data, Licensed data, Research resource, Dataset sharing.

1. Introduction

For the research in informatics and the related fields, large scale datasets obtained in the real world are indispensable as the research resource. However, building such datasets solely for research purposes is in most cases unrealistic because of time and cost.

Fortunately, private companies are recently more willing to provide their data generated in their business for academic purposes with motives for social contributions as well as their own interests. However, such data cannot be distributed as open data because of various risks on the companies.

For mediating such companies and researchers, National Institute of Informatics (NII) set up the Center for Dataset Sharing and Collaborative Research (DSC) [1], having them operate the IDR service [2] ("IDR" stands for "Informatics Data Repository").

This paper briefly describe the IDR's activities first and then the measures for reducing risks on both the companies and the researchers from legal and social aspects.

[1] http://www.nii.ac.jp/en/research/centers/dsc/
[2] http://www.nii.ac.jp/dsc/idr/en/index.html

2. IDR's Activities

The major role of IDR is to accept datasets from private companies and to provide them to academic researchers [1-2]. Keeping the dataset fixed with a well-defined specifications and letting researchers use the same dataset, transparency, reproducibility and comparability of the research results are secured.

IDR negotiates with each company on the conditions for providing the dataset to researchers, on the procedure of application from and contract with the researchers, etc.

Many of the datasets are built from data collected via the companies' business over the Internet, such as Q-and-A service, e-commerce, and SNS. They tend to include, in various formats, personal information, information that potentially intrude third parties' rights, etc., and also may have negative effects on the companies' economical profits if abused.

Since NII belongs to the academia, knowing well about researchers, IDR can give a company advices on the researchers' interests, undesirable uses they tend to make, etc., and collaborate in making the terms of use.

On the researchers' side, IDR not only distributes the datasets but also acts as a

consolidated point of contact to all the companies. Almost all procedures including application, contract conclusion and usage report, as well as announcement from the companies and inquiry from the researchers are handled by IDR.

IDR also provides occasions for the companies and the researchers to gather and to exchange problems and ideas [3][3].

As of October 15, 2017, IDR is handing 16 datasets provided by 9 private companies[4], which have been licensed to, in total, approximately 700 university laboratories, etc. (or 450 distinct ones) and more than 2,300 individuals. According to the reports collected from the users every year, the number of publications is more than 600 as of the end of March 2017 although still incomplete.

3. Risks and Countermeasures

When a company provides a dataset to third parties, even created by themselves, they have to follow the laws as well as the terms of the service applicable when they accepted the data. This is still the case even when provided via IDR for academic research under respective contracts. If the dataset holds any problem, not only the company but also IDR and the dataset users are exposed to risks of lawsuits and social criticism. Therefore, when contacted by a company, IDR checks the following issues and assesses the risks:

- If personal information is included, and, if so, is obtained under the consent on providing to researchers for academic purposes;
- If copyright of the posts is properly processed and if the company is authorized to provide the data to third parties;
- If the data is free from invasion of privacy, slander and defamation, or the company has at least a system to eliminate doubtful posts; and
- If the data includes any other type of information which is susceptible to social criticism, especially on the network.

It is not rare that a company is too optimistic in providing a dataset for academic research purposes. For instance, they may offer a dataset for which they have not obtained necessary agreement from the service users due to their misunderstanding that publicly accessible data is free from personal information regulation.

In case the dataset does not meet the requirements, IDR may propose to revise their terms of services, to process or to remove some part of data, and so on. Since all risks cannot be eliminated even with such effort, however, IDR sets up a contact network in which IDR act as the hub, so that when somebody finds a problem in the dataset, IDR and the company can know of it soon, prepare the countermeasure, and disseminate it to all the users.

4. Conclusions

The authors first introduced the activities of IDR and some statistics on the datasets and the users. In providing the datasets created via real service businesses on the Internet, they pointed out some critical issues to be considered, risks accompanied and the countermeasures they take. IDR is contributing to the promotion of research in informatics and the related fields in that many researchers are using the datasets and have produced a lot of research results.

References

1. Oyama, K., Ohsuga, T., Sharing dataset for informatics research at NII. *Journal of Information Processing and Management*, 59(2), 105-112, 2016 (in Japanese) (http://doi.org/10.1241/johokanri.59.105)
2. Oyama, K., Ohsuga, T., Shared use of data sets as resources for informatics research. *Journal of the Japanese Society for Artificial Intelligence*, 31(2), 254-261 , 2016 (in Japanese)
3. Kiyota, Y., NII-IDR User Forum 2016. *Journal of Information Processing and Management*, 59(12), 867-871, 2016 (in Japanese) (http://doi.org/10.1241/ johokanri.59.867)

[3] IDR User Forum 2017 (in Japanese) http://www.nii.ac.jp/dsc/idr/userforum/

[4] Dataset List is available at http://www.nii.ac.jp/dsc/idr/en/datalist.html

The Evolution of Data Publication and the Role of Persistent Identifiers and Linked Open Data in Dynamic Data Mobilization

Peter L. Pulsifer[1*]

[1*] *National Snow and Ice Data Center, University of Colorado, 449 UCB, Boulder, Colorado, 80309, USA*
Email: pulsifer@nsidc.org

Summary. Data publication and citation models have existed for millennia. In recent decades, Information and Communications Technologies have transformed the way that we publish and cite data. A Linked Open Data approach combined with a variety for different persistent identifier schemes and accessed through a semantic web model is becoming a powerful system for mobilizing data. This type of approach will be necessary as data systems evolve and must be considered in our discussions about sharing data across disciplines.

Keywords. Linked Open Data, Persistent Identifiers, Semantic Web, Interoperability, History of Science.

1. Introduction

Since before the advent of the printing press in the fifteenth century, humans have collected, organized and disseminated knowledge to better understand their environment and society. For millennia, Indigenous Peoples have synthesized empirical observations and shared knowledge orally, and more recently using writing and digital technologies [1]. Before Current Era, the Library at Alexandria held and managed thousands of documents and it loss is recognized as an example of the loss of an era of knowledge. The International Polar Year (1882-83) marks a milestone in formal data management and publication for the Polar Regions, as was the formation of the World Data System emerging from the International Geophysical Year (1957-58). These systems have served us well, however in the era of digital information and communications technologies, there is a movement towards a new form of data publication based on identifiers such as the Digital Object Identifier (DOI) [2]. The appropriateness of "publication" as a conceptualization for referencing and accessing published knowledge has been debated [3]. In this paper we examine established and emerging mechanisms for publishing and citing data and argue that "Linked Open Data" (LOD) models will continue to evolve as a foundational layer in the way that humans manage knowledge.

2. Linked Open Data

Linked Open Data is a term used to describe exposing, sharing, and connecting information resources on the Web using persistent identifiers (PIDs) and the Semantic Web model [4]. DOIs are well known PIDs, however there are many others being used to positively identify data resources [5]. The LOD and Semantic Web model provide a framework for documenting, understanding and combining information resources with PIDs that goes well beyond simple linking.

3. Linked Open Data, Data Publication and Citation

In contemporary research and society, data are often dynamic as are the information products generated from the transformation and mediation of data resources. An effective data publication and citation model will need to consider how to deal with dynamic data (including metadata) that are represented and identified in different ways. Relying solely on manual, deliberate, human-authored data publications will result in a significant paucity in

available information. This is particularly important as we see the emergence of sensor webs and The Internet of Things as sources of observational data.

4. Conclusions

Ultimately, our data publication and citation model will be based on LOD based on PIDs, the concepts of the Semantic Web (however they evolve), and ultimately techniques such as machine learning. This will allow for the mobilization of dynamic data through various forms of mediation, ultimately serving a myriad of applications of value to society. Failure to consider these emerging trends now may limit the success of the movement towards data publication and citation as it is often defined.

Acknowledgments. The author acknowledges the support of the U.S. National Science Foundation.

References

1. Pulsifer, P., Gearheard, S., Huntington, H. P., Parsons, M. A., McNeave, C., McCann, H. S.. The role of data management in engaging communities in Arctic research: overview of the Exchange for Local Observations and Knowledge of the Arctic (ELOKA). *Polar Geogr.,* 35, 271–290, 2012
2. Lawrence, B., Jones, C., Matthews, B., Pepler, S., Callaghan S., Citation and peer review of data: Moving towards formal data publication. *International Journal of Digital Curation,* 6(2), 2011
3. Parsons, M. A., Fox, P. A., "Is data publication the right metaphor?", *Data Science Journal* , 12, 2013.
4. Pulsifer, P. L., Brauen, G., Geo-semantic Web. Kitchin, R., Lauriault, T. P. Wilson, M. W. (*Eds.),* in *Understanding Spatial Media*. Sage, London, isbn: 9781473949683, 136-148, 2017
5. Duerr, Ruth, E., et al., "On the utility of identification schemes for digital earth science data: an assessment and recommendations." *Earth Science Informatics*, 4.3, 2011, 139, 2011

Vocabulary Broker Application Connecting Data, Information and Literature Across Various Scientific Domains

***Bernd Ritschel*[1]*, *Günther Neher*[2], *Toshihiko Iyemori*[3]

[1]* *Graduate School of Science, Kyoto University,*
Oiwake-cho, Kitashirakawa, Sakyo-ku, Kyoto, 606-8502, Japan (until 2016)
[2]* *Potsdam University of Applied Sciences, 14469 Potsdam, Kiepenheuerallee 5, Germany*
[3]* *Graduate School of Science, Kyoto University,*
Oiwake-cho, Kitashirakawa, Sakyo-ku, Kyoto, 606-8502, Japan
Email: berndritschel@yahoo.de

Summary. The semantic Web approach contains the potential to connect data, information and related scientific paper across domains, based on qualified and semantically enriched metadata and context information. Keywords taken from controlled vocabularies are used to tag data but also literature and other appropriate resources in the Web. Matching semantic relationships within terminological ontologies connect equal or similar concepts and therefore appropriate data, information and scientific paper. The Vocabulary Broker framework and application provide the idea, methods and tools for a cross-discipline mashup of scientific and social gained data as well as related paper and decisions based on that resources.

Keywords. Metadata, Terminological Ontology, Semantic Web, Simple Knowledge Organization System.

1. Introduction

To understand each other, a common vocabulary and a coherent interpretation of concepts behind terms of that vocabulary are necessary. This general statement is valid for human language but also scientific understanding and collaboration within a specific domain and also cross-domain. Concepts describing properties of resources are summarized by appropriate keywords. Depending on personal interests, scientific disciplines, data usage, etc., keywords are used for the qualification of related concepts. Often keywords are structured within a specific vocabulary or treated as freely selectable terms. Within the semantic Web approach, concepts and related keywords are modelled as terminological ontologies. Furthermore appropriate methods and semantic Web language constructs are used to address dependencies and relationships between concepts. Relationships of concepts between vocabularies, finally provide the potential to mash up different but related resources, such as data and literature which are tagged by appropriate keywords.

2. Vocabulary Broker Framework

2.1.Data resources and descriptive metadata

Metadata and context information is used to describe data resources in the Web. For the description of properties different metadata standards, such as FGDC or ISO are existing [1]. As part of a metadata standard or in addition concept-based keywords are used for further qualification of resources. Appropriate keywords from a controlled vocabulary serve for data tagging. Examples of controlled vocabularies in the geo and space science domain are NASA Global Change Master Directory or SPASE Allowed Values [1]. Within specific scientific fields, if no other vocabularies match the data properties, an own structured vocabulary must be created. Especially such relationships as “broader term“, ”narrower term” or “same as” between concepts respectively keywords should be qualified.

2.2.Terminological ontologies

Within the semantic web, structured vocabularies, such as taxonomies or thesauri are

modeled as terminological ontology. The Ontology Web Language (OWL) and especially the Simple Knowledge Organization System (SKOS) provide properties for the qualification of relationships between concepts respectively keywords. Such relationships and appropriate properties are "skos:related", "skos:broader", "skos:narrower" or "skos:closematch" and others. These properties describe the structure, the direction and the strength of relationships between concepts respectively keywords within one or between different terminological ontologies [1]. Systematically structured terminological ontologies containing a high level of appropriate properties, are the basis for mapping or merging activities.

2.3.Ontology mapping

Mapping or merging of terminological ontologies is based on analysis processes of underlying concepts. In order to simplify the mapping process only the keywords are used, instead of the whole concept descriptions. Before the real mapping procedure, keywords must be prepared. Such preparations are term stemming, done by different algorithms and others. The level of performance can be controlled by precision and recall parameters. Furthermore the keyword-comparison or mapping procedure depends on the depth of the related broader or narrower terms which are used for the input. The VB modul Skotheme provides the mapping capabilities of the whole framework [1].

2.4.Ontology mapping

Semantically mapped keywords and appropriate concepts provide a mediation functionality within and between vocabularies. The visualization of mapped structures and properties shows dependencies and relationships of appropriate concepts accross different scientific domains. The brokerage functionality provides a network of relationships of similar but also different concepts, which supports the finding, and understanding of new and undetected correlations within data and other resources, such as appropriate literature, decisions and beyond.

3. Vocabulary Broker Application

The visualization of dependencies and properties of relationship between concepts across vocabularies, such as GCMD or SPASE vocabulary, is a typical academically shaped application of the VB. Depending on access links to data and literature, wich are described and tagged by concepts and keywords from mapped vocabularies, the Vocabulary Brocker automatically connects the appropriate data and scientific paper from different resources. An example for such an application is the detection and provision of related data and literature connecting geoscience data from IUGONET with appropriate scientific paper in the Web. Starting point is the input of a certain keyword from a suitable vocabulary, such as SPASE, GCMD, GEMET or UAT [1]. Mapped relationships provide a network of related keywords across the vocabularies. All resources, such as data and literature which are described by the input keyword, and because of semantic relationships also mapped keywords are displayed. Finally, associated links provide the direct access from the Vocabulary Broker (VB) to all appropriate resources in the Web.

4. Conclusions

The VB provides both, a framework with methods and tools for the design and mapping of vocabularies, and an application which connects data and appropriate resources, such as scientific paper. Terminological ontologies, which are used for the description of resources, such as data repositories or scientific paper, are the basement for the VB. The VB provides functionality for the mediation and the appropriate data access within a knowledge network, such as the planned WDS knowledge network [2].

References

1. Ritschel, B., et.al., Experiments using Semantic Web technologies to connect IUGONET, ESPAS and GFZ ISD data portals. *Earth, Planet and Space,* 68, 181, doi:10.1186/s40623–016–0542-x, 2016
2. Hugo, W., New ideas for communities of practice - network of networks. *SAJG,* 2(2), ISBN:2225–8531, 2013

SeaDataNet, a Network of Distributed Oceanographic Data Centres Now Going to the Cloud

Serge Scory[1]*, ***Dick M. A. Schaap***[2], ***Michèle Fichaut***[3]

[1]* *Royal Belgian Institute of Natural Sciences, Gulledelle 100, 1200 Brussels, Belgium*
[2] *Marine Information Service MARIS B.V., Koningin Julianalaan 345A, 2273 JJ Voorburg, The Netherlands*
[3] *Ifremer, Centre Bretagne - ZI de la Pointe du Diable - CS 10070, 29280 Plouzané, France*
Email: Serge.Scory@naturalsciences.be

Summary. The SeaDataNet infrastructure comprises a network of interconnected oceanographic data centres and a central portal. In this talk we present the services currently provided by the infrastructure and the technologies making them possible. We then describe how new technologies, in particular cloud storage and cloud computing, will be used to offer new and better services.

Keywords. Marine data, metadata catalogues, standards, distributed infrastructure, cloud.

1. Introduction

SeaDataNet finds its roots at the turn of the century in a concerted action gathering 13 European marine data centres (Euronodim, 1998–2001) and aiming at:

- developing, maintaining and publishing jointly four meta-data catalogues to keep track of ocean and marine data;
- exchanging experience and cooperating in development, promotion and implementation of data & information management practices and methods;
- developing and organizing an overall capability for handling, processing, quality-controlling and archiving a variety of oceanographic and marine data types.

This exploratory phase outlined the concept of an integrated network of marine data centres, accessible through a unique portal.

Thanks to successive R&D projects funded by the European Commission (Sea-Search, 2002–2005; SeaDataNet, 2006–2011; SeaDataNet II, 2011–2015), this network is now a pan-European infrastructure for managing marine and ocean data undertaken by more 100 National Oceanographic Data Centres (NODC's) and oceanographic data focal points from 34 coastal states in Europe.

The presentation will give information on the services presently provided by SeaDataNet infrastructure and services. It will highlight the new technological challenges that distributed data infrastructures are now facing.

2. Current deployment

The SeaDataNet infrastructure comprises a network of interconnected data centres and a central portal. The portal provides users a harmonised set of metadata directories and controlled access to the large collections of datasets, managed by the interconnected data centres. The population of directories increases steadily in cooperation with and involvement in many associated EU projects and initiatives such as EMODnet (an initiative for significantly increasing marine knowledge in Europe by 2020) .

SeaDataNet at present gives overview and access to more than 2 million data sets for physical oceanography, chemistry, geology, geophysics, bathymetry and biology. SeaDataNet is also active in setting and governing marine data standards, and exploring and establishing interoperability solutions to connect to other e-infrastructures on the basis of standards of ISO (19115, 19139), and OGC (WMS, WFS, CS-W and SWE). Standards and associated SeaDataNet tools are made available at the SeaDataNet portal for

wide uptake by data handling and managing organisations.

3. On–going developments

SeaDataCloud marks the third phase of developing the pan-European SeaDataNet infrastructure for marine and ocean data management.

This project (2016–2020) aims at further developing standards, innovating services & products, adopting new technologies, and giving more attention to users. Moreover, it is about establishing a strong cooperation between the SeaDataNet consortium of marine data centres and the EUDAT consortium of e-infrastructure service providers.

SeaDataCloud aims at considerably advancing services and increasing their usage by adopting cloud and High Performance Computing technology. SeaDataCloud will empower researchers with a packaged collection of services and tools, tailored to their specific needs, supporting research and enabling generation of added-value products from marine and ocean data. Substantial activities will be focused on developing added-value services, such as data subsetting, analysis, visualisation, and publishing workflows for users, both regular and advanced users, as part of a Virtual Research Environment (VRE).

4. Conclusions

SeaDataNet is an efficient marine data infrastructure that keeps evolving to fit the needs of a wide range of user categories.

New technologies, in particular cloud storage and cloud computing, offer new opportunities to develop new and better services. Addressing those is the purpose of the on-going SeaDataCloud project.

In this overview of this 20+ years long endeavour, we show what made SeaDataNet a success story so far and how it plans to address next technological challenges for providing new and better services.

Development of Data Sharing and Archiving on International Relations

Kiyohisa Shibai[1]*

[1]* *Joint Support-Center for Data Science Research (DS), Organization of Information and Systems, 10-3, Midori-cho, Tachikawa-shi, Tokyo 190-0014, Japan*
Email: kshibai@ism.ac.jp

Summary. The Institute of Statistical Mathematics has carried out the longitudinal survey called "The Japanese National Character survey" since 1953, and now it was extended as cross-national surveys including the Japanese overseas. It has clarified the nationals have a variety of characteristics and there are many diversities and similarities on social values and general response patterns among people over the world. However, all the characteristics are not necessarily unchanging. Some can be changed with the passage of time and social changes. Therefore, in order to research national characters, we need to collect many survey data cross-nationally and longitudinally. However, we meet difficulties of collecting reliable data because of restrictions of laws and political institutions of each country and sometimes the relationships between the researcher's state and research object's state. The legal and political restrictions are serious problems which cannot be resolved by a single researcher's knowledge and technology. Thus, we necessarily consider the topic of data sharing and archiving very important for the development of data science on international relations. International society is rapidly changing just now because of immigration and democratization. It is impossible for a researcher by oneself to correspond the social changes and collect the data, so we have to collaborate each other and share data.

Keywords. Cross national survey, international relations, immigration, and democratization.

1. Cross National Survey

The Institute of Statistical Mathematics has carried out the longitudinal survey called "The Japanese National Character survey" since 1953, and now it was extended as cross-national surveys including the Japanese overseas: Surveys on Japanese Americans of Hawaii & of the West Coast, and Japanese Brazilian (1971-99), Japanese ancestry Americans in the West coast of U.S.A, Seven Country Survey (Japan, USA, Britain, France, Germany, Italy, and the Netherlands) (1987-93), East Asia Values Survey (EAVS) (Japan, China [Beijing, Shanghai], Hong Kong, Taiwan, South Korea & Singapore) (2002-05), Pacific-Rim Values Survey (PRVS) (Japan, China [Beijing, Shanghai], Hong Kong, Taiwan, South Korea, USA, Singapore, Australia, and India) (2004-09), Asia-Pacific Values Survey (APVS) (Japan, China [Beijing, Shanghai], Hong Kong, Taiwan, South Korea, USA, Singapore, Australia, India, and Vietnam) (2010-14) [1].

We need to collect many survey data cross-nationally and longitudinally for comparative analysis of national characters, and we have found some characters of the nationals and their similarities and differences on social values and general response patterns among people over the world. It depends on the passage of time and social changes whether the cultural links are strengthened or weakened [2-3].

2. Importance of Data Archiving

It is important to archive a variety of national data continuously because they are needed to grasp the similarities and differences.

For example, according to APVS, there are some similarities and differences in confidence to the social organizations. While Singaporean,

Indian, and Vietnamese have strong trust in "religious organizations," many of Japanese and Chinese living in Beijing don't trust them: contrast South East Asia and South Asia with North East Asia; only American and Australian don't have trust in the media, "press and television": contrast Western culture with Asian culture; "Federal bureaucracy" and "congress" are untrustworthy in democratic states: contrast of political systems; "Science and technology" is much trustworthy in every state: universal value.

We can find when and how the characters are changed by archiving many data. In Vietnam, there are some generations: experience of the Vietnam War or not, born in the North, South or unified Vietnam. We know the hard battles and violence in the Vietnam War and the political and economic systems were different between the North Vietnam and the South Vietnam. In a question of APVS "Which one of the following countries or regions would you like to see develop the friendliest relationship for our own national interest," the largest percentage of answers is the United States. Especially people living in the south area, former South Vietnam, did it more than people living in the north area. In contrast, more Vietnamese in the north area answered Russia than people in the south area. In addition, the post-Vietnam War generation likes the United States than the Vietnam War generation even in not only the south area but also the north area [4].

3. Importance of Data Sharing

There are about 200 countries in the world and there are also many languages, cultures, and legal and political systems. Plenty of preparations and budget are needed for getting data of such people. In developing countries, it may be impossible to do the web survey which is much easier and cheaper than the face-to-face survey.

Consequently, data sharing is much useful for the development of cross national survey because we meet difficulties of collecting reliable data because of restrictions of laws and political institutions of each country. For example, Chinese government watches all nations, so it is much difficult for us to include questions concerning to the Communist Party and anti-Japanese feeling even though we don't intend to research them. The legal and political restrictions are serious problems which cannot be resolved by a single researcher's knowledge and technology.

4. Rapid Changes in International Society

We have to develop the cross national survey and archive more data because the international society is rapidly changing just now. The large number of immigrants and refugees have moved to the Europe and the North America, but many of them have not assimilated culturally. In the result, constructions of the societies are being changed and serious conflicts between the residents and immigrants have occurred. In addition, especially in Asia and Africa, many people are demanding on democratization of their own countries. The political system has an effect on nationals' values as mentioned above.

It is impossible for a researcher by oneself to correspond the social changes and collect the data, so we have to collaborate each other and share data.

References

1. *Cross-National Comparative Survey on National Character* http://www.ism.ac.jp/~yoshino/index_e.html [accessed 2017/10/12]
2. Yoshino, R., Shibai, K., et al., The Asia-Pacific Values Survey 2010-2014: Cultural Manifold Analysis (CULMAN) of National Character. *Behaviormetrika,* 42, 2, 99-120, 2015
3. Inglehart, R., Welzel, C., Changing Mass Priorities: The Link Between Modernization and Democracy. *Perspectives on Politics,* 8, 2, 551-567, 2010
4. Shibai, K., Vietnamese Characteristics of Social Consciousness and Values: National Character, Differences between North and South, and Gaps between the Vietnam War Generation and the Post-war Generation. *Behaviormetrika,* 42, 2, 167-189, 2015

Data and Metadata Management at DIAS: Toward More Open Earth Environmental Information Platform

Toshiyuki Shimizu[1*]

[1*] *Graduate School of Informatics, Kyoto University, Yoshida-Honmachi, Sakyo-ku, Kyoto 606-8501, Japan*
Email: tshimizu@i.kyoto-u.ac.jp

Summary. Data Integration and Analysis System (DIAS) has continuously collected and managed earth observation data, and works as one of core research data platforms in Japan. As we have various kinds of data in the storage of DIAS, we are managing data by creating metadata in the dataset granularity. From 2016, phase III of DIAS started with more focus on open science, and we have started assigning Digital Object Identifiers (DOIs) to subset of datasets stored in the DIAS repository including the datasets we accepted from outside of DIAS. We are also focussing on metadata quality for better findability of datasets.

Keywords. DIAS, metadata, open science, metadata quality.

1. Introduction

Data Integration and Analysis System (DIAS) collects and stores data related to earth observation, and provides research platform for analysing data [1]. DIAS started from 2006, and after the successful completion of phase I and phase II of DIAS, the current project is in phase III which has started from 2016 with the aim of its practical operation.

Various kinds of datasets such as satellite data, ocean observation data, reanalysis data, land use data, etc. are stored in the DIAS repository, and many applications have been developed on them. Since we need to handle variety of datasets, we have developed the systems for managing metadata in the dataset granularity.

Also, with the movement of open science, the DIAS open science special interest group was launched, and we are discussing to make DIAS more open platform.

2. Data and Metadata Management

2.1 Data registration procedure

The requests of data deposit to DIAS are done with the submission of the application forms. After the review process, the data are stored in the DIAS repository if they are approved.

All datasets stored in the DIAS repository must have metadata, and data providers need to create metadata using our system.

2.2 Metadata management

The dataset metadata of DIAS is designed to be applicable to various kinds of datasets stored in the DIAS repository. It contains the fields for dataset name, contact point, creator, abstract text, topic category, spatiotemporal information, etc. The internal metadata format is ISO 19139, and each field is mapped to the corresponding element.

Once metadata is created using our system, the dataset becomes open to the public, and searchable from the DIAS dataset search and discovery system [2]. Users can specify the desired dataset and download data files through the system (Figure 1).

3. Open Science Activities

In our current project, phase III of DIAS, we focussed more on open science. One of the outcomes of the activities of the DIAS open science special interest group is assigning Digital Object Identifiers (DOIs) to datasets. The first assignment of a DOI in DIAS was done on March 2017 [3], and now we have several datasets with DOIs.

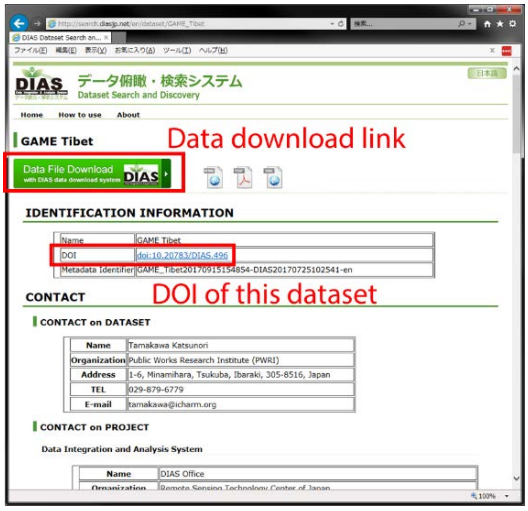


Figure 1. An example of a dataset metadata on the DIAS dataset search and discovery system [2].

The DOI of a dataset is stored in the metadata of the dataset, and we can find it at the dataset page in the DIAS dataset search and discovery system as shown in Figure 1.

Our other topics include getting official certificates of trustworthy data repositories so that DIAS can be considered as trustworthy from stakeholders.

4. Current and Future Prospects

4.1 DIAS as a national repository

Thus far, the stored datasets in the DIAS repository were mainly datasets from inside of DIAS or the collaborative projects. However, we have accepted some datasets outside of DIAS recently.

Since more and more journals request evidence data for papers, the demands for the data repositories are also increasing. However, data repositories which can store large volume of data are limited. Since DIAS has a very large storage space, it can be a candidate of a data repository especially for the valuable research data in Japan.

4.2 Focussing on metadata quality

For the end users of DIAS, findability of datasets stored in the DIAS repository are important, and thus we are focussing on metadata quality for providing better services.

As one of the attempts of improving metadata quality, we developed keyword recommendation method [4] since we considered keyword information is important for searching and categorizing datasets.

Besides, since we have not only metadata of DIAS datasets but also metadata from data centres outside of DIAS in the DIAS dataset search and discovery system, integrated management of different types of metadata is another topic concerning metadata quality.

5. Conclusions

DIAS is not only a data repository, but also an information platform for data science. We are managing various kinds of datasets through the metadata. About open science, the DIAS open science special interest group leads the discussion, and started assigning DOIs to datasets. We would like to make continuous efforts in order to make DIAS more open platform both for data providers and users.

Acknowledgments. I thank people in the DIAS open science special interest group, Dr. Asanobu Kitamoto, Dr. Masafumi Ono, Dr. Hiroko Kinutani, Dr. Masatoshi Yoshikawa, and Mrs. Yoko Nakahara for helpful discussion.

References

1. Kawasaki, A., Yamamoto, A., Koudelova, P., Acierto, R.A., Nemoto, T., Kitsuregawa, M., Koike, T., Data Integration and Analysis System (DIAS) Contributing to Climate Change Analysis and Disaster Risk Reduction. *Data Science Journal*, 16, 41, 1–17, 2017
2. DIAS dataset search and discovery system, http://search.diasjp.net/en [accessed on: Oct. 2017]
3. DIAS First DOI Registration, http://www.diasjp.net/infomation/press-release-dias-first-doi-registration/ [accessed on: Oct. 2017] (in Japanese)
4. Ishida, Y., Shimizu, T., Yoshikawa, M, A Keyword Recommendation Method Using CorKeD Words and Its Application to Earth Science Data. *11th Asia Information Retrieval Societies Conference*, 96–108, 2015

Data Sharing at the National Research Institute for Earth Science and Disaster Resilience

Katsuhiko Shiomi[1]*

[1]* *Network Center for Earthquake, Tsunami and Volcano, National Research Institute for Earth Science and Disaster Resilience, 3-1 Tennodai, Tsukuba-shi, Ibaraki-ken, 305-0006, Japan*
Email: shiomi@bosai.go.jp

Summary. After the 1995 Kobe Earthquake, National Research Institute for Earth Science and Disaster Resilience, NIED, established the nation-wide seismograph networks as a part of Japanese government policy. Ground motions observed at each station are automatically sent not only to the NIED data management centre at Tsukuba but also to the Japan Meteorological Agency, JMA, universities, and other related research institutes in real-time. By using the data, JMA announces the earthquake early warning (EEW). JMA also constructs the detailed hypocentre catalogue and open it to the public. NIED has responsibility to accumulate seismic waveform data and to open them to the public via the Internet. In order to check how much of these data are used for research activities, NIED has just started to discuss to apply the Digital Object Identifier (DOI) to them.

Keywords. Seismograph networks, data exchange, quality check, data citation, Digital Object Identifier.

1. Introduction

National Research Institute for Earth Science and Disaster Resilience, NIED, is pursuing research activities in a wide range of natural hazards, disaster mitigation, and effective disaster response and recovery to improve the level of science and technology for disaster risk reduction. To achieve these research subjects, NIED is operating several kinds of nation-wide observation networks for monitoring earthquake, tsunami, volcanic and meteorological activities [1]. Most of stations were constructed with premise to open the observed data to the public. Earthquake observation networks are one of them. In this report, NIED's recent activities about sharing the seismic observation data are introduced mainly.

2. NIED Seismograph Networks

NIED is currently operating three kinds of seismograph networks in the land area [2] and two networks at the ocean bottom. To detect weak signal from micro-earthquake and estimate precise locations of hypocentres, the high-sensitivity seismograph network, Hi-net, is operated. F-net is the broad-band seismograph network to observe ground motions in broad frequency range, and used for the research of earthquake mechanisms. K-NET and KiK-net are strong motion seismograph networks that accurately observes strong ground motions. These networks were newly constructed or expanded after the 1995 Kobe Earthquake to monitor earthquake activity and ground motion for the whole nation. Recently, two ocean bottom seismograph and Tsunami observation networks are established: S-net for the Pacific coast of the eastern Japan and DONET for the Nankai area. According to the policy determined by the Headquarters for Earthquake Research Promotion, NIED has responsibility to accumulate all seismic waveform data observed by these networks, to store them for long time and to open them to the public via the Internet. Everyone can download the data from the NIED's websites for non-commercial use.

3. Data Exchange in Real-Time

In Japan, not only NIED but also the Japan Meteorological Agency, JMA, universities, and other related research institutes operates their own seismograph networks. To maximize the use of limited data, all related organizations are sharing the observed high-sensitivity and broadband seismograph data in real-time. JMA uses these data to monitor the earthquake activity and to construct the precise nation-wide hypocentre catalogue. Moreover, JMA announces the earthquake early warning (EEW) by using both Hi-net and JMA data. Universities use the data for their research and educational activities. As each K-NET station has a function of the JMA-scale seismic intensity meters, observed seismic intensity is sent to JMA immediately once a large earthquake occurs. JMA integrates them with its own data, and announces as earthquake information to the public.

4. Introduction of DOI and Problems

About 20 years have passed since the operation began, NIED seismograph networks have become essentials for earthquake disaster mitigation and seismological research. In order to maintain the networks in the future, it is important to show the significance of the networks, especially in the research field.

In 2017, NIED organized a new group named 'Administration Office for Information Integration'. This group discusses how to open the NIED data to the research group and/or the public efficiently and effectively, and how to show the importance of the NIED data. For the latter point, number of research papers in which the data are used would be a useful index. Therefore, the group concluded that introduction of the Digital Object Identifier (DOI) to the data is one of the good answers to visualize cited performance of the data.

NIED seismograph networks are working now and their data are increasing every second. Each seismograph network is composed by 100 - 1000 stations. Each station has observation parameters and repair histories individually. For the field of earthquake observation, it is difficult to classify the waveform data to "good" or "bad" as a result of quality check. For example, fine three-component seismograms are needed to get the JMA-scale seismic intensity, although even only one-component data is appropriate for use to know the arrival time of earthquake motion. Unit for the DOIs to the seismic waveform data should be determined while considering efficient metadata management and user's convenience.

5. Conclusions

NIED operates several kinds of observation networks. Especially seismograph network data are open to the public since the network has established. Waveform data by NIED, JMA, universities, and other institutes are shared in real-time and used for their own activity. NIED receives and stores all of them.

To maintain the NIED networks, it is required to visualize effectiveness of their data. To answer this requirement, it is necessary to show the cited performance of the data based on the reliable database. Introduction of the data DOI would be one of the key answers in the future.

The seismograph networks have to continue to provide effective information for earthquake disaster mitigation. Although there are some difficulties to apply the data DOIs to waveform data of the NIED seismograph networks, the DOIs would play an important role to maintain the nation-wide networks in long-term.

Acknowledgments. The author would like to appreciate all people concerned the observation networks operated by NIED.

References

1. National Research Institute for Earth Science and Disaster Resilience, http://www.bosai.go.jp/e/about/brochure_nied.pdf [accessed on: October, 2017]
2. Okada, Y., Kasahara, K., Hori, S., Obara, K. Sekiguchi, S., Fujiwara, H., Yamamoto, A., Recent progress of seismic observation networks in Japan—Hi-net, F-net, K-NET and KiK-net—. *Earth, Planets and Space*, 56, xv–xxviii, 2004

Data paper of JAMSTEC Report of Research and Development

Daisuke Suetsugu[1]*

[1]**Japan agency for Marine-Earth Science and Technology, 2-15, Natsushima-cho, Yokosuka-city, Kanagawa, 237-0061, Japan*
Email: dai@jamstec.go.jp

Summary. The JAMSTEC-R editorial committee initiate new category of paper called "Data paper" for publication on JAMSTEC-R from April, 2017. The data paper is defined as a paper that describes data contents, acquisition methods, data formats, and access for data obtained by observations, experiments, measurements, or computer simulation. A data paper does not include analysis, interpretation, or scientific conclusions. JAMSTEC-R will publish data papers on data obtained by JAMSTEC members working for scientific research and technical development, or data obtained by observation/research facilities of JAMSTEC. Upon publication, the data are accessible via the JAMSTEC-R data repository. We hope that publication of data papers enables access to valuable but unrecognized data and help promote Open Data.

Keywords. Data paper, Marine science, Database.

1. Introduction of JAMSTEC and JAMSTEC-R

Japan Agency for Marine-Earth Science and Technology (JAMSTEC) has the main objective to contribute to the advancement of academic research in addition to the improvement of marine science and technology by proceeding the fundamental research and development on marine, and the cooperative activities on the academic research related to the Ocean for the benefit of the peace and human welfare.

JAMSTEC publishes the peer-reviewed journal "JAMSTEC-R" (JAMSTEC Report of Research and Development) biannually (Sep / Mar) that reports research and development results on marine-earth science and technology. Its electronic journal is also available on J-STAGE. Those who are involved in research and technology development at JAMSTEC, and those who conduct a survey or research activity using JAMSTEC's survey equipment, research facilities or data and samples etc., are all eligible for paper submission regardless of organization to which they belong. We have four types of submission: "Original Paper", "Review", "Report" and the new category "Data Paper". All manuscripts have been peer-reviewed. Extensive coverage of ocean and the earth related fields. Submissions can be either in English or in Japanese.

2. Data paper of JAMSTEC-R

The JAMSTEC-R editorial committee has initiated new category of paper called "Data paper" for publication on JAMSTEC-R from April, 2017. The data paper is defined as a paper that describes data contents, acquisition methods, data formats, and access for data obtained by observations, experiments, measurements, or computer simulation. A data paper does not include analysis, interpretation, or scientific conclusions. JAMSTEC-R will publish data papers on data obtained by JAMSTEC members working for scientific research and technical development, or data obtained by observation/research facilities of JAMSTEC. Upon publication, the data are accessible via the JAMS
TEC-R data repository.

Before the official initiation of data paper, the JAMSTEC-R editorial committee asked J. Yoshimitsu and M. Obayashi to submit the first data paper for JAMSTEC-R as a trial. The paper was on a database of travel times of seismic waves propagate in the Earth. The travel times

were measured from seismograms. In the trial review process of the data paper, description of data measurement method, data quality, and data format were most important. After the review process, the paper was published in March, 2017 [1]. The trial publication was useful for us to establish a format of data paper and rules of writing and submission, and review process.

References

1. Yoshimitsu, J., Obayashi, M., A database of global seismic travel times. *JAMSTEC Rep. Res. Dev.*, 24, 23-29, 2017

Interdisciplinary Online Data Sharing Service on ADS

***Takeshi Sugimura*[1]*, *Takeshi Terui*[1], *Hironori Yabuki*[1]**

[1]* *National Institute of Polar Research, 10-3, Midori-cho, Tachikawa-city, Tokyo, 190-8518, Japan*
Email: sugimura.takeshi@nipr.ac.jp

Summary. Sharing of data has progressed at various fields in recent years, but in scientific fields the sharing of data between the researchers has not progressed so much. This is because specialized knowledge and great effort are required to extract information from scientific data. A data provider needs to provide the service to reduce such costs, and should make effort to promote the sharing of data. We have developed the web application with which anyone can retrieve information from data easily.

Keywords. Sharing of Data, Web Application, Visualization, ADS, Satellite Data.

1. Introduction

In recent years, anyone can acquire various data from online because of advance of networking infrastructure. It also became easy for a researcher to get the data of the out of own specialized fields. On the other hand, it is necessary for a great effort and technical knowledge to find out the appropriate data from a huge kind of dataset. Even if the appropriate data can be acquired, it is very difficult to process the data and to judge about the contents of the data for a person with different specialized field. Data providers should make an effort to build the system that can reduce such costs. Arctic Data-archive System (ADS), which we are constructing, aims to build the system that facilitates a sharing and understanding the data of the various fields.

2. Web Application

In order to retrieve the information from dataset, complex work and technical knowledge is often needed. We build web application that can reduce such costs. Because the web application works on browser, this application does not need to install special software. Furthermore, since data is downloaded automatically on background, even a user that does not have technical knowledge can access data easily. By arranging intelligibly graphical user interface (GUI), it is devising so that application can be operated only by mouse operation. The application can provide a file of the various formats which secondary use is possible. By satisfying the above functions, we make the applications which not only a researcher but ordinary persons can use.

2.1. VISION

We developed GUI-based online grid data visualization application named "VISION". The VISION offers several kinds of visualization and analysis function: shaded colour map, contour map, cross-section data plot, and time-series data plot. We implemented some grid dataset: AMSR2 satellite product, SSM/I satellite product, NCEP reanalysis data, Climate Research Unit, and some simulation model output data.

One-dimensional data can be graphed using VISION-Graph. This application automatically loads the data of automatic observation station and graphs such data. Thus we can acquire near-

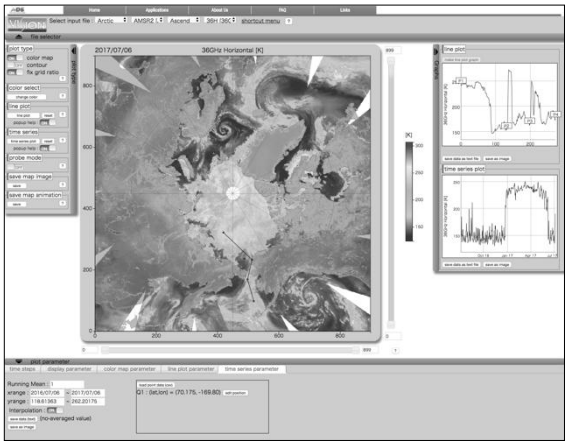


Figure 1. Visualization result using VISION

real time information from observation station around Polar region.

2.2. VISHOP

VISHOP is the application for releasing various information using image and interactive graph. There are many access counts of this page because everyone can easily learn about the state of Polar Region from satellite data. The VISHOP also offers a various data: sea ice extent graph, sea ice forecast image, camera photograph of stationary observation station, quick-look image of observation data.

2.3. Arctic Sea Route Search System

This application provides an optimum shipping route information in Arctic Ocean. This was constructed based on the research products by Arctic Sea Route research group: Yamaguchi Laboratory in the University of Tokyo [1-3]. The route search is performed in consideration of the following.

- Priority between time and distance along route path
- Time-spatial variations of sea ice thickness and height
- Ship size, Ice class and Ice-breaking capability
- The draft of ships and sea-floor topography on the route

This application is still improving with developing a search algorithm.

3. Conclusions

The ADS facilitates a sharing and understanding the data using web application. Because this application is designed so that anyone can use easily, it is possible for everyone to use data of different specialized field.

Acknowledgments. This research is supported by the Arctic Challenge for Sustainability (ArCS) project promoted by the National Institute of Polar Research.

References

1. Choi, M., Chung, J., Yamaguchi, H., Nakagawa, K., Arctic sea route path planning based on an uncertain ice prediction model. *Cold Regions Science and Technology*, 109, 61-69, 2014
2. Nakano, Y., Study on route optimization in the Northern Sea Route. *Master thesis and abstract of Graduate school of frontier science*, The University of Tokyo, 2015 (in Japanese)
3. Yamaguchi, H., Nakano, Y., Navigation support system in icy water. *GRENE Arctic Climate Change Research Project 2nd Special seminar* "Towards the sustainable use implementation of the Arctic Ocean", Nov 6th, 2015 (in Japanese)

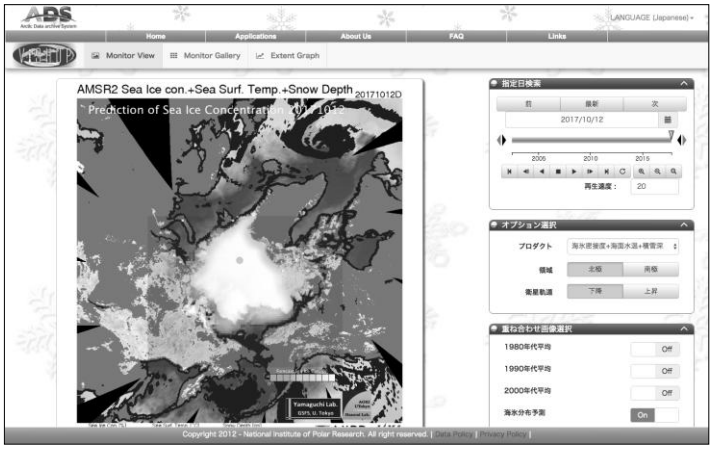


Figure 2. Arctic information using VISHOP

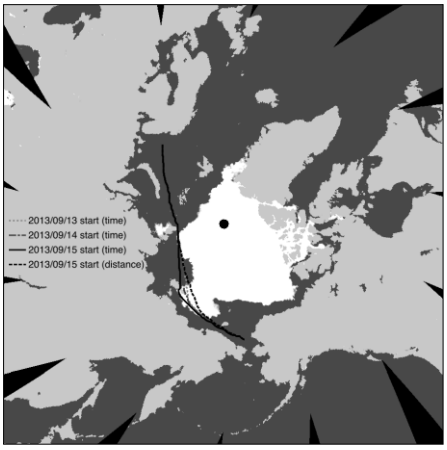


Figure 3. Results of Arctic Sea Route Search.

Inter-university Upper Atmosphere Global Observation Network (IUGONET) Metadata Database

Yoshimasa Tanaka[1*,2,3], ***Norio Umemura***[4], ***Atsuki Shinbori***[4], ***Shuji Abe***[5], ***Masahito Nose***[6], ***Satoru UeNo***[7], ***IUGONET project team***

[1*] *Joint Support-Center for Data Science Research, Research Organization of Information and Systems, 10-3, Midori-cho, Tachikawa-shi, Tokyo 190-8518, Japan*
[2] *National Institute of Polar Research, ROIS, 10-3, Midori-cho, Tachikawa-shi, Tokyo 190-8518, Japan*
[3] *The Graduate University for Advanced Studies (SOKENDAI), Shonan Village, Hayama, Kanagawa 240-0193, Japan*
[4] *Institute for Space-Earth Environmental Research, Nagoya University, Furo-cho, Chikusa-ku, Nagoya 464-8601, Japan*
[5] *International Center for Space Weather Science and Education, Kyushu University, Motooka, Nishi-Ku, Fukuoka 819-0395, Japan*
[6] *Data Analysis Center for Geomagnetism and Space Magnetism, Graduate School of Science, Kyoto University, Kitashirakawa-Oiwake Cho, Sakyo-ku, Kyoto 606-8502, Japan*
[7] *Kwasan and Hida Observatories, Graduate School of Science, Kyoto University, Kurabashira, Kamitakara-cho, Takayama, Gifu 506-1314, Japan*
Email: ytanaka@nipr.ac.jp

Summary. We present characteristics of the metadata database for upper atmospheric data, called IUGONET Type-A, which was developed by Inter-university Upper atmosphere Global Observation Network (IUGONET) project and was newly released in October, 2016. The IUGONET metadata format was designed based on the Space Physics Archive Search and Extract (SPASE) metadata model developed by the SPASE consortium, with some modifications for the upper atmospheric data. The IUGONET Type-A provides users one-stop web service to search data, get the information of the data including quick-look plot, find scientifically interesting events, plot the data interactively, and lead to more detailed analysis of the data using dedicated analysis software, SPEDAS.

Keywords. IUGONET, metadata database, upper atmosphere, interdisciplinary study.

1. Introduction

The upper atmosphere has characteristics as follows: (a) Both vertical coupling between the multiple spheres and global horizontal circulation are important. (b) There are a variety of data sets for plasma and neutral gas, electric and magnetic fields obtained with various instruments. (c) The long-term variation is so important that it is necessary to analyse long-term monitoring data. Thus, data sharing and collaborative research are essential to understand the mechanism of various phenomena in the upper atmosphere.

Inter-university Upper atmosphere Global Observation Network (IUGONET) is a Japanese inter-university project that started in FY2009 to share various ground-based observational data of the Earth's upper atmosphere, Sun and planets archived by Japanese universities and institutes since International Geophysical Year (IGY; 1957-1958) and promote interdisciplinary study using the data [1]. So far, we have mainly developed two tools; one is a metadata database that enables to cross-search various kinds of the upper atmospheric data across the IUGONET members, and the other is an analysis software that can analyse such data in an integrated fashion.

2. IUGONET Metadata Database (IUGONET Type-A)

Since the upper atmospheric data have usually been archived and opened to public individually

by each observer (i.e., university or institute), it is difficult for researchers, who are not related to the observation or belong to different research fields, to find, get, and analyse the data. To overcome this issue, we released the first version of IUGONET metadata database in FY2011, which can search various kinds of the upper atmospheric data distributed across the IUGONET members and provide users with the basic information of the data (e.g., contact persons, access URL, data use policy, etc.) . The IUGONET metadata format was designed based on the Space Physics Archive Search and Extract (SPASE) metadata model with some modifications for the upper atmospheric data. However, this version was developed on the basis of DSpace (http://www.dspace.org/), which is an open source repository software typically used for the academic repository, so it provided only simple metadata information in text format and could not link to analysis software [2]. In the next step, therefore, it was required to have the capability to show quick-look plots of data, interactively plot data, and analyse data or link smoothly to dedicated analysis software.

We developed a new metadata database and released it as IUGONET Type-A (http://search.iugonet.org/) in October, 2016. The IUGONET Type-A provides users one-stop web service to search data, get the information of the data (i.e., metadata) including quick-look plots, find scientifically interesting events, plot the data interactively, and lead to more detailed analysis of the data using the dedicated analysis software, Space Physics Environment Data Analysis Software (SPEDAS; http://themis.ssl.berkeley.edu/software.shtml).
The functions newly added to IUGONET Type-A are summarized as follows:

- Easy data search from lists of instruments/ projects and observation region
- Display of quick-look plots of various upper atmospheric data that have been created by SPEDAS in advance
- Display of search results by quick-look plots to assist users to compare many kinds of data and find scientifically interesting events
- Interactive data visualization using SPEDAS
- Display of procedures for data analyse to lead users to more detailed analysis with SPEDAS

In addition, we regularly have data analysis workshops in Japan and sometimes in other countries, especially in Asia and African region, to explain how to use our data and tools.

3. Conclusions

We released a new metadata database for the upper atmospheric data, IUGONET Type-A, in October, 2016. It allows users to experience a sequence of research steps, such as searching data, getting information of data including quick-look plots, finding interesting events, interactively plotting data, and proceeding to more detailed data analysis using SPEDAS.

Acknowledgments. The IUGONET project was supported by the Special Educational Research Budget (Research Promotion) [FY2009] and the Special Budget (Project) [FY2010-2014] from the Ministry of Education, Culture, Sports, Science and Technology (MEXT), Japan.

References

1. 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., Motoba, T., IUGONET project team, Inter-university Upper Atmosphere Global Observation NETwork (IUGONET). *Data Sci. J.*, 12, WDS179-WDS184, doi: 10.2481/dsj.WDS-030, 2013
2. Abe, S., Umemura, N., Koyama, Y., Tanaka, Y.-M., Yagi, M., Yatagai, A., Shinbori, A., Ueno, S., Sato, Y., Kaneda, N., Progress of the IUGONET system - metadata database for upper atmosphere ground-based observation data. *Earth, Planets and Space,* 66, doi:10.1186/1880-5981-66-133, 2014

Possibility and prevention of data tampering in the referee process of data journal

Takeshi Terui[1]*, ***Yasuyuki Minamiyama***[1], ***Kazutsuna Yamaji***[2]

[1]* *National Institute of Polar Research, 10-3, Midori-cho, Tachikawa-shi, Tokyo 190-8518, Japan*
[2] *National Institute of Informatics, 2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo 101-8430, Japan*
Email: terui.takeshi@nipr.ac.jp

Summary. The risk of data tampering exists on the peer review process in the data journal. When the Polar Data Journal was launched, we analyzed this possibility and preventive measures. There is a possibility that data tampering is done under the review process whether it is fault or not. What is most important to the data journal is whether the authenticity of the data at posting and publication is preserved. Therefore, Polar Data Journal calculates the hash value of data immediately after posting and immediately before publication, and detects data falsification by comparing hash values. It is expected that this method can easily prevent data tampering.

Keywords. Data Tampering, Review Process, Hash Value, Security.

1. Introduction

In the data journal, the submitted data itself is reviewed. The correctness and validity of the data are evaluated by the editorial board and reviewer than the description contents on the paper. Because data is only a simple digital information asset, there is a risk of information security. The risk that you must pay attention to is the data tampering. Data tampering significantly reduces the trustworthiness of data journals. It is very important to take this preventive measure. This paper introduces measures to prevent the above data falsification in Polar Data Journal.

2. Data Tampering

Data tampering in the data journal is one of the incidents that become apparent when the data published after publication is different from the peer-reviewed data. This incident can happen regardless of purpose or negligence. In the data journal, published data is subject to peer review by experts. However, when data falsification becomes apparent, the peer review process will be overwhelmed. Ultimately, the data journal is considered untrusted and has the property of retreating open science. Therefore, efforts must be made to prevent data falsification.

There is a hash value comparison as a technology for detecting data tampering. In Polar Data Journal, hash values of data at posting and reposting are calculated and stored in the paper submission system. Calculate the hash value of the data released just before publication of the paper again and compare it with the stored hash value. When data alteration is performed, the hash value has different results. The office of Polar Data Journal detects data alteration and reports to editorial committee.

In this research presentation, we introduce a process that is highly likely to be falsified and measures to prevent it.

Management of Marine-Earth Science Data and Samples in JAMSTEC

Seiji Tsuboi[1*]

[1*] *Center for Earth Information Science and Technology, JAMSTEC, Yokohama, Kanagawa, 236-0001, Japan*
Email: tsuboi@jamstec.go.jp

Summary. JAMSTEC possesses a large number of leading-edge facilities and equipment and has obtained Data and Samples of extremely high academic value. These Data and Samples are the common property of the human community, and it is important that they are made open for research and educational purposes and be available for use into the future on a global basis. We set up Basic Policies on the Handling of Data and Samples and manage these Data and Samples for use by research organizations and researchers for scientific and educational purposes, and try to provide them promptly and smoothly.

Keywords. Marine data and samples, Data and Sample policy.

The Data Management and Engineering Department of CEIST receives, collects and archives various data and samples collected by JAMSTEC through its research activities related to marine and earth surveys and observations. We also control data quality, and develop and operate a data release system. Based on the Basic Policies on the Handling of Data and Samples by JAMSTEC, enacted in FY2007, we are managing and archiving data and samples obtained using JAMSTEC vessels, research submersibles and autonomous underwater vehicles carried by JAMSTEC vessels. Researchers may access to the Data and Sample Research System for Whole Cruise Information (DARWIN) and find information for data, rock samples, and sediment core samples obtained by research vessels and submersibles, and links to related databases. JAMSTEC E-library of Deep-sea Images (J-EDI) provides an access to variety of unique deep-sea images. We have joined the Tohoku Ecosystem-Associated Marine Sciences (TEAMS)—a program launched in February 2012 for surveys and research on marine ecosystems around 2011 Tohoku earthquake region—and are developing a data management system for disseminating and providing research results and information obtained in this program and building data sharing/release systems. Release of biodiversity information is also our major task. We comprehensively distribute JAMSTEC-owned biological information using the Biological Information System for Marine Life (BISMaL), which handles information on diversity and distribution of marine organisms, collect information on diversity of marine organisms around Japan, and widely release it after systematically organizing it. Since 2010, JAMSTEC has been serving as a Japanese node in the Ocean Biogeographic Information System (OBIS), and providing BISMaL data to OBIS. In light of the fact that OBIS was placed under the umbrella of the International Oceanographic Data and Information Exchange (IODE), a system managed by the Intergovernmental Oceanographic Commission of UNESCO (UNESCO/IOC), in January 2015, JAMSTEC was recognized as an IODE Associate Data Unit (ADU).

The Antarctic Master Directory, sharing Antarctic (meta)data from multiple disciplines

Anton P. Van de Putte[1]*, ***SCADM members***

[1]* *BEDIC, OD Nature,Royal Belgian Institute for Natural Sciences, Rue Vautierstraat 29, Brussels, B-1000, Belgium*
Email: avandeputte@naturalsciences.be

Summary. The Scientific Committee on Antarctic Research (SCAR, www.scar.org) is an inter-disciplinary committee of the International Council for Science (ICSU). SCAR is charged with initiating, developing and coordinating high quality international scientific research in the Antarctic region (including the Southern Ocean), and on the role of the Antarctic region in the Earth system. SCAR promotes free and unrestricted access to Antarctic data and information by promoting open and accessible archiving practices. SCAR aims to be a portal to data repositories of Antarctic scientific data and information. SCAR's Standing Committee on Antarctic Data Management (SCADM) facilitates co-operation between scientists and nations with regard to scientific data, and advises on the development of the Antarctic Data Directory System. Here we provide an overview of some of the SCAR data sharing platforms an products. The most important of which is the Antarctic Master Directory (AMD, the largest collection of Antarctic data set description in the world, holding over 7700 dataset descriptions from 25 countries.

Keywords. SCAR, SCADM, GCMD, Antarctic Master Directory.

1. Introduction

The Scientific Committee on Antarctic Research (SCAR, www.scar.org) is an inter-disciplinary committee of the International Council for Science (ICSU). SCAR is charged with initiating, developing and coordinating high quality international scientific research in the Antarctic region (including the Southern Ocean), and on the role of the Antarctic region in the Earth system. SCAR promotes free and unrestricted access to Antarctic data and information by promoting open and accessible archiving practices. SCAR aims to be a portal to data repositories of Antarctic scientific data and information. SCAR's Standing Committee on Antarctic Data Management (SCADM) facilitates co-operation between scientists and nations with regard to scientific data, and advises on the development of the Antarctic Data Directory System.

Data and information are valuable and irreplaceable resources. Proper management of data and information is not an "add-on" or an additional task; it is a fundamental aspect of modern science.

In the pursuit of various scientific objectives, it is often necessary to use data and information collected by scientists from many countries. SCAR recognizes the critical and essential importance of the stewardship of data and information within national and international programmes and its accessibility to all.

SCAR has adopted a Data and Information Management Strategy (DIMS), developed by the SCAR Standing Committee on Antarctic Data Management (SCADM), to ensure that the scientific user community has adequate access to data and information.

2. SCADM

The Scientific Committee on Antarctic Research (SCAR) and the Council of Managers of National Antarctic Programmes (COMNAP) established the Joint Committee on Antarctic Data Management

(JCADM) in 1997 to manage Antarctic data. In December 2008 the formal linkage with COMNAP ceased and JCADM became SC-ADM from January 2009.

SC-ADM helps facilitate co-operation between scientists and nations with regard to scientific data. It advises on the development of the Antarctic Data Directory System and plays a major role in the International Polar Year data system (IPYDIS).

Members of SC-ADM are usually managers of the National Antarctic Data Centres or a relevant national contact.

3. SCAR Data Policy

In accordance with the Twelfth WMO Congress, Resolution 40 (Cg-XII, 1995); the Thirteenth WMO Congress, Resolution 25 (Cg XIII, 1999); the ICSU 1996 General Assembly Resolution; the ICSU Assessment on Scientific Data and Information (ICSU 2004b); Article III-1c from the Antarctic Treaty; the Intergovernmental Oceanographic Commission Data Exchange Policy and in order to maximize the benefit of data gathered under the auspices of SCAR Projects, the SCAR Executive Committee (EXCOM) requires that SCAR data, including operational data delivered in real time, are made available fully, freely, openly, and on the shortest feasible timescale.

The only exceptions to this policy of full, free, and open access are:

- where human subjects are involved, confidentiality must be protected;
- where data release may cause harm, and where specific aspects of the data

may need to be kept protected (for example, locations of nests of endangered birds).

ICSU defines "Full and open access" as equitable, non-discriminatory access to all data preferably free of cost, but some reasonable cost-recovery is acceptable. WMO Resolution 40 uses the terms "Free and unrestricted" and defines them as non- discriminatory and without charge. "Without charge", in the context of this resolution means at no more than the cost of reproduction and delivery without charge for the data and products themselves.

Metadata are essential to the discovery, access, and effective use of data. All SCAR data should be accompanied by a full set of metadata that completely documents and describe the data. In accordance with the ISO standard Reference Model for an Open Archival Information System (OAIS) (CCSDS 2002), complete metadata may be defined as all the information necessary for data to be independently understood by users and to ensure proper stewardship of the data. Regardless of any data access restrictions or delays in delivery of the data itself, all SCAR Projects should promptly provide basic descriptive metadata of collected data to the Antarctic Master Directory (AMD) system.

4. The Antarctic Master directory

The Antarctic Master Directory is the largest collection of Antarctic data set description in the world, holding over 7700 dataset descriptions from 25 countries. It is hosted by the Global Change Master Directory (GCMD) of the CEOS-IDN network to minimise duplication of resources and metadata.

In addition to the AMD portal, the GCMD has an IPY portal which highlights data that have been collected over the International Polar Year.

5. Other data Products

Besides the AMD SCAR has a number of other data products that includes a host of mapping resources, access to biological data (biodiversity. Aq, SO-CPR, SO-diet), environmental data (READER) and seismic data (SDLS).

The Canadian Consortium for Arctic Data Interoperability: An Emerging Initiative for Sharing Data Across Disciplines

Shannon Christoffersen Vossepoel[1*], ***Maribeth S. Murray***[1]

[1*] *Arctic Institute of North America, University of Calgary, ES 1040, 2500 University Dr. NW, Calgary, Alberta T2N 1A4, Canada*
Email: shannonv@ucalgary.ca

Summary. The Canadian Consortium for Arctic Data Interoperability (CCADI) is a consortium of Canadian-based polar institutions that aims to advance information sharing, nationally and internationally, through the development of an integrated Canadian arctic data management system. This proposed system will facilitate information discovery across disciplines and data types, establish metadata and data sharing standards, enable interoperability among Canada's existing data infrastructures, and will be accessible to a broad audience of users. Established in 2015, the CCADI (http://ccadi.ca/) has made great progress toward this vision and has laid the groundwork for future development.

Keywords. Canadian Consortium for Arctic Data Interoperability, information sharing, interoperability, standards, ethically open data.

1. Introduction

The Canadian Consortium for Arctic Data Interoperability (CCADI) consists of partners from academic, Inuit, government, and non-profit institutions. Current membership includes the Arctic Institute of North America, University of Calgary; GeoSensor Web Lab, University of Calgary; Innovis Lab, University of Calgary; Geomatics and Cartographic Research Centre, Carleton University; Centre d'études nordiques, Université Laval; Centre for Earth Observation Science, University of Manitoba; Faculty of Law, University of Ottawa; Canadian Cryospheric Information Network and Polar Data Catalogue, University of Waterloo; Inuit Tapiriit Kanatami; Inuvialuit Regional Corporation; Natural Resources Canada, Polar Knowledge Canada; Cybera Inc.; Polar View; and Sensor Up Inc [1]. This diverse partnership has established an excellent forum for working with the unique multidisciplinary data of the Arctic.

Created in 2015, the CCADI is committed to advancing collaboration, both nationally and internationally, around arctic information and data sharing [1].

2. The CCADI and Information Sharing

The CCADI is working to develop an integrated Canadian arctic data management system that will facilitate the discovery of information across numerous data types, both qualitative and quantitative; enable interoperability among existing arctic data infrastructures, both Canadian and international; establish metadata and data sharing standards for Canadian arctic data that will facilitate international data sharing; and that is accessible to a broad audience of users [1].

The CCADI is also working with Inuit organizations to ensure that Inuit Knowledge is appropriately represented; that Inuit are involved in the design and development of cyberinfrastructure involving their data and that they have stewardship over its distribution; and that Inuit Knowledge and western science can be explored for synergies and areas of interoperability. In support of this work, the CCADI follows the International Arctic Science Committee`s precepts for "ethically open data" as

outlined in their *Statement of Principles and Practices for Arctic Data Management* [2].

3. Progress Towards the Vision

Since its establishment, the CCADI has been making strides toward implementing its goals for information sharing. CCADI members are active contributors to the Arctic Data Committee (ADC) of the International Arctic Science Committee and Sustaining Arctic Observing Network, the International Study of Arctic Change (ISAC), the Open Geospatial Consortium (OGC), the Research Data Alliance (RDA), World Data System, and the Polar Libraries Colloquy [1]. CCADI members also meet regularly and have been successfully working on human interoperability within CCADI and growing our network of partners. Additionally, CCADI has been actively working on funding applications to begin development on the cyberinfrastructure required to support these goals.

4. Conclusions

The status of the CCADI initiative and its stated goals for advancing arctic information sharing in Canada and internationally are summarized in this paper. While still a new initiative, the CCADI has made some good progress towards its goals. Pending funding support, the CCADI is well-positioned to make significant advancements for arctic data and information sharing in the next five years.

Acknowledgments. The authors would like to express their appreciation and thanks to the fellow members of the Canadian Consortium for Arctic Data Interoperability as well as to the members of the Inuit National Data Management Committee.

References

1. The Canadian Consortium for Arctic Data Interoperability, http://ccadi.ca. [accessed on: October 2017]
2. International Arctic Science Committee. *Statement of Principles and Practices for Arctic Data Management*. April 16, 2013 https://iasc.info/data-observations/iasc-data-statement. [accessed on: October 2017]

Outline of the Arctic Data Archive System (ADS)

Hironori Yabuki[1*,2], ***Takeshi Sugimura***[2], ***Takeshi Terui***[2]

[1*] *Polar Environment Data Science Center, DS, ROIS, 10-3 Modori-cho, Tachikawa, Tokyo, 190-8518, Japan*
[2] *International Arctic Environment Research Center, NIPR, ROIS, 10-3 Modori-cho, Tachikawa, Tokyo, 190-8518, Japan*
Email: Yabuki.hironori@nipr.ac.jp

Summary. Arctic Data archive System(ADS), through proceed with the visualization and the development of online analysis system of integrated big data, aiming for integrated analysis information platform, not only as a mutual distribution platform of data, we have developed a system that enables open access research data and scientific knowledge obtained in the Arctic research. Various applications and services developed by ADS should not be used only in the Arctic but should be used as a bipolar data publish platform. The ADS team is currently preparing to publish not only Arctic data but also Antarctic data.

Keywords. Arctic Global warming, ArCS, Data Management.

1. Introduction

The easy access use is made possible from the industrial and the social public using research results(thesis and research data, etc.) using a public research fund, and a concept as open science aiming at linking it to creation of innovation by opening the new way as well as promoting a scientific technical research effectively is showing a rapid expanse to creation of worldwide. And the principle opening to the research result and data by a public research fund by GRC (Global Research Council), OECD (Organization for Economic Cooperation and Development) and G8 in 2013 etc.

Under these background, even in Arctic research, open access of a variety of variation mechanism and scientific knowledge, such as future prediction result brought about by actual grasp their environment change has been demanded.

In order to clarify the environmental variation system of complex Arctic with a variation of the time-space scale that is different consisting of air-land- marine, and human sphere, through interdisciplinary research, through interdisciplinary research, a wide variety of observational data, simulation data, satellite data, and even there is a need for the creation of A New knowledge of using the big data that integrates the research results. Also those integrated with the big data, scientific knowledge by using these, it is necessary to continue to properly publish to society.

2. Development of Arctic Data archive System(ADS)

Arctic Data archive System(ADS: https//ads.nipr.ac.jp), through proceed with the visualization and the development of online analysis system of integrated big data, aiming for integrated analysis information platform, not only as a mutual distribution platform of data, we have developed a system that enables open access research data and scientific knowledge obtained in the Arctic research.

ADS has been doing the systems development of following up to now.

- Metadata Management System by own metadata schema.(KIWA: Fig.1)
  - Metadata exchange system by using OAI-PMH, GI-cat.
  - Currently, this service is carried out in cooperation with GCW in WMO. Also this service have done coordination with GEO-Portal.

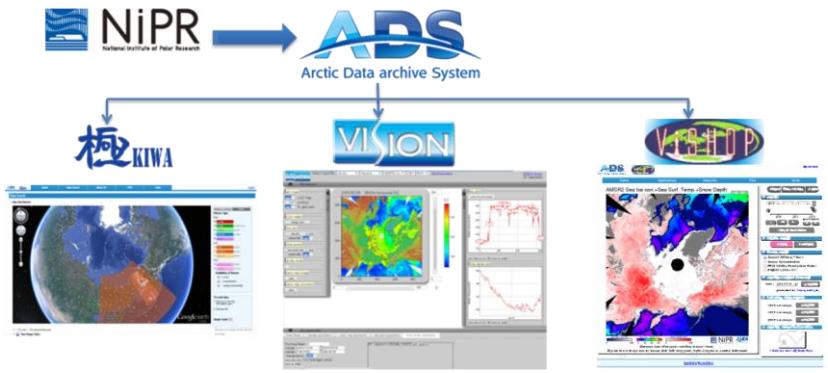


Fig.1 : Structure of ADS, Research data registration system and Metadata search service(KIWA), Online visualization application for Climate, Satellite and Simulation data(VISION) and Semi-real-time polar environ. obs. Monitor and Sea Ice prediction(VISHOP)

  - A system for space-time search using GoogleEarth collected data
  - DOI (Digital Object Identifier) registration system
- Visualization and analyzed system for the satellite data and grid data by online(VISION: Fig1)
- System to Semi-real-time polar environ. obs. monitor and sea ice prediction in the Arctic, Antarctic by using the satellite data (AMSR2) that is delivered in near-real-time from JAXA.(VISHOP:Fig1)
- System for visualizing numerical data such as time-series data(VISION-Graph)

3. Future development and challenges

ADS is not only a system that provides the data to various data users and stakeholders, in order to promote joint research and international cooperation in the Arctic region, anyone that is aimed at developing integrated analysis platform through the available Web interface. Furthermore in ADS, to developing of the information providing service of push-type in accordance with the needs of stakeholders.

By widely publish the technology developed in ADS, to promote the technology transfer system construction, to help the same technical problem solved in other areas. In ADS future, to carry out research and development the following items.

- Advancement of data and meta-data registration and retrieval system
- Promotion of data registration and data usage
- Enhanced of international cooperation of data and metadata
- Advancement of visualization, basic data analysis and the like of software and Web applications in order to provide an integrated analysis platform
- System construction of the push-type information service
- Advancement of small and medium-sized data server linkage function by ADS grid
- System Technology publishing, which is research and development in the ADS and technology transfer promotion to other systems.

4. Conclusions

The share of research data and scientific knowledge in the Arctic and non-Arctic nations, there are need for coordination of data repository and data center in a various country.

Important to drive the open-science, it is important data published and data cited, it is necessary to promote these data published and data cited. We, through the development of ADS activity, believe that can contribute to the sharing of research data and scientific knowledge in the Arctic and non-Arctic nations.

Polar Science and Data Journal by "Elsevier Publisher"

Takashi Yamanouchi [1*]

[1*] *National Institute of Polar Research, ROIS, 10-3 Midoricho. Tachikawa-shi, Tokyo 190-8518, Japan*
Email: yamanou@nipr.ac.jp

Summary. *Polar Science* is an international, peer-reviewed journal published by the National Institute of Polar Research and Elsevier to promote the Arctic and Antarctic sciences. Since the beginning of the first issue in 2007, more than 500 submissions were received and 371 manuscripts were published. The journal is believed to contribute greatly to the polar research communities. *Data in Brief* provides a way for researchers to easily share and reuse each other's datasets by publishing data articles that describe data, facilitating reproducibility. This increases traffic towards associated research articles and data, leading to more citations. *Data in Brief* is open access and covers data from all research disciplines. *Polar Science* offers authors the possibility to co-submit one or more data articles alongside their research article.

Keywords. polar journal, data journal, publication.

1. Introduction

Polar Science is an international, peer-reviewed quarterly journal published by the National Institute of Polar Research (NIPR) and Elsevier (https://www.journals.elsevier.com/polar-science/). It originated from the previous Proceedings series of symposia held by NIPR on Polar Upper Atmosphere Science, Polar Meteorology and Glaciology, Polar Geoscience, Antarctic Meteorite Research and Polar Bioscience. The first issue of *Polar Science* was published in August 2007 as a regular international scientific journal. The initiation of new journal was intended to promote the transmission of the results of polar research (especially of Japanese and Asian scientists) to the international communities and to produce a higher level of the circulation. Now, we have submissions from 39 nations, all over the world.

The journal, *Polar Science,* will be giving authors the opportunity to co-submit brief articles describing their data and methods alongside the main research article to dedicated open access cross disciplinary journals *Data in Brief* (http://www.journals.elsevier.com/data-in-brief) and *MethodsX* at Elsevier. This enables authors to get more exposure and credit for their work that otherwise can't be published in a traditional article format, and readers to get much more value out of the research data and methods.

In order to publish original research data/dataset, the new data journal, *Polar Data Journal*, has also been commissioned by NIPR.

2. Polar Science

Polar Science is dedicated to publishing original research articles for sciences relating to the polar regions of the Earth, as well as other planets. It aims to cover 13 disciplines that cover most aspects of physical, geo and life sciences. *Polar Science* also has an Open Archive whereby articles are made freely available from ScienceDirect after an embargo period of 24 months from the date of publication.

To encourage future research in the polar regions, restructuring of the disciplines is planned, especially the inclusion of social/humanity sciences. This direction was chosen due to the current trends of Arctic research to be much broader, not only in the field of natural science but also to include discussions with stake-holders (Indigenous people living in the Arctic, policymakers and citizens). Also in relation to Antarctic science (e.g., SCAR: Scientific Committee on Antarctic Research/ICSU) a new need for research in humanities and history has been recognised.

3. Data in Brief

Data articles are brief, peer-reviewed publications about research data. Sharing data makes it accessible and enables others to gain new insights and make interpretations for their own research. Thanks to a detailed dataset description, the data published in data articles can be reused, reanalyzed and reproduced by others.

Data articles are easy to submit and subject to a quick and transparent peer review process. These articles ensure that data and the metadata to understand it are actively reviewed, curated and formatted. In addition, the articles will be indexed and made available immediately upon publication.

Data in Brief is an open access journal that publishes data articles. Not all the data collected during the research cycle makes it into a final publication; data articles complement full research papers, providing an easy channel for researchers to publish their datasets and receive proper credit a recognition for the work they have done. This is particularly true for replication data, negative datasets or data from intermediate experiments, which often go unpublished.

Sharing data helps reproduce results, increases transparency and can enable other researchers to build on previous work. Data articles provide a peer reviewed, quick and easy way to share and get cited for the work that goes into collecting data, in addition to making data available.

Data in Brief provides a way for researchers to easily share and reuse each other's datasets by publishing data articles that describe data, facilitating reproducibility. This increases traffic towards associated research articles and data, leading to more citations. *Data in Brief* is open access and covers data from all research disciplines. A growing number of Elsevier journals in all research areas offer authors the possibility to co-submit one or more data articles alongside their research article.

At the research article revision stage, authors will be given the option to convert some of their supplementary data and/or methods-related supplementary material into one or multiple *Data in Brief/MethodsX* articles, a new kind of articles that describe research data and details of customized research methods. This gives authors a chance to publish additional peer reviewed articles, which will be more discoverable and citable than supplementary data at the end of their research paper. The research article in *Polar Science* will be linked with the co-submitted *Data in Brief* and/or *MethodsX* articles once they are accepted and get published so that these articles are easy to find.

The NIPR foresees the trends of data sharing and publication. To provide a home for publishing the original data/data set of polar sciences and technologies, the NIPR has recently launched a new data journal titled "Polar Data Journal" (https://pdr.repo.nii.ac.jp/).

4. Conclusions

Polar Science is an international, peer-reviewed and quarterly journal. It is dedicated to provide original research articles for science related to the polar regions of the Earth and other planets. *Polar Science* includes 13 disciplines; they cover most aspects of physical, geo and life sciences. Those articles should be of interest to the broad polar science community, not limited to interests of those who only work under specific research subjects.

Authors can submit a brief article describing their data alongside their main research article to a dedicated OA journal *Data in Brief* for peer review and publication. This enables authors to get more exposure and credit for their work that otherwise can't be published in a traditional article format, and readers to get much more value out of the data, methods and software.

Alternative platforms are available for publishing original research data/data sets. The new *Polar Data Journal* launched by the NIPR is one of these platforms, which encourage more Polar science researchers to share, publish and further cite data,

Acknowledgments. We would appreciate the help we received from several Elsevier Publishers, David Parsons, Emilie Wang and Young Wu, and also the efforts to edit and publish the paper by Ms. Mayumi Asano of the Polar Science Editorial Office.

International Workshop on Sharing, Citation and Publication of Scientific Data across Disciplines

AUTHORS INDEX

Joint Support-Center for Data Science Research (DS),
Tachikawa, Tokyo, Japan

5–7 December 2017

Authors Index :

(Alphabetical Order; First Author Only; Indicate ("Poster" Presentation-Number))