



Polar Data Forum IV – An Ocean of Opportunities

PRACTICE PAPER

ANNEMIE ROSE JANSSEN

PHILIPPA BRICHER

KAREN PAYNE

RENUKA BADHE

NICOLE BIEBOW

TACO DE BRUIN

RUTH DUERR

PJOTR ELSHOUT

ALLISON GAYLORD

ØYSTEIN GODØY

PATRICK GORRINGE

DAMIEN GUIHEN

JOHNATHAN KOOL

JAN RENE LARSEN

JOSEPH NOLAN

ANTONIO NOVELLINO

WILLIAM MANLEY

ANDREAS MAROUCOS

MOLLY MCCAMMON

MARIBETH MURRAY

JENN PARROTT

JAY PEARLMAN

HELEN PEAT

PETER PULSIFER

LLOYD SYMONS

MARTEN TACOMA

SOULAINE THEOCHARIDES

STEIN TRONSTAD

CHANTELLE VERHEY

ANTON VAN DE PUTTE

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

ubiquity press

CORRESPONDING AUTHOR:

Anton Van de Putte

Royal Belgian Institute of Natural Sciences, Brussels, Belgium; Université Libre de Bruxelles, Brussels, Belgium

avandeputte@naturalsciences.be

KEYWORDS:

Polar data management; open access data; data policy; interoperability; polar federated search; semantic data; best practices

TO CITE THIS ARTICLE:

Janssen AR, Bricher P, Payne K, Badhe R, Biebow N, de Bruin T, Duerr R, Elshout P, Gaylord A, Godøy Ø, Gorringer P, Guihen D, Kool J, Larsen JR, Nolan J, Novellino A, Manley W, Marouchos A, McCammon M, Murray M, Parrott J, Pearlman J, Peat H, Pulsifer P, Symons L, Tacoma M, Theocharides S, Tronstad S, Verhey C and Van de Putte A. 2023. Polar Data Forum IV – An Ocean of Opportunities. *Data Science Journal*, 22: 18, pp. 1–13. DOI: <https://doi.org/10.5334/dsj-2023-018>

ABSTRACT

This paper reports on the Hackathon Sessions organised at the Polar Data Forum IV (PDF IV) (20–24 September 2021), during which 351 participants from 50 different countries discussed collaboratively about the latest developments in polar data management. The 4th edition of the PDF hosted lively discussions on (i) best practices for polar data management, (ii) data policy, (iii) documenting data flows into aggregators, (iv) data interoperability, (v) polar federated search, (vi) semantics and vocabularies, (vii) Virtual Research Environments (VREs), and (viii) new polar technologies. This paper provides an overview of the organisational aspects of PDF IV and summarises the polar data objectives and outcomes by describing the conclusions drawn from the Hackathon Sessions.

1. INTRODUCTION

The Polar Data Forum (hereafter shortened to PDF) is a place where polar data holders get together and make more use of data. The Forum has two main components: the Conference, where the border between funding, policy, and data is explored through presentations and posters; and Hackathon Sessions, where the polar data community opens the dialogue to make progress on their shared objectives.

Since its first edition in 2013, the PDF has grown in terms of participation numbers and diversity of sessions. This year, PDF IV was hosted by the Royal Belgian Institute of Natural Sciences (RBINS) and the European Polar Board (EPB) and organised in close collaboration with the 2nd Southern Ocean Regional Workshop for the United Nations Decade of Ocean Sciences for Sustainable Development (hereafter referred to as the UN Ocean Decade). This fusion enabled cross-fertilisation of ideas and highlighted data management issues that supported UN Ocean Decade activities in the Southern Ocean, such as those detailed in the Southern Ocean Action Plan (Janssen et al. 2022), as well as broader polar data management issues. The data needs for the Southern and Arctic Oceans have many commonalities and many science organisations collect data at both polar regions, there is particular value in developing solutions that work for both the Arctic and Antarctic.

The PDF IV hosted lively discussions on the emerging field of Virtual Research Environments (VREs), new polar technologies, federating metadata search, and how to document data flows into aggregators. Progress was made on documenting best practices and refining data policies for all three polar data committees (the SOOS Data Management Sub-Committee – DMSC, the Standing Committee on Antarctic Data Management – SCADM, and the Arctic Data Committee – ADC) (Tronstad et al. 2021).

The main objective of this paper is to familiarise readers with the latest developments in polar data management by summarising and providing key future actions discussed at the PDF Hackathon Sessions. This paper is organised as follows: the first three sections will acquaint readers with the context of the PDF and characterise the organisation of Hackathon Sessions and their participants. The remainder of the paper will provide a summary of the objectives and outcomes of each Hackathon Session. Finally, we will reflect on the experience of hosting an online Forum and highlight the importance of the PDF for Open Data Science by describing the conclusions drawn from the Hackathon Sessions.

2. HISTORY OF THE POLAR DATA FORUM

The Polar Data Forum (PDF) is a place where polar data holders get together and make more use of data. The Forum has two main components: the Conference, where the border between funding, policy and data is explored through presentations and posters; and Workshop Sessions and Hackathons, where the Polar Data Community opens the dialogue to make progress on their shared objectives (PDF IV Scientific Steering Committee, 2021).

Since its first edition in 2013 in Tokyo, Japan, the PDF has grown in terms of both participation numbers and diversity of sessions. Polar Data Forum I (International Forum – Polar Data Activities in Global Systems 2013) identified issues and made observations and recommendations on polar data management. PDF I highlighted the need to improve ways in which people and systems share data in a meaningful way in order to develop open and connected systems based on a culture of trust and acknowledgement of data production and use.

In 2015, PDF II selected similar themes to PDF I and focussed on discussing the significant progress made since 2013. During PDF II, new priorities for polar data management were identified as well as key new themes that have evolved. The community also planned a set of action-oriented recommendations and activities.

The Marine Data Workshop organised at PDF III (PDF III Scientific Steering Committee, 2021) in 2019 in Helsinki, Finland discussed and agreed on a strategy to unlock marine data not freely available via well-established interconnected data portals such as the Copernicus Marine Service, the In Situ Thematic Centre (INS TAC), the European Marine Observation and Data Network (EMODnet), and others. The workshop focussed on (i) the implementation of the FAIR

principles by organisations collecting and managing polar observing networks, (ii) negotiations with international and national data centres with restrictive data policies, (iii) the involvement of private companies, and (iv) the establishment of support systems for research organisations with limited data management facilities.

3. POLAR DATA FORUM IV

With the aim of further broadening the polar data community, sharing knowledge and experience, and collaboratively working towards finding solutions to common objectives, the fourth edition of the PDF was organised from the 20th to the 24th of September 2021. The Forum welcomed 351 participants from 50 different countries, spread across all continents of the globe.

3.1. FORUM ORGANISATION

This edition welcomed a total of 46 presentations spread across the 10 selected themes for the Conference component of the Forum: (i) FAIR data (Findable, Accessible, Interoperable, Reusable) (Wilkinson et al. 2016; Tanhua et al. 2019), (ii) federated metadata search, (iii) vocabularies and semantic interoperability, (iv) data for modellers and remote sensing, (v) co-production of data, information and knowledge, (vi) new ships and real-time data in low connectivity locations, (vii) logistical information management, (viii) knowledge mobilisation and decision making, (xi) data policy, and (x) barriers to data sharing and user needs. Recordings of PDF IV conference presentations are available on the [EPB's YouTube channel](#).

Polar Data Forum IV also hosted eight Hackathon sessions (detailed in Section 3, see [Table 1](#)) revolving around (i) best practices for polar data management, (ii) data policy, (iii) documenting data flows into aggregators, (iv) data interoperability, (v) polar federated search, (vi) semantics and vocabularies, (vii) Virtual Research Environments (VREs), and (viii) new polar technologies. These events gathered participants in an intensive collaborative work environment dedicated to finding solutions towards specific challenges. Most of the Hackathons Sessions in this edition were led by already well-established hacking teams, which allowed us to follow-up the work of past meetings and further boost long-term collaborations, while also welcoming new members and early-career polar professionals.

Table 1 Overview of the Hackathon Sessions organised at PDF IV.

HACKATHON NAME	LEAD(S)
Best practices for Polar Data Management – an interactive session bringing together polar researchers and polar data managers	Helen Peat and Johnathan Kool
Data Policy	Stein Tronstad, Peter Pulsifer, and Maribeth Murray
Documenting Data Flows into Aggregators	Philippa Bricher, Antonio Novellino, Jay Pearlman, and Patrick Gorringer
Interoperability for polar observing networks and logistical resources	Joseph Nolan, Bill Manley, Jan Rene Larsen, and Allison Gaylord
Polar Federated Search	Philippa Bricher, Chantelle Verhey, Ruth Duerr, and Taco de Bruin
Semantics and Vocabularies	Chantelle Verhey
VREs for Polar Oceans	Soulaine Theocharides, Anton Van de Putte, and Karen Payne
Ocean technology supporting polar science for the next decade	Andreas Marouchos, Damien Guihen, and Lloyd Symons

3.2. HACKATHON ORGANISATION

In order to comply with the open data principles, Hackathon Sessions were accessible to all registered participants. Hackathon themes were pre-selected by the PDF Science Committee and organised in collaboration with external polar data experts. Hackathon conveners were tasked with overseeing the organisation of their session and providing participants with a clear scope and agenda.

Hackathons were hosted online using Zoom, a communication platform for online meetings. Every Hackathon team used Google Docs to report on their progress towards their shared challenges. These reports served as a basis to write this paper. A wrap-up video summarising the objectives and outcomes of each Hackathon Session is available on YouTube.

4. HACKATHON SESSIONS

4.1. BEST PRACTICES FOR POLAR DATA MANAGEMENT – AN INTERACTIVE SESSION BRINGING TOGETHER POLAR RESEARCHERS AND POLAR DATA MANAGERS

Summary

Members of the Arctic and Antarctic data communities held a workshop to interactively engage with researchers and program leaders. They discussed how data management expertise can help collate datasets and improve the trust that researchers can have in data compilations and data products. This workshop enabled interested participants to come and discuss their project together with a range of data management experts and acted as a forum to provide information, offer advice, and help to build networks.

Hackathon proceedings and outcomes

This Hackathon Session brought together data managers and researchers to explore ways that they can facilitate sharing datasets for both data providers and users. One of the key challenges that was highlighted by the community was to balance how much effort to put into managing raw data and producing analysis-ready products. A further challenge, especially relevant in an interdisciplinary community, is that data formats chosen by a data provider may not be readily usable by scientists from other backgrounds. Thus, data managers in the session highlighted the increasing trend towards separating data formats from back-end storage, which improves both machine- and human-readability by making datasets available in a range of formats. Furthermore, there was discussion on the best practices for storing code to support the reproducibility of data processing, analysis, and modelling. Finally, it was agreed that the handling of highly bespoke data types will continue to be a challenge. However, this challenge can be minimised by scientists standardising collection processes and formats through developing and following best practices wherever possible. It can also be minimised by making data management part of the experimental design, rather than an afterthought.

4.2. DATA POLICY

Summary

At the Polar Data Forum III a process was initiated to develop a basis for alignment of polar data policies. The process, involving data managers and experts from international polar data committees in both the northern and southern hemispheres, resulted in a report recommending ten fundamental principles for polar data policies. During this Hackathon Session, those principles were examined as well as other recent developments relating to international data policies. During this Hackathon experts discussed the next steps towards further alignment across polar and global scientific communities and observation systems. Representatives of the relevant organisations were invited to present their views on the continued alignment process.

Hackathon proceedings and outcomes

This session began with an update on the progress of a community-agreed set of recommendations for aligning polar data policies. Some recent developments were highlighted for incorporation in the policy alignment document, including the UNESCO Recommendation on Open Science (UNESCO 2021). Certain aspects of Indigenous data management were also discussed, *inter alia* the issue of sensitivity and how to allow metadata on sensitive data to be made available without issues. The following discussion on the data policy recommendations led to the finalisation of the reporting for the policy statements (Tronstad et al. 2021).

There were discussions on the data policy revision processes of the World Meteorological Organization (WMO) (WMO 2021) and the Intergovernmental Oceanographic Commission (IOC) (UNESCO-IOC 2019) of UNESCO. Planning discussions focussed on the processes the group needs to follow to update the data policies of the Arctic Data Committee (ADC), Scientific Committee on Antarctic Research (SCAR), and Southern Ocean Observing System (SOOS), so that all three polar data committees can produce policies that follow the common set of recommendations. Alignment of these processes with the initiative to develop a data policy for the Arctic Council was also discussed. This initiative is being coordinated by the Arctic SDI Group and was planned

to be pursued at an Arctic data policy workshop during the Arctic Science Summit Week (ASSW) in Tromsø in March 2022. However, following the Russian invasion of Ukraine in February 2022 and subsequent ongoing war, the Arctic Council has suspended its activities.

4.3. DOCUMENTING DATA FLOWS INTO AGGREGATORS

Summary

With several countries building new polar research vessels and the associated data systems, there is a growing need to standardise data handling processes. One example is the development of an event logger deployment vocabulary to standardise description of actions and processes for ship instrument deployments. To minimise duplication of efforts, it makes sense to discuss development of standardised terminology that can be broadly applied on any marine platform.

Hackathon proceedings and outcomes

Organised as a joint session with the UNESCO-IOC Ocean Best Practices System's Fifth Annual Community Workshop, this Hackathon Session brought together participants from oceanographic data centres, data aggregation portals, and the Ocean Best Practices System. All participants acknowledged the need for more transparency in data sources and how the data is being aggregated. Such transparency would support users of portals to improve their assessment of the completeness of what they're accessing and how data has been treated up to the portal presentation. It would also help portal administrators to better identify potential gaps and duplications in their holdings, and more readily test for and fix errors in data feeds.

Participants developed a list of core fields to describe a harvesting relationship between a single source data centre and one aggregator. There was extensive discussion on the best ways to document multiple steps of Quality Control/Quality Assurance, when there are nested aggregations that involve individual datasets going through multiple data centres before they reach a particular data aggregator. It was agreed that each aggregator should pass on all metadata and processing descriptions from all previous steps.

The group concurred to engage the Global Ocean Observing System (GOOS) and explore establishing a new task team to agree on standards for documenting these relationships, as there is a distinct community need for these relationships to be publicly documented.

4.4. INTEROPERABILITY FOR POLAR OBSERVING NETWORKS AND LOGISTICAL RESOURCES

Summary

Within the polar and oceanographic communities, there is an increasing interest to integrate databases of observing assets, networks, and logistical resources, including research stations, projects, vessels, and expeditions. Integrating these databases requires crosswalks of metadata models and the agreement of best practice methods for sharing the information between catalogues in order to achieve a unified view within and beyond the polar regions.

Hackathon proceedings and outcomes

This session began with updates on progress from a number of projects that aim to improve the interoperability of databases holding information on polar observing networks and polar observing assets. The group reported on Polar dex – a new online application for discovery of polar research infrastructures and observing assets, logistical planning, and coordination, developed by the European Polar Board (EPB) in cooperation with several partners ([European Polar Board, 2021](#)). Polar dex served as a use case, demonstrating the need and value of improved interoperability in polar observing networks and logistical resources, and the need for a metadata standard when describing these. Furthermore, the Sustaining Arctic Observing Networks (SAON) Polar Observing Assets Working Group (POAwg) discussed their efforts to improve integration by focussing on interoperability parameters about research and monitoring assets ([SAON-POAwg, 2021](#)). Finally, there were discussions on an EU-PolarNet 2 initiative to develop a procedure for ongoing collection and collation of European polar observing capacities and activities. The session concluded with a broad invitation for participants to join one or more of the polar asset cross walking activities.

4.5. POLAR FEDERATED SEARCH

Summary

Federated metadata search for the polar regions is dependent on the data centres that host polar-relevant data being able to present discovery metadata in a common way. The Polar Federated Search Working Group (POLDER) is currently developing a Best Practices guidance for the implementation of schema.org as a potentially lightweight discovery metadata standard serving long-tail data in particular. [Schema.org](https://schema.org) is a collaborative, community activity with a mission to create, maintain, and promote schemas for structured data on the Internet, on web pages, in email messages, and beyond. The polar guidance contributes to and draws on similar efforts in related science communities. This session made progress on the development of that Best Practices documentation and informed on the development of the pilot federated search tool (POLDER, 2021) that is currently being supported by the World Data System (WDS).

Hackathons proceeding and outcomes

During this session, POLDER continued to develop its Best Practices guidance for using schema.org as a potentially lightweight interchange standard for federating discovery metadata. This is an ongoing effort that goes beyond the Polar Data Forum, through the Polar to Global (P2G) Hackathons Sessions.

A team from Carleton University in Canada introduced the Mapping the Polar Data Ecosystem (MPDE) – a tool for storing, updating, and sharing information on metadata harvesting relationships among polar data centres that POLDER collated during 2018 and 2019. The new tool takes information that took significant community time to collate and puts it in a data structure that readily supports filtering and updating of the data, as well as providing tools to visually explore the harvesting relationships among metadata catalogues.

The final section of the Hackathon discussed the direction of the activities of the World Data System – International Technology Office (WDS-ITO) pilot federated search tool. The time allotted to POLDER's Polar Pilot Federated Search (PPFS) was used to discuss recent developments in the project. This included the host domain, potential repositories to be included, and to use the space to announce the second advisory team meeting where these topics could be discussed in further detail. Additionally, this was the first time the polar community had the opportunity to meet the dedicated web developer hired for this project, and were able to voice any comments, questions, and concerns in their direction. Overall, this session represented the official beginning of development for the PPFS tool.

4.6. SEMANTICS AND VOCABULARIES

Summary

The Semantics and Vocabularies Hackathon worked on populating a Gap Analysis regarding the current state of the Polar Semantic Landscape. Stemming from the work of the Semantic harmonisation Cluster from Earth Science Information Partners (ESIP) where the sea ice vocabularies of Semantic Web for Earth and Environmental Terminology (SWEET) and Environment Ontology (ENVO) were aligned, this session was used to identify where resources can be allocated moving forward. The Hackathon Session identified the most common vocabularies, and how they are being used in polar settings, and prioritised which would be useful to align next. This can be used to ensure inclusion of Indigenous semantic resources, and/or outline a plan for how to include these in future.

Hackathon proceedings and outcomes

The community was able to collaboratively identify its strengths, weaknesses, opportunities, and threats and determined what an ideal minimum state would look like. This information was then converted into a SWOT Analysis visualisation (Verhey 2022). The group expanded upon the ADC Semantic Working Group's compilation of polar specific ontological resources, with Hackathon attendees reviewing the current state of the list and adding any additional resources they were not already included. Additionally, it was decided that the resource would benefit from inclusion of information such as a description of each resource, the label of resource hierarchy, and simple use cases. The further development of this work will be carried out in the space of the ADC Semantic Vocabulary Working Group. Finally, the Hackathon participants

brainstormed an ideal state of polar semantic resources. This entailed identifying a need for a set of controlled vocabularies and building tools around a small subset; continued collaboration on an international level and sustainable funding to continue this work; and continued outreach to scientific community and Indigenous communities to ensure semantic priorities support specific practical use cases. Moving forward, the community intends to utilise semantics to further enhance meta(data) interoperability. By incorporating standardised vocabularies, with structured identifiers in metadata records, the community is able to utilise existing infrastructure to enhance the Findability, Accessibility, and Interoperability of the FAIR principles.

4.7. VRES FOR POLAR OCEANS

Summary

Virtual Research Environments (VREs), also known as Science Gateways or Virtual Labs, are digital platforms or programs that provide varying degrees of collaboration, computational power and services designed to support collaboration and communication in teams of researchers. The goal of this session was to articulate the needs of a VRE for the polar data community. It was both a needs assessment and visioning exercise that asked participants who represent many different roles what features or components they need in a VRE. Additional discussion points included compiling existing VREs used by the polar research community, and what makes polar data unique.

Hackathon proceedings and outcomes

Virtual research environments are an emerging concept in polar marine environments. Discussions in this session revealed highly varied understandings of the concept of virtual research environments. Participants' backgrounds included software developers, database administrators, and researchers (biology, meteorology, geosciences, and oceanography, etc.).

Participants had highly diverse needs in terms of platforms, data types, and analysis tools. The group identified a crucial need for stakeholder-specific solutions for each VRE. This included discussions about the accessibility of web applications. While these applications were described as more democratic and accessible for many users, bandwidth requirements can restrict use for those in remote communities and in the field. It was acknowledged that the applications' analytic capabilities should cater to the intended audience. The group recognised that public-facing visualisation tools should be simple and intuitive to encourage use by people outside the polar data community, while community-specific tools may expect more knowledge of back-end calculations. Participants also identified a need for the polar marine community to focus on developing analysis-ready data products. To ensure reproducibility, the metadata associated with these products should be made complete (specifically, metadata associated with version control).

4.8. OCEAN TECHNOLOGY SUPPORTING POLAR SCIENCE FOR THE NEXT DECADE

Summary

Observing and understanding the Southern Ocean presents significant challenges due to its remoteness and extreme weather conditions. To address these challenges, changes are needed with regards to the science community's approach to developing technology and deploying infrastructure. Therefore, it is essential to discuss how the changing landscape of infrastructure investment focused on Southern Ocean observing will enable a seed change in observations for the next decade. Topics for this Hackathon Session included trends in infrastructure investment, autonomy, maximising asset use and coordination, and interoperability.

Hackathon proceedings and outcomes

This session began with a discussion on the problems that emerging polar technologies need to solve, including being interoperable, able to operate at scale and able to provide more flexibility in multi-mission parameters. These pressures are being increasingly driven by the need to improve return on investment for funders, who need to collect more data at lower costs. There was general agreement that the future of ocean observing will involve a mix of monolithic and distributed assets. These solutions are more likely to be found when communication and engagement are well established between scientists and the engineers/technicians who are creating the new technologies.

The group allocated significant discussion time to focus on cheap, capable, and sacrificial assets which can provide observations in hard-to-access areas (e.g., under ice), where there are higher risks of equipment being lost. Consequently, this raised questions about the impact of polar technology on the equity of access to polar science; with new technologies often being expensive to develop and purchase. This inequity may leave less-affluent or less-developed polar programmes behind. The group also addressed the challenges for sacrificial assets, where agreement needs to be struck between the technical providers and science programs about what are considered as acceptable risks.

Finally, the group highlighted some of the common issues between providers and scientists, such as requirements for data obtained from polar technology to be made FAIR. This issue can be problematic for companies that provide technology in terms of maintaining the confidentiality of the data where their for-profit structure requires. Additionally, vastly larger data streams will increase demands on data managers. Engineers in the session discussed the need for clearly-defined Essential Ocean Variables (EOVs) or key variables in the ocean sciences, to enable them to better prioritise sensor designs and allocation of payloads on observing platforms.

5. DISCUSSION

The discussion revolves around three main topics. First, we reflect on the experience of hosting an online Forum. Second, we review the outcomes of the PDFs and highlight the importance of the PDF for Open Data Science. Finally, we reflect on past PDFs legacies and discuss ways to refine the polar community's efforts in order to calibrate a new, more powerful, flexible, and inclusive model for future PDF meetings.

5.1. A VIRTUAL POLAR DATA FORUM

Due to the COVID pandemic, like many scientific events during this time, PDF IV was organised as a virtual online meeting. This brought a few logistical challenges such as providing a dynamic and enjoyable meeting environment, setting a manageable schedule in order to ensure a balanced and representative participation across the globe, and to maintain a high level of participation during the entire event.

By advertising the event and session themes early and sending frequent reminders to the polar community, the PDF IV Organising Committee and Hackathon Chairs were successful in keeping momentum and strong engagement throughout the whole Conference and Hackathon Sessions. Compared to previous editions of the Polar Data Forum, PDF IV welcomed higher numbers and more diverse participants: compared to 110 in-person participants at PDF II in 2015, PDF IV gathered 351 participants from 50 different countries, spread across all continents of the globe. The online format of PDF IV enabled the waiving of costs for participants, which would normally entail flights, hotels, registration fee, etc.). Therefore, the 'no-cost' benefit of PDF IV was also a factor which increased the accessibility of the event to a wider community. The attendance was relatively stable during the entire event with an average of 214 participants per day from Monday 20th until Wednesday 22nd September. Thursday 23rd and Friday 24th welcomed 186 and 133 participants, respectively.

As noted, participant numbers for PDF IV were high in comparison to previous editions, with a wide geographic distribution. However, notable was the modest number of participants from some key countries, regions and institutions in the polar community (e.g., Russia, China, Japan, South Korea, Finland (hosts of PDF III), etc.). This may have been a result of limitations to the extent of organisers' networks for announcements in these communities, language barriers or other issues.

Despite the benefits of an online PDF, there were also (well-known) drawbacks of such a format. These included 'Zoom fatigue' of participants attending multiple sessions back-to-back, often with other tasks taking part of their attention simultaneously. Additionally, the globally distributed locations of participants across time zones meant unsociable working hours were necessary for many participants to attend relevant sessions. Furthermore, despite the provision of a virtual 'coffee break area' by organisers, opportunities for informal, community-building interactions between participants were limited and under-utilised. These drawbacks considered, feedback from participants indicated a well-organised and engaging event, with appreciation noted for efforts to make PDF IV accessible and inclusive.

5.2. RELEVANCE OF POLAR DATA FORUM FOR OPEN DATA SCIENCE

Since its first edition in 2013, the PDF has been a place where polar data holders gather and discuss how to make more and better use of data. This includes advocating for open data access and finding ways to do this effectively by applying the FAIR principles (Wilkinson et al. 2016). This fourth edition of the PDF gathered participants in an intensive collaborative work environment dedicated to finding solutions towards specific challenges (see Table 1). Most of the Hackathon Session in this edition were led by already well-established hacking teams, which allowed them to make progress with regards to previous PDF meetings. Polar Data Forum IV Hackathons also therefore allowed hacking teams to maintain and strengthen long-term collaborations, while welcoming new members, including early-career polar professionals. Polar Data Forum IV Hackathon Sessions enabled each group to make progress on shared objectives:

- I. The Best Practices Group held broad and high-level discussions regarding polar data management, including the need to balance preservation of raw data and analysis-ready products.
- II. The Data Policy Group (SOOS-DMSC, SCADM, and ADC) is seeing its alignment of data policies slowly coming to fruition.
- III. The Data Flows Group was one of the new groups attending PDF IV. This group acknowledged the need for transparency and preservation of metadata with regards to documenting what is being aggregated (and where from) by developing a list of core fields.
- IV. The Interoperability Group discussed the need to develop standards about information sharing for polar observing assets such as projects, sites, and more.
- V. POLDER is one of the long-standing groups and has been meeting bi-monthly during the P2G Hackathon Sessions. This group used PDF IV to make further progress on developing the Best Practices guidance for implementing schema.org.
- VI. The Semantics and Vocabularies Group did a suite of gap analyses on the Polar Semantic Landscape and expanded the compilation of polar specific ontological resources.
- VII. The VRE Group had a broad discussion. One of the major outcomes of this session was that there are highly varied understandings of the concept of VREs stemming from the various needs of the polar community.
- VIII. The Ocean Technology Group, like the Data Flows Group, was also new to the PDF and enabled to make the link with the UN Ocean Decade by discussing ways that ocean technology can support polar science in the next decade, and the implications for data managers due to expected shifts in monolithic observing systems that will increase data streams and, in turn, escalate demands on data managers.

Overall, PDF IV enabled the polar research and data communities to identify many improvements to be made for data management and policy. Regarding data policy, much progress has been made towards the alignment of the three polar data committees' policies which follow a common set of recommendations (Tronstad et al. 2021).

Polar Data Forum IV underlined that including data management as part of the experimental design is instrumental to facilitate the standardisation of collection processes and formats. Implementing this, alongside the use of controlled vocabularies and making data available in a wide variety of formats, will support the reproducibility of data processing, analysis, and modelling globally. Improved reproducibility would improve the quality of scientific research and its impacts. Furthermore, having analysis-ready data, with complete, transparent, and accessible metadata will not only improve the interoperability of datasets, but it will also help the polar community to better identify potential gaps and avoid duplication of efforts. Furthermore, the community acknowledged the need to support the development of data sharing tools that store information and place it in a data structure that readily supports filtering and updating of the data, while also offering the possibility to visually explore the harvesting relationships among polar metadata catalogues. For data users, it is important that such tools are user-friendly and cater to their varied needs. Identifying and understanding the needs of data users is essential to improve already-existing resources, or to develop new platforms and analysis tools for stakeholder-specific solutions.

The inequity in access to polar science needs to be addressed. Since solutions are more likely to be found when communication and engagement are established, there is a need to strengthen collaborations on an international level to keep the dialogue open between scientific, Indigenous and data communities as well as the engineers and technicians who develop necessary new technologies. Additionally, there is also an unequal access to funding for polar research around the world, with the bulk of polar research currently being funded by governmental agencies. Therefore, to improve access to the polar regions, there is a need to attract a diversity of funding, and a better communication of the return of investment.

5.3. LESSONS LEARNED AND THE WAY FORWARD

As a culmination of a series of meetings, PDF IV taught the organisers and the polar data community as a whole the value of persistent coordinating bodies, broad collaboration, sharing of human and financial resources, learning and innovation through iteration, and being open to adaptation to meet challenges. Polar Data Forum I in 2013 came at an important time in the evolution of the polar data community. The successful Fourth International Polar Year (IPY), which saw a surge in polar data activities and community building, came to an end in 2012. Thus, PDF I, initiated by the World Data System, SCAR, IASC, CODATA, and other established coordinating bodies in 2013, was an important event in terms of maintaining the momentum established during the IPY. This highlighted the importance of building and maintaining persistent coordinating bodies that go beyond time-limited projects and programs like the IPY, or mandate-specific working groups. These persistent bodies can provide important continuity.

The success of PDF I provided a platform for PDF II when the newly formed Arctic Data Committee joined with the original aforementioned conveners and well-established groups like the Standing Committee on Arctic Data Management, SAON, AMAP, and others to innovate on the previous success. Polar Data Forum II introduced three important innovations. First, organisation and funding of the event was shared between two primary organisers and many other supporters. The meeting was held in Waterloo Canada, while much of the funding was provided by the US National Science Foundation through the University of Colorado. This demonstrated to the community that the PDF could be organised through the partnership of many contributors rather than the traditional model that sees a 'host organisation' take on most of the planning and execution. Second, the format of the PDF was expanded from a two day plenary presentation focused event, to a six day event that included, posters and lightning talks, a number of business meetings of groups such as SCADM, ADC, Polar View, and the first joint meeting of SCADM and ADC. This allowed community members to become more active in their collaboration. Third, PDF II included representation from Arctic Indigenous organisations resulting in discussions of critically important topics such as Indigenous community engagement, and Indigenous data sovereignty.

The Third Polar Data Forum continued in the tradition of working under the leadership of persistent organising bodies while organising at a local level (led by the Finnish Meteorological Institute). The major innovation at PDF III was the addition of a series of 'hackathons' that focused on producing specific, collaborative results. This set the stage for the P2G Hackathon series established in June of 2020 and ultimately, the PDF IV hackathons. While knowledge sharing through presentations, posters etc. is a very important part of the PDFs, these hackathons have become a core method of collaboration in the polar data world and are a major achievement of the PDF series.

The combination of persistent coordinating bodies, broad collaboration among organisations and individuals, sharing of human and financial resources, and incrementally modifying and improving the PDF meetings provided a foundation for PDF IV. The major innovations and achievements and innovations of PDF IV are the focus of this paper and will not be repeated here, however, it is important to note two very significant lessons learned. For the first time, PDF IV included sufficient resources to add the support of a dedicated, experienced professional (the lead author) who could focus attention on all aspects of this large undertaking. This took the PDF to 'the next level'. Second, the PDF IV organisers were faced with bringing the community together during a global pandemic. While PDFs I–III provided valuable foundational elements on which to build, the ability of the PDF IV organisers to adapt and innovate through this unprecedented time set it apart from the previous PDFs and has set a new, more powerful, flexible, inclusive model for future PDF meetings.

6. CONCLUSION

Despite being the first online version, PDF IV demonstrated how much the polar data community has grown, not only in numbers and geographical diversity, but also in depth, diversity, and in the richness of discussions, particularly in themes such as VREs, logistics, and supporting decision-making. This fourth edition of the PDF enabled the polar data community to make concrete progress towards developing projects by getting more practical work done during Hackathon Sessions and allowing for more communication and cross-fertilisation of ideas between the poles.

Although PDF IV was organised in close collaboration with the 2nd Southern Ocean Regional Workshop for the UN Ocean Decade, the Arctic community dominated in the PDF discussions compared to the Antarctic community. However, this unevenness reflects how much larger the Arctic community is. Although the Arctic-Antarctic asymmetry has been fairly consistent throughout all PDF editions, one important thing to mention is that the Antarctic community was in higher numbers than during previous editions. This demonstrates that, although the polar data community is relatively small compared to other regions, it is extremely successful in linking partners together and building a community that shares data and information. The Poles bring us together and allow the polar data community to act as a silo breaker for the global data community. Overall, the Southern Ocean Decade and Polar Data Forum Week 2021 can be considered as a successful event. Participants provided results pertaining to all the Hackathons Sessions. However, it is important to mention that the objectives of each Hackathon Session are the result of continuous discussions and are considered as ongoing. Polar Data Forum is increasingly establishing itself as the essential venue for the polar data community to gather and drive forward shared progress for the benefit of data providers, managers and users throughout the international Arctic and Antarctic communities.

DATA ACCESSIBILITY STATEMENTS

Recordings of PDF IV conference presentations are available on the [EPB's YouTube channel](#).

ACKNOWLEDGEMENTS

We thank the PDF IV Science Committee, the PDF IV Organising Committee, PDF IV Hackathon conveners, PDF IV Conference Chairs and participants of the Southern Ocean Decade and Polar Data Forum Week 2021 for all the inspiring discussions and visions.

FUNDING INFORMATION

This work has been supported by the Belgian Science Policy Office (BELSPO) [contract numbers FR/36/AN1/AntaBIS and BL/36/FWI34_PUNDOS-PPODS] in the Framework of EU-Lifewatch.

COMPETING INTERESTS

The authors have no competing interests to declare.

AUTHOR AFFILIATIONS

Annemie Rose Janssen  orcid.org/0000-0002-6745-6542

Royal Belgian Institute of Natural Sciences, Brussels, Belgium; School of Natural Sciences, Macquarie University, Sydney, Australia

Philippa Bricher  orcid.org/0000-0001-7975-5307

Southern Ocean Observing System, Hobart, Australia

Karen Payne

International Technology Office – World Data System, Victoria, Canada

Renuka Badhe  orcid.org/0000-0001-5255-744X

European Polar Board, The Hague, The Netherlands

Nicole Biebow

European Polar Board, The Hague, The Netherlands; Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research (AWI), Bremerhaven, Germany

Taco de Bruin  orcid.org/0000-0001-9149-2095

NIOZ Royal Netherlands Institute for Sea Research, Texel, the Netherlands; International Oceanographic Data and Information Exchange of IOC/UNESCO, Ostend, Belgium

Ruth Duerr

Ronin Institute, Montclair, NJ, USA

Pjotr Elshout  orcid.org/0000-0002-7936-0441

European Polar Board, The Hague, The Netherlands

Allison Gaylord  orcid.org/0000-0002-3254-6791

Nuna Technologies, Homer, AK, USA

Øystein Godøy  orcid.org/0000-0001-6410-3488

Svalbard Integrated Arctic Earth Observing System, Longyearbyen, Norway; Norwegian Meteorological Institute, Oslo, Norway

Patrick Gorringe

Swedish Meteorological and Hydrological Institute, Norrköping, Sweden

Damien Guihen  orcid.org/0000-0002-5786-6599

University of Tasmania, Hobart, Australia

Johnathan Kool  orcid.org/0000-0002-9902-3522

Australian Antarctic Division, Canberra, Australia

Jan Rene Larsen

Sustaining Arctic Observing Network, Arctic Monitoring and Assessment Programme, Tromsø, Norway

Joseph Nolan  orcid.org/0000-0003-1892-8328

European Polar Board, The Hague, The Netherlands; European Global Ocean Observing System (EuroGOOS), Brussels, Belgium

Antonio Novellino  orcid.org/0000-0002-3263-5583

ETT S.p.A., Genova, Italy

William Manley  orcid.org/0000-0001-7922-9753

University of Colorado, Boulder, USA

Andreas Marouchos  orcid.org/0000-0003-3625-921X

Commonwealth Scientific and Industrial Research Organisation (CSIRO), Hobart, Australia

Molly McCammon

Alaska Ocean Observing System, Anchorage, AK, USA

Maribeth Murray  orcid.org/0000-0003-2821-9381

University of Calgary, Calgary, AB, Canada

Jenn Parrott

Inuvialuit Regional Corporation, Inuvik, NT, Canada

Jay Pearlman

Institute of Electrical and Electronics Engineers, Paris, France

Helen Peat  orcid.org/0000-0003-2017-8597

British Antarctic Survey, Cambridge, UK

Peter Pulsifer  orcid.org/0000-0001-9561-3640

Carleton University, Ottawa, ON, Canada

Lloyd Symons

Australian Antarctic Division, Canberra, Australia

Marten Tacoma  orcid.org/0000-0002-3733-7278

NIOZ Royal Netherlands Institute for Sea Research, Texel, the Netherlands

Soulaine Theocharides  orcid.org/0000-0001-7573-4330

International Technology Office – World Data System, Victoria, Canada

Stein Tronstad  orcid.org/0000-0002-9743-7211

Norwegian Polar Institute, Tromsø, Norway

Chantelle Verhey  orcid.org/0000-0002-0047-7875

International Technology Office – World Data System, Victoria, Canada

Anton Van de Putte  orcid.org/0000-0003-1336-5554

Royal Belgian Institute of Natural Sciences, Brussels, Belgium; Université Libre de Bruxelles, Brussels, Belgium

- European Polar Board.** 2021. *Polardex – Interactive Polar Infrastructure Database*. Available at: <https://polardex.org/> [Last Accessed 21 June 2023].
- International Forum – Polar Data Activities in Global Systems.** 2013. Polar Data Forum I website. 15–16 October 2013, National Museum of Nature and Science, Tokyo, Japan. Available at: <https://sites.google.com/a/icsu-wds.org/archive-international-forum-on-polar-data-activities-in-global-data-systems/> [Last Accessed 13 February 2023].
- Janssen, AR,** et al. 2022. *Southern Ocean Action Plan (2021–2030) in support of the United Nations Decade of Ocean Science for Sustainable Development*. Switzerland: Zenodo. DOI: <https://doi.org/10.5281/ZENODO.6412191>
- PDF III Scientific Steering Committee.** 2021. *Polar Data Forum III – Website*. Helsinki, Finland: Zenodo. Available at: <https://zenodo.org/record/4784210#.Y-nCquxBw-R> [Last Accessed 13 February 2023].
- PDF IV Scientific Steering Committee.** 2021. *Polar Data Forum IV – Website*. Online: Zenodo. Available at: <https://zenodo.org/record/7804814> [Last Accessed 21 June 2023].
- POLDER.** 2021. *POLDER Federated Search*. Available at: <https://search.polder.info/> [Last Accessed 21 June 2023].
- SAON-POAwg.** 2021. *POAwg – Polar Observing Assets Working Group*. Available at: <https://www.polarobservingassets.org/> [Last Accessed 21 June 2023].
- Tanhua, T,** et al. 2019. What We Have Learned From the Framework for Ocean Observing: Evolution of the Global Ocean Observing System. *Front. Mar. Sci.*, 6: 471. DOI: <https://doi.org/10.3389/fmars.2019.00471>
- Tronstad, S,** et al. 2021. *Alignment of Polar Data Policies – Recommended Principles*. DOI: <https://doi.org/10.5281/ZENODO.5734900>
- UNESCO.** 2021. *UNESCO Recommendation on Open Science*. <https://en.unesco.org/science-sustainable-future/open-science/recommendation> [Last Accessed 19 April 2022].
- UNESCO-IOC.** 2019. *IOC Oceanographic Data Exchange Policy (2019)*. https://www.iode.org/index.php?option=com_content&view=article&id=51&Itemid=95 [Last Accessed 19 April 2022].
- Verhey, C.** 2022. *PDF4 – Semantics and Vocabularies Hackathon SWOT Visualization (Version 1)*. Switzerland: Zenodo. DOI: <https://doi.org/10.5281/zenodo.6344558>
- Wilkinson, MD,** et al. 2016. The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* 3, 160018. DOI: <https://doi.org/10.1038/sdata.2016.18>
- World Meteorological Organization.** 2021. *Unified Policy for the International Exchange of Earth System Data*. Available at <https://public.wmo.int/en/our-mandate/what-we-do/observations/Unified-WMO-Data-Policy-Resolution> [Last Accessed 19 April 2022].

TO CITE THIS ARTICLE:

Janssen AR, Bricher P, Payne K, Badhe R, Biebow N, de Bruin T, Duerr R, Elshout P, Gaylord A, Godøy Ø, Gorringer P, Guihen D, Kool J, Larsen JR, Nolan J, Novellino A, Manley W, Marouchos A, McCammon M, Murray M, Parrott J, Pearlman J, Peat H, Pulsifer P, Symons L, Tacoma M, Theocharides S, Tronstad S, Verhey C and Van de Putte A. 2023. Polar Data Forum IV – An Ocean of Opportunities. *Data Science Journal*, 22: 18, pp. 1–13. DOI: <https://doi.org/10.5334/dsj-2023-018>

Submitted: 05 October 2022

Accepted: 08 May 2023

Published: 23 June 2023

COPYRIGHT:

© 2023 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

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