<table>
<thead>
<tr>
<th>Presenting author name</th>
<th>Additional authors’ information</th>
<th>Abstract Title</th>
<th>Abstract Text</th>
</tr>
</thead>
<tbody>
<tr>
<td>Afonso Ferreira</td>
<td>Prof Brito, Ana, C. she/her MARE – Marine and Environmental Sciences Centre, Faculdade de Ciências, Universidade de Lisboa, Lisbon, Portugal Prof Mendes, Carlos, R.B. he/him Laboratório de Fitoplâncton e Microorganismos Marinhos, Instituto de Oceanografia, Universidade Federal do Rio Grande, Rio Grande, Brazil</td>
<td>OC4-SO algorithm: improving satellite retrievals of chlorophyll a in the Antarctic Peninsula</td>
<td>Remote sensing of chlorophyll a (chl-a; a proxy of phytoplankton biomass) is as an essential tool for monitoring marine ecosystems, particularly for polar regions where in-situ data availability is very limited. For the Southern Ocean, however, the current global chl-a algorithms tend to underestimate chl-a, preventing accurate retrievals of chl-a. This work sought to develop a new chl-a algorithm for the Antarctic Peninsula and to assess potential drivers of chl-a underestimation in this region. OC4-SO, the algorithm developed and presented here, exhibits the following advantages: i) developed using a 20+ years and spatially thorough in-situ HPLC-derived chl-a dataset; ii) based on a multi-sensor product, which are expected to be increasingly popular as the number of operating satellites increases; iii) adequately corrects for chl-a underestimation, now allowing for the assessment of the Antarctic seasonal cycle which was not previously possible using global algorithms; and v) exhibits a good performance for Antarctic waters outside of the Antarctic Peninsula. Low particulate backscattering coefficients in the green region of the light spectrum were seen to be main factor linked to chl-a underestimation in high chl-a waters, corroborating previous works. Overall, the OC4-SO can be a valuable tool for ocean colour remote sensing in Antarctic waters, enabling more accurate estimates of chl-a and contributing to a better understanding of this ecosystem’s role to the global biogeochemical cycle.</td>
</tr>
<tr>
<td>Oceanos e Clima, Instituto de Oceanografia, Universidade Federal do Rio Grande, Rio Grande, Brazil</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| Dr. Guerreiro, Catarina  
  she/her  
  MARE – Marine and Environmental Sciences Centre, Faculdade de Ciências, Universidade de Lisboa, Lisbon, Portugal |
| Dr. Sá, Carolina  
  she/her |
| Dr. Jackson, Thomas  
  he/him  
  Plymouth Marine Laboratory, Prospect Place, Plymouth PL1 3DH, United Kingdom. |
<table>
<thead>
<tr>
<th>Alex J. Tate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Continued development of a data management system for the new UK polar research vessel, the RRS Sir David Attenborough</td>
</tr>
</tbody>
</table>

The new UK polar research vessel, RRS Sir David Attenborough, is currently undergoing tests and trials before deploying to the Antarctic in late 2021. The vessel hosts a wide array of scientific instrumentation and as a marine platform will deploy a range of remotely-operated data gathering equipment, both marine and airborne. The resulting datasets are both complex and voluminous and their management and availability need to fulfil the requirements of a wide range of end users from cruise participants to global data archiving facilities.

The UK Polar Data Centre, based at the British Antarctic Survey, is leading on the development of the scientific data management systems aboard the new vessel. Major areas of focus include; data and event logging, data visualisation and access, automated quality control routines and sensor metadata.

The presentation will provide an update on these major focus areas as well as an up-to-date overview of the scientific instrumentation currently being commissioned on the vessel and the types of data that are now being routinely collected.
<table>
<thead>
<tr>
<th>Alice Fremand</th>
<th>Mr Bodart, Julien he/him</th>
<th>Improving FAIRness of aerogeophysics datasets at the UK Polar Data Centre</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>UK Polar Data Centre - British Antarctic Survey <a href="mailto:julart@bas.ac.uk">julart@bas.ac.uk</a></td>
<td>The UK Polar Data Centre (PDC, <a href="https://www.bas.ac.uk/data/uk-pdc/">https://www.bas.ac.uk/data/uk-pdc/</a>) is the focal point for Arctic and Antarctic environmental data management in the UK. Part of the Natural Environmental Research Council’s (NERC) network of environmental data centres and based at the British Antarctic Survey (BAS), the PDC coordinates the management of polar data from UK-funded research and supports researchers in complying with national and international data legislation and policy. Reflecting the multidisciplinary nature of polar science, the datasets handled by the data centre are extremely diverse. Aerogeophysics data, including aeromagnetics, aerogravity and radar echo-sounding data, are a key asset to the geoscience and glaciology community as they provide crucial information about the Earth’s geological structure and ice thickness, and thus contribute directly to our understanding of the sea-level-rise potential of the Antarctic Ice Sheet. Improving Findability, Accessibility, Interoperability and Re-use (FAIRness) of these data is thus at the core of PDC’s mission. In the last two years, significant progress has been made to improve the management of BAS aerogeophysics data, a challenging task considering that BAS is one of the largest acquirers of airborne geophysics data over Antarctica. In 2020, we published bedrock elevation data for fourteen airborne radar surveys and more than thirty airborne gravity and magnetics datasets over Antarctica. This year, we will release large swaths of airborne radar data collected by BAS over the last two decades. This includes extensive surveys over Pine Island and Thwaites glaciers, two catchments that have received considerable attention in recent years due to their potential to affect sea-level-rise on a global scale. In order to encourage the reuse and increase the value of these data, we also participate in a large variety of scientific research projects internationally. For instance, the wide coverage of airborne radar echo sounding over Antarctica is crucial for the SCAR Action Group BEDMAP3 project which aims to produce new maps of Antarctic ice thickness and bed topography. We hope this work will greatly benefit the geophysical and glaciological community and serve as a successful example of an effective partnership between scientists and data managers.</td>
</tr>
</tbody>
</table>
Spotlighting Arctic Data Collections with the DataONE Portals Infrastructure

Federation across large numbers of diverse, geographically dispersed, data repositories supporting Arctic research data provides significant advances for discovery, accessibility and reuse. However, Arctic data are preserved across a broad collection of repositories including domain specific repositories, generalist repositories and institutional repositories that also hold significant collections of data not related to the Arctic. Consequently, while federated networks enable data to be discovered from a single location, the level of precision and specificity required for specialist fields cannot be met using a generic search.

At DataONE, we provide portal infrastructure to enable organizations, research groups, and collaborations to build customized data discovery portals that collate and present a curated subset of the full data catalog. For example, researchers have created portals to organize data on the circumpolar active layer (CALM), for settlement and historical church surveys in Iceland, and to collate research products from the Toolik Field Station in Alaska. These data portals incorporate user defined search queries across all metadata fields, enabling portal creators to build faceted search terms that have direct relevance to their research community and the data they are exposing. Portals include customizable branding options to effectively represent the organization or research collaboration, and free text pages for descriptive content. In addition to the rich search and discovery capabilities, portals include FAIR assessment of metadata quality and aggregated data usage and citation metrics.

The current release of data portals reflects the outcome of ongoing usability assessment and community engagement. Portal development is informed by user experience feedback and recent focus groups have provided additional insights in how the community use, and want to extend, these services. In this talk we will provide an overview of the portal features and future functionality, in addition to describing our design process and opportunities for interaction.
Ms. McLean, Erin.
she/her/hers
National Center for Ecological Analysis and Synthesis
University of California, Santa Barbara
1021 Anacapa St
Santa Barbara, CA, USA
mclean@nceas.ucsb.edu
<table>
<thead>
<tr>
<th>Antonio Novellino</th>
<th>Dr Colombo Federica (she), ETT SpA, Via Enrico Albareto 21, 16153 Genova, Italy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mr Alba Marco (he), ETT SpA, Via Enrico Albareto 21, 16153 Genova, Italy</td>
<td></td>
</tr>
<tr>
<td>Dr Sallée Jean-Baptiste (he), Sorbonne Universités, UPMC University, Paris 06, UMR 7159, LOCEAN-IPSL F-75005, Paris, France</td>
<td></td>
</tr>
<tr>
<td>Dr Lecornec Amelie (she), Sorbonne Universités, UPMC University, Paris 06, UMR 7159, LOCEAN-IPSL F-75005, Paris, France</td>
<td></td>
</tr>
</tbody>
</table>

**SO-CHIC: a data infrastructure for the Antarctic community**

The project SO-CHIC (Southern Ocean Carbon and Heat Impact on Climate) aims at improving our understanding of the heat and carbon exchanges between the atmosphere and the deep ocean, by quantifying their fluxes at the air-sea-ice interface and estimating interannual variability of heat and carbon storage in the Southern Ocean. SO-CHIC brings together old and new observations, and makes these observations openly accessible to maximise impact on climate reports, services, and models.

New observations are being collected by means of several platforms (e.g., autonomous surface vehicles, ships and floats and, under ice floats, fixed stations, gliders and deep gliders, etc.). These observations also include parameters, such as radiative and turbulent surface flux, latent heat, fCO2, etc. that should be available together with classical parameters and accessible from a single one stop shopping system. Therefore SO-CHIC is developing a specific data management infrastructure and a data portal to manage and offer these variables, making data and metadata available without any restriction to all interested third parties, including the integration into projects and initiatives such as EMODnet, CMEMS, SeaDataNet. SO-CHIC data management infrastructure also candidates itself to be the new SOOS data management backend.

This new infrastructure consists of a data buffer that links/harvests data from operational sources and where data can be pushed by data providers and a cluster of metadata and data publishing tools (GeoNetwork, GeoServer and ERDDAP) to expose metadata and data according FAIR principles. On top of this backend a mapviewer allows general users to search for data by parameter, platform, supporting project, together with applying temporal and spatial filters.

**Acknowledgments:** SO-CHIC project (http://www.sochic-h2020.eu/) has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement #821001.
Climate change presents a major and growing challenge to the Arctic region, producing risks as well as opportunities. Access to accurate data and information has therefore never been more important to better understand critical physical, biogeochemical and biological processes.

The spatial coverage of observing systems in the Arctic is scattered in time and space. Existing observations are of varying quality and made available on different timescales. Furthermore, a major part of marine observations are funded via fixed-term research projects, of which only one third openly share their data. In this context, EMODnet (Physics, Data Ingestion), Copernicus Marine Service In Situ Thematic Center (CMEMS INSTAC), Copernicus In Situ Coordination Group and EuroGOOS have joined efforts to establish a dedicated marine Arctic data portal with the purpose to:

- Ease access to open and free marine Arctic in-situ data
- Foster integration in European data aggregator infrastructures (CMEMS, EMODnet and SDN)
- Unlock existing data from a variety of projects and initiatives not yet freely and openly exchanged
- Provide support to share data through existing European data aggregators
- Provide visibility to existing observing system platforms sharing data in the Arctic region

In November 2020, the Arctic Data Portal (https://arctic.emodnet-physics.eu/), implemented under EMODnet Physics, was launched. It provides straightforward access to circumpolar datasets and metadata records from European data aggregators (CMEMS, SDN, EMODnet, Pangea) and other sources. The portal is composed of a map interface displaying the spatial distribution of observing platforms, a tool selecting by parameter and type of observing platform and plotting tools for displaying the datasets. The Arctic portal is similar to the SOOS-map portal that EMODnet Physics set up in collaboration with the Southern Ocean Observing System (SOOS). This has greatly improved access to key Southern Ocean datasets.
| Christina Bienhold | Mr. Matomo Niwano | From locally to globally FAIRified (meta)data management: Reflections and outcomes from an Arctic Molecular Observatory | High-quality (meta)data management begins long before deposition in large-scale archiving services. All research and operations personnel must share custodianship of (meta)data as part of contemporary research culture, allowing its value to be secured and rapidly realized. To this end, we have established a (meta)data management approach which links the Frontiers in Arctic Marine Monitoring (FRAM) Programme's Molecular Observatory to global digital ecosystems. The Observatory monitors microbial community dynamics across latitudinal, depth, and temporal gradients, providing valuable information from a remote and rapidly changing region. It is thus crucial to preserve high-quality and rich (meta)data, such that it can be discovered, understood, and re-used by future generations. Our approach centers on a simple relational metadatabase, which structures and validates environmental, contextual, and procedural metadata generated from the field, through the lab, to the in-silico processing and archiving of FRAM's multi-omic sequence data. Above the technology, we emphasize the metadatabase's design, which aligns our local metadata – from its creation – with institutional archives, community standards, and data stored in international archives (e.g. the European Nucleotide Archive and PANGAEA). This local solution supports operational-grade data stewardship within FRAM, while supporting global interoperability by exporting (meta)data compliant with specifications such as the MiXS (Minimum Information about any (x) Sequence) developed by the Genomic Standards Consortium (GSC). To boost interoperability further, we are leading an effort to align MiXS to the Darwin Core specification. We also support our personnel's pursuit of compliance and accuracy by pre-selecting appropriate Arctic and deep-sea content from more complex components of such standards. We stand ready to produce additional exports to connect our holdings with other global standards and interoperability frameworks, such as UNESCO's Ocean Data and Information System. Here, we present the status and reflections on our multi-pronged approach to practically FAIRifying data in an Arctic molecular observatory. |
| Mr. Matthias Wietz | HGF-MPG Group for Deep Sea Ecology and Technology, Alfred Wegener Institute Helmholtz Center for Polar and Marine Research, Bremerhaven, Germany | Matthias.Wietz@awi.de | |
| Ms. Raïssa Meyer | HGF-MPG Group for Deep Sea Ecology and Technology, Max Planck Institute for Marine Microbiology, Bremen, Germany | rmeyer@mpi-bremen.de | |
| Dr. Pier Luigi Buttigieg | Helmholtz Metadata Collaboration, GEOMAR Helmholtz Center for Ocean Research, Kiel, Germany | Pier.Buttigieg@awi.de | |
The Permafrost Discovery Gateway portal is building scalable cyberinfrastructure to help researchers understand the extent, spatio-temporal variations, and impact of pan-Arctic permafrost thaw and other phenomena. We present the PDG portal, an online application that provides interactive visualization, exploration, and access to geospatial permafrost data products derived from high-resolution satellite imagery (Big Imagery). The portal leverages many existing software technologies, and builds upon them using standard and emerging geospatial processing tools and techniques. The PDG portal is built using the open source DataONE Portal system to allow easy customization of the application. It provides multiple interactive viewers including the Fluid Earth Viewer, which enables global and regional visualization of Arctic data products over time, and a Cesium-based Imagery Viewer that facilitates exploration of pan-Arctic, sub-meter map products over time. A Plot Viewer is also planned that will facilitate interactive data analysis where researchers can produce their own 2D and 4D graphs from the derived Big Data layers being produced by the PDG machine learning workflows. We discuss our initial data layer example of pan-Arctic ice-wedge polygon data derived from sub-meter imagery and processed into both raster and vector tiles that allow for performant panning and zooming using the Imagery Viewer. We present the challenges of creating a generic workflow that can be used to prepare tile layers across multiple input data formats, the indexing involved in making the data accessible for analysis online, and the workflow needed to provide access to the original layer data for download and further analysis through a customized search interface for the NSF Arctic Data Center archive.
<table>
<thead>
<tr>
<th>Daniela</th>
<th>ISOBatA: Italian Pilot project for the exploitation of underway seafloor datasets in the Antarctic region and surrounding areas.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accettella</td>
<td>ISOBATA is a recently approved PNRA (Italian National Antarctic Reasearch Program) Following the example of USA and German research vessels example (Wölf et al.2019), Isobata designed a systematic approach to efficiently exploit seafloor data sets collected within the Antarctic region, during transit times of the renewed Icebreaker R/V Laura Bassi (former E. Shackleton). International scientific exchanges will be critical in defining best practices and standardized protocols for collecting high-quality multibeam and underway data during transit time in the Ross Sea and surrounding Area. Optimizing opportunistic data collection will aim to provide supranational scientific products in order to highlight the key role of underway data in remote regions. The project takes into account the need to store all acquired data and metadata, a final project output will be the establishment of an Italian Antarctic Bathymetric Data Repository integrated with the PNRA National Antarctic Data Center (NADC). This will constitute the national reference for the implementation of the “Hydrographic Surveying of Antarctic Waters Working Paper”, in support of the Seabed 2030 project (Jakobsson et al 2017, Mayer et al. 2018). High-standard protocols and best practices will primarily inform data acquisition, processing and interpretation. Data collection plan has been designed promoting the optimization of resources with an effective saving in the use of infrastructural and logistic support. Mapping the Southern Ocean by 2030 will require long-term projects; at his natural end, ISOBatA research units will work to ensure that all the gathered know-how, database and international collaborations will continue as a Multidisciplinary Observatory Underway Project (MIMOSA – Multidisciplinary Italian Marine Observatory For the Study of Antarctic). ISOBatA fulfil SCAR, CONMAP, IHO and Seabed 2030 guidelines.</td>
</tr>
</tbody>
</table>
David Arthurs  Peter Pulsifer  Sandy Starkweather  Arctic Data Systems in the ROADS Process

The crafting of the Sustaining Arctic Observing Networks’ (SAON) Roadmap for Arctic Observing and Data Systems (ROADS) has made significant progress. It sets out a multi-phase process and governance structure for defining Shared Arctic Variables (SAV) that will provide benefits across a wide range of stakeholders and scales. However, the focus so far has been on observing systems and work remains to define the data systems portion of ROADS. This presentation will examine how the ROADS guiding principles the have evolved for Arctic Observing Systems might be applied to Arctic Data Systems. Elements of those principles include:

§ Indigenous Peoples’ equitable partnership and funding for their active participation is critical to ROADS from its inception through its implementation;
§ All aspects of the ROADS process should support broadly shared benefit from the observing and data systems;
§ The ROADS process should complement and integrate, without wasteful duplication, the current planning approaches used by existing networks (regional to global), activities and projects;
§ ROADS should support stepwise development through a flexible, federated and evolving structure that allows grassroots identification of themes, infrastructures and regional foci.

Challenges for Arctic Data Systems that will considered in this context include:

§ Improving coordination among funders.
§ Enhancing global data communities and governance structures.
§ Supporting data community building, coordination, and engagement.
§ Ensuring long-term support for data management and curation.
§ Engaging with and enhancing existing activities rather than creating new initiatives.
§ Facilitating a change in attitude from proprietary data to data as a common good.
§ Improving education and training in data science.
§ Building on interoperable standards and ethically open and FAIR data principles.
§ Involving and respecting the perspective of Indigenous peoples in data collection and management.
§ Embracing cloud platforms and new analytical techniques (e.g., AI).
<p>| Ekaterina Kim | Data Use Cases: Applying AI and Data Science Tools to Optical Images and AIS Data from the Arctic | There are extreme challenges unique to the Arctic, from human activities and impacts in remote Arctic locations to Arctic data acquisition, sharing, and quality. The amount of Arctic data is growing. In fact, this growth is faster than the capacity of experts to process, adequately validate, and evaluate all uncertainties in the data. Ocean sciences and maritime industry-oriented applications in the Arctic regions can benefit from learning on previous data and from cross-disciplinary expert knowledge. Despite rapid progress in artificial intelligence (AI) and data science, AI applications to Arctic science, engineering, and technology (e.g., automated validation of remote sensing data from the Arctic, learning from past humans’ activities in the ice infested waters) have received less scientific attention in comparison to the fields of finance, logistics, medicine, advertisement, etc. This worldwide trend may be attributed to the data quality, availability, and the expertise that is needed to process these data. There is a strong need to direct AI applications towards solving Arctic challenges. In this talk I would like to present two examples on how artificial intelligence can help in automated interpretation of ice imagery from ground operations (e.g., imagery from surface vessels, shore stations, in-situ campaigns) and how it can help us in learning from historical AIS data (e.g., from the Kara Sea region). Data processing algorithms underlying the presented examples have been made publicly available on GitHub (i.e., ice image classification, ice image segmentation, and AIS data processing). |</p>
<table>
<thead>
<tr>
<th>Frank, O, Nitsche</th>
<th>Dr. Tinto, Kirsteen (she/her) Lamont-Doherty Earth Observatory of Columbia University Palisades, NY, USA <a href="mailto:tinto@ldeo.columbia.edu">tinto@ldeo.columbia.edu</a></th>
</tr>
</thead>
<tbody>
<tr>
<td>Dr. Neville, Shane (he/him) Lamont-Doherty Earth Observatory of Columbia University Palisades, NY, USA <a href="mailto:nshane@ldeo.columbia.edu">nshane@ldeo.columbia.edu</a></td>
<td></td>
</tr>
<tr>
<td>Dr. Carbotte, Suzanne (she/her) Lamont-Doherty Earth Observatory of Columbia University Palisades, NY, USA <a href="mailto:carbotte@ldeo.columbia.edu">carbotte@ldeo.columbia.edu</a></td>
<td></td>
</tr>
<tr>
<td>Improving findability and accessibility of long-tail data through the US Antarctic Program data Center</td>
<td>The US Antarctic research community has developed disciplinary repositories with excellent tools for specific datasets. Many datasets do not fit into these disciplinary repositories, but provide valuable, often unique data that would otherwise be difficult to preserve or find. The US Antarctic Program Data Center (USAP-DC) is funded by the National Science Foundation (NSF) and provides services to help scientists document, preserve, and disseminate all Antarctic research data including those for which no disciplinary repository exists. Thus, these datasets are preserved for future access. We provide DOIs for these datasets, and increase findability and visibility of the dataset by sharing through various data portals including the Antarctic Master Directory, DataOne and serving of schema.org in dataset landing pages. We recently improved our search interface to include a better map-based search option on the USDAP-DC web page. USAP-DC also maintains project pages for NSF awards and for a groups of collaborative research awards that include basic information of these projects and provide links to datasets hosted by different repositories including those hosted at USAP-DC. This improves findability of data linked to specific projects even if they are distributed across different repositories.</td>
</tr>
<tr>
<td>Giulio Verazzo</td>
<td>Chiara Ripa</td>
</tr>
<tr>
<td>Author title (e.g. Ms., Mr, Prof.): Dr. Last Name, First Name, Middle Initial: Schildhauer, Mark P. Author pronouns: he, him, his Department (optional): Arctic Data Center, NCEAS Institution (required): University of California, Santa Barbara Street Address (optional): City, State, Country (required): Santa Barbara, CA, USA e-mail address (optional): <a href="mailto:schild@nceas.ucsb.edu">schild@nceas.ucsb.edu</a></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td></td>
</tr>
<tr>
<td>Author title: Ms. Last Name, First Name, Middle Initial: Csik, Samantha, R. Author pronouns: she/her Department: National Center for Ecological Analysis and Synthesis Institution: University of California, Santa Barbara Street Address (optional): City, State, Country: Santa Barbara, CA, USA e-mail address: <a href="mailto:scsik@nceas.ucsb.edu">scsik@nceas.ucsb.edu</a></td>
<td></td>
</tr>
<tr>
<td>Author title (e.g. Ms., Mr, Prof.): Mr. Last Name, First Name, Middle Initial: Mecum, Bryce, D. Author pronouns: He/him Department (optional) Institution (required): National Center for Ecological Analysis and Synthesis Street Address (optional): City, State, Country (required): Santa Barbara, CA, USA e-mail address (optional):</td>
<td></td>
</tr>
</tbody>
</table>
| Building a custom ontology to enable advanced search capabilities for MOSAiC expedition datasets archived in the Arctic Data Center In October 2020, the Multidisciplinary drifting Observatory for the Study of Arctic Climate, “MOSAiC” expedition, completed a year-long data collection effort, contributing to our understanding of climate processes in the Central Arctic, and more generally, global climate change. The Arctic Data Center (ADC) has since worked with MOSAiC researchers to archive, in a custom data portal, all of the USA’s National Science Foundation-funded data products and corresponding metadata records collected during the expedition (https://arcticdata.io/data-portals/).

To facilitate greater data findability and interpretability, the ADC team built a MOSAiC ontology (i.e. a controlled vocabulary expressed in RDF/OWL) to formally define and inter-relate terms, that were then used to semantically annotate the “native” terminology used by researchers in describing MOSAiC metadata. The ontology reduces search ambiguity and increases the precision of search efforts, enabling users to better understand the meaning of terms, and refine their searches.

The MOSAiC ontology serves as a case study for building custom, project-specific ontologies to improve data transparency and findability. In addition, due to its standard format and open schema, the MOSAiC ontology can be easily shared, revised, and extended to accommodate new types of information, or new interpretations of the data contents; and used for annotation without requiring changes to the underlying data or data structures. As a standardized vocabulary, the MOSAiC ontology could be used to annotate all MOSAiC datasets, allowing for enhanced interoperability across the three international repositories (PANGAEA, ARM and ADC) holding MOSAiC data. Alignment of MOSAiC terms with broader community ontologies, such as the Environment Ontology, EnvO, or the NERC vocabulary, would further increase the findability and reusability of these invaluable MOSAiC data resources. |
| Jean Baptiste Barré | Ms. Anne Chapuis  
| | she/her  
| | IGE  
| | Grenoble, France  
| | anne.chapuis@univ-grenoble-alpes.fr  
| Ms. Amélie Bataille  
| | she/her  
| | IGE  
| | Grenoble, France  
| | amelie.bataille@univ-grenoble-alpes.fr  
| Mr. Gaël Durand  
| | he/him  
| | IGE  
| | Grenoble, France  
| | gael.durand@univ-grenoble-alpes.fr  
<p>| Customized data workflow for an interdisciplinary research project. | The EU-funded PROTECT project aims to assess and project changes in the land-based cryosphere in order to produce robust global, regional and local projections of Sea Level Rise (SLR) on a range of timescales. The project brings together SLR and cryosphere communities, whose contributions are strongly interdependent. This implies an effective data workflow that has been designed specifically for PROTECT although it can be reproduced in other projects. The workflow relies on a sequence of two processes. The internal collaboration process corresponds to the successive steps followed by the workdata: preparation, processing and sharing between partners. The preparation and processing cycles can be managed on the project server, providing an available and secure space for the partners data. The sharing cycle is supported by the project data portal developed for that purpose. The data portal is built on top of the open-source data platform CKAN. Partners have in particular the ability to share their datasets using metadata, avoiding unnecessary large file transfers. Access to working data is limited to the consortium through access rights. The second process corresponds to the data opening to enhance exploitation of the PROTECT results. It follows closely the FAIR principles and is supervised by the PROTECT scientific steering committee. It proposes among others to make the data available on the EU Zenodo platform and to attach the datasets to the PROTECT Zenodo community to ensure better visibility. This workflow has been set up by the project data manager. |</p>
<table>
<thead>
<tr>
<th><strong>Jennifer, P, Sevadjian</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Dr. Wilson, Cara she/hers</td>
</tr>
<tr>
<td>NOAA Southwest Fisheries Science Center Monterey, CA, USA</td>
</tr>
<tr>
<td>Dr. Robinson, Dale he/his</td>
</tr>
<tr>
<td>University of California, Santa Cruz Santa Cruz, CA, USA</td>
</tr>
<tr>
<td><strong>PolarWatch</strong></td>
</tr>
<tr>
<td><strong>Overcoming common data access challenges to provide oceanographic satellite remote sensing data for the polar regions</strong></td>
</tr>
<tr>
<td>PolarWatch is the NOAA CoastWatch node focused on the distribution of oceanographic satellite data for Arctic and Antarctic waters. PolarWatch serves both near real-time and historical satellite data, including measurements of sea ice cover, ocean temperature, ocean color products, and winds. Important complementary in-situ datasets are also provided through PolarWatch, including data from BioGeoChemical-Argo floats and the International Arctic Buoy Programme. Additional sea ice data sets and SAR data will be added in the coming year. Data served through PolarWatch are easily previewed with Arctic, Antarctic or global projections using the online PolarWatch data catalog. Data are easily accessible through the PolarWatch ERDDAP data distribution system. Like all CoastWatch nodes, PolarWatch is a value-added provider, not just providing data, but also helping users to access data by developing tools and tutorials for easier data access, providing training and hands-on assistance, finding or creating data products to address user needs, and working directly with users on projects. In this presentation we will provide an overview of the data and services provided by PolarWatch.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Jill S. Prewitt</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A data tool for planning, prevention, and response to international maritime incidents in the Bering Strait</strong></td>
</tr>
<tr>
<td>Vessel traffic through the Bering Strait is increasing, and is expected to continue increasing as sea ice decreases and the open water season lengthens. With this increased vessel traffic comes increased risk of marine accidents and oil or other contaminant spills in the area. However, this area also has a lack of infrastructure and resources to respond to incidents of this nature, and has transboundary communication and data sharing challenges between Russian and US authorities. For example, the US Arctic Environmental Response Management Application (Arctic ERMA), the US online and field data tool that is used for both planning, prevention and response actions, is stored on US Government servers (i.e., NOAA) and not currently accessible by Russian authorities. For these reasons, there is a need for an oil spill response tool that is accessible by both US and Russian authorities in the event of an oil spill - or some other contaminant spill - in the Bering Strait area. The Bering Strait Transboundary Incident Response Tool will co-locate relevant data from the Russian and the US sides of the Bering Strait, as well as the northern Bering and southern Chukchi Seas. The final product will allow for both Russian and US authorities and scientists to access the same data and to share the same visualizations of the area within the Bering Strait region. This information is valuable not only for a potential response to an emergency event such as an oil spill, but will be useful as a public information resource to a broader group of interested parties on both sides of the EEZ, including coastal communities, conservation groups, resource managers and academic researchers.</td>
</tr>
<tr>
<td>Kenton McHenry</td>
</tr>
</tbody>
</table>
FAIR and consistent data management in a large multidisciplinary project: The Nansen Legacy

The Nansen Legacy project is a collaboration of over 200 biologists, oceanographers, atmospheric scientists and geologists from 10 research institutions across Norway. They investigate the impact of climate change on the northern Barents Sea and adjacent Arctic Basin through multiple research expeditions into often ice-covered waters. Designing a system that ensures good and consistent data management across such a large, multidisciplinary project is challenging, as multiple documentation standards must be considered to encompass the wide range of data involved. Yet success in achieving this is particularly important to facilitate multidisciplinary collaborations, both across the project and beyond, as researchers are often less aware of the data available to them outside of their own field, which may compliment or strengthen their work.

To ensure good data management across the project, we have developed a standardized spreadsheet template generator based on Darwin Core, also including terms from NetCDF-CF, that simplifies data publication. The populated templates are fed into a searchable metadata catalogue available through the SIOS (Svalbard Integrated Arctic Earth Observing System) website, that provides an overview of the data collection activities. Documented and published sampling protocols ensure consistent data collection procedures between different researchers, and to clarify how the data were collected for those not involved. The Nansen Legacy project is committed to the FAIR data management principles proposed by Wilkinson et al. (2016), and published datasets are harvested and made accessible via a single access point on the SIOS website. Frequent data management training is available to educate all project participants on the best practices for all of these procedures.
<table>
<thead>
<tr>
<th>Name</th>
<th>Department and Location</th>
<th>Title</th>
<th>Abstract</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maria D. Ananicheva</td>
<td>Department of glaciology of IG RAS, Moscow, Russia</td>
<td>New data on the glaciers of the Koryak Highlands: assessment of the state using satellite images and field studies</td>
<td>As a result of deciphering the Sentinel-2 images (2018 and 2019), 189 glaciers were discovered in the Koryak Highlands. The reduction in glacier areas in comparison with the USSR Glacier Inventory (late 1960s) varies widely from 20 to 74%. The error of glacier size determination from space images was estimated basing on ground-based survey of glaciers from a drone (UAV) during the expedition in September 2020. Contemporary glaciers of the Highlands are mainly corrie glaciers of the northern, NE and NW aspects. The change of equilibrium line altitude in the glacier groups within the river basins, identified in the Glacier Inventory for this area, vary from a minimum rise of several tens of meters to maximum values of 400-500 m. Spatial patterns of trends in mean summer and annual temperatures, total precipitation and precipitation for the cold period from 1966 to 2019 were compiled: the values of temperature trends increase towards the Kamchatka Peninsula, and precipitation trends - from the continent to the coastal zone. The obtained ratios of these trends with the degree of glacial groups reduction makes it possible to link the trends of the regional climate with the glacial systems' change that is the basis for predicting the reduction of glaciers for a period of several decades.</td>
</tr>
<tr>
<td>Masaki Kanao</td>
<td>Data and metadata sharing for polar sciences as the NADC in Japan</td>
<td>Data and metadata sharing for polar sciences as the NADC in Japan</td>
<td>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 and publish the data involving Japanese research activities as one of a National Antarctic Data Center (NADC). The data policy of PEDSC was established in February 2007, based on the requirements of the Standing Committee on Antarctic Data Management (SCADM) of the Scientific Committee on Antarctic Research (SCAR). At the International Polar Year (IPY2007-2008), a significant number of multi-disciplinary data have been compiled. These collected data/metadata have a tight collaboration with the Global Change Master Directory (GCMD), the Polar Information Commons (PIC), as well as several data centers belonging to the World Data System (WDS). In terms of data activities in polar communities of the Scientific Committee on Antarctic Research (SCAR) and the International Arctic Science Committee (IASC), tighter linkages of data/metadata sharing within the Asian Forum for Polar Sciences (AFoPS) countries has been discussed and should be further promoted by the involved Asian countries, in particular China, India, South Korea, Malaysia and Japan. In this presentation, a decade of history of polar data management is demonstrated, in particular focusing on data/metadata sharing, international collaboration among global data bodies and initiatives, data publication and citation, as well as data journal issues (Polar Data Journal).</td>
</tr>
</tbody>
</table>
We present an open, well-tested system for federated Arctic data discovery that leverages schema.org metadata for datasets. Through the POLDER effort, international Arctic data repositories have focused on adapting and adopting the Science on schema.org guidelines for publishing datasets. This lightweight vocabulary for representing dataset-level metadata is being adopted across data repositories worldwide, driven largely by Google's promotion. Many repositories also hold more detailed metadata records using a diverse set of well-established metadata languages, including the ISO 19115 family of specifications, Ecological Metadata Language, and many others. Schema.org metadata provides a common, lightweight mechanism for spanning these communities of practice alongside the more detailed original metadata. DataONE's metadata harvesting system handles and harmonizes all of these metadata formats.

The DataONE harvest engine is a scalable, flexible open source system that handles multiple protocols and specifications. A simple route for repositories to be discoverable through DataONE is to provide a sitemap listing schema.org entries. DataONE also supports other harvest protocols and the DataONE API. Regardless of the transfer protocol, once metadata is harvested, DataONE's indexing system validates the content against published schemas, harmonizes the vocabulary into a consolidated search index, and provides a suite of services such as assessing FAIR metadata quality and reporting on data access and citation. Indexed metadata is available through the DataONE search service for programmatic access, through augmented schema.org entries, and through the DataONE web site. The search interface provides thematic, spatial, temporal, and project-based searches across the Arctic. We leverage DataONE's system for building custom data portals to provide a federated Arctic Data search portal.

We present an overview of the DataONE harvester for schema.org along with challenges encountered in scaling harmonized metadata indexing across a diverse group of dozens of Arctic repositories.
| Author title (e.g. Ms., Mr, Prof.): Mr. | Mecum, Bryce |
| Author pronouns: he, him, his | |
| Department (optional): DataONE | |
| Institution (required): University of California Santa Barbara | |
| Street Address (optional): | |
| City, State, Country (required): Santa Barbara, CA, USA | |
| e-mail address (optional): | mecum@nceas.ucsb.edu |

| Author title (e.g. Ms., Mr, Prof.): Mr. | Nenuji, Rushiraj |
| Author pronouns: he, him, his | |
| Department (optional): DataONE | |
| Institution (required): University of California Santa Barbara | |
| Street Address (optional): | |
| City, State, Country (required): Santa Barbara, CA, USA | |
| e-mail address (optional): | nenuji@nceas.ucsb.edu |
| Matthew Davey | Dr Andrew Gray University of Edinburgh  
Dr Pete Convey BAS  
Prof Alison Smith University of Cambridge | Metabolic and Genomic Data Production from Antarctic Snow Algae Blooms | Snow algae are found in snowfields across cold regions of the planet, forming highly visible red and green patches below and on the snow surface. In Antarctica, they contribute significantly to terrestrial net primary productivity due to the paucity of land plants, but our knowledge of these communities is limited.

Here I provide information on how we collect, analyse and store the metabolic and genomic data that we collect not only in the lab but in the field and all the logistical challenges that presents. I will describe the metabolic and species diversity of green and red snow algae communities from locations in Ryder Bay (Adelaide Island, 68°S), Antarctic Peninsula.

These data show the complexity and variation within snow algae communities in Antarctica and provide initial insights into the contribution they make to ecosystem functioning. I also discuss logistics of sampling and data collection in remote regions for metabolomic studies, current and future research on linking these and new metabolic profiles to satellite images and how we use citizen science to enhance our sampling location and mapping planning for current and future seasons. |
The Antarctic Meteorological Research and Data Center (AMRDC) aims to bring Antarctic meteorological datasets, with an emphasis on observational data sets, to the broader community. Meteorological data presents unique challenges in storage, sharing, and description. To meet the needs of its extensive archival holdings, the research team explored several software platforms and metadata standards. The recently launched AMRDC Repository was ultimately built on open-source data platforms that encourage extensibility and interoperability. The AMRDC Repository’s information architecture seamlessly accommodates a diverse array of meteorological data formats, while also offering opportunities for collecting usage analytics as well as encouraging metadata harvesting and/or sharing with other discipline-specific repositories and databases. As a discipline-specific data center, the AMRDC aims to be an active host of real-time and archived data. Data servers and services will be a part of this effort including formal support for the Antarctic-Internet Data Distribution system. The research team’s future goals for the repository extend beyond data transfer and storage into implementing open-data standards which will allow users of meteorological data processing software to efficiently access the AMRDC Repository’s data holdings. The AMRDC will be a fully recognized Antarctic meteorological data center, for and by the community. It has the potential to be a participant in the WMO’s Antarctic Regional Climate Center network establishment efforts. This presentation will summarize the AMRDC Repository’s progress from concept to fully-fledged data repository, while outlining the research team’s commitment to FAIR repository standards (Findability, Accessibility, Interoperability, and Reusability) as well as plans for development and institutional partnerships.
The Mapping the Polar Data Ecosystem (MPDE) project aims to better document and understand the polar components of the global information ecosystem in order to target gaps in information resources. Its purpose is also to guide the ongoing development of the increasingly interconnected global information system in support of governance, research, livelihoods and myriad other applications. Under the guidance of the Polar Data Discovery Enhancement Research (POLDER) Working Group and related bodies, the Geomatics and Cartographic Research Centre at Carleton University has used the Nunaliit software framework to develop documentation tools for the MPDE project.

Nunaliit is used to display and collect data in a customizable multimedia atlas and visualization application. The primary focus of this atlas iteration was to compile existing data and provide tools for further documenting polar data catalogs into the MPDE atlas, as well as to link the activity and relationships between and among catalogs.

The Nunaliit MPDE atlas uses schemas, which act as templates for the data documents in the atlas, and modules, which dictate how the atlas narrative and visualizations are displayed. A schema for the documenting catalogs was created, based on the columns from tabular data from an existing survey. A custom module was also built to display the data in a table form, alongside a map module, which provides an alternate spatial view of the data.

We discuss the interoperability challenges resolved when building the MPDE atlas (e.g. importing and reformatting tabular data, geocoding etc.). The presentation concludes with an overview of next steps, including the development of a network visualization graph and publication of Linked Open Data.
Large tracts of Canada’s Arctic coastline are subject to a wide range of novel stresses, driven by varying combinations of warmer atmospheric and oceanic temperatures, increasingly extensive open waters, a longer open-water season, and more energetic storms. These impacts are amplified where the coast comprises un lithified sediments containing large volumetric fractions of ice. Near-shore ecosystems are also at risk, due to increased influxes of sediment and organic matter.

Nearly every settlement in the Canadian High Arctic is located in a coastal setting. Direct impacts of coastal geomorphological change include increasing risk of loss of property, infrastructure and cultural heritage through erosion of the land-base, exacerbated by rising sea-level and higher-amplitude storm surges. Such threats relate at least equally to sites of high significance for local culture and heritage. Any potential for disruption of near-shore food webs raises major concerns for communities’ food security, which depends heavily on marine biological resources.

These observations reveal a pressing need to quantify signs and rates of geomorphological change along Arctic coastlines, focusing on areas close to settlements and sites of cultural and archaeological importance. To inform initiatives towards mitigation and remediation, as well as to support further scientific investigation, the results of such an assessment must be made available in accessible form to local communities, heritage resource managers, those responsible for local infrastructure and planning, associated administrative bodies, and northern colleges.

This presentation describes a project led by the Inuvialuit Regional Corporation to accomplish these goals for the Inuvialuit Settlement Region, as a use-case under the umbrella of the Canadian Consortium for Arctic Data Interoperability. We will discuss how we collaboratively identified a suitable approach to conducting the assessment; outline progress towards implementing a discoverable, searchable and interoperable database through which to disseminate results; and assess options for validating modelled results through ‘citizen-science’ participation.
| Mikkel H Bojesen | Closing the gap of navigational risk in unchartered Arctic waters with novel satellite-based data products | Maritime activity across the Arctic is increasing alongside an increased economic activity and geopolitical awareness in the region. In this context, supporting safety at sea, safeguarding national sovereignty, and offering logistical support to the civil society, efficient naval and first responding capacities are being important as ever before across Arctic societies. But arctic maritime operations, are associated with high level of navigational risk due to missing or inaccurate nautical information and vast areas are still left uncharted. Consequently, each year the harsh Arctic environment claims its victims. In situations of distress, first responders must very often search enormous areas with little opportunity for efficiently identifying the exact location of those in need, since means of communication is very limited, outside the populated areas. In this presentation we showcase examples of how co-production between navies and industry research can bring about satellite-based decision support for both efficient object identification and navigational risk-minimization in terms of mapping of submerged rocks, intertidal zones and precise delineation of coast lines. This data production relies on recent developments within application of multi-temporal bathymetric retrieval models and high resolution Synthetic Aparture Radar (SAR) imagery. Demonstration work in Greenland is presented together with a discussion on future applications across the Arctic. |
| Noor Johnson | Dr. Matthew Druckenmiller, he/him, National Snow and Ice Data Center, University of Colorado Boulder  
Prof. Peter Pulsifer, he/him, Carleton University | Looking back, looking ahead: The ELOKA Program at 15 years | The Exchange for Local Observations and Knowledge of the Arctic (ELOKA) partners with Indigenous organizations and researchers to facilitate the collection, preservation, exchange, and use of Indigenous Knowledge and community-based observations. ELOKA was initiated in 2006/2007 during the International Polar Year to address a gap in resources and support for data management focusing on Indigenous Knowledge and Arctic community data. Since then, Arctic research has undergone a period of significant change, with a growing emphasis on community-led research and co-production of knowledge. Fifteen years since its founding, ELOKA is beginning a new five-year period of collaborative work with new and existing partners from across the Arctic funded by the National Science Foundation. This presentation will summarize ELOKA’s progress and challenges over the past fifteen years, situating them in the larger context of a changing research environment. We will share an overview of the ELOKA community data management (CDM) system and protocol as well as plans for the next phase of ELOKA’s work focusing on understanding and enhancing the use and usability of CDM infrastructures. This includes co-developing use cases of specific data products with partners, convening two thematic working groups to exchange approaches and ideas, and collaborative development of a use and usability framework for community data management. ELOKA’s collaborative effort to develop a usability framework will broadly consider how cross-cultural sharing, storytelling, data sovereignty protocols, and capacity building may strengthen the use and dissemination of community data. |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Paul Arthur Berkman</td>
<td>Arctic Sustainability and the Satellite Record of Maritime Ship Traffic</td>
<td>The oldest continuous satellite record of maritime ship traffic north of the Arctic Circle will be discussed with implications for sustainable development in the Arctic Ocean. The satellite Automatic Identification System (AIS) big-data are being analyzed with extreme speed and cost-effectiveness in the cloud, applying 'space-time cube' methodologies. As a case study, emphasis of this oral presentation will be on global precedents that are being established with the &quot;precautionary approach&quot; that is entering into force for Arctic and non-Arctic States with the Agreement to Prevent Unregulated Commercial Fishing on the High Seas of the Central Arctic Ocean.</td>
<td></td>
</tr>
<tr>
<td>Peter L. Pulsifer</td>
<td>The CCADI Technical and Management Teams</td>
<td>Towards an Ecosystem of Interoperable Data Infrastructures: contributions from the Canadian Consortium for Arctic Data Interoperability</td>
<td>Polar data are required by the Arctic community to support research on a wide variety of topics and domains (physical, social, economic, etc.). Mobilizing and maximizing the value of Arctic data requires data infrastructures that can serve and mediate data for a wide range of different users and applications. Developing useful and usable infrastructures requires attention to system design and emergence at many different levels: foundational data storage, management, and preservation; methods and technologies for transforming and mediating data; representation and portrayal for different audiences; use of emerging technologies such as online platforms, machine learning, semantics, and natural language processing; ensuring respectful and ethical use of data and infrastructure. In this paper, we present a review and analysis of selected activities and experiences of the international Arctic and polar data communities, primarily through the activities of the Arctic Data Committee, but including partnered initiatives with the Antarctic and broader global data communities. More specific examples and conclusions are drawn from the work of the Canadian Consortium on Arctic Data Interoperability and its partners. From these experiences, we share a number of lessons learned and discuss future directions.</td>
</tr>
<tr>
<td>Petra ten Hoopen</td>
<td>UK Polar Data Centre</td>
<td>Polar biodiversity data: from a national marine platform to a global data portal.</td>
<td>Global access to accurate biodiversity data is a prerequisite to our progress in understanding biodiversity dynamics in an ecosystem. Despite recent major advancement in sharing data on the world's species, there are still many challenges. One of them relates to the mechanics of guiding data systematically from its provenance to end users. This is not surprising considering that behind every species occurrence data point in a global biodiversity data portal lies a huge effort. It takes years to orchestrate a successful sampling campaign and requires significant resources to manage samples obtained in often extreme remote conditions and to secure preservation of and access to the acquired data. Here we briefly describe biodiversity data flow from a polar ship to a national data repository and to a global data portal and highlight few crucial points in this process which aims to systematically share information pieces into the mosaic of our polar species biodiversity knowledge.</td>
</tr>
</tbody>
</table>
| Petra Ten Hoopen | Dr. Bricher, Philippa, Southern Ocean Observing System, Hobart, Tasmania, Australia  
Dr. Pfeil, Benjamin  
he/him  
Bjerknes Climate Data Centre  
University of Bergen  
Bergen, Norway  
Benjamin.Pfeil@uib.no  
and the SOOS Data Management Sub-Committee | Building a science community through data tools | The Southern Ocean Observing System is building a community out of large numbers of scientists, data managers, and institutions that don’t usually work together. It spans more than sixty countries, National Antarctic Programs, and oceanography programs across a multitude of scientific disciplines. Creating a cohesive and collaborative community from such disparate actors is a challenging task, but a necessary one if the SOOS vision of a sustained system of coordinated ocean observations is to be achieved.

For SOOS, our data systems are one of the most visible ways in which we bring together activities from across the community, as well as providing an incentive for our scientists to engage with formal data management activities within their institutions and nations.

SOOSmap is a map-based portal of highly curated and standardised datasets, developed for SOOS by EMODnet Physics, and it provides not just access to datasets but also a very visible way to identify gaps in observing and data sharing efforts in the Southern Ocean, and a tool to support and improve observing system design.

In this presentation, we will also describe our metadata portal, hosted by NASA, and our engagement through POLDER with efforts to develop federated metadata search for polar regions, which will make it much easier for researchers to find and access the wide variety of data types which haven’t yet been incorporated into a standardised and QCed data management system.

Such tools are crucial for creating a tangible shared outcome of SOOS’ efforts and the desire to get their data “on the map” is a direct incentive for our science community to engage with data managers to achieve the SOOS vision. |
| Pip Bricher                     | Mr. Keeble, Simon  
He/him  
Blue Lobster  
Dr. Badhe, Renuka  
She/her  
European Polar Board  
Joseph Nolan  
He/Him  
European Polar Board | Who’s going where and how? What can they do when they get there?  
Towards an integrated database for polar research logistics and infrastructures  
As polar research and observing systems come of age, there is increasing interest in sharing information about logistical resources, which in turn makes it possible for the resources themselves to be shared across institutions and nations, facilitating multi-agency collaboration.  
In contrast with the extensive efforts that have been made to align metadata standards for discovering scientific datasets, attempts to standardise the way logistical resources and infrastructures are documented are in their infancy.  
A new collaboration between the Southern Ocean Observing System (SOOS) and the European Polar Board (EPB) is a prime example of the challenges inherent in aligning these infrastructure and logistics databases to maximise their use and share resources.  
DueSouth, SOOS’ database of upcoming expeditions to the Southern Ocean, holds information about planned expeditions by Antarctic and oceanographic researchers, as well as fishing and tourist vessels. However, it holds very little information about the vessels, aircraft, and research stations associated with those expeditions, and has struggled to collect information about the projects associated with them. Meanwhile, the European Polar Board’s infrastructure database, developed collaboratively with EU-PolarNet, INTERACT, EuroFleets, COMNAP, SIOS and others, contains detailed information about European Arctic and Antarctic research stations, vessels, aircraft and other assets and facilities. The EPB recently agreed to host DueSouth for SOOS, which will provide an opportunity to integrate the databases and align core fields. The combination of the databases in a modern, cloud-based, serverless technology provides high availability and high performance, with a scalable platform to be made available to the polar communities. This facilitates easy access to search and discovery of expedition, logistics and infrastructure resources, maximising use and collaboration.  
In this presentation, we will discuss the challenges and lessons learned so far in integrating these databases, our engagement with similar efforts from EU-PolarNet 2 and the Polar Observing Assets Working Group (POAwg), and our recommendations for the way forward. |
| Pip K Bricher  | Ms Duerr, Ruth  | Progress on federating metadata search for the polar regions | Polar-relevant metadata is stored in catalogues across the globe and includes a wide variety of data types, such as measurements taken from instrumentation, maps and atlases, photographs, and recordings of oral histories. For many data types, there is no realistic prospect of standardising and aggregating the data itself; therefore, federated metadata search is the only viable way to make these datasets easily discoverable, and so maximise their value. POLDER is a working group of the Southern Ocean Observing System, Standing Committee on Antarctic Data Management, and the Arctic Data Committee, and it is extending some of the existent guidance and tools needed to make federated metadata search possible for the polar research community. A key element of the POLDER effort is generating a best practice guide to implementing schema.org, as a potentially lightweight discovery metadata standard that will sit alongside data centres’ existing metadata standards. We hope that schema.org metadata will thus help navigate between the multitude of existing, richer metadata standards that are commonly used in data centres, and allow discovery of the long tail of data that’s currently hard to find. Once the best practices guide is completed, POLDER intends to implement a small pilot federated search project through the implementation of the POLDER Schema.org mark-up recommendations, building on and contributing to the tool development in allied communities. This federated search is intended to serve the polar research community by creating a single user interface for researchers, community members and data managers to search across numerous polar repositories in a single query. In this presentation, we will describe the process to date on creating a template for schema.org metadata and a best practice guide to accompany that template, as well as the lessons learned so far in the process of developing the tools to support federated metadata search. |
| Ms Ingram, Rebekah  | she/her  |  |  |
| Mr de Bruin, Taco  | he/him  |  |  |
| Ms Verhey, Chantelle  | she/her  |  |  |
| Mr Godøy, Øystein  | he/him  |  |  |
| and the POLDER working group  |  |  |  |
| Rebekah R. Ingram | Linking attributes, linking systems: advancing knowledge translation in the Canadian Consortium for Arctic Data Interoperability | The Canadian Consortium for Arctic Data Interoperability (CCADI) aims to develop a research data infrastructure that facilitates information discovery, promotes the sharing of standards, and enables interoperability among existing Arctic data infrastructures at universities and institutions across Canada. To achieve the goal of interoperability, several issues regarding semantics and terminology must be considered. These include: the wide variety of data pertaining to the Arctic, ranging from climate and forecasting data and cryosphere and oceanographic data to social science data centring upon Inuit experiences of the landscape; the differences in terminological use between different data centres and institutional repositories; differences in the use and type of instrumentation; and the framework used for metadata cataloguing. This paper outlines our semantic strategies for the CCADI project. Through interaction with other organizations working with Arctic Data such as Arctic Data Committee-Sustaining Arctic Observing Networks (ADC-SAON), Polar Data Discovery Enhancement Research (POLDER) Working Group, and Earth Science Information Partners (ESIP), we focus on aligning ourselves with and leveraging work already in progress with respect to new and emerging standards and their vocabularies. Close collaboration with domain experts also helps us maintain balance between terminology that is too narrow and too broad while capturing the relationships between different groups’ terms without overly restricting them. Through a discussion of these issues and the steps taken to resolve them, we seek to document and share a usable workflow as an available roadmap for others working toward semantic interoperability in data-based initiatives. |
| Rob E Jennings | Digital Antarctica – A FAIR approach to Australian Antarctic research data | The Digital Antarctica project is an initiative under the Australian Antarctic Program Partnership that aims to increase the FAIRness of the collaborating partners, with a specific focus on interoperability. The initiative was based on recommendations coming out of the Clarke review of the Australian Antarctic Science Program, which called for a federated data architecture to facilitate a world-class centre for Antarctic data analytics.

The Australian Antarctic Program Partnership consists of predominantly science-focussed organisations performing research across the Antarctic and Southern Ocean regions. Each organisation collects and serves data to meet the unique focus and purpose of the organisation and its data users.

Because of the disparate nature of the organisations and their needs, there is no definitive and simple way to find and integrate Australian Antarctic data. Any researcher or policy maker looking for Australian Antarctic data must first know where to look.

The Digital Antarctica project is addressing this by: working with the partnership organisations to define agreed standards of data storage, data delivery, and interoperable services; assisting organisations in implementing those standards; and constructing a prototype to demonstrate how those services and standards can be used to create a federated data model to help find, access, and integrate Australian Antarctic research data.

This presentation will delve into what makes Digital Antarctica unique. It will also cover how we have addressed the challenges which arise from working closely with stakeholders from a small but diverse group of partner organisations, including: agreement on various aspects of the initiative (for example, terms, scope, goals, and standards); engaging with organisations of varying sizes; and finding approaches that suit a wide range of data levels, data types and breadth of data offerings. |
<table>
<thead>
<tr>
<th>Ms. Kaylin Bugbee</th>
<th>Advancing Discovery in NASA’s Science Mission Directorate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ms. Mark Parsons</td>
<td></td>
</tr>
<tr>
<td>Mr. Daniel Berrios</td>
<td></td>
</tr>
<tr>
<td>Mr. Ashish Acharya</td>
<td></td>
</tr>
<tr>
<td>Dr. Emily Foshee</td>
<td></td>
</tr>
<tr>
<td>Dr. Anirudh Prabhu</td>
<td></td>
</tr>
<tr>
<td>Mr. Ahmed Eliesh</td>
<td></td>
</tr>
</tbody>
</table>

NASA’s Science Mission Directorate (SMD) has established a long-term goal of developing and implementing capabilities that enable open science across NASA’s scientific disciplines. The SMD has five divisions including Earth Science, Planetary Science, Heliophysics, Astrophysics, and the Biological and Physical Sciences. Each division within SMD produces, examines and catalogs tens of petabytes of data to fulfill scientific objectives and provide scientific findings to millions of people.

Currently, access to these data tends to be discipline and division specific. There currently are no mechanisms that provide discovery and access to heterogeneous scientific data from across all five divisions. Working towards such mechanisms and helping to cultivate a strong community of practice across SMD to share recommended practices and discuss common strategies is the goal of the project described here.

This project is using semantics to prototype cross-divisions discovery methods based on user submitted use cases that are explored in a series of workshops involving domain, data, and semantic expertise. Two use cases will be explored over the summer. The first attempts to discover biological sample data acquired during periods when the International Space Station was flying through relatively high radiation environments such as the South Atlantic Anomaly. The second involves discovering atmospheric data from planets both within and outside our solar system. Between these two use cases, data from all five divisions is needed.
<table>
<thead>
<tr>
<th>Author</th>
<th>Affiliation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Buttigieg, Pier Luigi</td>
<td>HGF MPG Joint Research Group for Deep-Sea Ecology and Technology, Bremen, DE</td>
</tr>
<tr>
<td>Berg-Cross, Gary</td>
<td>Ontolog Forum, Washington, DC area USA</td>
</tr>
<tr>
<td>Blumberg, Kai L</td>
<td>University of Arizona, Tucson, AZ USA</td>
</tr>
<tr>
<td>Schildhauer, Mark</td>
<td>UCSB, Santa Barbara, CA USA</td>
</tr>
<tr>
<td>Whitehead, Brandon</td>
<td>Landcare Research, Manaaki Whenua, New Zealand</td>
</tr>
<tr>
<td>Nancy Wiegand</td>
<td>University of Wisconsin-Madison, Madison, Wisconsin, USA</td>
</tr>
<tr>
<td>Verhey, Chantelle</td>
<td>International Science Council, World Data Systems, Victoria, BC Canada</td>
</tr>
</tbody>
</table>

The Earth Science Information Partners (ESIP) federation, is a community of more than 170 organizations including “USA federal data centers, government research laboratories, research universities, education resource providers, technology developers, and various nonprofit and commercial enterprises” with a vision of “promoting the collection, stewardship and use of Earth science data, information and knowledge that is responsive to societal needs”. Over the years, with the recognition that the role of human and machine actionable semantics, as represented by terminologies in the form of controlled vocabularies, glossaries, thesauri, and ontologies, is becoming increasingly important to these goals, a Semantics Harmonization cluster formed. This group has spent several years grappling with methods for harmonizing existing diverse semantics resources, with cryospheric and polar terminology as its first target. Here we describe the resulting outputs and the leading practices developed and used.
<table>
<thead>
<tr>
<th>Sabina Di Franco</th>
<th>Mr. Roberto Salzano</th>
<th>Metadata for a spectral library on snow and ice and FAIR principles</th>
</tr>
</thead>
<tbody>
<tr>
<td>he Institute on Atmospheric Pollution (IIA) - CNR, Italy Florence, Italy <a href="mailto:roberto.salzano@cnr.it">roberto.salzano@cnr.it</a></td>
<td>he Institute on Atmospheric Pollution (IIA) - CNR, Italy Florence, Italy <a href="mailto:enrico.boldrini@cnr.it">enrico.boldrini@cnr.it</a></td>
<td>To share a collection of field-collected hyperspectral data of snow (Snow and Ice Spectral Library - SISpec), a metadata profile, specific to cryosphere properties was created to describe the dataset and increase interoperability. The spectral data were collected in Polar areas and were intended for analysing and processing multispectral satellite images to study the seasonal evolution of snow surface. The metadata profiles of various spectral libraries on rock, soils, and vegetation were examined. In the absence of specific standards for the cryosphere, the international classification of snow was adopted as a guideline (Fitz, 2009). To ensure compliance with the &quot;open access&quot; rules, we searched for a balance between the ERC guidelines, the FAIR principles defined by the Research Data Alliance, and the GEO Data Sharing Principles. The ISO 19115 standard and the INSPIRE guidelines were chosen as the standard framework to describe SISpec. When the available metadata schema was not sufficient or well-suited, metadata extensions or new detailed metadata components were created.</td>
</tr>
</tbody>
</table>
Controlled vocabularies for polar activities

Controlled vocabularies are useful tools for organizing information. They help to have the specific terminology of an area of knowledge to catalogue and retrieve information and enhance semantic interoperability. In a controlled vocabulary are collected variants and synonyms of concepts linked together in a logical order, or sorted into categories.

SnowTerm is an example of a structured reference multilingual scientific and technical vocabulary, covering the terminology of a specific knowledge domain such as the polar and the mountain environment. The thematic areas, at present, deal with snow and ice physics, snow and ice morphology, snow and ice radiometry, remote sensing and GIS in cryosphere environment, sea ice, avalanches and glaciers.

BiodivThes represents a vocabulary of terms covering the field of environment, ecology and biological diversity. It includes both biotic and abiotic concepts.

For both vocabularies the identification, acquisition and harmonisation of controlled multilingual terminologies brought to the development of a complete basic reference list of terms in English, partially multilingual.

The terminology of these sources was analysed with respect to the degree of semantic relevance in the field excluding both terms too generic or considered as non-pertinent.

At present, SnowTerm database contains around 3,700 terms mainly in Italian and English. BiodivThes contains around 1,800 terms in English and Italian. For the vertical structure of the vocabularies, we adopted the Classification Scheme already in use for the development of the CNR EARTH Thesaurus. The hierarchical setup is based on facets; according to its intrinsic features, the structure can be used as a semantic reference system, stable and partially independent from the context.

We are implementing a glossary and thematic structure. The possibility of applying different thematic schemes could allow the exploration of concepts according to different perspectives, which may emphasize particular and contingent aspects.
<table>
<thead>
<tr>
<th>Sandy Starkweather</th>
<th>Mr. Jan Rene Larsen he/him AMAP Secretariat, Tromso, Norway</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linking many centers – Progress on defining and engaging around Sustaining Arctic Observing Network’s Roadmap for Arctic Observing and Data Systems (SAON-ROADS)</td>
<td>SAON-ROADS has matured as a concept since it was formally introduced to the 2020 Arctic Observing Summit (AOS 2020) as a framework for planning, partnership development and integration of the needed enhancements to the future Arctic observing and data infrastructures. AOS 2020, with its 350 attendees from 28 countries, provided valuable opportunities to dialog about the planning and partnership constructs, including assessment strategies organized around societal benefit, planning mechanisms organized around essential variables and governance for an integrated advisory/expert process. A subsequent series of SAON-led governance discussions have enhanced our understanding of how SAON will use ROADS to draw together its diverse network of partners into a coherent structure of advising and subject matter experts to develop recommendations. Collectively, these partners have turned to SAON to guide Arctic observing and data system development, yet it is important to recognize that SAON’s ability to influence partner actions through collaborative governance is non-hierarchal and therefore constrained. Ostrom (2010) would describe the SAON governance model as polycentric; as the name implies this describes governance systems through which multiple centers of authority are working toward a common goal. This talk will present the role of polycentric concepts in informing the shape of the planning tools and governance constructs within the ROADS process and provide updates and status on the process itself, including how the data community might already begin to engage.</td>
</tr>
</tbody>
</table>
Data Life-Cycle Management to encompass an automated Data Pipeline at Ocean Networks Canada: Community Fishers Program

Ocean Networks Canada (ONC), an initiative of the University of Victoria, operates world-leading ocean observatories and data repository services. Marine infrastructure installed on the West, East, and Arctic coasts of Canada by ONC and partner organizations deliver data from hundreds of instrument types deployed on a variety of platforms with different data acquisition systems. ONC’s data infrastructure, Oceans 2.0, serves a growing array of instruments and platforms from data collection, assurance, and description, to data preservation, discovery, integration, analysis, and distribution. Community Fishers is ONC’s program that supports Citizen Science, oceanographic observations collected by individuals on a small marine vessel and facilitated by a custom in-house developed android application. This program started back in 2015, with ONC’s partnership with the Pacific Salmon Foundation’s Salish Sea Marine Survival Project. Here, we feature the evolution of the original framework into a full-fledged automated data life-cycle; it begins with instrument integration and sample planning, and the cycle concludes with a near real-time display (and archival of multiple data products) of quality-controlled data for the sampled water column. This data management pipeline was further extended to support community-based monitoring in the Canadian Arctic at Frobisher Bay with the Indigenous Community at Iqaluit through a partnership with the Fisheries and Oceans Canada (DFO). Some of the initial challenges, such as reduced tablet battery life in the low-temperature Arctic, and rendering the Geo-spatial map portal for near real-time display of the data in low-bandwidth remote locations, were overcome with continued development. During this data life-cycle, ONC adheres to research data community standards and best practices in data management including FAIR (findable, accessible, interoperable, and reusable) data principles. As a member of the World Data System since 2014, it continues its long-standing commitment to meeting rigorous data repository certification criteria.
As of writing the Argo project has a good coverage of the open oceans, and provides an unique open dataset for the oceans physical state. However, the closer we go to the poles, the scarcer the coverage of Argo floats gets. The main reason for this is that operating the floats is more challenging with the constant risk of ice. The practices to operate in such conditions are various, and for a new operator the risks and procedures can be daunting.

In the Euro-Argo Rise project, one of our targets is to further develop the Argo operations in both Arctic and Antarctic regions. For this end, we work on two approaches: First is to develop and improve ice-avoidance algorithms and review their performance and limitations on both regions. The second approach is to gather up experiences and lessons learned from float operations in these regions. The new developments within the project and the historical information will be combined to form an extensive best-practices guidance for lowering the barrier for current and future operators to enter in the polar regions.

Within the project we have deployed Argo floats with Ice Sensing Algorithms (ISA) in both polar regions, and in the Baltic Sea. The Baltic Sea is added, as it has ice cover for only a part of the year, and for its size, it is easy to retrieve the floats after missions for further study. In this presentation we will show the results of our first experiments and the findings that will be the basis of the best-practices proposed by the project.
Data policies are important tools to set expectations among the observing community, other rights holders and users about how and what data to share and how to treat data shared by others. As a primary resource for science and cross-disciplinary collaboration data should be managed according to widely recognised principles. Data policies based on shared, fundamental principles will clarify obligations and stipulate norms with respect to data sharing, access, management, preservation, and acknowledgment. Agreement on such principles will facilitate collaborative research and serve to increase the productivity of data for scientific, operational, management and decision-making purposes. This is particularly important in polar regions, where data collection tends to be prohibitively expensive.

Polar data sharing and open data policies go back to the First International Polar Year (1882-1883). The fourth International Polar Year (2007-2008) provided a major impetus to improving data management at both poles and introduced a seminal data policy specific to polar research. Later on, several polar science bodies introduced their own data policies modelled on the IPY data policy. While these share major ideas and obligations, they were not written to be explicitly aligned and differ in important aspects. In addition, they pre-date the FAIR principles and other key principles for data management, as well as other modern technological and institutional changes that significantly impact the ways in which scientific data are managed and shared.

In this paper we examine such developments along with the data policy recommendations and policies of important global and regional organisations. We conclude by recommending ten fundamental principles for adoption in polar data policies.
<table>
<thead>
<tr>
<th>Gothenburg, Sweden</th>
<th>Mr. Riel, Simon, Natural Resources Canada, Ottawa, Ontario, Canada</th>
</tr>
</thead>
<tbody>
<tr>
<td>Goa, India</td>
<td>Mr. Samy, VS, National Centre for Polar and Ocean Research, Goa, India</td>
</tr>
<tr>
<td>Den Helder, Netherlands</td>
<td>Mr. Tacoma, Marten, Royal Netherlands Institute for Sea Research, Den Helder, Netherlands</td>
</tr>
<tr>
<td>Sion, Switzerland</td>
<td>Ms. Thomas, Jen, Swiss Polar Institute, Sion, Switzerland</td>
</tr>
<tr>
<td>Pretoria, South Africa</td>
<td>Dr. Treasure, Anne, South African Environmental Observation Network (SAEON) Information Systems, Pretoria, South Africa</td>
</tr>
</tbody>
</table>
Neural networks and transfer learning for glacial lake outburst flood monitoring

The severity and frequency of glacial lake outburst floods (GLOFs) continues to increase as anthropogenic climate change is spurred on by increased carbon emissions in the atmosphere. As glacial melting and permafrost melting increase in intensity, regions with glaciers experience higher rates of flooding, which can cause immense economic loss and hundreds of lives lost in these events. In order to monitor the increasingly large number of glacial lakes, especially those that are located near human population centers, it is important to have computational mechanisms in place for automated real-time assessment. We propose a machine learning framework to address this issue. By training a convolutional neural network (CNN) for this problem on multitemporal satellite imagery, we enable deployable technologies that predict GLOF events and impacts on surrounding areas. In particular, we collect high-resolution satellite imagery data from previous GLOFs around the world, such as in Iceland, Alaska (United States), Pakistan, and Tibet, utilizing repositories provided by ESA and NASA. We curate a dataset based on paired images (pre- and post-GLOF). In this way, we can train the CNN on the change detected between these two instances, which can further aid in predictions in the form of an output from 0 to 10 indicating the severity of damage caused due to the glacial outburst. However, because machine learning algorithms require a large quantity of data (hence “big data”), we must also employ transfer learning. We propose a Markov logic network framework to achieve this, incorporating data from events that were not necessarily GLOFs but included glacial movement and/or flooding. When deployed, models like the one we propose can allow for both the monitoring of GLOFs in action as well as predict GLOFs in the near future by assessing changes using data collected from satellites in real time.
Walking On Thin (Arctic) Ice: Negotiating the Digital Landscape of Protected and Sensitive Arctic Social Science Research Data

Federally-funded US Arctic social science researchers working with sensitive/protected datasets face a number of unique data preservation and sharing challenges. When appropriate, the National Science Foundation (NSF) requires data generated from projects to be made publicly available via a data repository such as the NSF Arctic Data Center (ADC). Yet researchers can find it challenging to respond to sometimes conflicting norms, guidelines and regulations, including, for example, Institutional Review Board (IRB), consent forms and data management plans; implementation of CARE (Collective benefit, Authority to control, Responsibility, and Ethics) and FAIR (Findable, Accessible, Interoperable, Reusable) data principles; and Indigenous data sovereignty agreements that uphold Indigenous rights to manage their own data. This study provides an overview of the landscape of Arctic social science data management and describes the development of a survey of Arctic NSF award recipients designed to identify which factors assisted, delayed, or hindered researchers from sharing their data. These results, coupled with the review of current literature and the state of the art of repositories handling sensitive and protected data, will inform the development of proposed features at the NSF ADC incorporating changes to data submission processes. Specifically, these features will include metadata entry fields for cultural heritage and enable flagging of sensitive and protected data and the different levels that will govern their accessibility and reuse. Preliminary data will be gathered from deployment of the new features testing the degree of increased data sharing assurance to PIs who may otherwise be apprehensive about sharing data and fulfilling their NSF grant data management requirements.
A fundamental challenge exists for assessment, planning, and synthesis of Arctic observing. Assets such as sites, transects, observatories, projects, and programs are deployed in a diverse and distributed fashion across numerous observing systems. At this time, it is difficult to strategically assess status, overlap, and gaps because most inventories and portals are limited in scope. Furthermore, only a fraction of observing systems share information about observing assets in a way that can be accessed, harmonized, and aggregated for a comprehensive perspective. To help address this challenge, a new “Polar Observing Assets Working Group” has been formed under the SAON Committee on Observations and Networks (CON; see https://polarobservingassets.org). This group builds upon steps taken by the Polar data community for the interoperability of “dataset-level” metadata, but in this case for discovery-level details in “asset-level” metadata. The group will identify and promote best practices for the use of relevant metadata standards, controlled vocabularies, crosswalks, federated search, and linkages to scientific datasets. For more details on three identified tasks, see https://www.polarobservingassets.org. Outcomes will make it easier for networks and planners to avoid duplicated effort while optimizing resources. Participation is open and encouraged, and will help to showcase and integrate the summed contributions of multiple systems.