



HORIZON 2020
Research and Innovation action
Grant Agreement No. 730965



ARICE: Arctic Research Icebreaker Consortium:
A strategy for meeting the needs of marine-based research
in the Arctic

Deliverable 7.2. Report on user and stakeholder
feedback on the current status of their data
management and potential gaps

Submission of Deliverable

Work Package	WP7
Deliverable no. & title	D7.2 Report on user and stakeholder feedback on the current status of their data management and potential gaps
Version	2
Creation Date	01.07.2018
Last change	05.05.2019
Status	<input type="checkbox"/> Draft <input checked="" type="checkbox"/> WP lead accepted <input checked="" type="checkbox"/> Executive Board accepted
Dissemination level	<input checked="" type="checkbox"/> PU-Public <input type="checkbox"/> PP- Restricted to programme partners <input type="checkbox"/> RE- Restricted to a group specified by the consortium <input type="checkbox"/> CO- Confidential, only for members of the consortium
Lead Beneficiary	AP
Contributors	<input checked="" type="checkbox"/> 1 – AWI, <input checked="" type="checkbox"/> 2 – SPRS, <input checked="" type="checkbox"/> 3 - NPI, <input checked="" type="checkbox"/> 4 - ULAVAL, <input checked="" type="checkbox"/> 5 – UAF/CFOS, <input type="checkbox"/> 6 – AP, <input checked="" type="checkbox"/> 7 – CSIC-UTM, <input type="checkbox"/> 8 – CNR, <input type="checkbox"/> 9 - WOC, <input type="checkbox"/> 10 – IOPAN, <input type="checkbox"/> 11 – FMI, <input type="checkbox"/> 12 - CNRS, <input checked="" type="checkbox"/> 13 – NERC-BAS, <input type="checkbox"/> 14 – DTU-AQUA <input checked="" type="checkbox"/> 15 – ARCTIA
Due date	30.09.2018
Delivery date	17.05.2019

Table of Content

1 Introduction.....	4
1.1 Abstract.....	4
1.2 Objectives of the report	4
2 Communication with users and other stakeholders	5
2.1 ARICE Kick-off Meeting and General Assembly 2018.....	5
2.2 Collaboration with other WP’s within the project	5
2.3 Questionnaires.....	5
2.4 Collaboration with potential users, NODCs and data repositories	5
3 Data Stakeholders	6
3.1 Data creators	6
3.2 Data management systems	8
3.2.1 Primary data repositories.....	9
3.2.2 Other national data centres	14
3.2.3 Data repositories and aggregators.....	14
3.3 Data users	16
Summary and Potential gaps	17
Final Recommendations.....	17

1 Introduction

1.1 Abstract

The dramatic changes taking place in the Arctic Ocean have generated an extremely high interest not only in the scientific community, but also in the citizens and policy makers. Arctic climate is now warming rapidly and much larger changes are projected. Evidence of the recent warming of the Arctic is provided by records of increasing temperatures, melting glaciers, sea ice, and permafrost, and rising sea levels. Changing Arctic brings not only concerns but also possibilities. The Arctic is estimated to hold as much as 30% of the world's unused resources. Sea ice in the Arctic is diminishing much faster than predicted, due to climate change, making year-round Trans-Arctic shipping a real possibility in the near future. Data from the Arctic Ocean are thus continuously needed, not only to evaluate the current state of the Arctic Ocean but as well to observe, quantify and evaluate those environmental changes and their ecological impact.

Icebreaker-based research is crucial to this. Data from the ice-covered areas of the Arctic Ocean is very scarce, and thus ARICE will contribute to obtain a better picture of the ARCTIC by funding cruises in this area, on board of European and international research icebreakers. This deliverable will focus on the feedback from users and stakeholders on the current status of their data management and potential gaps.

The deliverable “D7.2 – Report on user and stakeholder feedback on the current status of their data management and potential gaps” is the second deliverable to fulfil in the “Enhancing virtual and remote access to data” Work Package (WP7).

1.2 Objectives of the report

The main goal of the Deliverable 7.2 is to collect user and stakeholder feedback through the organisation of online workshops, participation in relevant events organised by other WPs, as side events at major relevant conferences and through questionnaires and partner surveys, in order to:

- identify who are the potential data users and stakeholders;
- To describe the current data management systems of data providers;
- To identify the gaps in current data management systems and describe how the ARICE data management system will contribute to improving the services currently offered.

2 Communication with users and other stakeholders

Working on this deliverable required a broad and intensive communication with potential users and stakeholders. We used every opportunity to talk to the users to determine their data requirements and how ARICE data would fit into their data needs. Besides personal communication a questionnaire has been undertaken to find out how the stakeholders plan to redistribute their data. Polar data centres, who are also among the stakeholders, have been questioned in order to determine their preferences.

2.1 ARICE Kick-off Meeting and General Assembly 2018

The Kick-off and project annual meeting took place on the 6-7 February 2018 at The German Maritime Museum in Bremerhaven, Germany, and was open to all project partners, external collaborators and interested colleagues.

A data management workshop was organized during the event providing users and data managers an opportunity to address project's data management issues. Results of the discussion were documented and incorporated in the data management plan (DMP). DMP was published as project Deliverable 7.1.

2.2 Collaboration with other WP's within the project

WP7 closely collaborated with WP6 in order to obtain knowledge about the environmental parameters which are collected by automatic measurements during navigation in ice covered seas.

Outcome of the WP6 – “Expanding the monitoring and observation capacities in the Arctic Ocean” was extensively used for developing the DMP. WP6 conducted a survey on underway and automatic data collection in the Arctic Ocean. This survey is not a subject of this report. Detailed information on it can be found in the Survey on underway and automatic data collection in the Arctic Ocean (Deliverable 6.1) – see the project website www.arice.eu.

2.3 Questionnaires

A survey was developed and redistributed among the operators of the participating icebreakers. Its scope and content is shown in the following section of the report.

2.4 Collaboration with potential users, NODCs and data repositories

An extensive communication with potential users, data repositories and representatives of national data centres has been undertaken. We conducted several custom surveys to identify the users who could be interested in ARICE data. We determined and contacted a broad list of data repositories and aggregators to find out more about them and their data preferences.

3 Data Stakeholders

Data stakeholders are all the individual or groups that could affect or be affected by the data under discussion.

In our case, we can divide the data stakeholders in three groups:

- those who create data,
- those who set rules and requirements for data,
- and those who use data.

3.1 Data creators

Data creators are the research icebreakers involved in the ARICE project. We undertook a questionnaire to determine the current status of their data management. The main questionnaire scope was to determine how exactly the data collection and processing is presently done, and how it would fit to the research cruises funded through the ARICE project.

The questionnaire was redistributed among the contact persons of the icebreakers involved in the ARICE project:

- PRV Polarstern, Germany;
- IB Oden, Sweden;
- RV Kronprins Haakon, Norway
- RRS Sir David Attenborough, United Kingdom;
- CCGS Amundsen, Canada;
- RV Sikuliaq, United States of America;
- MSV Fennica, Finland.

Every respondent was asked the following set of questions:

1. Which parameters do you measure?
2. Are your data openly available? If yes, which data?
3. What is the data embargo period (if there is any)?
4. Which data repositories you use to store the data?
5. How soon will the data be uploaded to the repository after the cruise?
6. Does your data support any unique identifiers (DOI, for example)?
7. What is the license for the data?
8. Can we automatically access the data?
9. Can we automatically access your icebreaker's position?

We have grouped the outcome of the survey to simplify the analysis.

Parameters

The following list of parameters was combined from the respondent's answers. This is what the research vessels routinely measure:

- Atmospheric parameters: air temperature, wind speed and direction, air pressure, relative humidity and precipitation;
- Oceanic parameters: bathymetry, water temperature, water salinity, currents, sea state (wave height), pH, nutrients, carbon, oxygen and chlorophyll;

- Sea ice and snow variables: sea ice thickness, sea ice concentration, snow/ice surface topography, snow thickness and albedo.

The respondents emphasized that the list of parameters is basic and can be significantly broaden if there is such need.

Data availability

All the respondents confirmed that their research data is openly available. But PIs are able to put embargo period on their data if there is such need.

Embargo

The embargo period also depends on the PIs. According to the answers it usually takes no more than two years.

Repositories

The table below shows the repositories being used by the icebreakers.

Icebreaker	Repository
PRV Polarstern	PANGAEA (https://pangaea.de/) DShip-System (https://dms.awi.de/)
IB Oden	Environment Climate Data Sweden (https://ecds.se/)
RV Kronprins Haakon	Norwegian Marine Data Center (https://nmdc.no/) National Infrastructure for Research Data (https://www.sigma2.no)
RRS Sir David Attenborough	British Oceanographic Data Centre (https://www.bodc.ac.uk/); UK Polar Data Centre (http://www.dcc.ac.uk)
CCGS Amundsen	Polar Data Catalogue (https://www.polardata.ca/)
RV Sikuliaq	National Centers for Environmental Information (http://www.nodc.noaa.gov/)
MSV Fennica	Baltic Icebreaking Management (http://baltice.org)

Time lag

The time period when the respondents usually submit data to the corresponding data repositories varies significantly. Some submit their data right after the cruise while the other do it within two years.

Unique identifiers

Neither of the respondents provide data with unique identifiers. Unique identifiers (e.g. DOI) are provided by the data repositories.

License

Creative Commons attribution license for data (v.4.0)

(<https://creativecommons.org/licenses/by/4.0>) is the most commonly used licence.

According to the license data can definitely be re-used by third parties after the end of the projects. There is no restriction on such usage. Nevertheless, the PI's can apply different licenses.

Automatic access to data

Neither of the respondents provide direct access to the data. Such service is provided by the repositories.

Automatic access to vessel position

Most of the icebreakers provide access to their position. Some provide constant tracking information while the others only when vessels are out in the sea.

	PRV Polarstern	IB Oden	RV Kronprins Haakon	RRS Sir David Attenborough	CCGS Amundsen	RV Sikuliaq	MSV Fennica
Position	✓	✓	✓	✓	✓	✓	✓
Speed	✓	✓	?	?	✓	✓	✓
Heading	✓	✓	?	?	✓	✓	✓
Wind speed	✓	✓	?	?	✓	✓	✓
Wind direction	✓	✓	?	?	✓	✓	✓
Atmospheric pressure	✓	✓	?	?	✓	✓	✓
Air temperature	✓	✓	?	?	✓	✓	✓
Humidity	✓	✓	?	?	✓	x	✓
Water temperature	✓	✓	?	?	✓	✓	x
Salinity	✓	x	?	?	✓	x	x
Oxygen	✓	x	?	?	x	x	x
ph	✓	x	?	?	x	x	x
Fluorescence	✓	x	?	?	✓	x	x
Sound velocity	x	x	?	?	✓	x	x
Transmittance	x	x	?	?	✓	x	x
Precipitation	✓	x	?	?	x	x	x
Visibility	✓	x	?	?	x	x	x
Direct radiation	✓	x	?	?	x	x	x
CO ₂	✓	x	?	?	x	x	x

Complete information on RV Kronprins Haakon and RRS Sir David Attenborough was unavailable at the time of writing. The table will be updated as soon as such data becomes available.

3.2 Data management systems

3.2.1 Primary data repositories

The Icebreakers involved in ARICE are ought to send their data to the corresponding national data centres. In this section we will describe these repositories in detail.

PANGAEA

The information system PANGAEA is operated as an Open Access library aimed at archiving, publishing and distributing georeferenced data from earth system research. The system guarantees long-term availability of its content through a commitment of the hosting institutions.

Most of the data are freely available and can be used under the terms of the license mentioned on the data set description. There are a few protected data sets under moratorium from ongoing projects. The description of each data set is always visible and includes the principle investigator (PI) who may be asked for access.

Each dataset can be identified, shared, published and cited by using a Digital Object Identifier (DOI). PANGAEA also allows data to be published as supplements to science articles (example) or as citable data collections in combination with data journals like ESSD, Geoscience Data Journal, Scientific Data, or others.

PANGAEA offers a wide range of web services (SOAP / REST). This includes OAI-PMH for metadata harvesting. The API allows to retrieve any set of numerical and textual data. All PANGAEA datasets also ship with Schema.org/Dataset metadata.

PANGAEA is open to any project, institution, or individual scientist to use or to archive and publish data.

DSHIP-System

DSHIP-V3 system is a complex measurement data management system, developed by Werum Software & System CIS AG. DSHIP provides a variety of configuration, diagnosis, and control features to record data from different instruments, distribute data on board the vessel, visualize data, and to export data from the database. With the Data Acquisition and Management System for technical, nautical and scientific data is acquired, managed, visualized and archived on a number of research platforms.

The acquired raw data is archived in the onshore Archive System and made available via its Web-Interface. For each platform, a link is offered in the navigation area directing to the extraction interface, the data storage for extracted data and the documentation.

Environment Climate Data Sweden (ECDS)

Environment Climate Data Sweden, ECDS, was an infrastructure project to improve Swedish researchers' access to environmental and climate data. ECDS was established by the Swedish Meteorological and Hydrological Institute (SMHI) with the support of the Swedish Research Council (VR) in 2009. The Swedish National Data Service (SND) became responsible for the ECDS data portal in June 2016.

SND is a national resource that facilitates access to new and existing Swedish research data within and outside of Sweden. SND also provides support to researchers in Sweden throughout the process of data management. SND's primary role is to support the

accessibility, preservation, and re-use of data and related materials. Data and metadata are submitted to SND using a web form. Data receive a detailed metadata description and a persistent identifier, and are added to our search catalogue. Re-use of data could entail secondary analysis by researchers in the same or other fields, or students using them for essays or papers. A researcher that deposits data with SND retains the right to decide how and by whom their data are used. SND's mandate is provided by the Swedish Research Council. SND is located in Gothenburg and hosted by the University of Gothenburg.

As a member of DataCite, SND is able to provide Swedish authorities and institutions that work with preserving and making research data accessible, with their own DOI-prefix.

The SND system is compliant with different metadata standards such as Data Documentation Initiative, Dublin Core, MARC-XML, DataCite and ISO 19115. The ISO standard OAIS (ISO 14721:2012) is used as a reference for the daily SND operations and work.

Norwegian Marine Data Center (NMDC)

The Norwegian Marine Data Centre (NMDC) is by nature a distributed data management system with a number of physically separated data centres that are united to a virtual data centre utilising interoperability protocols allowing users to access data wherever located physically. NMDC is primarily affecting interfaces to the contributing data centres, not the internal structure of the data centre.

Being a distributed system, NMDC relies on data storage facilities at the participating institutions. NMDC coordinates the development work, and provide the infrastructure for a common, unified gateway to all distributed data.

NMDC has a common data directory, consisting of a metadata system comprising local metadatabases; services and interfaces for metadata exchange, and a central, searchable data catalog. The metadata system is based on the protocols OAI PMH (prioritised) and OGC CSW. All metadata must comply with the ISO 19115 and DIF standards, with XML as the exchange format. The data centres participating in NMDC operate their own metadatabases in compliance with the NMDC specifications, while NMDC implements the metadata exchange infrastructure and a search interface for end users.

NMDC is still being under active development. The next phase will be the implementation of a data distribution and exchange infrastructure, based on a central node at NMDC and distributed data storage facilities. The central node will include functionality to aggregate and transform data from the distributed data sources, and expose the data products to users through a RESTful API. Initially data formats such as NetCDF / CF, ESRI shapefiles and Darwin Core Archive will be served. The NMDC data providers will expose their data in several ways, including OPeNDAP, HTTP / JSON, OGC WFS / WCS and FTP. Data integrity, version control, access control and IPR management constitute important considerations during the implementation. In addition to the data delivery infrastructure, NMDC will work to connect its data exchange systems to international data sharing infrastructures such as SeaDataNet, Inspire, GEOSS, WIS, GCMD and others. In later phases NMDC will also implement more advanced solutions for extraction, transformation and visualization of data across multiple sources.

National Infrastructure for Research Data (NIRD)

The new Norwegian data infrastructure, named NIRD (National Infrastructure for Research Data), provides storage resources with yearly capacity upgrades, data security through geo-

replication (data stored on two physical locations) and adaptable application services, multiple storage protocol support, migration to third-party cloud providers and much more. Alongside the national high-performance computing resources, NIRD forms the backbone of the national e-infrastructure for research and education in Norway, connecting data and computing resources for efficient provisioning of services.

Any employee and student from Norwegian universities and university colleges, as well as researchers from independent research institutes with projects funded by public grants. The NIRD storage system consists of SFA14K controllers, 10TB NL-SAS drives with a total capacity of 12PiB in addition to a centralized file system (IBM GridScaler) supporting multiple file, block and object protocols.

NIRD provides published datasets with persistent identifiers (PIDs) using the DataCite DOI service.

British Oceanographic Data Centre (BODC)

The British Oceanographic Data Centre is a national facility for looking after and distributing data concerning the marine environment. BODC deals with biological, chemical, physical and geophysical data. Its databases contain measurements of nearly 22,000 different variables. BODC maintains and develops the National Oceanographic Database (NODB). The NODB is a collection of marine data sets originating mainly from UK research establishments.

NODB adhere to the relational model of database design and use the Oracle Relational Database Management System. The NODB stores metadata in a relational database in the form of a data series. Metadata includes the data originator, position and collection start/end dates and times. The actual data measurements are stored outside the database in binary data files. These data are available by online delivery via our NODB data series facility. Each data series has one row in the master NODB table. The table links to other NODB tables containing metadata about parameters, restrictions, documentation and storage.

BODC follows ISO19139/ISO19115, SeaDataCloud, MEDIN, DataCite, ISO19156, Esri ASCII raster and INT16 GeoTiff metadata standards.

UK Polar Data Centre

UK Polar Data Centre is an organisation coordinating the management of data collected by UK-funded scientists in the polar regions. The UK polar data centre relies on the Digital Curation Centre (DCC) is an internationally-recognised centre of expertise in digital curation with a focus on building capability and skills for research data management. The DCC provides expert advice and practical help to research organisations wanting to store, manage, protect and share digital research data. DCC incorporates CF (Climate and Forecast) Metadata Conventions – a standard for climate and forecast “use metadata” that aims both to distinguish quantities (such as physical description, units, or prior processing) and to locate the data in space–time. It also incorporates Directory Interchange Format (DIF) which is an early metadata initiative from the Earth sciences community, intended for the description of scientific data sets. DIF includes elements focusing on instruments that capture data, temporal and spatial characteristics of the data, and projects with which the dataset is associated. FGDC/CSDGM - Federal Geographic Data Committee Content Standard for Digital Geospatial Metadata is also used. BGDC policy states that non-Federally authored standards that are endorsed by the FGDC have the same status as FGDC developed standards. Since ISO 19115, an internationally-adopted schema for describing geographic information and services, and the associated standards are endorsed by the FGDC, DCC also

uses this format. ISO 19115 provides information about the identification, the extent, the quality, the spatial and temporal schema, spatial reference, and distribution of digital geographic data.

Polar Data Catalogue (PDC)

The Polar Data Catalogue (PDC) is the data repository for the ArcticNet Network of Centres of Excellence, the Government of Canada Program for the International Polar Year (IPY), the Northern Contaminants Program (NCP), the Beaufort Regional Environmental Assessment (BREA), the Nunavut General Monitoring Plan (NGMP), the Canadian High Arctic Research Station (CHARS), and other related Canadian and international research programs.

The Polar Data Catalogue provides an interface for registered users to enter metadata and data that are searchable by the general public through the PDC Search Tool. This is an easy to use mapping interface that allows for fast and simple retrieval of spatial data in the Arctic and Antarctic. To upload data, contributors must be researchers with our partner organizations as well as registered users of our website. The uploaded data will be available for public search after our Approvers conduct Quality Assurance measures to inspect and approve files.

The repository describes and provides access to diverse data sets generated by Arctic and Antarctic researchers. The metadata records follow ISO 19115 and Federal Geographic Data Committee (FGDC) standard formats to provide exchange with other data centres. The records cover a wide range of disciplines from natural sciences and policy, to health and social sciences. The PDC Geospatial Search tool is available to the public and researchers alike and allows searching data using a mapping interface and other parameters.

PDC has its own unique records and supports harvesting from other repositories. PCS supports ISO 19115 (North American Profile) metadata standard and has a custom keyword Library.

National Centers for Environmental Information (NCEI)

NCEI manages the world's largest collection of publicly available oceanographic data. NCEI does not conduct any data collection programs of its own; it serves solely as a repository, dissemination, and analysis facility for data collected by others.

NCEI maintains two ocean profile databases. The first is the Global Temperature-Salinity Profile Program (GTSP) and the second is the World Ocean Database (WOD). The GTSP database is comprised of real-time data transmitted over the Global Telecommunication System (GTS) and delayed-mode data from national oceanographic data centers which participate in the International Oceanographic Data and Information Exchange (IODE) system of the Intergovernmental Oceanographic Commission (IOC). Most profiles contain temperature and salinity as a function of depth. The GTSP focuses on managing data collected after 1990. At the request of the oceanographic community, each GTSP profile has an associated history file that contains information about any changes made to the profiles or their associated metadata for errors that were discovered during processing. GTSP is updated on a daily basis and serves the needs of the real-time oceanic and atmospheric forecasting communities.

NCEI also supports the World Ocean Database (WOD), which is the world's largest collection of ocean vertical profile data. The WOD contains profiles dating back to the 19th century and includes data from over 20 variables such as temperature, oxygen, nutrients, plankton, and

tracers. Every three months, the WOD includes data from GTSP by applying additional quality control procedures to the latest GTSP collection. Errors or problems discovered through the WOD quality control processes are reported back to GTSP for inclusion in that database and its methods, if applicable. The WOD also includes historical data recovered as part of the IOC/IODE Global Oceanographic Data Archaeology and Rescue project. Substantial amounts of additional data received at NODC and its collocated World Data Center for Oceanography, are also included in the WOD. The WOD contains more detailed quality control flags than GTSP does. The WOD is widely used for ocean climate diagnostic analyses, for long-term ocean data assimilations in numerical models, and for comparisons with satellite data.

NCEI contains both unique and aggregated records. Presently NCEI supports ISO 19115-2 metadata standard, and moving towards ISO 19115-1. Each of the archives within NCEI has traditionally maintained its own controlled vocabularies but these are cross-walked with GCMD where possible. The GCMD vocabulary is missing some terms from NCEI. Data discovery is possible primarily through NOAA OneStop service (<https://data.noaa.gov/onestop>).

Baltic Icebreaking Management (BIM)

Baltic Icebreaking Management is an organisation with members from all the Baltic Sea states. BIM is a result of the annual meetings of the Baltic states icebreaking authorities which have assembled for more than 25 years.

The overall objective of BIM is to ensure a well-functioning, year-round maritime transport system in the Baltic Sea through enhancing the strategic and operational cooperation between the Baltic Sea countries in the area of winter navigation assistance.

BIM makes its data available via Baltice.org which is a single access point to reliable and up to date information related to winter navigation in the Baltic Sea area. This site gathers information and instructions from icebreaking authorities from all the Baltic Sea countries.

3.2.2 Other national data centres

Primary investigators may be obligated to send their data in their specific national data centres. We have identified some of the data centres which may be used by PI for publishing their data.

Country	Data center	Link
Belgium	Royal Belgian Institute of Natural Sciences, Operational Directorate Natural Environment, Belgian Marine Data Centre	http://www.mumm.ac.be/datacentre/
Canada	Marine Environmental Data Section (MEDS), Oceans Science Branch (DFO - OSB), Fisheries and Oceans Canada	http://www.dfo-mpo.gc.ca/
China	National Marine Data and Information Service	http://www.nmdis.org.cn/nmdisenglish/
Denmark	International Council for the Exploration of the Sea (ICES)	http://ices.dk
France	French Institute for the Exploitation of the Sea, IFREMER Centre de Brest	http://wwz.ifremer.fr
Japan	Japan Oceanographic Data Center (JODC)	http://www.jodc.go.jp
Netherlands	NIOZ Royal Netherlands Institute for Sea Research	http://www.nioz.nl/
Russian Federation	All-Russian Research Institute Hydrometeorological Information - World Data Center	http://www.meteo.ru/nodc/index.html

3.2.3 Data repositories and aggregators

Data repositories and aggregators are governmental and private entities providing services for data storage, processing and retrieval. We identified more than 40 national and global repositories. Some would like to host the data; some would incorporate it into their services. Below is a list of some of them.

#	Data repository	Website
1	Alaska Ocean Observing System	https://www.aos.org
2	Arctic Data Explorer	http://arctic-data-explorer.labs.nsidc.org
3	Arctic Institute of North America	https://arctic.ucalgary.ca
4	Arctic Spatial Data Infrastructure	https://arctic-sdi.org
5	ArcticConnect	http://arcticconnect.org
6	Arctic Research Mapping Application	http://armap.org
7	Atmospheric Radiation Measurement Data Center	https://www.arm.gov
8	Biological and Chemical Oceanography Data Office	https://www.bco-dmo.org

9	Centre Etude Nordique	http://www.cen.ulaval.ca
10	Chinese National Arctic and Antarctic Data Centre	http://www.chinare.org.cn
11	CLIVAR and Carbon Hydrographic Data Office	https://cchdo.ucsd.edu
12	Data.gov	https://www.data.gov
13	DataONE	https://www.dataone.org
14	Dept. Fisheries and Oceans Canada	http://www.pac.dfo-mpo.gc.ca
15	Dryad	https://datadryad.org
16	EarthChem	http://www.earthchem.org
17	Environment Canada	https://weather.gc.ca
18	Environment Climate Data Sweden	https://snd.gu.se
19	Environmental Data Initiative	https://portal.edirepository.org
20	Group on Earth Observations System of Systems	http://www.geoportal.org
21	European Global Ocean Observing System	http://eurogoos.eu
22	International Arctic Research Center - University of Alaska (IARC)	https://uaf-iarc.org
23	Joint Technical Committee of Marine Meteorology Platform Support Centre	http://www.jcommops.org
24	National Centers for Environmental Information	https://www.ncei.noaa.gov
25	National Institute of Polar Research	http://www.nipr.ac.jp
26	National Snow and Ice Data Center	https://nsidc.org
27	Netherlands Polar Data Center	https://npdc.nl
28	Norwegian Polar Institute	https://data.npolar.no
29	NSF Arctic Data Center	https://arcticdata.io
30	NWT Discovery Portal	http://nwtDiscoveryportal.enr.gov.nt.ca
31	OBIS	http://www.iobis.org
32	Ocean Data and Information System	http://odis.incois.gov.in
33	Open Data	https://open.canada.ca
34	The Rolling Deck to Repository	http://www.rvdata.us
35	Research Data Archive	https://rda.ucar.edu
36	Scholars Portal Dataverse	https://dataverse.scholarsportal.info
37	SeaDataNet	https://www.seadatanet.org
38	SESAR	http://www.geosamples.org

39	Svalbard Integrated Arctic Earth Observing System	https://sios-svalbard.org
40	Swiss Polar Institute	https://polar.epfl.ch
41	tDAR	https://www.tdar.org
42	The Norwegian Meteorological Institute	https://www.met.no
43	World Meteorological Organisation Integrated Global Observing System	https://www.wmo.int

3.3 Data users

Data users are those individuals and organisations that use data for a specific purpose.

In general, users of Arctic environmental data need a variety of data pan Arctic distributed, ranging from hydrometeorological and other environmental data, to visualization of the data and data models.

The most frequent requirements users apply to data and metadata are:

- Accuracy –must be accurate enough for analysis,
- Standardization –must follow international standards,
- Findability –must be easily discoverable,
- Accessibility –must be easily accessible,
- Interoperability –must be present in interoperable formats.

We identified the following groups of users interested in Arctic data:

Scientists and research centres

Scientists are the main users of Arctic environmental data for their research Data is then used in several ways, i.e. to feed data models and to publish scientific papers that evaluate the actual state and evolution of the Arctic Ocean.

Educators

Educators from schools and universities use data, data models and data visualization as a valuable source of environmental data for teaching of oceanography, meteorology and environmental change in the polar regions.

Federal, State and Local Government Agencies, policy makers

A number of federal and state regulations require agencies to curate and manage hydrometeorological information. National and regional governing bodies are also both concerned and interested about the problems and opportunities facing the Arctic because of the climate change.

Private companies

Insurance companies have the need to increase their knowledge on various Arctic issues in order to manage the risk of increased activity by their customers in the area. The shipping industry is expected to increase their activity substantially in the Arctic. This is both due to increased shipping within the Arctic itself because of increased activity, as well as increased trans-Arctic shipping - with the opening of the Northern Sea Route, through the Barents Sea, as well as the “over the North Pole” route. The oil industry is increasing the activities in the

Arctic waters. The Arctic conditions are harsh and the increased activity has to be thoroughly prepared, which demands more precise and extensive information. Increased mining activity in the Arctic will demand extensive information, as well.

We identified the following private entities interested in Arctic data:

- Insurance companies;
- Banks, investments firms and business analysts;
- Consulting companies;
- Shipping companies;
- Mining, oil and gas.

4. Summary and Potential gaps

In summary, data creators perform observations and send raw data to the national data centers we identified in the chapter 3.2.1 of this report. Primary investigators may also send their data and metadata in the corresponding national data centers. Using well-established national repositories provides data security and long term preservation. Also, national data centers provide interfaces for data discovery and access. The national data centers perform initial quality control, make sure that metadata are into compliance with the ISO 19115 metadata standard, transform data into interoperable formats, apply DOI and publish the complete datasets online. Such data flow has been proven by years and is considered as a robust way of transferring data from research vessels to the data storage and final users.

The users, in turn, look for easily accessible data. The users do not want to know all the transformations which have been done to the data. They need an open-access accurate dataset in a popular format with a full set of documentation they can find online, download and use.

National data centers do this job well and improvements are being constantly implemented. Nevertheless, there are some issue to be addressed. Looking at the outcomes of the workshop and questionnaires it is possible to recognize a set of areas that may deserve a follow up and/or a discussion:

- Data findability and accessibility needs to be improved;
- NODCs are often harvested by other repositories. This could be a source of data duplication. Users may get confused which source is the original one;
- Users may be overwhelmed with the amount of information available on the NODCs;
- NODCs provide data in a variety of different formats. This problem is being addressed by some of the NODCs and common interoperable formats are being implemented;
- It is presently unclear how much freedom PIs will have in choosing data licenses and embargo periods.

5. Final Recommendations

The ARICE data management system and its data visualization are intended to improve the current status of data management systems. It will provide a user-friendly platform to visualize and explore data collected in the Arctic Ocean. The data management system will not store actual data in order not to create duplicates but instead will provide an easy and

understandable way to find the data and access it. The 3D Virtual Icebreaker will be utilized to simplify data dissemination.

Users will be able to access the 3D interface of the Virtual icebreaker and to navigate through the ship discovering equipment on board, the data, and details of the ship's mission. Access to the 3D interface will be available via the project website. This approach to the data dissemination is an innovative way of presenting scientific results. The majority of the data centers are oriented to the scientific community as the main data consumer and not friendly enough for regular users. This approach is currently not implemented anywhere else and will make ARICE project results visible for a much broader audience.