HORIZON 2020

Research and Innovation action

Grant Agreement No. 730965

ARICE: Arctic Research Icebreaker Consortium:

A strategy for meeting the needs for marine-based research in the Arctic

Deliverable 7.1. ARICE Data Management Plan

Submission of Deliverable

Work Package	WP7
Deliverable no. & title	7.1 ARICE Data Management Plan
Version	1
Creation Date	June 2018
Last change	July 2018
Status	☒ Draft ☒ WP lead accepted ☐ Executive Board accepted
Dissemination level	☒ PU-Public ☐ PP- Restricted to programme partners ☐ RE- Restricted to a group specified by the consortium ☐ CO- Confidential, only for members of the consortium
Lead Beneficiary	AP
Contributors	☒ 1 – AWI, ☐ 2 – SPRS, ☐ 3 - NPI, ☐ 4 - ULAVAL, ☐ 5 – UAF/CFOS, ☒ 6 – AP, ☒ 7 – CSIC-UTM, ☐ 8 – CNR, ☒ 9 - WOC, ☐ 10 – IOPAN, ☒ 11 – FMI, ☒ 12 - CNRS, ☐ 13 – NERC-BAS, ☐ 14 – DTU-AQUA
Due date	31.06.2018
Delivery date	07.09.2018

Table of Contents

Executive summary

This Data Management Plan (DMP) is based on the H2020 FAIR Data Management Plan template designed to be applicable to any H2020 project that produces, collects or processes research data. The purpose of the DMP is to describe the data that will be produced, collected or processed during the project, as well as the plans for sharing and preservation of the data. This plan is a living document that will be updated during the project life.

ARICE follows a metadata-driven approach where datasets are documented in a standardized manner for efficient sharing via virtual access to the data and open access preservation. This approach is aligned with relevant activities in the context of OGC, GEOSS, WMO, JCOMM, GCOS, GOOS, SAON, ICSU-WDS, CODATA, RDA, and INSPIRE.

ARICE promotes free and open access to data. Within this plan an overview of the data collection procedures is provided as well as an initial outline of dissemination.

Introduction

The purpose of the DMP is to document how the data generated by the project is handled during and after the project. It describes the basic principles for data management within the project. This includes standards and generation of discovery and use metadata, data sharing and preservation and life cycle management. This DMP is a living document that will be updated during the project life in time with the periodic reports. ARICE is following the principles outlined by the Open Research Data Pilot and The FAIR Guiding Principles for scientific data management and stewardship (Wilkinson et al. 2016).

This DMP is based on the H2020 FAIR Data Management Plan template designed to be applicable to any H2020 project that produces, collects or processes research data.

1. Data summary

Purpose of the data collection/generation and its relation to the objectives of the project

Data collection will be generated in the frame of the research cruises funded through the ARICE project on board the six ARICE Research Icebreakers:

Types and formats of data the project will generate/collect

In this frame, the cruises funded through ARICE will generate a variety of data which could include:

1. Atmospheric parameters:
   - Air temperature;
   - Wind speed and direction;
   - Air pressure;
   - Water vapor or humidity;
   - Precipitation;
   - Cloud fraction or cloud base height;
   - CO_2, methane or other greenhouse gases;
   - Ozone or aerosols;
   - Radiation budget.

2. Oceanic parameters:
   - Bathymetry;
   - Water temperature;
   - Salinity;
   - Currents;
   - Sea state (wave height);
   - Ocean surface heat flux;
   - Ocean color;
   - pH;
   - Nutrients;
   - Carbon (inorganic or organic), oxygen or nitrous oxide (N_2O);
   - Primary production / Chlorophyll;
   - Fish abundance, marine mammals or birds.

3. Sea ice and snow variables:
   - Sea ice thickness;
   - Sea ice concentration;
   - Sea ice extent / sea ice edge;
   - Sea ice drift;

© ARICE Consortium   31/07/2018

- Ice load (on the hull);

- Floe size;

- Ice ridge frequency;

- Snow/ice surface topography;

- Snow thickness;

- Albedo;

- Melt ponds (fraction or appearance).

The list of parameters depends on the research cruises that will be granted access to the ARICE Research Icebreakers and also on the research equipment each of the Research Icebreakers makes available.

Although a high variety of data formats could be generated within the ARICE project, the most popular formats are expected to be Microsoft Excel (XLS); Comma-separated values (CSV), Text (TXT) and netCDF (NC). Data format standardization is described in the section 2.3.

Re-use of existing data

ARICE is primarily focused on gathering new data. For the time being there is no intend to re-use any existing data.

Origin of the data

Data will be collected on board the ARICE Research Icebreakers during ARICE funded cruises. A major goal of the ARICE data management is to make data generated within the project visible and useful for regional and global monitoring programs, Arctic researchers, Arctic communities and individuals. Therefore, a data portal will be established, which will provide a unified view on the data produced by the ARICE project. This approach is essential in order to increase the visibility of the ARICE project and benefit from the generated data.

Expected size of the data

The precise of the data cannot be determined at this stage. This information will be provided in the following versions of the DMP.

Data utility

Researchers demand continuous and all year round observations in Arctic waters, with an emphasis on winter observations.

2 FAIR data

ARICE aims at its research data to be findable, accessible, interoperable and reusable (FAIR).

2.1 Making data findable, including provisions for metadata

Identification and location of data by means of a standard identification mechanism

All metadata will be available through a web-based search interface available through the central project website (www.ARICE.eu). Providing datasets with DOI depends on the PI's decision.

Naming conventions used by ARICE

ARICE is following the ISO 19115 naming convention.

Search keywords optimization

The keyword search will be provided to speed up the search.

Provision of version numbers

This question will be answered in the future editions of the DMP.

Types of metadata created through ARICE

ARICE will provide all necessary metadata according ISO 19115 standard including information about the identification, the extent, the quality, the spatial and temporal schema, spatial reference, and distribution of the data. The project will also provide Cruise Summary Reports on icebreakers' cruises, formulated using tags from the ISO19115 metadata standard, according to the National Oceanographic Data Centres' required format for reporting.

2.2 Making data openly accessible

Data produced and/or used in the project that will be made openly available as the default and data sharing restrictions

Data or metadata generated on research cruises funded through ARICE, not affected with moratorium period, will be made openly available and accessible. Access to the datasets, apart from metadata, depends on the PI. Some data will be affected by a moratorium period, with access restricted to the scientific party and its designated partners for a period of 1 to 2 years after the cruise takes place. Requests of external users for data access during this time will be forwarded to the data originators for their decision.

Data accessibility

The Principal Investigator of an ARICE cruise must submit the data generated together with sufficient metadata (ISO 19115) to the respective IODE National Oceanographic Data Centre (NODCs) directly after the cruise. The NODC will then make sure that the data sets are quality controlled, archived in the NODCs and linked to the metadata of the respective cruise. All datasets will also be integrated into the ARICE Data Management System, the ARICE project's database and disseminated in an interoperable open format through the 3D Icebreaker outreach tool.

Methods and software tools needed to access the data

31/07/2018

Data will be accessible via a Web interface with a possibility to download. For the most data formats users do not need any specific software. Data in netCDF format can be read with open source software provided by Unidata (https://www.unidata.ucar.edu/software/netcdf/).

Documentation about the software needed to access the data

Such documentation will be developed.

Possibility to include the relevant software

The project will include links to freely available software into the documentation.

Place where the data and associated metadata, documentation and code be deposited

All the data will be deposited at the respective IODE National Oceanographic Data Centre (NODC).

Explored appropriate arrangements with the identified repository

Such arrangements will be made in the nearest future.

Access to the data in case of restrictions on use

Some data may have temporal access restrictions (embargo period). After the embargo period, data should be made accessible.

Need for a data access committee

At present, we do not envisage a need for a "data access committee". Requests for cruise data subject to moratorium will be forwarded to the cruise PIs. If the need of a "data access committee" arises during the project, the Steering Board of ARICE will take this role.

Conditions for access

Conditions of access will be developed during the project. PIs can use different licenses for accessing the data. If there is no such requirement, RV's conditions for access will be used.

Identification of the person accessing the data

Such person will have to accept terms of agreement before getting an access to the data. Doing that requires to create an account.

2.3 Making data interoperable

Interoperability of the data produced in the project

ARICE will distribute datasets in model ready format and web-services interoperable to Earth Sciences platforms. It will contribute to advanced data and computing services by producing highly structured datasets compatible to the current climate and earth system modelling programmes. In order to be able to reuse data, standardization is important. This implies both standardization of the encoding/documentation, as well as the interfaces to the data.

Data and metadata vocabularies, standards and methodologies

The project's metadata will follow the ISO 19115 standard. The data will be encoded as netCDF files. NetCDF files following the CF convention is self-describing and interoperable. Application of the CF conventions implies requirements on the structure and semantic annotation of data (e.g. through identification of variables/parameters through CF standard names).

Standard vocabularies for all data types present in data sets

The same standard vocabularies will be used for all the data types.

Mappings to more commonly used ontologies in case it is unavoidable that you use uncommon or generate project specific ontologies or vocabularies

If such case occurs at any stage of the project, ARICE shall provide mappings to ISO 19115.

2.4 Increase data re-use (through clarifying licenses)

Data licensing

ARICE promotes free and open data sharing. Each dataset needs a license attached. The recommendation in ARICE is to use Creative Commons attribution license for data (https://creativecommons.org/licenses/by/3.0/ for details). Nevertheless, the PIs are able to apply different licensing.

Time frames for data availability for re-use

The embargo period for the data depends on the PIs. According to the project survey such delay may take up to 1-2 years. ARICE will do its best to make the data available as soon as possible.

Data usability for third parties

The data can definitely be re-used by third parties after the end of the projects. There is no restriction on such usage.

Time frames for keeping data re-usable

There is no certain date as of yet. It depends on the each specific NODCs.

Data quality assurance processes description

Data quality procedures depend on the corresponding NODC and must be described in its documentation.

31/07/2018

3 Allocation of resources

Costs for making data FAIR

In the current situation it is not possible to estimate the cost for making ARICE data FAIR. Part of the reason is that this work is relying on existing functionality at the contributing data centres and that this functionality has been developed over years.

The way to cover the costs

The cost of preparing the data in accordance with the specifications and initial sharing is covered by the ARICE project.

Responsibility for data management

The Principal Investigator of an ARICE cruise must submit the data generated together with sufficient metadata to the respective IODE National Oceanographic Data Centre (NODCs) right after the cruise. Arctic Portal is responsible for integration all the datasets into the ARICE Data Management System, the ARICE project's database and disseminating them in an interoperable open format through the outreach tools including the 3D Icebreaker.

Long term preservation resources

Maintenance of this over time is covered by the business models of the data centres. In the current situation there is no overview of the costs of long-term preservation of data as this is the responsibility of the contributing data centers and the business model for these differs. This information will be updated in further versions of the DMP.

4 Data security

Data security provisions

The generated data together with sufficient metadata are stored in the respective IODE National Oceanographic Data Centre (NODCs). The NODCs make sure that the data sets are stored properly and secure. A copy of the data is stored in the ARICE data management system which uses secure HTTP for communicating between the data management system and users. Concerning the internal security, ARICE uses the best practices from OAIS. The technical solution uses automated check sums and replication.

Long term preservation and curation of the data

IODE National Oceanographic Data Centre (NODCs) are governmental organisations with ensured long term support.

5 Ethical aspects

Ethical or legal issues that can have an impact on data sharing

ARICE project will not raise any highly sensitive data. Nevertheless, the research groups involved will follow, where applicable and available, international and national ethical research principles, national legal regulations on personal data as well as any applicable local legislation. Data will be collected, stored, protected and disposed of according to the applicable national and local regulations. In addition, all research, independent from the field site location, will comply with the EU General Data Protection Regulation (GDPR) approved by the EU Parliament on 14 April 2016 and enforced on 25 May 2018 on the protection of individuals with regard to the processing of personal data and on the free movement of such data. All publication and dissemination of the projects data will be done in a manner respecting the research participants right to privacy and no link to actual persons will be included in such materials.

An Ethical Advisor will be engaged in the project to oversee the ethical aspects ARICE might have.

Informed consent for data sharing and long term preservation

Informed consent for data sharing will be included in questionnaires dealing with personal data. Personal data will not be stored beyond the project life.

6 Other

Other national/funder/sectorial/departmental procedures for data management

Through its partners, ARICE will be cooperating with ongoing and new EU-Projects for Arctic research, such as ICE-ARC, INTAROS, APPLICATE and Blue Action as well as for infrastructures optimization, including the polar areas: INTERACT, SIOS-PP, EMSO, ICOS, and FixO3. ARICE will cooperate with relevant data initiatives like e.g. the SAON Committee on Information and Data Services (CDIS), the "International Polar Data Forum", an initiative of both SCAR and IASC, and the YOPP Data Portal.