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 7.3. Vessels' capabilities and limitations to adopt the ARICE system

Submission of Deliverable

Work Package	WP7
Deliverable no. & title	D7.3 Vessels' capabilities and limitations to adopt the ARICE system
Version	1
Creation Date	03.07.2019
Last change	26.07.2019
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	CSIC-UTM
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 ☐15 – ARCTIA
Due date	31.12.2018
Delivery date	26.07.2019

1. Abstract

Data from the ice-covered areas of the Arctic Ocean is very scarce. To improve this situation, the ARICE project will contribute to obtaining a better picture of the ARCTIC by funding scientific cruises in the area, on board of six European and international research icebreakers.

As pointed out in D7.1 “Data management plan”, 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.

As has been summarized in the D7.2 “Report on user and stakeholder feedback on the current status of their data management and potential gaps.”, the research icebreakers involved in the ARICE project act as data creators, performing observations and sending raw data to the national data centres […]. Primary investigators involved in ARICE surveys may also send their data and metadata to the corresponding national data centres. […] In addition, national data centres will provide interfaces for data discovery and access, performing initial quality control, making 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 deliverable “D7.3 – Vessels’ capabilities and limitations to adopt the ARICE system” is the third deliverable to fulfil in the “Enhancing virtual and remote access to data” Work Package (WP7).

The objective of this deliverable is to provide an analysis of the on-board infrastructure needed to implement the recommendations pointed in the D7.2 and then highlight the necessary capacities to undertake the aforementioned recommendations as well as an estimate of the limitations that may compromise them.

2. Necessary key elements to implement Data Management Plan in ARICE

In order to implement appropriately the Data Management, Plan the following elements should be stablished at the ARICE infrastructure

1.1. Policies and Standards

The necessary data policies have been outlined at the former deliverables of the work package, however it is necessary to highlight that the success on the application of the Data Management plan is highly dependent on the “tempos” of data and metadata creation.

In several previous EU projects, such as Eurofleets(1) and Eurofleets2, SeaDataNet 1 and 2 and ODIP 1 and 2, where data management of research vessels has been a key issue, it has been shown that data and metadata should be created as close as possible to the moment of data acquisition. This means that the main production of data and metadata products must be on board and, therefore, a great effort in the implementation of the data management plan should be focused on the ship's facilities.

For Metadata is recommended the use of **ISO 19115 content model** as it is the common accepted metadata model to follow the European INSPIRE directives. According to Article 5(1) of Directive

2007/2/EC, Member States shall ensure that metadata are created for the spatial data sets and services corresponding to the themes listed in Annexes I, II and III, and that those metadata are kept up to date. According to Article 5(4) of Directive 2007/2/EC, Implementing Rules shall be adopted taking account of relevant, existing international standards and user requirements. In the context of metadata for spatial data and spatial data services, the standards EN **ISO 19115, EN ISO 19119, and ISO 15836** (Dublin Core) have been identified as important standards.

Following the SeaDataNet approach, as it is an EU reference project in marine data management, the ARICE XML Schema's should be based upon the ISO 19115 DTD. As in **SeaDataNet**, two main Metadata schemas should be considered. For the cruise description the **Cruise Summary Report (CSR**) should be considered, and for Data sets description, the **Common Data Index (CDI).** Both CSR and CDI are an upgrade to **ISO 19139** format has been applied, especially for making these INSPIRE compliant and for making the exchange between the national providers and the central portal operators more efficient by automatic harvesting.

Moreover, following the SeaDataNet approach, ARICE metadata formats should use of the SeaDataNet Common Vocabularies and the EDMO directory (European Directory of Marine Organizations) which underpins the relations and semantic coherency between the different metadata services.

The SeaDataNet vocabulary services are technically managed and hosted by the British Oceanographic Data Centre (BODC) by means of the NERC Vocabulary Server (NVS2.0). The vocabularies are made available as web services for machines and by means of client interfaces for end-users. The client interfaces provide end-users options for searching, browsing and CSV-format export of selected entries. The machine interfaces are provided via a SOAP Application Programming Interface (API) for exchanging structured information across computer networks as the result of calls. It relies upon XML (eXstensible Markup Language) documents for passing messages.

Following the same approach as SeaDataNet, data sets should be accessible via download services. Delivery of data to users requires common data transport formats, which interact with other standards (Vocabularies, Quality Flag Scale). Therefore, the following data transport formats have been recommended:

- SeaDataNet ODV4 ASCII for profiles, time series and trajectories;
- SeaDataNet NetCDF with CF compliance for profiles, time series and trajectories;
- NetCDF with CF compliance for 3D observation data such as ADCP.

1.2. Software and Tools suggested

Even though CSR and CDI metadata file formats have been designed to be completed and finalized at the shore facilities of the correspondent National Data Centres, the initialization of these files on board before the end of the campaigns is highly recommended, or what is better (in the case of the CDI) leave the different sets of data as definitive, raw data, during the campaign. The purpose is not to loose detail in the data description and, at the same time, to fully describe the acquisition environment. Important in this stage is the creation of logbooks where all the incidents and events occurred during the acquisition of the data are recorded, enriching their context and being a first analysis of their quality.

As ISO metadata files are not user friendly (large XML files with many nested elements) and following the idea that they should be easily created on board, the use of specific software tools developed under SeaDataNet umbrella for metadata creation is highly recommended. In this sense, the use of

Mikado software tool is recommended on board as a part of the well stablishes data management on board routines.

Mikado is a software tool, developed by IFREMER, written in Java, which enables data centres to prepare XML metadata files for the SeaDataNet directories EDMED, CSR, EDMERP, CDI and EDIOS. Data centres can prepare XML entries manually or can generate XML entries automatically by interfacing with local databases. MIKADO also enables data centres to manage XML entries as a local collection.

Mikado uses the latest SeaDataNet XML Schema's and makes use of the Common Vocabularies, EDMO and EDMERP directories. The latter are each time synchronised by online connections to the respective Web Services on user demand.

The use of **Mikado** implies and facilitates the use of the BODC Common Vocabularies.

To support data centres in using common Data Transport files, SeaDataNet is providing **Nemo** software as a useful tool (written in java). **Nemo** enables the conversion from any type of ASCII format to the SeaDataNet ODV ASCII formats as well as the SeaDataNet NetCDF (CF) format (for time series, profiles and trajectory observations). **Nemo** makes use of the Common Vocabularies and EDMO directory. The latter are each time synchronised by online connections to the respective Web Services on user demand.

1.3. Computing and Communications Facilities needed

The generation of data and metadata on board ships does not really require a very specific infrastructure of computing and communications resources. The software developed under the umbrella of the SeaDataNet project can be run on any user computing platform, does not require access to centralized computer or storage services on ships, nor does it require permanent access to the Internet.

In terms of communications resources this is an accurate response as data and metadata can be sent to the data centres once the vessel has docked in port.

In case of choosing to send data and metadata during the surveys, it is sufficient to establish a daily periodicity and file compression, which allows the transmission via satellite.

3. Necessary capabilities

As it is mentioned before all the necessary elements needed to implement the Data Management Plan already exist, and there isn't anything that should be implemented particularly or adapted to the ARICE project's needs. Computing and Communications facilities currently present on-board ARICE ships should be enough to run the suggested tools to produce Data and Metadata products and send them to data centres in real time or once every survey will be closed at the destination port.

Technicians/crew members, that will take the responsibility to produce metadata and the associated data sets on board should be trained in the use of the suggested software tools, that are quite user friendly and easy to install and use. It is recommended as well to have a short training on metadata and data transport formats standards and general use of vocabularies, for a better understanding of the process.

A formal instruction about the responsibility, specifications to use and process of metadata and data creation on board should be prepared and distributed among the different scientific parties and crew/technicians members.

4. Foreseen limitations

Broadband satellite transmission systems are usually based on geostationary constellations (VSAT, Inmarsat) whose geographical coverage is limited between 70ºN and 70ºS. This is probably the only one technical limitation expected to propagate to shore daily data and metadata reports, is when the ARICE vessels sail outside of this limit. . However, any other satellite facilities (LEO constellations like Iridium) with less capacity but with polar coverage could be used instead.

