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 7.4. Scheme of the data management system and its specifications

Submission of Deliverable

Work Package	WP7
Deliverable no. & title	7.4 Scheme of the data management system and its specifications
Version	1
Creation Date	March 2019
Last change	
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	31.03.2019
Delivery date	

Table of Contents

1 Introduction

1.1 Abstract

An ARICE data management system (DMS) aims to provide seamless access to the project's data to researchers, government agencies, and the general public. The system is based on FAIR principles and takes into account the state-of -the-art approaches to oceanographic data management. The system implements interoperability, standards compliance, user accessibility, metadata discovery, and open access, aiming to be a referential system among the polar and maritime community.

1.2 Objectives of the report

The main goal of the Deliverable 7.4 (D7.4) is to design the ARICE data management system and formulate its specifications. The deliverable is based on the Deliverables 7.1, 7.2 and 7.3, and covers the following topics:

- data flow,
- data management system design,
- data management scheme,
- data management system specifications.

2 Data management system design

2.1 Introduction

As it is described in the Deliverable 7.2, the ARICE vessel operators and researchers are obliged to send their data into corresponding national data centers. Hence there is no need in creating a separate project data repository as it was originally planned as it could raise the problem of data duplication and long-term support. To avoid this problem, the data management system will function as a federated search platform rather than a data repository. It will provide fast and easy access to the project data as well as other relevant data through the projects 3D-Icebreaker interface.

2.2 Data products

According to the Deliverable 6.1, automatic environmental observations on board vessels navigating in the Arctic are currently very few and limited to basic atmospheric and oceanic variables. In addition to those, some visual observations of sea ice are recommended in D6.2.

Based on D6.2, we assume that the data management system will include automatic, and thus possibly real-time observations of:

- Air temperature
- Air pressure
- Wind speed and direction
- Relative humidity or dew point
- Ocean surface layer temperature
- Ocean surface layer salinity

As the needed technology for automatic sea ice observations does not exist yet, the D6.2 recommends visual observations of:

- Location of ice edge
- Sea ice concentration
- Sea ice types (new ice, first year ice, multiyear ice, deformed ice)
- Sea ice thickness
- Floe size
- Icebergs

These above listed variables can be measured on board the majority of the vessels and thus they form the basis of the data in the DMS. However, the data to be included is not restricted to these variables. For instance, some research vessels are instrumented for automatic measurements of greenhouse gases, precipitation, cloud base height and radiation budget and those will also be included in the DMS as technically possible.

2.3 Data flow and data management system scheme

The data flow consists of two parts:

- near real-time data,
- delayed mode data.

Near real-time data are streamed by the icebreakers to their corresponding online platforms. Such data include coordinates, speed, heading, wind speed and direction, atmospheric data, humidity, air temperature, and sea surface temperature. Some additional parameters can be used (see D7.2). The

data management system will automatically harvest such online platforms and show the data on the project's data portal.

Other data will be transferred by the vessel operators and by primary investigators (PIs) to their respective national oceanographic data centers (NODCs). This process can take from several weeks up to two years (according to D7.1). NODCs are responsible for data quality control/quality assurance, interoperability and accessibility. The data management system will harvest metadata from all the NODCs where ARICE data is stored. Such metadata will be used to provide users with a fast and reliable search over all the data. The data management system scheme is shown on fig. 1.

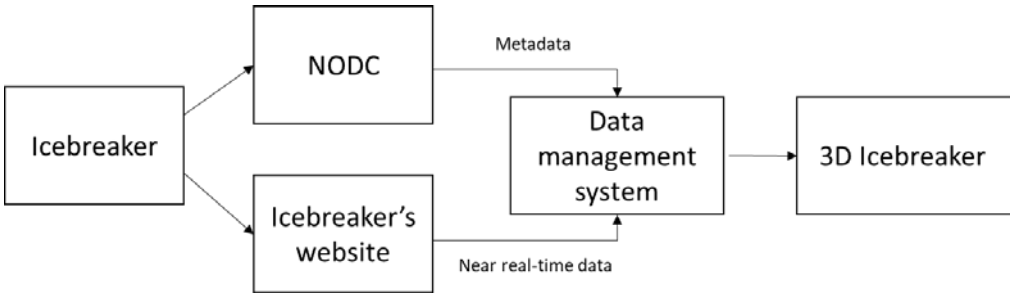

Figure 1. Data management system scheme

3. Data management system specifications

3.1 Server architecture

The DMS will run on two physical servers. Having two identical servers is essential in security and seamless operation terms. The servers will by synchronized to maintain constant uptime and for data security . In case of the main server failure the second one will immediately take over. Each server will have two virtual servers (see chapter 3.2).

Each server will have as minimum two Xeon Silver 4110 processors to secure adequate task performance. Every server will have at least of 192 GB of RAM (6 pieces of TS 32GB TruDDR4). This amount of rapid memory is essential for smooth data processing. Also, having to run several virtual servers requires a considerable amount of RAM. Every server will have an array controller, LAN, Emulex card and a power supply. Additionally, each server will have two fast SSDs with system files.

For security reasons the DMS will have a firewall and a separate router. Having a firewall is crucial for protecting the IT infrastructure from potential external attacks.

3.2 Data management system software

The DMS software will consist of two virtual servers run under VMware software. The first server will run geoserver software and the second one will be a webserver. The main DMS software will run on the webserver. Its main functions will be:

- to harvest and visualize data streamed from the icebreakers,
- to harvest and index NODCs metadata,
- to provide users with data search and conversion tools,
- to provide data through the 3D-Icebrakers interfacing tools.

Both servers will run on the Ubuntu Server 18.04 operation system. The geoserver software will be used for storing some additional relevant data layers. Its functionality will be described in detail in the section 3.6.

The DMS software will be:

- Responsive (work on desktops, tablets and mobile devices regardless of operating system),
- Standards compliant,
- Developed using the latest interoperable open source technologies (Django, OpenLayers),
- Server and client-side optimized,
- Regulary updated through out the project.

The DMS software will consist of a webserver, a database, server-side programming tools, client-side programming tools, and geospatial rendering frameworks. All the software components should be open sourced. The following software products have been chosen:

- Database: PostgreSQL (+ PostGIS extension), GeoServer, MySQL,
- Server-side programming: Python,
- Web server: WSGI,
- Client-side programming: JavaScript + AngularJS + Bootstrap,
- Geospatial rendering: OpenLayers + OpenStreetMap.

3.3 Data storage

The DMS will only store limited data locally. The main reason for that is not to duplicate datasets. The data should be stored by NODCs in their own formats. Most of the NODCs ARICE works with are FAIR principles compliant and provide their data in interoperable formats.

3.4 Metadata formats

Most of NODCs support ISO 19115 as the main metadata format. The DMS will be primarily harvesting this format but other formats (ISO flavors, schema.org, DIF etc.) will be supported if required by the ARICE project.

3.5 Conversion algorythms

This feature will allow users to convert data into common interoperable formats. The DMS will automatically download the selected dataset, convert it, and provide the user with a download link. After conversion and user download the data will be removed from the server. This feature may not be frequently required as NODCs usually provide users with data in interoperable formats with no conversion needed.

3.6 Mapping system

The DMS will include an interactive mapping system. It will show current position of the icebreakers and data streamed by them (e.g. speed, air and water temperature, wind velocity, wave index etc.). This part of the DMS will be located on the webserver and consist of JavaScript modules harvesting streamed data and publishing it on a 2D and3D mapping interface.

3.7 Additional data

The mapping system will provide users with the viewing of and access to a broad collection of Arctic-related data (e.g. infrastructure, atmospheric data, sea ice extent, climate models etc.) for a comprehensive comparative data analysis.

The data will be stored on the geoserver. Currently, this data library consists of more than 600 data layers. It will be significantly extended during the ARICE project for the benefit of the project partners and external users.

