




NERSC Technical Report no. 429

Smart Ocean: Metadata and data formats for oceanographic and acoustic data products from field experiments in Bjørnafjorden and Langenuen

by

**Torill Hamre
Frode Monsen
Espen Storheim
Hanne Sagen
Tor I. Olaussen
Stein Sandven**

10 August 2023

	Nansen Environmental and Remote Sensing Center (NERSC) Jahnebakken 3 N-5007 Bergen, Norway Phone: + 47 55 20 58 00 E-mail: post@nersc.no http://www.nersc.no
---	---

TITLE: Smart Ocean: Metadata and data formats for oceanographic and acoustic data products from field experiments in Bjørnafjorden and Langenuen	REPORT IDENTIFICATION: NERSC Technical Report 429 https://doi.org/doi:10.5281/zenodo.10843661
CLIENTS: Research Council of Norway European Commission	CONTRACT: SFI Smart Ocean - Flexible and cost-effective monitoring for management of a healthy and productive ocean (project no. 309612) INTAROS - Integrated Arctic Observation System (Grant Agreement no. 727890) CAPARDUS - Capacity-building in Arctic standardization development (Grant Agreement no. 869673)
CLIENT REFERENCE:	AVAILABILITY: Public
INVESTIGATORS: Torill Hamre Frode Monsen Espen Storheim Hanne Sagen Tor I. Olaussen Stein Sandven	AUTHORISATION:
	DATE: 10 August 2023
SUMMARY: This report describes the delivery chain for ocean mooring and vessel mounted observing platforms used to collect, generate, and publish oceanographic and acoustic data products in the Smart Ocean project. Focus is on the steps in the delivery chain where metadata is compiled, data is formatted using standard formats and published in the Norwegian Marine Data Centre (NMDC). A PDF template has been created to enable easy registration of metadata, before applying open-source tools such as Rosetta and Thredds Data Server to support data formatting and publishing. Each data product is assigned a unique identifier, a Digital Object Identifier (DOI), which enables data consumers to cite the use of data products in presentations, papers, and services. With continued and increasing amounts of data from fields campaigns that are planned in the Smart Ocean and other research projects, it is important to operationalise the delivery chain. This entails using and as needed customising the metadata template described in this report, further enhancing Rosetta and associated tools for metadata generation and validation, as well as automating the publication process in Thredds for NMDC. Assignment of DOIs from NMDC will ensure that all data products can be cited. Furthermore, using the metadata template will make sure all staff involved in the delivery chain will be acknowledged along with research projects and funding bodies that have supported data collection, processing, quality control, formatting and publication.	

TABLE OF CONTENT:

1. Introduction	3
2. Metadata and data standards	3
2.1. Standards applied for Smart Ocean	4
2.2. Metadata for oceanographic and acoustics data products	5
3. Metadata template for Smart Ocean data products	6
4. Data file conversion using Rosetta	7
5. Publishing data products in NMDC.....	9
6. Future use and development of the data delivery chain.....	10
7. References	11
Appendix A Example of global metadata for a mooring dataset	12
Appendix B Example of metadata for a variable in a mooring dataset.....	13

1. Introduction

SFI Smart Ocean is a centre for research-based innovation, hosted by the Department of Physics and Technology at the University of Bergen (UIB, 2023). Smart Ocean aims to create a wireless observation system for multi-parameter monitoring of underwater environments and installations. The system, based on autonomous smart sensors, will serve as an enabling fundament to realize flexible, distributed, robust, energy efficient, cost-effective, and safe marine measurements and big-data handling, to solve societal and industrial challenges.

This report describes the data delivery chain for oceanographic and acoustics data products used at NERSC (Fig.1). The metadata and data standards used are identified, with brief description of major elements and pointers to further information. Tools used to support the publication process is briefly presented to show how the data is transformed into a standard data format with ample metadata and then made publicly available in an open data infrastructure, the Norwegian Marine Data Centre (NMDC) (NMDC, 2023). By publishing data products in NMDC, the data products will be available through standard interfaces in data portals and applications. Products can further be disseminated in European data infrastructures provided that these include the metadata recognizing all actors in the data delivery chain as well as the data centre hosting the data product. Dissemination is the responsibility of the data centre.

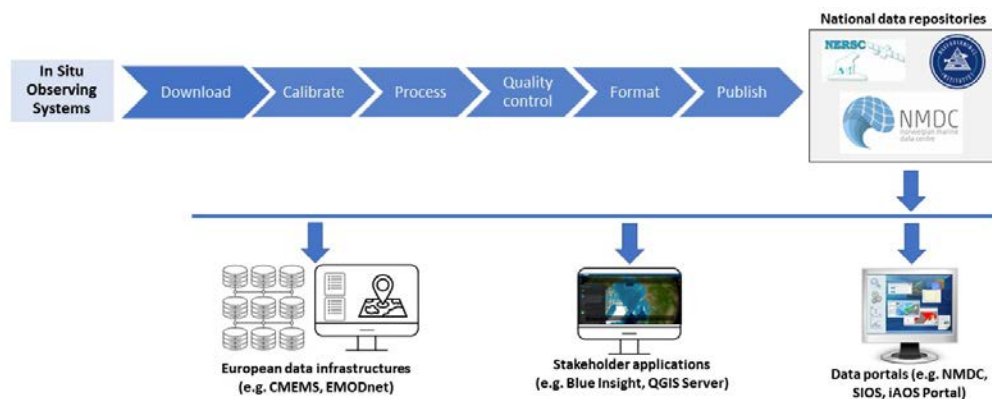


Figure 1. NERSC data delivery chain used for oceanographic and acoustics data products.

The remainder of this report is organized as follows. Section 2 provides a brief description of the metadata and data standards that are used, and Section 3 the metadata template developed at NERSC to capture the required metadata. Section 4 presents Rosetta, a tool for converting data in ASCII files to NetCDF. Section 5 outlines the major steps in publishing a dataset, while Section 6 concludes with the plans for future development of the data delivery chain.

2. Metadata and data standards

The use of standard metadata and data formats is paramount to enable reuse. Each dataset must include sufficient description, metadata (also called “data about data”), to enable scientists and other actors not involved in the data collection and processing to understand the content, assess whether it is suited for their planned use, and if so, to incorporate the data in their workflow in an appropriate manner. The NERSC team in Smart Ocean has adopted leading standards used in marine data management, namely the Climate and Forecast (CF) convention (Eaton et al., 2022) and Attribute Convention for Data Discovery (ACDD) (ESIP, 2020) for metadata and Network Common Data Form (NetCDF) (UniData, 2022) for data. The ACDD specifies metadata elements that are used to search for data. Such metadata elements are called discovery metadata. The CF convention includes both discovery metadata and use

metadata, where the latter describes characteristics of the data that are needed to use it properly, e.g., in calculations or comparison of data from different data system.

The choice of data and metadata standards is based on developments in the Horizon 2020 project Integrated Arctic Observation System (INTAROS) (Sandven et al., 2022). An INTAROS working group of experts in ocean observing and data managers defined a common metadata profile for different categories of ocean mooring data (Hamre et al., 2021). This profile is based on NetCDF/CF and the metadata structure developed by the OceanSITES program (OceanSITES, 2010). The metadata template will be reused field experiment data collected by NERSC and partners in several projects, including the EU funded project High Arctic Ocean Observation System (HiAOOS), and the Useful Arctic Knowledge: Training, Collaboration and Innovation in ocean observing (UAK-2) project funded by the Research Council of Norway.

2.1. Standards applied for Smart Ocean

NetCDF/CF defines a set of required elements, including descriptive and usage metadata, that should be stored in a NetCDF file to allow users to both discover datasets and decide whether they are fit for a particular purpose. With NetCDF/CF each variable can be described in detail using standard names and units, allowing computer applications to extract, process (e.g., re-grid) and display user selected values. In addition to parameter names and units, the CF also enables definition of, amongst others, time (point or interval), coordinate axis, depth axis and map projection. To describe the data set as a whole, CF includes metadata elements for data set title, name of institution producing the data originally, data source, history (of processing), references (e.g., scientific, or technical literature) and comments. The Attribute Convention for Data Discovery (ACDD) defines a set of attributes recommended to make a dataset stored in NetCDF easier to discover by data systems (ESIP, 2020). These attributes can be used by data servers such the Thematic Real-time Environmental Distributed Data Services (THREDDS) (UniData, 2020) to extract metadata from the datasets that are published and exporting those metadata to other metadata standards (e.g., Dublin Core (ISO, 2017), DIF (NASA, 2010), ISO 19115 (ISO, 2014a)). This will help both data systems (e.g., data catalogues, portals and digital libraries) and users in finding and using the published data more efficiently.

In Smart Ocean, the NERSC team publishes their data in the Norwegian Marine Data Centre (NMDC), a Norwegian data infrastructure coordinated by the Institute of Marine Research (NMDC, 2023). NERSC is a partner of NMDC and makes data available through its NERSC Data Node. From this node, the NMDC Data Portal harvests metadata to facilitate a common entry point to all data provided by NMDC partners.

Standard vocabularies such as GCMD Science Keywords and SeaDataNet-2 vocabularies (e.g., for parameter and sensor names, units, keywords) are used to mark up the metadata in a manner that facilitates machine readable search and retrieval. Metadata following GCMD DIF must be encoded using DIF-9 or DIF-10 XML schemas; metadata following ISO 19115 must be encoded in XML using the ISO 19139 standard. To facilitate reuse of a dataset, a more detailed description of its content, including among others, parameters and units used, is needed. Such usage metadata are well defined for some types of data, e.g., physical oceanography data from CTDs and EO data from satellites. However, for other types of data, there is a lack of standards. For instance, while GCMD and ISO19115 can represent general descriptive metadata for acoustic data from scientific experiments, there is currently no standard that can fully represent the usage metadata. The lack of standardized representation of metadata for passive acoustic data from distributed ocean acoustics observation networks, has been addressed by NERSC in other ongoing research projects.

To describe usage metadata in a consistent and machine-actionable manner, it is imperative to use standard vocabularies. These offer unambiguous definitions of key concepts and terminology within a

scientific domain, with each term having its unique identifier (URIs). This enables computer programs to distinguish between syntactical similar terms and to deduce the semantics (i.e., meaning) of the metadata. As an example, the GCMD keyword vocabulary¹ describes Science and Services Keywords (Olsen et al., 2013), Data Centers, Projects, Instruments, Platforms, Locations, Horizontal Data Resolution, Vertical Data Resolution, Temporal Data Resolution and URL Content Types. The NERC Vocabulary Server (NVS) Common Vocabulary² holds an extensive set of vocabularies for parameters, sensors, platforms, ships, organisations, projects, to name a few. These vocabularies have been developed by the British Oceanographic Data Centre in collaboration with the scientific community in a series of projects for the past decades. Currently, the development of the NVS Common Vocabulary is carried out within the frame of SeaDataNet³, a distributed Marine Data Infrastructure with more than 50 partners and sub-contractors from Europe, Russia, Africa and Australia. A third highly relevant vocabulary is the list of standard parameters names and units in the NetCDF Climate Forecast (CF) conventions (Eaton et al., 2011).

2.2. Metadata for oceanographic and acoustics data products

Table 1 shows the metadata elements that are used for NERSC oceanographic and acoustic data products in Smart Ocean. The metadata profile is based on earlier work in INTAROS (Hamre et al., 2021), extended with new elements to provide credit to the project funding the data collection, scientists and project managers with administrative roles that facilitate for data collection and more information on scientists planning and conducting the specific field experiments.

Table 1. Metadata elements for NERSC oceanographic and acoustics data products.

Metadata field	Description (convention)
	NetCDF file description
Conventions	Name of conventions followed by the dataset, e.g., “CF-1.6” (CF)
featureType	Type of sampling geometry for the dataset. For CTD/XBT stations, the value shall be “profile”. (CF)
data_set_language	The language used for metadata attributes in the dataset, e.g., “english”. (Rosetta)
	Overview of dataset
title	Informative title of the dataset. This should contain the name of the project and/or location, what kind and when the data was collected. (CF)
summary	A short paragraph describing the dataset, e.g., what parameters it contains, time period, location, and in what context the data was collected. (ACDD)
area	The geographic area where the data was collected. (Rosetta)
source	The method used for producing the original data. For in situ observations, it should categorise how data was collected. (CF)
platform	Name of platform(s) holding the sensor used to generate the data. (ACDD)
cruise_id	Name of the cruise where data was collected. (INTAROS)
id	A unique identifier of the dataset, provided by the naming authority. (ACDD)
naming_authority	Name of organisation assigning the id. Recommend using reverse DNS naming. (ACDD)
keywords	A set of keywords describing the dataset, using terms from a controlled vocabulary. (ACDD)
keywords_vocabulary	Name of keyword vocabulary. (ACDD)
iso_topic_category	A high-level classification of topics for which the data can be relevant. (INSPIRE)

¹ <http://gcmd.nasa.gov/learn/keywords.html>

² <https://www.seadatanet.org/Standards/Common-Vocabularies>

³ <https://www.seadatanet.org/>

Metadata field	Description (convention)
processing_level	A short text describing the processing (or quality control) level of the data. (ACDD)
	<i>Project and funding</i>
project	Name of the project funding data collection, optionally with project number. (ACDD)
project_id	Project or contract number for the project. If multiple projects separate by comma. (Smart Ocean)
funding_agency	Name of body funding the project(s). If multiple projects separate by comma. (Smart Ocean)
project_lead	Name of person leading the project(s). If multiple projects separate by comma. (Smart Ocean)
project_lead_email	Email address of person leading the project(s). If multiple projects separate by comma. (Smart Ocean)
	<i>Lead for scientific work in the project</i>
principal_investigator	Name of principal investigator of the project(s). (OceanSITES)
principal_investigator_email	Email address of the principal investigator(s). (OceanSITES)
	<i>Responsible investigator for generating the dataset</i>
investigator	Name and affiliation of the responsible investigator. (Smart Ocean)
investigator_email	Email address of the responsible investigator. (Smart Ocean)
	<i>Cruise responsible</i>
cruise_responsible	Principle Investigator(s) of the cruise and their organisation. Names, separated by semi-colon if multiple investigators. (INTAROS)
cruise_responsible_email	Email address of each cruise responsible, separated by semi-colon if multiple investigators. (INTAROS)
	<i>Data management and publishing</i>
data_assembly_center	Data assembly centre in charge of the dataset. (OceanSITES)
publisher_name	Name of publisher. If type of publisher is “institution”, name of organisation that published the data. (ACDD)
publisher_institution	The organisation that published the data. (ACDD)
publisher_type	Type of publisher, either “person”, “group”, “institution” or “position”. (ACDD)
publisher_email	Email address of publisher. (ACDD)
publisher_url	URL of person or institution that published the data. (ACDD)
	<i>Contributors to assembling and publishing data products</i>
contributor_name	Names of all contributors, separated by semi-colon. (ACDD)
contributor_role	Roles of all contributors, in the same order as names, and separated by semi-colon. (ACDD)
contributor_email	Email addresses of all contributors, separated by semi-colon. (ACDD)
	<i>Acknowledgement and citation</i>
citation	Statement to be used when citing the dataset in publications. (OceanSITES)
acknowledgement	Statement for acknowledging the projects funding data collection. (ACDD)
license	URL to a standard of specific data license, or a text describing any restrictions in data access and use. (ACDD)

**Add fields for geo-spatial coverage

3. Metadata template for Smart Ocean data products

To support data publication, we generated a PDF file that the scientists can use to enter the metadata needed for publication of Smart Ocean field experiment data products. This PDF provides a set of pages for entering:

- Overview of dataset.
- Project and funding.

- Lead for scientific work in the project.
- Responsible investigator for the data collection.
- Cruise responsible.
- Data management and publishing.
- Contributors to assembling and publishing data.
- Acknowledgement and citation.

The different parts enable registration of general information about the dataset, giving recognition to all actors involved in its generation. This covers all activities from providing funding and defining the scientific objectives, planning and conducting the field experiment, processing and publishing data after the cruise, informing about data license, and to providing statements for citation and acknowledgement when data are used in e.g., scientific papers or other publications.

The PDF form contains four pages for entering metadata. Figure 2 shows the first page in the form filled in for a dataset from a mooring deployed in Stokksundet near Bømlø in March 2022. The form contains several free text fields and a few fields where values are chosen from a drop-down list. An example of the latter is used to select the vocabulary defining keywords for the dataset.

Global Attributes

Data set Language

Overview of Dataset

Title	SFI Smart Ocean: Oceanographic and acoustic data collected in Stokksundet, March-April 2022
Summary	Oceanographic data (temperature, conductivity, pressure, currents) measured by instruments in a mooring placed next to a seafood farm in Stokksundet, Norway. The instruments collected data from 17 March to 28 April 2022 as part of a
Area	Stokksundet, Norway
Source	Stokksund-M2022
Platform	Mooring
Cruise ID	Bømlø2022_March-April
ID	Bømlø2022_instrument1
Naming Authority	no.nersc.sfismartocean
Keywords	EARTH SCIENCE, OCEANS, OCEAN TEMPERATURE
Keywords Vocabulary	GCMD Science Keywords
Iso Topic Category	oceans
Processing Level	Converted to text.

Figure 2. Screenshot of the first page of the PDF form for global metadata for ocean mooring data products.

4. Data file conversion using Rosetta

Rosetta is a web application that converts a data file in comma separated value (CSV) ASCII text format to a standard NetCDF file. The user is guided through a series of steps to select input data file and then subsequently describe the structure and content of this file, including additional information describing the data (Figure 3). The definitions and descriptions (metadata) are stored in separate file, called a template, that can be reused to convert additional data files with a similar structure. If a template is available, it can be loaded before the data file is chosen.

After selecting the header lines in the input file (Figure 3(a)), and defining which character separate the columns in the data file, the next step is specifying which the data columns to include in the conversion

and their content. A column can contain a coordinate variable such as latitude, longitude, altitude, or time (Figure 3(b)). Other columns contain the parameters measured or computed (Figure 3(c)). Each column to be included in the output NetCDF must be specified with a unique name, data type, and unit. Additional metadata can be added depending on whether the column contains a coordinate variable, an observation, or a derived quantity.

The next two steps complete the description of the observing platform and additional information for the dataset. For the latter, the PDF form described in the previous section contains the needed metadata. When all needed metadata has been provided, Rosetta converts the selected parts of the input data file. The resulting NetCDF as well as the template can then be downloaded.

Rosetta can be accessed at <https://rosetta.nersc.no/>. A detailed user manual is available from the site (Hamre et al., 2022).

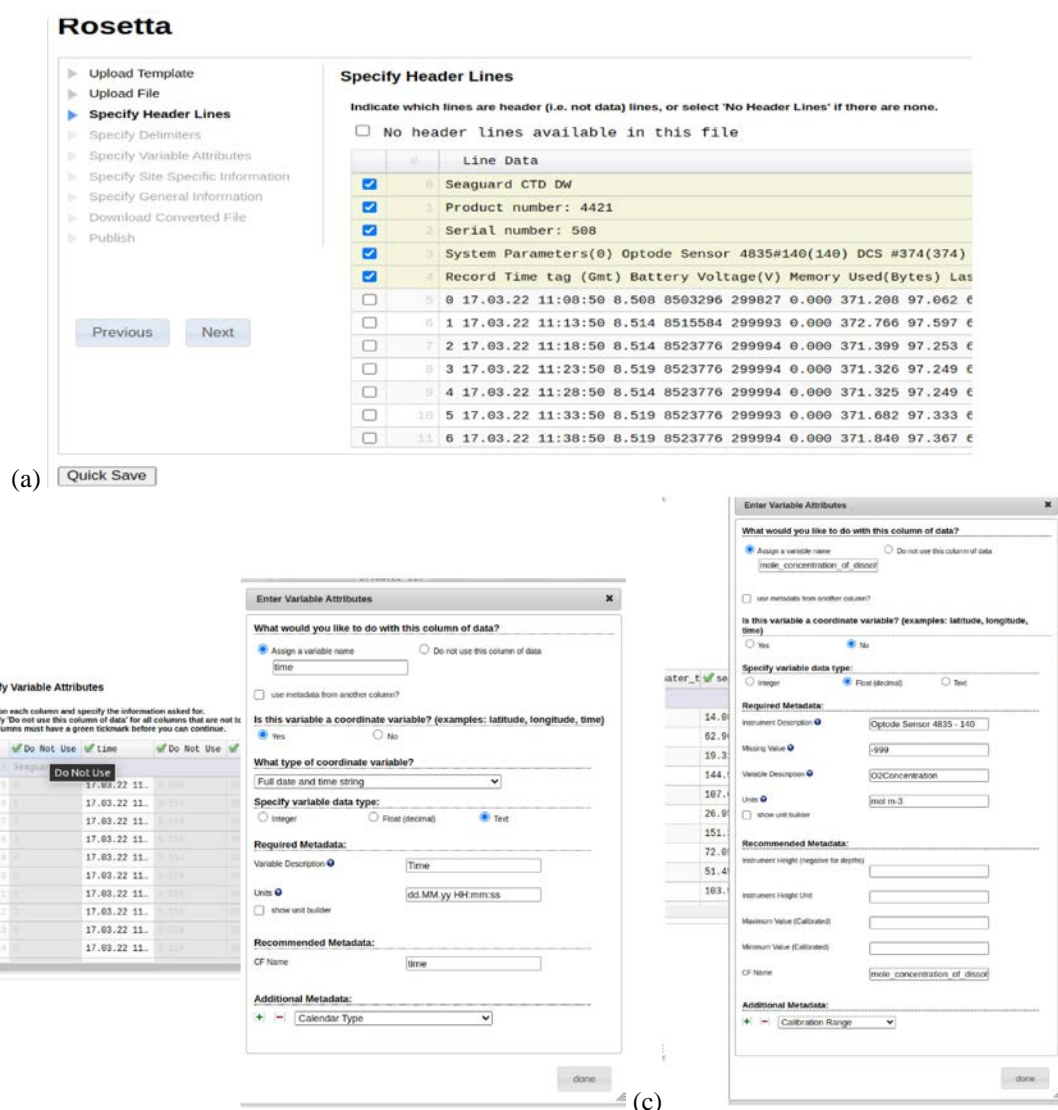


Figure 3. Rosetta screenshots showing the steps of file conversion (a) and examples of web forms for specifying metadata for variables in the data file for a coordinate variable (b) and a measured parameter (c).

5. Publishing data products in NMDC

When a NetCDF file has been generated by Rosetta and validated using a compliance checker, the dataset is published on the NERSC Data Node of the Norwegian Marine Data Centre (NMDC). This entails uploading the file to a directory on the Thredds Data Server and creating a small configuration file enabling the server to provide access through one of more standard interfaces. The Thredds data server supports several standard interfaces including Open-source Project for a Network Data Access Protocol (OPeNDAP) and Open Geospatial Consortium Web Map Service (OGC WMS), enabling client software and GIS tools to connect and extract (parts of) the dataset as a file, or generate a raster version (map) of the data to overlay with other data.

The Thredds Data Server offers access to datasets and their data files through an easy-to-use graphical interface in a web browser (Figure 4). Depending on the configuration for the dataset, the user can navigate to individual datafiles for download of full/parts of files, or display of selected parameters on a map in the web browser.

(a)

(b)

Figure 4. NERSC Thredds Data Server (TDS) offering access to the Stokksundet mooring data 2022, comprised of three NetCDF files holding data for two CTD instruments and one currentmeter mounted on this mooring. Each file is shown in the folder for the dataset (a). Clicking on a file name opens a data access page (b) for the selected file, with more information about the data file and links for access through a standard interface.

Figure 4. NERSC Thredds Data Server (TDS) offering access to the Stokksundet mooring data 2022, comprised of three NetCDF files holding data for two CTD instruments and one currentmeter mounted on this mooring. Each file is shown in the folder for the dataset (a). Clicking on a file name opens a data access page (b) for the selected file, with more information about the data file and links for access through a standard interface.

To make the new dataset visible through the NMDC, a metadata record is created in the NERSC data catalogue. The central data portal of NMDC harvests these metadata into its own data catalogue, making it visible for all users (Figure 5(a)). NMDC further generates a web page (called a landing page) with metadata and access links for the dataset (Figure 5(b)). This page includes the Digital Object Identifier (DOI), which is a unique reference for citing the dataset when used in presentations, papers, and

services. The DOI assigned by NMDC contains “NERSC” as part of the reference, branding the organisation as a data provider. The DOI can be pasted into the browser address bar and will open the dataset’s landing page.

(a)

(b)

Seamless access to Norwegian marine data

Nansen Environmental and Remote Sensing Center
<https://doi.org/doi:10.21335/NMDC-NERSC-1124437080>

SFI Smart Ocean: Oceanographic and acoustic data collected in Stokksundet, March-April 2022

Recommended citation:
 Espen Storheim (Nansen Environmental and Remote Sensing Center) Bjørnar Hallaråker Røsvik (University of Bergen) Hanne Sagen (Nansen Environmental and Remote Sensing Center) Frode Monsen (Nansen Environmental and Remote Sensing Center) Tor I. Olaussen (Nansen Environmental and Remote Sensing Center) Torill Hamre (Nansen Environmental and Remote Sensing Center) (2023) SFI Smart Ocean: Oceanographic and acoustic data collected in Stokksundet, March-April 2022
<https://doi.org/doi:10.21335/NMDC-NERSC-1124437080>

To cite this dataset use the following:
 Text citation

Usage:
 Users must display the following citation in any publication or product using this dataset.
 "Storheim, Espen, Røsvik, Bjørnar Hallaråker, Sagen, Hanne, Monsen, Frode, Olaussen, Tor I., Hamre, Torill (2022). SFI Smart Ocean: Oceanographic and acoustic data collected in Stokksundet, March-April 2021."

Abstract
 Oceanographic data (temperature, conductivity, pressure, currents) measured by instruments in a mooring placed next to a seafood farm in Stokksundet, Norway. The instruments collected data from 17 March to 28 April 2022 as part of a Master study. The data is presented in Røsvik (2022). Røsvik, Bjørnar Hallaråker, 2022. Multidisciplinary Moorings in the Arctic and Coastal Areas. Master Thesis. The University of Bergen. Available at: <https://hdl.handle.net/11250/3001633> (Accessed: 22 September 2022).

Scientific keywords:
 Key words:

Data downloads

Type	Description	URL
GET DATA	Description	https://thredds.nersc.no/thredds/fileServer/SFI-Smart-Ocean/Stokksundet-mooring-data-2022/SBE37SM-RS232_03709105_2022_04_28_converted.nc
GET DATA	Description	https://thredds.nersc.no/thredds/fileServer/SFI-Smart-Ocean/Stokksundet-mooring-data-2022/SBE37SM-RS232_03709104_2022_04_28_converted.nc
GET DATA	Description	https://thredds.nersc.no/thredds/fileServer/SFI-Smart-Ocean/Stokksundet-mooring-data-2022/AADI_RCM.nc
GET DATA THREDDS CATALOG	Description	https://thredds.nersc.no/thredds/catalog/SFI-Smart-Ocean/Stokksundet-mooring-data-2022/Hydrophone-recordings/catalog.html

Figure 5. NMDC central data portal home page with search criteria entered (a) and a dataset landing page for a NERSC Smart Ocean dataset published in NMDC (b).

6. Future use and development of the data delivery chain

The metadata template (PDF form) described in section 3 will be used for future ocean mooring and other in situ datasets generated by NERSC within SFI Smart Ocean. As part of Smart Ocean there are plans for deploying two moorings north of Svalbard in 2025, to collect oceanographic and acoustic observations for a period of one year.

To efficiently handle the growing amounts of data from ocean moorings it is important to operationalise the data delivery chain. This entails using and as needed customising the metadata template described in this report, further enhancing Rosetta and associated tools for metadata generation and validation, as well as automating the publication process in Thredds for NMDC. This will contribute to increased availability of oceanographic and acoustic data products north of Svalbard. The data products will be well documented with metadata describing, among others, all staff that have contributed to the generation and publication of the data, acknowledgement of funding agencies supporting data collection, processing and dissemination, and standard names, units and valid ranges for observed and derived quantities. All datasets will be assigned a DOI, ensuring proper citation, and a data license defining how the data can be used by data consumers.

7. References

- Eaton, B., J. Gregory, B. Drach, K. Taylor and S. Hankin, NetCDF Climate and Forecast (CF) Metadata Conventions, Version 1.6, 5 December 2011.
- Eaton, Brian, Jonathan Gregory, Bob Drach, Karl Taylor, Steve Hankin, Jon Blower, John Caron, Rich Signell, Phil Bentley, Greg Rappa, Heinke Höck, Alison Pamment, Martin Juckes, Martin Raspaud, Randy Horne, Timothy Whiteaker, David Blodgett, Charlie Zender, Daniel Lee, David Hassell, Alan D. Snow, Tobias Kölling, Dave Allured, Aleksandar Jelenak, Anders Meier Soerensen, Lucile Gaultier, and Sylvain Herlédan, CF convention, Version 1.10, 31 August 2022. <http://cfconventions.org/> (accessed 9 February 2023).
- ESIP. 2020. Attribute Convention for Data Discovery 1-3. https://wiki.esipfed.org/Attribute_Convention_for_Data_Discovery_1-3 (accessed 10 February 2023).
- Hamre, Torill, Sagen, Hanne, Sandven, Stein, Danielsen, Finn, Ottersen, Geir, Beszczynska-Moller, Agnieszka, Morvik, Arnfinn, Yamakawa, Asuka, Schewe, Ingo, & Enghoff, Martin. (2021). Deliverable 1.8 Data Management Plan V2. Zenodo. <https://doi.org/10.5281/zenodo.7015039>
- Hamre, Torill, Monsen, Frode, Yamakawa, Asuka and Olaussen, Tor (2022). Rosetta User Manual 13 January 2022.
- ISO, 2014a. ISO 19115-1:2014 Geographic information — Metadata — Part 1: Fundamentals.
- ISO, 2014b. ISO 19115 Topic Category vocabulary. <https://apps.usgs.gov/thesaurus/thesaurus-full.php?thcode=15> (accessed 10 February 2023).
- ISO, 2017. ISO 15836-1:2017 - Information and documentation - The Dublin Core metadata element set - Part 1: Core elements".
- NASA, 2010. Directory Interchange Format (DIF) Standard. <https://www.earthdata.nasa.gov/esdis/esco/standards-and-practices/directory-interchange-format-dif-standard> (accessed 13 February 2023).
- NMDC, 2023. The Norwegian Marine Data Centre, <https://nmhc.no/> (accessed 9 February 2023).
- OceanSITES, 2010. OceanSITES User's Manual NetCDF Conventions and Reference Tables. Version 1.2. June 29, 2010.
- Olsen, L.M., G. Major, K. Shein, J. Scialdone, S. Ritz, T. Stevens, M. Morahan, A. Aleman, R. Vogel, S. Leicester, H. Weir, M. Meaux, S. Grebas, C. Solomon, M. Holland, T. Northcutt, R. A. Restrepo, R. Bilodeau, 2013. NASA/Global Change Master Directory (GCMD) Earth Science Keywords. Version 8.0.0.0.0.
- Sandven, Stein, Sagen, Hanne, Pirazzini, Roberta, Beszczynska-Möller, Agnieszka, Danielsen, Finn, Gonçalves, Pedro, Ottersen, Geir, Zona, Donatella, Buch, Erik, Gustavson, David, Voss, Peter, Iversen, Lisbeth, Hamre, Torill, Sejr, Mikael, & Higgins, Ruth. (2022). Deliverable 1.11 Final synthesis report. Zenodo. <https://doi.org/10.5281/zenodo.7033824>
- UIB, 2023. SFI Smart Ocean, <https://sfismartoocean.no/> (accessed 9 February 2023).
- UniData, 2020. TDS Online Tutorial. <https://docs.unidata.ucar.edu/tds/current/userguide/index.html> (accessed 10 February 2023).
- UniData, 2022. NetCDF Users Guide v1.1. <https://docs.unidata.ucar.edu/nug/current/index.html> (accessed 10 February 2023).

Appendix A Example of global metadata for a mooring dataset

The following global metadata was extracted from a NetCDF file generated by Rosetta. The metadata elements have been ordered in the same sequence as in Table 1.

```
:Conventions = "CF-1.6";
:featureType = "timeSeries";
:data_set_language = "english";
:title = "SFI Smart Ocean: Oceanographic and acoustic data collected in Stokksundet, March-
April 2022";
:summary = "Oceanographic data (temperature, conductivity, pressure, currents) measured by
instruments in a mooring placed next to a seafood farm in Stokksundet, Norway. The
instruments collected data from 17 March to 28 April 2022 as part of a Master study. The
data is presented in Røsvik (2022).\nRøsvik, Bjørnar Hallaråker, 2022. Multidisciplinary
Moorings in the Arctic and Coastal Areas. Master Thesis. The University of Bergen.
Available at: https://hdl.handle.net/11250/3001633 (Accessed: 22 September 2022).";
:area = "Stokksundet, Norway";
:source = "Stokksund-M2022";
:platform = "Mooring";
:cruise_id = "Bømlo2022_March-April";
:id = "Bømlo2022_instrument1";
:naming_authority = "no.nersc.sfismartoocean";
:keywords = "EARTH SCIENCE, OCEANS, OCEAN TEMPERATURE";
:keywords_vocabulary = "GCMD Science Keywords";
:iso_topic_category = "oceans";
:processing_level = "Converted to text.";
:project = "SFI Smart Ocean (RCN project no. 309612)";
:project_id = "309612";
:funding_agency = "Research Council of Norway";
:project_lead = "Marie Bueie Holstad, University of Bergen";
:project_lead_email = "marie.holstad@uib.no";
:principal_investigator = "Hanne Sagen, Nansen Environmental and Remote Sensing Center";
:principal_investigator_email = "hanne.sagen@nersc.no";
:investigator = "Espen Storheim, Nansen Environmental and Remote Sensing Center";
:investigator_email = "espen.storheim@nersc.no";
:cruise_responsible = "Espen Storheim, Nansen Environmental and Remote Sensing Center";
:cruise_responsible_email = "espen.storheim@nersc.no";
:data_assembly_center = "NERSC";
:publisher_name = "Nansen Environmental and Remote Sensing Center";
:publisher_institution = "Nansen Environmental and Remote Sensing Center";
:publisher_type = "institution";
:publisher_email = "datamanager@nersc.no";
:publisher_url = "https://www.nersc.no/";
:contributor_name = "Espen Storheim;Bjørnar Hallaråker Røsvik; Hanne Sagen; Frode Monsen;
Tor I Olaussen; Torill Hamre";
:contributor_role = "Data collection and processing; Data collection; Principal
investigator; data formatting and curation; publishing data; data manager ";
:contributor_email = "espen.storheim@nersc.no;bjornar.rosvik@student.uib.no
;hanne.sagen@nersc.no; frode.monsen@nersc.no; tor.olausen@nersc.no;
torill.hamre@nersc.no";
:citation = "Users must display the following citation in any publication or product using
this dataset. \n“Storheim, Espen, Røsvik, Bjørnar Hallaråker, Sagen, Hanne, Monsen, Frode,
Olaussen, Tor I., Hamre, Torill (2022). SFI Smart Ocean: Oceanographic and acoustic data
collected in Stokksundet, March-April 2021.””;
```

```
:acknowledgement = "Data were collected and processed in the RCN project SFI Smart Ocean  
(project no. 309612), and made freely available by NERSC under SFI Smart Ocean (project  
no. 309612).";  
:license = "https://creativecommons.org/licenses/by/4.0/. Users must include the  
acknowledgement and citation for these data.";
```

Appendix B Example of metadata for a variable in a mooring dataset

The following variable definition was extracted from the NetCDF file generated by Rosetta for the same dataset as in Appendix A.

```
float sea_water_temperature_3(dateTime=11302);  
  :_userSuppliedName = "sea_water_temperature";  
  :_coordinateVariable = "non-coordinate";  
  :missing_value = -999.0f; // float  
  :standard_name = "sea_water_temperature";  
  :units = "degree_C";  
  :source = "Optode Sensor 4835 - 140";  
  :long_name = "Temperature";  
  :_columnId = "8";  
  :coordinates = "time lat lon alt";
```