Norwegian Marine Data Centre

Notes Before Completing the Application

We have read and understood the notes concerning our application submission.

True

Reviewer Entry

Reviewer 1

Comments:

Reviewer 2

Comments:

CORE TRUSTWORTHY DATA REPOSITORIES REQUIREMENTS

Background & General Guidance

Glossary of Terms

BACKGROUND INFORMATION

Context

R0. Please provide context for your repository.

Repository Type. Select all relevant types from:

National repository system; including governmental

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:

Brief Description of Repository

The Norwegian Marine Data Centre (hereafter named NMD) at the Institute of Marine Research (hereafter named IMR), was established as a national data centre dedicated to processing and long-term storage of marine environmental and fisheries data, as well as making data available to the public through data products and publishing data sets.

NMD aims to be the national hub for all marine data collected by IMR and other institutions, and has been a “National Oceanographic Data Centre” (NODC) since 1971. NMD is the coordinator of the national research infrastructure NMDC, comprising of 16 Norwegian partners collaborating to make marine data available for research.

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:

Brief Description of the Repository’s Designated Community.

The repository’s Designated Community is researchers within Norwegian marine research. The research is within marine sciences such as management of fish stock, salmon lice development, causes of increase in ocean temperature.

In addition to the research community, NMD also makes data available to the public.

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:

Level of Curation Performed. Select all relevant types from:

D. Data-level curation – as in C above; but with additional editing of deposited data for accuracy

Reviewer Entry

Reviewer 1

Comments:
Accept - using this as a lens to evaluate subsequent relevant requirements.

Reviewer 2

Comments:

Comments

NMD maintains data from IMR and other institutions and universities. The data curation level on IMR data is category D.

Due to international cooperation, some of the data also undergo a secondary quality control at partner repositories and to ensure high quality data feedback is given to NMD.

All original data is stored at NMD and changes are only made on copies of original data.

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:

Insource/Outsource Partners. If applicable, please list them.

Reviewer Entry

Reviewer 1

Comments:

Reviewer 2

Comments:

Summary of Significant Changes Since Last Application (if applicable).

When submitting the first CTS application in 2018, NMD was part of “MØR” - Marine Ecosystems and Resources group at IMR. However, in 2021, a restructuring at the IMR resulted in the new department "HI Digital", comprising of NMD, the IT section and a Digital Development group.
NMD are currently 15 employees dedicated to managing and publishing data. Funding of NMD has not been affected by

the restructure.

Two ongoing major project application processes involving NMD:
- A continuation of CMEMS (Copernicus Marine Environment Monitoring Service) for 2022-2026 is in progress.
- Phase two of NMDC (NMDC-II), within the aim of this project is upgrading the national research infrastructure NMDC for 10 more years.

NMD is continuously evolving and improving standards and over the three-year period since the first NMD CTS application, some of the developments are as follows:

As NMD is coordinating the NMDC infrastructure, IMR has chosen to use the NMDC data portal for publishing data. Approximately 600 datasets are available, and this number is continuously growing.

The use of Digital Object Identifiers (DOI) started in 2017 and is now an implemented routine, further described in section R7.
The data deposit portal mentioned in the previous CTS application is now in the planning process, and further described in section R7.
A new system for monitoring the data flow of the data collected on cruises at IMR, Dataset Tracker, was also mentioned in the previous application. This is now in the process of developing specifications for the system.

Another significant change since last application is that NMD now have a documented preservation plan, further described in section R10, and a Data Management Plan referred to in section R9.

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:

Other Relevant Information.

The NMD is national data centre for marine environmental and fish data.

The IMR/NMD is a partner of SeaDataCloud: https://www.seadatanet.org/About-us/SeaDataCloud/Partners.
IMR/NMD is also a member of EMODnet in the categories Biology and Chemistry:
http://www.emodnet.eu/partners-portal/4 and http://www.emodnet.eu/partners-portal/5.

NMD is responsible for the Arctic in-situ Thematic Assembly Centre (TAC) serving data in the Copernicus Marine environment monitoring service: http://marine.copernicus.eu/services-portfolio/access-to-products/?option=com_csw&task=results&advancedsearch-geographical_area[]=advancedsearch-geographical_area-arctic.

NMD is responsible for the data portal arctic ROOS with near real-time in-situ data: http://arctic-roos.org/In-situ.

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:

ORGANIZATIONAL INFRASTRUCTURE

1. Mission/Scope

R1. The repository has an explicit mission to provide access to and preserve data in its domain.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:
4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:

Response:

The main tasks of NMD is to:

- Collect, quality-assure and store aqua cultural, marine environmental, and fisheries data
- Make the data available for research

Please see the description of NMD for reference:
https://www.hi.no/en/hi/forskning/research-groups-1/the-norwegian-marine-data-centre-nmd.

NMD was established as a national data centre for handling marine environmental and fisheries data. The group's main activity is to collect and store secured and quality controlled marine environmental and fisheries data, and make it available for scientists: https://www.hi.no/en/hi/forskning/research-data-1.

Each year, the Ministry of Trade, Industry and Fisheries sends an award letter to the Institute of Marine Research: https://www.regjeringen.no/contentassets/76a9cd4d079d415dbd6d7faa0ebae0e9/havforskningsinstituttet-tildelingsbrev-2021.pdf.

The purpose of NMD is stated in the letter. As the document is available in Norwegian only, two of the relevant sections are translated below:

Page 4: «Objective 3 – To generate, acquire, collect, manage and make available relevant data of high quality, for research, management and the industry»

Page 5: "The institute shall prepare plans for data acquisition and data processing within all relevant fields, and work towards standardize data collection and develop common methods internationally. The institute has a national main responsibility of collect, manage and publish marine data in Norway. The data management shall follow the FAIR principles: Findable, Accessible, Interoperable, and Reusable."

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:

2. Licenses

R2. The repository maintains all applicable licenses covering data access and use and monitors compliance.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:

4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:

Response:

Data collected under the auspices of IMR is the property of the institute. IMR data is managed by NMD. This means that NMD has the rights to copy, transform, store and provide access to the data. NMD has the responsibility of the data from IMR and thus has the responsibility for preservation of the data. This responsibility covers data availability, integrity, consistency, safe storage and protection of privacy.

The data deposited at NMD is freely available under the conditions of the Norwegian License for Open Government Data (NLOD): https://data.norge.no/nlod/en/2.0 and the license Creative Commons Attribution 4.0 International (CC BY 4.0): https://creativecommons.org/licenses/by/4.0/legalcode.

Furthermore, the source must be acknowledged as stated in the Data Policy by IMR: https://www.hi.no/resources/imr/Data-policy-IMR.pdf.

The licenses above still apply if NMD is managing data that is not property of IMR.

NMDC harvest metadata from 5 partner institutions in addition to IMR. If metadata entries from the partners does not contain a specific license information, this will not be displayed in the entry.
All IMR data entries in nmdc.no should include license information. The NMDC Project encourages all partners to document a policy on data sharing and re-use. The current status is documented here: https://nmdc.no/nmdc/datapolicy.

Reviewer Entry

Reviewer 1

Comments:
Accept.

Reviewer 2

Comments:

3. Continuity of access

R3. The repository has a continuity plan to ensure ongoing access to and preservation of its holdings.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:
4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:

Response:

The NMD is implemented as an integral part of the IMR infrastructure. The IMR is a subordinate institute at the Ministry of Trade, Industry and Fisheries https://www.regjeringen.no/en/dep/nfd/organisation/etater-og-virksomheter-under-narings--og-fiskeridepartementet/Subordinate-agencies-and-institutions/id115215/.

The IMR has existed as an independent institute since 1947 and it is very unlikely that the IMR will be closed. NMD is organized as a permanent section within IMR. Please see the organizational chart, where NMD is listed as an individual section: https://www.hi.no/en/hi/about-us/organisation.

NMD is also acting as the National Oceanographic Data Centre (NODC) of Norway since 1971. The head of NMD, Helge Sagen, is member of International Oceanographic Data and Information Exchange (IODE) National Coordinator for Oceanographic Data Management: http://iode.org/index.php?option=com_oe&task=countryReports&report[countryID]=161.

In the unlikely event of the closure of IMR the data at NMD will be transferred to the National Archives of Norway: https://www.arkivverket.no/en/about-us/the-national-archives-of-norway. A yearly copy of the data at NMD is deposited here.

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:

4. Confidentiality/Ethics

R4. The repository ensures, to the extent possible, that data are created, curated, accessed, and used in compliance with disciplinary and ethical

norms.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:
4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:

Response:

The Data Policy for the Institute of Marine Research states that IMR requires that all its research data is processed ethically both internally and by external users: https://www.hi.no/resources/imr/Data-policy-IMR.pdf.

The NMD is following the Guidelines for research ethics in science and technology by the Norwegian National Research Ethics Committees: https://www.etikkom.no/en/ethical-guidelines-for-research/guidelines-for-research-ethics-in-science-and-technology/, as well as the Animal Welfare Act: https://www.regjeringen.no/en/dokumenter/animal-welfare-act/id571188/.

Whenever new data is handed over to NMD, the "Procedure to review disclosure risk in data" is being followed: https://www.hi.no/forskning/forskningsgrupper/norsk-marint-datasenter-nmd/kvalitetsdokumentasjon/pub/docs/6238.pdf.

Data with disclosure risk are stored on disks with specific limited access. This type of data at NMD has a low disclosure risk and an access list is managed by NMD. To get access to these data, a confidentiality statement must be signed. Also, NMD is following regulations from The Norwegian Data Protection Authority: https://www.datatilsynet.no/en/regulations-and-tools/regulations/.

We do have measures in place to handle and distribute data not owned by IMR. In these cases, agreements on data management of the data have been signed by all parties involved.

NMD is managing business sensitive data that might affect stock exchange. Such data is managed under certain access procedures by persons with a signed confidentiality statement, described in the procedure "Sensitive data at IMR": https://www.hi.no/forskning/forskningsgrupper/norsk-marint-datasenter-nmd/kvalitetsdokumentasjon/pub/docs/6237.pdf.

If cases of non-compliance with these conditions, the measures in place are the legal consequences that may apply according to national and international laws.

All employees of IMR are covered by the Public Administration Act, section 13, and have signed a confidentiality agreement. Please see Public Administration Act, section 13 for reference: https://lovdata.no/dokument/NLE/lov/1967-02-10

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:

5. Organizational infrastructure

R5. The repository has adequate funding and sufficient numbers of qualified staff managed through a clear system of governance to effectively carry out the mission.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:
4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:

Response:

NMD is an organizational part of the Institute of Marine Research (IMR), Norway’s largest centre for marine science. The main task of IMR is to advice Norwegian authorities on aquaculture and the ecosystems of Norwegian waters.

NMD has its main funding from its host institution IMR. Since 2021, NMD has been part of the new department “HI Digital”, together with the IMR IT expertise and developers. The NMD team is composed of 15 data managers and GIS experts.

Personnel from NMD are active internationally, e.g., EU funded projects, ICES and IOC/IODE communities. NMD held the

chair position in ICES Data Management for 10 years, in Working Group on Marine Data Management (WGMDM), Working Group on Data and Information Management (WGDIM) and Data and Information Group (DIG). Currently, HI Digital is chairing DIG.

IMR has more than 115 years of history in marine sciences and NMD has existed for nearly 50 years, and there are no plans for reducing funds for marine data management at IMR. NMD has been active partners since the mid 90’s and are currently partner in four EU funded projects in marine data management.

IMR has its own knowledge portal with various courses available for employees at IMR. The course portal is only available on the intranet. It is also possibly to attend International Council for the Exploration of the Sea (ICES) training courses and other courses if needed. In addition, NMD has its own budget for consumables and training.

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:

6. Expert guidance

R6. The repository adopts mechanism(s) to secure ongoing expert guidance and feedback (either inhouse or external, including scientific guidance, if relevant).

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:
4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:

Response:

The Institute of Marine Research encourages and supports on-going employee training and development to ensure the staff have appropriate knowledge and skills.
The NMD is a member of several international cooperation units and staff from NMD regularly attend meetings within these groups. The NMD seek advice through international cooperation in European Union (EU) projects, ICES and IODE.

Groups where NMD is a member:

EMODnet Biology: http://www.emodnet.eu/partners-portal/4
EMODnet Chemistry: http://www.emodnet.eu/partners-portal/5
EMODnet Data Ingestion: https://www.emodnet-ingestion.eu/about/who
ICES (International Council for the Exploration of the Sea): http://ices.dk/Pages/default.aspx
IODE (Intergovernmental Oceanographic Commission on UNESCO International Oceanographic Data and Information Exchange): https://iode.org/index.php?option=com_oe&task=countryReports&report[countryID]=161

User feedback from the Designated Community is requested biannually. The Designated Community is always welcome to send an email through the helpdesk service at NMD.

Reviewer Entry

Reviewer 1

Comments:
Accept. This response is unchanged from the previous certification, but that is fine.

Reviewer 2

Comments:

DIGITAL OBJECT MANAGEMENT

7. Data integrity and authenticity

R7. The repository guarantees the integrity and authenticity of the data.

Compliance Level:

3 – The repository is in the implementation phase

Reviewer Entry

Reviewer 1

Comments:
3 – The repository is in the implementation phase

Reviewer 2

Comments:

Response:

The NMD has a data ingestion policy with requirements that must be fulfilled when delivering data to NMD for storage: https://www.hi.no/forskning/forskningsgrupper/norsk-marint-datasenter-nmd/kvalitetsdokumentasjon/pub/docs/6156.pdf. The stored data must also be compliant with the data policy of IMR.

NMD is still in the planning process of developing a data deposit portal, where users can upload their data. This portal will have requirements for the content of the data (i.e., any disclosure data is not permitted) and mandatory metadata. The data managers at NMD are responsible for ensuring the metadata is complete before finalizing the ingestion into the data portal. The data portal will be developed in the near future.

NMD is responsible for importing data to databases and ensuring that this does not corrupt the data in any way. Usually, a dataset is verified by comparing the uploaded file to the database and an exported file of the same data. When the two files are similar, no errors occurred. NMD wish to implement the functionality to calculate checksums of files before upload and after download for comparison. This functionality will be implemented as part of the data deposit portal mentioned above.

Most of the data stored at NMD is acquired on cruises with IMR research vessels. Each cruise goes through an application process, where it is stated, which instruments will be used and hence what data will be collected before the cruises are approved and finally executed. All cruises are registered in the Cruise System at IMR, an internal system for application, approval, cruise planning and crew planning, where the applications are filled in.

After the cruise has ended, the cruise leader writes a Cruise Summary Report (CSR), which is uploaded to seadatanet.org and can be found here: https://seadata.bsh.de/Cgi-csr/retrieve_sdn2/start_sdn2.pl.
Cruise Summary Reports are the usual means for reporting on cruises or field experiments at sea. Together with the applications in the Cruise System at IMR, the CSRs are used by NMD to keep track of the data collected and handed over for storage.

When handed over to NMD, data from the cruises has been through a quality control process by the instrument personnel and scientists attending the cruise. It is the Cruise Leader’s responsibility to guarantee that the data is quality controlled. NMD is developing specifications of a system for monitoring the data flow of the data collected on cruises at IMR. The new system is known internally as the Dataset Tracker.

NMD also store data from other institutions. The data is made available through the data portal NMDC: www.nmdc.no, which also stands for the Norwegian Marine Data Centre (not to be confused with NMD). NMDC is a national infrastructure for marine data, which is a catalog of marine datasets and their associated metadata across all Norwegian marine regions. NMDC is a consortium of 16 national partners with NMD as the coordinator. The metadata associated with the data uploaded to NMDC is following internationally established standards like ISO19115, ISO19139 and DIF-9. By following these standards, the completeness of the data life cycle is strengthened.

The data stored at NMD is formatted in internationally established formats such as netCDF, ODV, and CSV.

NMD is, as part of IMR, subject to audit trails from the Office of the Auditor General of Norway both on data management and data related matters, see https://www.riksrevisjonen.no/en/Pages/Homepage.aspx.

The data stored at IMR are collected from cruises managed by the IMR or from collaboration partners working in projects with IMR, i.e., Knipovich Polar Research Institute of Marine Fisheries and Oceanography (PINRO) in Russia. Deposits from external depositors are stored manually by NMD staff and identified accordingly.

Reviewer Entry

Reviewer 1

Comments:
Accept at level 3, noting that the data deposit portal planning was also mentioned in the previous application. It will be helpful to see how this has evolved by the next application.

Reviewer 2

Comments:

8. Appraisal

R8. The repository accepts data and metadata based on defined criteria to ensure relevance and understandability for data users.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:
4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:

Response:

IMR data products are made available at https://www.hi.no/en/hi/forskning/research-data-1.
There are some inconsistencies between the Norwegian and English version of the webpage. More data is available at the Norwegian webpage http://www.imr.no/forskning/forskningsdata/datakatalog/.

Some of the datasets are offline and can be ordered through email. IMR data is also being made available at http://nmdc.no.

Uploaded data available at the data portal http://nmdc.no require metadata is following either ISO19139 or GCMD DIF-9. NMD is in a process of changing from DIF to ISO.

An agreed list of recommended formats has been established in the national research infrastructure NMDC. These recommendations can be found here: http://nmdc.no/resources/D3.1-Definition-of-data-formats-and-metadata-structure_V0.11.pdf.

When encountered with data that does not fall within the NMD mission profile, NMD refer the data owners to the relevant institution in Norway.

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:

9. Documented storage procedures

R9. The repository applies documented processes and procedures in managing archival storage of the data.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:
4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:

Response:

NMD undertakes data storage according to documented processes and procedures at IMR. The documents are available in the internal IMR Quality System “Kvalitetsportalen”, and externally at https://www.hi.no/forskning/forskningsgrupper/norsk-marint-datasenter-nmd/kvalitetsdokumentasjon/. Technical information on storage and backup procedures can be found in sections R15 and R16 of this application.

- NMD are following “Procedure for sensitive data at IMR”: https://www.hi.no/forskning/forskningsgrupper/norsk-marint-datasenter-nmd/kvalitetsdokumentasjon/pub/docs/6237.pdf
- NMD data management plan can be found here: https://www.hi.no/forskning/forskningsgrupper/norsk-marint-datasenter-nmd/kvalitetsdokumentasjon/pub/docs/7508.pdf

NMD provide guidance on data delivered to NMD for storage through the following procedures:

- IMR’s data policy: https://www.hi.no/resources/imr/Data-policy-IMR.pdf
- NMD data ingestion policy: https://www.hi.no/forskning/forskningsgrupper/norsk-marint-datasenter-nmd/kvalitetsdokumentasjon/pub/docs/6156.pdf
- Data management procedure for data collected on cruises with research vessels”: https://www.hi.no/forskning/forskningsgrupper/norsk-marint-datasenter-nmd/kvalitetsdokumentasjon/pub/docs/6239.pdf

The original data (Submission Information Package, SIP) is stored on tapes and loaded to the network system. The data (Archival Information package, AIP) is loaded to the staging area on the network system and validated. After the validation process the data (Disseminated Information Package, DIP) is moved to the correct folder path on the network system for longtime storage.

The data managers at NMD are responsible for the final validation of the data.

NMD stores all data relevant to the mission profile. Some organizations we trust will request copy of data, and this might be granted. Some marine data from IMR is exchanged with international projects and repositories. The original copy is still stored at the originator.

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:

10. Preservation plan

R10. The repository assumes responsibility for long-term preservation and manages this function in a planned and documented way.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:
4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:

Response:

As a National Oceanographic Data Centre, it is assumed that the NMD will take care of the data in its custody for an indefinite period of time.

The NMD preservation plan can be found here:
https://www.hi.no/forskning/forskningsgrupper/norsk-marint-datasenter-nmd/kvalitetsdokumentasjon/pub/docs/7507.pdf

The plan deals with all aspects of long-time preservation of data archived at NMD.

It defines the rights of the data originators and NMD in regards of deposits in the repository, and the contract between depositor and repository stating the responsibilities for data deposits.
The preservation plan also considers the risk of format obsolescence and technology change and describes the implemented IMR storage and backup systems and procedures.

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:

11. Data quality

R11. The repository has appropriate expertise to address technical data and metadata quality and ensures that sufficient information is available for end users to make quality-related evaluations.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:
4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:

Response:

Quality control of data acquired on cruises with research vessels from IMR is the responsibility of the cruise leader and performed before the data is delivered to NMD for long-time storage. NMD is responsible for checking the technical status of data, i.e., to determine if the data is readable and complete.

Routines are specific for each data type and collecting agency. For data published at nmdc.no, NMD accepts the data quality provided by the originator and provides no further QC.

NMD sends CTD data to Copernicus Marine Services in in near real time and in delayed mode. For these data, an automatic QC routine (https://archimer.ifremer.fr/doc/00251/36230/) is applied to flag the data. The data quality information is communicated by community standard QC flags, described in the link above.
Delayed mode CTD data undergoes an additional, manual QC and processing routine at NMD, using SAS and DIVA softwares (mentioned in section R15).

The data formats delivered to NMD is well known by the data managers, as they have years of experience and the appropriate expertise to handle these data types.

A landing page for each dataset is provided according to international requirements presenting the metadata in a manner of best practices. The mandatory metadata is helping the researchers to evaluate the quality of the data in the best possible way.

Citations to related work can be made available through the landing page for a dataset. NMD is part of the Norwegian

National Solution for minting DOIs to ensure such links are in place.

The Designated Community can provide feedback on data and metadata to NMD through IMR's internal OTRS5 system (Open Technology Real Services): https://www.otrs.com/otrs-free-help-desk/

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:

12. Workflows

R12. Archiving takes place according to defined workflows from ingest to dissemination.

Compliance Level:

3 – The repository is in the implementation phase

Reviewer Entry

Reviewer 1

Comments:
3 – The repository is in the implementation phase

Reviewer 2

Comments:

Response:

NMD has policies and procedures in place for proper managing of data delivered to NMD for long-time storage.

The ingestion begins with the data policy of IMR: https://www.hi.no/resources/imr/Data-policy-IMR.pdf

Data managers at NMD, who are responsible for finalizing the ingestion of the data to the archival system, are following several procedures:

- Procedure for review of disclosure risk in data:
https://www.hi.no/forskning/forskningsgrupper/norsk-marint-datasenter-nmd/kvalitetsdokumentasjon/pub/docs/6238.pdf

- Procedure for sensitive data at IMR:
https://www.hi.no/forskning/forskningsgrupper/norsk-marint-datasenter-nmd/kvalitetsdokumentasjon/pub/docs/6237.pdf
- Procedure for minting DOIs to data:
https://www.hi.no/forskning/forskningsgrupper/norsk-marint-datasenter-nmd/kvalitetsdokumentasjon/pub/docs/6123.pdf
- Data management procedure for data collected on cruises with research vessels:
https://www.hi.no/forskning/forskningsgrupper/norsk-marint-datasenter-nmd/kvalitetsdokumentasjon/pub/docs/6239.pdf

NMD handle any marine data. The data storage is specified in an agreed folder structure, to ensure structured storage. This is updated in case of new instruments and types of data and listed in the document Folder structure for marine field data at IMR: https://www.hi.no/om_havforskningsinstituttet/rederi/sms_systemet/nb-no/pub/docs/5405.pdf.

IMR internal case-handling system is the OTRS5 system, which gives each submission a unique ID and an automatic email sent to the user. The OTRS5 system is used by the Designated Community to submit a request for data archived at NMD. The system is also used in workflows for transferring cruise data from the research vessels to NMD.

NMD is planning a future data deposit portal mentioned in section R7, which will enable the user to submit data and metadata. The data and metadata will be evaluated against existing guidelines. The user will be asked to attach a user license (like CC BY 4.0) to the deposited data to ensure proper data citation and usage. The user will also be asked if they would like to have a persistent identifier attached to the data. Data outputs handled by NMD personnel are inspected to ensure they are readable and containing the expected data. However, user-initiated outputs are not inspected in any way.

NMD and IT personnel are monitoring the operational status of data systems on a regular basis. Automatic emails and warnings are initiated upon failure in the data systems.

Reviewer Entry

Reviewer 1

Comments:
Accepted at implementation phase.

Reviewer 2

Comments:

13. Data discovery and identification

R13. The repository enables users to discover the data and refer to them in a persistent way through proper citation.

Compliance Level:

3 – The repository is in the implementation phase

Reviewer Entry

Reviewer 1

Comments:
3 – The repository is in the implementation phase

Reviewer 2

Comments:

Response:

Cruise summary reports (CSR) containing an overview of data collected on cruises can be found here http://seadata.bsh.de/Cgi-csr/retrieve_sdn2/start_sdn2.pl.

Each year IMR publishes a report on cruises and data stations which can be found here (in Norwegian only) https://www.hi.no/hi/nettrapporter/fisken-og-havet.

NMD has developed a data catalog for IMR. Some of the data is available online, but the main part of the data is available through requests. The metadata is searchable in the data catalog: http://www.imr.no/forskning/forskningsdata/datakatalog/.

More detailed datasets are made available at NMDC. The data is accessible through metadata search www.nmdc.no. NMD, through NMDC, provides a DIF/ISO compatible searchable metadata catalog. Data is also made available through several international data portals like ICES, World Ocean Database (WOD), SeaDataCloud, Copernicus, Global Biodiversity Information Facility (GBIF) and European Marine Observation and Data Network (EMODnet).

Metadata from IMR available through NMDC can be harvested using technologies like OAI-PMH and IPT (Integrating Publishing Toolkit).

NMD has implemented the process of assigning the persistent identifier Digital Object Identifier (DOI) to datasets. Metadata regarding DOIs are following the DataCite Metadata Schema 4.0. http://schema.datacite.org. Not all data uploaded to NMDC is assigned with a DOI. This depends on if there is a request to have a DOI assigned, or if we find it beneficial. In the future more and more data uploaded to NMDC will be assigned a DOI.

For DOIs we are following the DataCite citation recommendations. See page 8 in DataCite Metadata Schema 4.0: http://schema.datacite.org/meta/kernel-4.0/doc/DataCite-MetadataKernel_v4.0.pdf.

The NMD procedure for minting DOIs to data is available here: https://www.hi.no/forskning/forskningsgrupper/norsk-marint-datasenter-nmd/kvalitetsdokumentasjon/pub/docs/6123.pdf.

Reviewer Entry

Reviewer 1

Comments:
Accepted as being in the implementation phase, with the expectation that a more comprehensive approach to dataset identifiers will emerge in the future as indicated, with related identifiers such as RORs and ORCIDs. The discovery interfaces seem straightforward for basic filtering functionality.
The DOI procedure provides a good description of the working vs completed dataset situation as well as guidance for versioning.

Reviewer 2

Comments:

14. Data reuse

R14. The repository enables reuse of the data over time, ensuring that appropriate metadata are available to support the understanding and use of the data.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:
4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:

Response:

At NMD, the metadata required are following the standards of Dublin Core, ISO, GCMD DIF and EDMED.

At NMD, we continuously adapt to new data streams from new instruments, and we try to standardize data from the data provider side. NMD accepts all documented formats and do not reject any data. Many different data formats are used at NMD. We follow the recommendations by NMDC, which is also used by the Designated Community, listed on page 33: http://nmdc.no/resources/D3.1-Definition-of-data-formats-and-metadata-structure_V0.11.pdf.

The same types of data have been collected over several years, well known by the Designated Community who in many cases also is the producer of the data.

As for special data formats, expert consultancy can solve the challenge. IMR has a great deal of expertise. At NMD, we regularly consults our international marine data respiratory network to ensure best practices and a good understanding of

the specialized data.

Reviewer Entry

Reviewer 1

Comments:
Accept.

Reviewer 2

Comments:

TECHNOLOGY

15. Technical infrastructure

R15. The repository functions on well-supported operating systems and other core infrastructural software and is using hardware and software technologies appropriate to the services it provides to its Designated Community.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:
4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:

Response:

HARDWARE:

NMD is using the infrastructure of IMR for long time storage, backup, and management of data.

All data is stored on the IMR storage system, running IBM Spectrum Scale (GPFS). Data is accessible internally through iSCSI or SMB/NFS protocols.
All data is archived on two sites and has GPFS Replication as an additional layer of data protection.

GPFS Replication provide an active/active, synchronous Disaster Recovery (DR) and is independent of the storage hardware solution between the two sites.

GPFS as a Software-defined storage gives a highly differentiated value:

- Virtually limitless scaling number of files and size of data
- Uses any combination of flash, spinning disk and tape
- Information lifecycle management (ILM) tools automatically move data based on policies to external storage (Tape pool)
- Possibility to add more storage capacity without affecting the application
- Good performance – more than 4 GBps (over 10 Gb iSCSI) - and simultaneous access to a common set of shared data from GPFS cluster nodes

Object storage has recently been implemented at IMR as an alternative solution for FTP protocol to give remote access to local data. Existing data on CES (File server) can be internally accessed by SMB/NFS/POSIX, and the same data after being objectized by system will be remotely available by object interface (Amazon S3 API).
The object storage experience for the end user is as simple as using an FTP client. When credential information is set up on S3 client tools, remote users can access the stored data from anywhere without VPN connection.

Active Archive solution:

The back-end storage for tiering data is the IBM Spectrum Archive EE tape storage, which means GPFS virtually extends the managed file system with the space provided by the Spectrum Archive EE service on LTFS tapes.

Tape library and Drive:

NMD data is located on two sites, and we are using IBM TS4500 tape library on each site. Licenses are activated on all slots on both libraries including 8* IBM TS1150 (Model 3592 EH8) tape drives plus two tape drives for scheduled verification tape tasks on library level. In addition, tape drive failover functionality is included in the library license. Control Path has been configured for two tape drives, so that in case one of the tape drives fail the other one can take the responsibility of the Control Path.

Tape Cartridges:

TS1150 JZ cartridges high performance of up to 360 MB/s and 10TB capacities.

Clustered NFS and SMB:

SMB and NFS clients can connect to any of the protocol nodes and get access to the shares defined. A clustered registry

makes sure that all nodes see the same configuration data. Therefore, clients can connect to any Cluster Export Services (CES) node and see the same data. Moreover, the state of opened files (share modes, open modes, access masks, locks, and so on) is also shared among the CES nodes so that data integrity is maintained. On failures, clients can reconnect to another protocol node and IP addresses are transferred to another protocol node.

Network operation is manned during opening hours and by a support team outside of opening hours. Network operations at IMR is up-and-running close to 100% of the time.

SOFTWARE:

At NMD we use a list of software tools, both for daily operations and other tasks.
An overview describing some of the main software tools used by NMD is listed below. Some are inhouse applications developed at NMD/IMR. Due to safety concerns the source codes are not published, but they are available by request. Some programs are developed as a part of international projects, while others again are commercial and license based, or open source.

Data ingestion tools:
S2D Data Manager - Inhouse software for uploading oceanographic and navigation from cruises
Biotic Editor - Inhouse software for uploading biological data from cruises
Cruise editor - Inhouse software for cruise metadata
EchosounderEditor - Inhouse software for uploading processed acoustic data

Data processing:
SAS Institute - An inhouse developed software VIKKHY (VIsual quality control of hydrographical data), SAS 9.4 https://support.sas.com/software/94/
DIVA/DIVAnd - Application for gridding data to be used in climatologies, developed by ULG/GHER Belgium.
NEMO,Mikado,Octopus -Software tools developed by projects in EMODnet Chemistry/SeaDataNet, further described at https://www.seadatanet.org/Software.
ODV - Ocean Data View, desktop program for visualizing data, developed by AWI Germany, odv.awi.de
SBE - CTD data processing tool, https://www.seabird.com/

Geospatial data:
Arcmap, Arcgis Pro - Licenced GIS software tool, https://www.esri.com/en-us/home
Qgis - Open source GIS tool, https://qgis.org/en/site/
Geoserver - Open source server written in Java, used for publishing geospatial data.
https://www.geosolutionsgroup.com/technologies/geoserver/

Some of the main languages used:
Python, R, C, Fortran, Java, Julia

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:

16. Security

R16. The technical infrastructure of the repository provides for protection of the facility and its data, products, services, and users.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:
4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:

Response:

Replication and Disaster Recovery

Physical protection:
The servers are stored in different buildings, and they are in different domains. Data on tape, including backups, are stored at different sites. To have an additional layer of data protection we are using GPFS Replication, providing an active/active, synchronous Disaster Recovery (DR) and is independent of the storage hardware solution between the two sites.
This means that applications can run simultaneously at both sites (as opposed to the active site/backup site). This allows either of the sites to experience a disaster that leads to a complete site loss (or a loss of the sites availability) without affecting operations on the second site: IO processing will continue after the disaster, after a short pause to handle node failures, without any downtime or loss of data.

Spectrum Scale Immutability:

After storing cruise data on GPFS storage, the files are set as Immutable. These files are write-once-read-many protected

(WORM). GPFS attributes can only be set by designated IT personnel. A file with the attribute “immutable” set to “yes” cannot be changed, renamed, or deleted.

Migration of files:

IBM Spectrum Archive EE enables the creation of a replica of each Spectrum Scale file during the migration process. The purpose of the replica function is to enable the creation of multiple LTFS copies of each GPFS file during the migration, which can be used for disaster recovery, also across two tape libraries at two different locations.

Backup system:

IBM Spectrum Protect (TSM) is used to make backups of the data. To have a redundant copy, the backup job is running over two tape libraries, where each file is backed up by two identical versions copied to two different tape locations. Backups are done daily. The system generates daily reports and will alert administrators of any errors or warnings.

Today, NMD has more than 5 Petabyte data, a number that is expected to increase rapidly in the future.

The IMR IT department is currently in the process of ISMS (Information Security Management System) certification, based on the ISO27001 standard. The documentation required for the certification is finished and is currently being revised internally at IMR.

For the NMD CTS application, the most relevant part of two of these documents has been translated to English and are included below: “Control of access to systems and applications” and “Roles and responsibilities within NMD (and directors) in relation to information security”.

Control of access to systems and applications:
Access to information and system functions in applications shall be restricted in accordance with the access control policy. Access restrictions should be based on individual requirements for business applications and be in compliance with the defined access control policy. The following should be considered to support access restriction requirements:

- Control over which data a particular user should have access to
- Control over the access rights of users, for example to read, write, delete and drive
- Control over access rights to other applications
- Limitation of the information in output data
- Acquisition of physical or logical access controls to isolate sensitive applications, application data or systems

Secure login procedures:

Where sub-policies for access control are required, access to systems and applications should be controlled through a secure login procedure.

The procedure for logging on to a system or application should be designed to reduce the possibility of unauthorized access to a minimum. The login procedure should therefore provide as little information as possible about the system or application, to avoid helping an unauthorized user unnecessarily. When logging in, the system or application must:

a) display a general warning that only authorized users should have access
b) do not display help messages during the login procedure that would help an unauthorized user
c) log failed and successful attempts with date and time
d) do not display passwords entered
e) do not transmit passwords in clear text over a network

Password management system:
Password management systems is interactive and ensures the quality of passwords. The central password management system:

• Allow users to select and change their own passwords
• Check that the passwords are of high quality
• Store passwords in a protected form
• Require two-factor authentication when used outside the corporate network
• Keep a record of previously used passwords and prevent passwords from being used again
• Do not display on-screen passwords when entering
• Store and transfer passwords in protected form

The internal document “Roller og ansvar innen informasjonssikkerhet (KS.ISMS-05)” lists all HI Digital employees roles related to information security. The security roles relevant for NMD is translated and listed below:

FUNCTION/ROLE - RESPONSIBILITY, AUTHORITY, AND TASKS

Director of IMR - Supreme decision-making authority in the business.
Ensure that the business complies with regulations and has a management system for ISMS.

Department director - Responsible for ensuring compliance with information security requirements in the department.
Pass on security-related information.
Department director HI Digital has an extended responsibility for information security through the IT section.

Head of department, NMD - Responsible for making sure the information security is maintained within the given program/section/group.
Pass on security-related information to their employees.
Give guidance to their employees and ensure that their behavior helps ensure safety.
Handle situations in the event of a breach of information security together with relevant personnel, depending on the severity of the situation.

Data steward - The role is designated to NMD. It includes gathering, quality control and storing all marine environmental data and
fish data, and to make these available for research.
Operational data management.
Operational research data.
Continuous data/time series.
Metadata (cruise information).
Software development.
Producing data products.
Reference data (metadata).
Define national and international standards and guidelines.
Contribute to international networks and research projects.
Contribute in «Norge digitalt» - a national collaboration on public geospatial data solutions.
Data storage.
Accommodating, quality controlling and processing of research data.
Ingesting data in the IMR research data bases TINDOR og Sea2Data.
Data delivery.
Function as a data center.

Data Protection Official/Officer - Inform and advice of the data stewards of the company’s duties in regards of handling personal. information.
Control that the privacy regulations are maintained.
Advice on data protection impact assessment (DPIA).
Cooperation with the Norwegian Data Protection Authority as a point of contact.
Prioritized effort where the privacy risk is higher.
Contribute to an overview of treatments in the company.
Gather information to identify treatment activities according to the instructions for Data Protection Official/Officer.

GDPR tasks: Process manager - The person responsible to the person we store / process information about.
Determines the purpose of the processing of personal data and the means to be used.

GDPR tasks: Data processor - Assignment from the data controller in accordance with the data processor agreement. Store and / or process personal data on behalf of the data controller.

Responsibility and authority for information security follow the ordinary line responsibility. Organizing the safety procedures is part of the company's ordinary organization.

Everyone holds a responsibility to ensure information security in connection with their own work tasks and to ensure that safety precautions are implemented.
The manager has a particular responsibility to ensure that work processes, tasks and services associated with his or her department are carried out in accordance with relevant requirements and established routines in the IMR quality system, including the information security system.

The information security instruction controls the user's behavior in regards of handling information security, and it is signed by the individual employee.

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:

APPLICANT FEEDBACK

Comments/feedback

These Requirements are not seen as final, and we value your input to improve the CoreTrustSeal certification procedure. Any comments on the quality of the Requirements, their relevance to your organization, or any other contribution, will be considered as part of future iterations.

Response:

Reviewer Entry

Reviewer 1

Comments:
There is a good demonstration of progress from the previous application, including plans for preservation and data management, and an increasingly comprehensive approach to dataset citations/DOIs.

Reviewer 2

Comments: