Assessment Information

[CoreTrustSeal Requirements 2017–2019](#)

Repository:	Norwegian Marine Data Centre
Website:	https://www.hi.no/en/hi/forskning/research-groups-1/the-norwegian-marine-data-centre-nmd
Certification Date:	04 October 2018

This repository is owned by: **Institute of Marine Research**

Norwegian Marine Data Centre

Notes Before Completing the Application

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

True

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

Comments

The Norwegian Marine Data Centre (hereafter named NMD) at the Institute of Marine Research (herafter named IMR) was established as a national data center dedicated to the professional processing and long-term storage of marine environmental and fisheries data and production of data products. NMD aims to be the national hub for all marine data, both data collected by the Institute of Marine Research and by others. NMD has been a "National Oceanographic Data Centre", NODC since 1971. NMD is the coordinator of the national research infrastructure NMDC which is 16 Norwegian partners working together to make marine data available for research.

See the homepage of the Norwegian Marine Data Centre
https://www.hi.no/en/hi/forskning/research-groups-1/the-norwegian-marine-data-centre-nmd

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 like management of fish stock, salmon lice development, causes of increase in ocean temperature, etc.
In addition to the research community NMD also serves data to the public.

Accept

Level of Curation Performed. Select all relevant types from:

A. Content distributed as deposited, B. Basic curation – e.g. brief checking; addition of basic metadata or documentation, C. Enhanced curation – e.g. conversion to new formats; enhancement of documentation, D. Data-level curation – as in C above; but with additional editing of deposited data for accuracy

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

Comments

The NMD stores many different data types. Many of these data types are stored as they are collected, others undergo a more extensive curation process.

All original data is stored at NMD and changes are only made on copies of original data.
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.
The data in NMD is mainly IMR data, but also other institutions and universities store data in the repository.

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

Outsource Partners. If applicable, please list them.

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

Other Relevant Information.

The repository has recently implemented and created procedures for data publication using Digital Object Identifiers (DOI).

The NMD is national data center for marine environment 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 TAC Thematic Assembly Center 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:
Accept

ORGANIZATIONAL INFRASTRUCTURE

I. 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 1

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

Reviewer 2

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

Response:

The most important tasks of the Norwegian Marine Data Centre at the IMR are

- to collect, quality-assure and store aqua cultural, marine environmental and fisheries data

- to make these data available for research

(Please see this link with a description of NMD for reference

https://www.hi.no/en/hi/forskning/research-groups-1/the-norwegian-marine-data-centre-nmd)

The Norwegian Marine Data Centre was established as a national data centre for handling of marine environmental- and fisheries data, and to present data products. 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/forskning/forskningsdata/en.

Each year the Ministry of Trade, Industry and Fisheries sends an award letter to the Institute of Marine Research where amongst others, the purpose of the Norwegian Marine Data Centre is stated. The document is only available in Norwegian, but on page three the purpose is stated (Translated from Norwegian): "The main tasks of the Institute of Marine Research is to make data and research results known and available for management, other research institutions, business partners and the society." Please see page 3 in this link https://www.regjeringen.no/contentassets/76a9cd4d079 d415dbd6d7faa0ebae0e9/statsbudsjettet-2018-tildelingsbrev-til-havforsknin-l2185137.docx

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

II. 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 1

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

Reviewer 2

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

Response:

Data collected under the auspices of IMR are the property of the institute. IMR data is managed by NMD. This means NMD has the rights to copy, transform, store the data and provide access to them. 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 are freely available under the conditions of the Norwegian License for Open Government Data (NLOD) and the license Creative Commons(CC BY 4.0): https://data.norge.no/nlod/en/1.0 and (link to CC BY 4.0: https://creativecommons.org/licenses/by/4.0/legalcode)
Furthermore, the source must be acknowledged.

The principles described above and NMD's responsibility towards data collected by IMR is described in the Data Policy document of IMR: 'IMR data is managed by the Norwegian Marine Data Centre (NMD)'
https://www.hi.no/filarkiv/2016/09/hi-datapolicy-revised2016-final-eng.pdf/en
and in the annual award letter from the Ministry of Trade, Industry, and Fishery which states that one of IMR's main responsibilities is to make collected and research data available to the public:
https://www.regjeringen.no/contentassets/76a9cd4d079d415dbd6d7faa0ebae0e9/statsbudsjettet-2018-tildelingsbrev-til-havforsknin-l2185137.docx

If NMD is managing data that is not the property of IMR, the data owner will be responsible for providing the correct license. If the data owner does not specify a license, the above licenses will apply.

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

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

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 this organizational chart "Organisasjonen" (only available in Norwegian. The NMD "Norsk Marint Datasenter" is listed as an individual section).
http://www.imr.no/om_havforskningsinstituttet/organisasjonen/organisasjonskart/nb-no.
The NMD is also acting as National Oceanographic Data Center (NODC) of Norway since 1971. The head of NMD, Helge Sagen, is member of the body IODE National Coordinator for Oceanographic Data Management. (IODE is an abbreviation of International Oceanographic Data and Information Exchange). Please see this link
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. A yearly copy of the data at NMD is deposited at the National Archives of Norway
https://www.arkivverket.no/en/about-us/the-national-archives-of-norway.

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

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

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/filarkiv/2016/09/hi-datapolicy-revised2016-final-eng.pdf/en.

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 we are following our procedure to review disclosure risk in data to determine what to do with the data:
ftp://ftp.imr.no/nmd/documents/Procedure%20for%20review%20of%20disclosure%20risk%20in%20data.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-and-decisions/
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 has 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. Please see our procedure "Sensitive data at IMR" here:
ftp://ftp.imr.no/nmd/documents/Procedure%20for%20sensitive%20data%20at%20IMR.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: https://lovdata.no/dokument/NLE/lov/1967-02-10

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept.

In future renewals please document further the impact on workflows of 'NMD is managing business sensitive data, that might affect stock exchange.'

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

Response:

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

NMD has its main funding from its host institution IMR, consisting of more than 20 regular positions, IT developers, data managers and GIS experts. NMD gets its IT resources through the IMR IT personnel and IT hardware resources. IMR has more than 115 years of history in marine sciences and NMD has excited 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'ies and are currently partner in four EU funded projects in marine data management.

IMR has its own knowledge portal with different 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.

NMD has three main fields of competence: IT development, Data management and Geographical Information System products distributed over approximately 20 man years.
IMR supports administrative services and IT services for NMD. 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, Working Group on Marine Data Management (WGMDM), Working Group on Data and Information Management (WGDIM) and Data and Information Group (DIG).
Currently NMD is represented in Arctic Data Committee http://arcticdc.org/

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

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

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

Reviewer 2

Comments:
Accept

DIGITAL OBJECT MANAGEMENT

VII. 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 1

Comments:
3 – The repository is in the implementation phase

Reviewer 2

Comments:
3 – The repository is in the implementation phase

Response:

The NMD has a data ingestion policy with requirements that must be fulfilled when data is delivered to NMD for storage: ftp://ftp.imr.no/nmd/documents/NMD%20data%20ingestion%20policy.pdf. Data stored at NMD must also be compliant with the data policy of IMR.

At NMD we wish to develop a data deposit portal, where it is possible for users to upload their data. The data portal will have requirements for the content of the data (i.e. any disclosure data is not permitted) and for mandatory metadata. It will be the responsibility of the data managers at NMD to ensure the metadata is complete before finalizing the ingestion into the data portal. The NMD data portal will be developed in the near future.

Most of the data stored at NMD is collected on cruises with research vessels from the Institute of Marine Research. Each cruise goes through an application process, where it is stated which instruments will be used and thereby what data will be collected, before it is approved and finally executed. All cruises are registered in the Cruise System at IMR (internal system for application, approval, cruise planning and crew planning), where the applications are filled in. During this process it must be stated which data types are planned to be collected during the cruise.

After the cruise has ended the cruise leader must fill in a Cruise Summary Report (CSR), which is uploaded to seadatanet.org. Cruise Summary Reports are the usual means for reporting on cruises or field experiments at sea. https://www.seadatanet.org/Metadata/CSR-Cruises. Together with the applications in the Cruise System at IMR the CSRs are used by NMD to have an overview of the data collected and handed over for storage.

Once the data from cruises is handed over to NMD the data has been through a quality control process by the instrument personnel and scientists attending the cruise, but it is the responsibility of the cruise leader to guarantee the data is quality controlled.

NMD is in the initial stage of describing the specifications of a new system for monitoring the data flow of the data collected on cruises at IMR. The new system currently known internally as the Dataset Tracker will be developed late 2018/early 2019.

The NMD also store data from others. The data is made available through the data portal NMDC, which also stands for the Norwegian Marine Data Centre (not to be confused to 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 IMR as the coordinator. The metadata associated with the data uploaded to NMDC is following internationally established standards like ISO19115, ISO19139 and DIF-9. https://gcmd.gsfc.nasa.gov/add/difguide/index.html. Following these standards ensure the completeness to a certain extent.

We have just implemented the process of creating Digital Object Identifiers (DOIs) for data uploaded to the data portal NMDC. Not all data uploaded to NMDC is assigned with a DOI. This depends 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. With regards to the version control of DOIs, the functionality has not yet been implemented on the Landing Pages of NMDC. This will happen within 2018. The procedure by NMD to assign DOIs is available here:

ftp://ftp.imr.no/nmd/documents/Procedure%20for%20minting%20DOIs%20to%20data.pdf

NMD is responsible for loading data to databases and ensuring that this does not corrupt the data in anyway. 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 during 2018, but the specifications have not yet been defined.

The data stored at NMD is formatted in internationally established formats like netCDF, ODV, ASCII, etc.

IMR/NMD is subject to audit trails from the Office of the Auditor General of Norway both on data management and data related matters. 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). Deposits from external depositors are stored manually by NMD staff and identified accordingly.

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:

Accept

VIII. Appraisal

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

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:
3 – The repository is in the implementation phase

Response:

IMR data products are made available at https://www.hi.no/forskning/forskningsdata/en. There are some inconsistencies between the Norwegian and English webpages. More data is available at the Norwegian webpage https://www.hi.no/forskning/forskningsdata/nb-no.

In 2004 a mapping of the IMR data were conducted resulting in an IMR catalogue, this data catalogue contains approximately 400 datasets mostly offline, but the data can be ordered through emails http://www.imr.no/forskning/forskningsdata/datakatalog/.
The future IMR data will be made available at http://nmdc.no/. NMDC is the national infrastructure of marine data coordinated by NMD. Unfortunately, in English NMDC and NMD are named the same, which is the Norwegian Marine Data Centre. NMDC is a cooperation between 16 Norwegian research institutions and is coordinated by NMD http://nmdc.no/en/projects/nmdc/om-nmdc/partners.

Today data uploaded and available at the data portal http://nmdc.no/ require metadata following the GCMD DIF https://gcmd.gsfc.nasa.gov/add/difguide/index.html standard. This also applies for data assigned a DOI and is stated in our procedure for DOIs ftp://ftp.imr.no/nmd/documents/Procedure%20for%20minting%20DOIs%20to%20data.pdf.

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

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept as in the implementation phase.

By the time the repository renews, we would expect further detail on how metadata/documentation are reviewed over time to ensure continued 'understandability' by the designated community.

IX. Documented storage procedures

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

Compliance Level:

3 – The repository is in the implementation phase

Reviewer 1

Comments:
3 – The repository is in the implementation phase

Reviewer 2

Comments:
3 – The repository is in the implementation phase

Response:

The NMD undertakes data storage according to documented processes and procedures at IMR. The documents are available in our internal application "Kvalitetssystemet" (quality system) and online. The procedures NMD are following are "Procedure for sensitive data at IMR" and "Data management procedure for data collected on cruises with research vessels":
ftp://ftp.imr.no/nmd/documents/Data_management_procedure_for_data_collected_on_cruises_with_research_vessels.pdf
and ftp://ftp.imr.no/nmd/documents/Procedure%20for%20sensitive%20data%20at%20IMR.pdf.

The NMD provide guidance on data delivered to NMD for storage through IMR's data policy, the NMD data ingestion policy and the "Data management procedure for data collected on cruises with research vessels". The original data (Submission Information Package (SIP)) is stored on tapes before loaded to the network system at IMR. Then the data

(Archival Information package (AIP)) is loaded to the staging area on the network system before 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.

X. Preservation plan

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

Compliance Level:

3 – The repository is in the implementation phase

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. Though it is obvious that NMD must take care of the data from IMR, also in the long-term perspective, we do not have a documented preservation plan, but we are working on it and will publish it within this year. The preservation plan will i.e. consider the risk of technology change, backups and file format obsolescence, where we plan to implement steps to review the data collection to check and update the file formats, if necessary. Currently the preservation plan is in work, but it has the following outline:

• A description of data collections

• A description of the rights of the NMD regarding deposits in the repository

• Contract between depositor and repository stating the responsibilities for data deposits

• Safe storage of the data deposits

• Measures in place to handle technology change and file format obsolescence

• A description of implemented backup procedures

• Maintaining the preservation plan

Currently we migrate data formats when required, but we do not have a regular migration plan in place, which we want to obtain with the preservation plan.

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept.

Preservation plan should be available within 1 year.

XI. 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:

3 – The repository is in the implementation phase

Reviewer Entry

Reviewer 1

Comments:
3 – The repository is in the implementation phase

Reviewer 2

Comments:
3 – The repository is in the implementation phase

Response:

Regarding data coming from cruises with research vessels from IMR it is the responsibility of the cruise leader that the data has been through a quality control process before it is delivered to NMD for long-time storage. The NMD is also responsible for the quality control of the data from cruises, but this is merely from a technical perspective i.e. to determine if the data is readably, etc.

The data delivered to NMD is well known by the data managers and they have the appropriate expertise to handle these data types as it has been done by NMD for many years.

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. Currently NMD is doing this on a manual basis, but future developments in data publication will do this in a semi-automatic way. 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 our internal OTRS5 system (Open Technology Real Services). https://www.otrs.com/otrs-free-help-desk/

XII. Workflows

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

Compliance Level:

3 – The repository is in the implementation phase

3 – The repository is in the implementation phase

Response:

The 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/filarkiv/2016/09/hi-datapolicy-revised2016-final-eng.pdf/en and the data ingestion policy of NMD
ftp://ftp.imr.no/nmd/documents/NMD%20data%20ingestion%20policy.pdf. Data managers at NMD, who are responsible
for finalizing the ingestion of the data to the archival system at NMD, are following several procedures:
Procedure for review of disclosure risk in data
ftp://ftp.imr.no/nmd/documents/Procedure%20for%20review%20of%20disclosure%20risk%20in%20data.pdf
Procedure for sensitive data at IMR
ftp://ftp.imr.no/nmd/documents/Procedure%20for%20sensitive%20data%20at%20IMR.pdf
Procedure for minting DOIs to data
ftp://ftp.imr.no/nmd/documents/Procedure%20for%20minting%20DOIs%20to%20data.pdf
Data management procedure for data collected on cruises with research vessels
ftp://ftp.imr.no/nmd/documents/Data_management_procedure_for_data_collected_on_cruises_with_research_vessels.pdf

Our internal case-handling system is the OTRS5 system, which gives each submission a unique ID and an automatic
email is send 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 the cruise data from the research vessels to the NMD.
NMD is planning a future data deposit portal (as mentioned in earlier requirements), which will enable the user to submit
metadata and data. 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 that are 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 operational
status of data systems on a regular basis. Automatic emails and warnings will be initiated upon failure in the data systems.

Reviewer Entry
Reviewer 1

Comments:
Accept
Reviewer 2

Comments:
Accept

XIII. 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

Response:

An overview of all data collected on cruises can be found on http://seadata.bsh.de/csr/retrieve/sdn2_index.html by searching for a specific cruise number from IMR.

Each year IMR publishes the report Report on cruises and data stations which can be found here

https://www.hi.no/publikasjoner/andre_publikasjoner/fisken_og_havet/nb-no

NMD has developed a data catalogue for IMR. Some of the data is available online, but the main part of the data is available through requests. The metadata is searchable at data catalogue

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 catalogue. 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).

Data from IMR available through NMDC can be harvested using the technologies 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 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. With regards to the version control of DOIs, the functionality has not yet been implemented on the Landing Pages of NMDC. This will happen within 2018. 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

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept.

Until all items have a DOI a Compliance Level of 3 is appropriate.

XIV. 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:

3 – The repository is in the implementation phase

Reviewer 1

Comments:
3 – The repository is in the implementation phase

Reviewer 2

Comments:
3 – The repository is in the implementation phase

Response:

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

Many different data formats are used at the NMD, but we follow the recommendations by NMDC, which is also used by the Designated Community. Please see recommendations listed here at chapter 4 "Data formats", 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, which is well known by the Designated Community, who in many cases also is the producer of the data. In case of special data formats expert consultancy can solve the problem.

IMR has a great deal of expertise and NMD can consult with these experts to ensure good understanding of the specialized data.

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept.

Until a preservation plan covering the approach to ensure data and metadata remain usable is in place, a Compliance Level of 3 is appropriate here.

TECHNOLOGY

XV. 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:

3 – The repository is in the implementation phase

Reviewer 1

Comments:
3 – The repository is in the implementation phase

Reviewer 2

Comments:
3 – The repository is in the implementation phase

Response:

Storage solution

All data is stored on our storage system, where we are running IBM Spectrum Scale (GPFS). Data is accessible internally

through iSCSI or SMB/NFS protocols.

For remote access we plan to implement Openstack Swift where we will utilize HTTP instead of SMB/NFS. Swift clients (users or applications) will first obtain a token from the authorization service (Keystone) after providing their credential information. The token is included in all requests to Swift, and the Swift proxy service verifies the token by comparing it with cached tokens or by contacting the authorization service. Back-end authenticator service for Keystone is Active Directory.

Authentication url and Swift storage url will be available through HTTPS for accessing the data from an external network. Openstack Swift will be implemented during 2018.

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

- Virtually limitless scaling number of files and size of data.

- Use any combination of flash, spinning disk and tape.

- Information lifecycle management (ILM) tools automatically move data based on policies to external storage (Tape pool).

- 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.

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

Our 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 our 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 almost 100% of the time.

XVI. Security

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

Compliance Level:

3 – The repository is in the implementation phase

Response:

Replication and Disaster Recovery

Our data is archived on two sites and to have an additional layer of data protection, we are using GPFS Replication. GPFS Replication provide 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 we set them as Immutable files. Those files are write-once-read-many protected (WORM). These attributes can be set by the user who has permissions to set GPFS attributes using the command "mmchattr". 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

We are using IBM Spectrum Protect (TSM) to make backups of our data. In addition, to have redundant copy of backup data, our 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.

"mmbackup" command in GPFS provides us much faster file system scanning times, which allows TSM backups to scale to many more objects compared to TSM progressive incremental (so far we have 120 million files and it will increase rapidly in future).

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept.

By the time the repository renews, there should be public evidence to support this Requirement. This should include specific measures for handling data designated as sensitive (ftp://ftp.imr.no/nmd/documents/Procedure%20for%20sensitive%20data%20at%20IMR.pdf)

APPLICANT FEEDBACK

Comments/feedback

These requirements are not seen as final, and we value your input to improve the core certification procedure. To this end, please leave any comments you wish to make on both the quality of the Catalogue and its relevance to your organization, as well as any other related thoughts.

Response:

Reviewer Entry

Reviewer 1

Comments:

Reviewer 2

Comments: