Content

1 Introduction 1
1.1 Background 1
1.2 Scope of the NMD Preservation Plan 1
2 NMD requirements 2
3 Rights of NMD regarding deposits in the repository 2
4 Transfer of assets 2
5 Contract between depositor and repository stating the responsibilities for data deposits .3
6 Measures in place to handle technology change and file format obsolescence....3
7 Storage and backup....4
8 Maintaining the preservation plan 5

1 Introduction

1.1 Background

The Norwegian Marine Data centre (NMD) at the Institute of Marine Research (IMR) is a national centre for professional processing and long-term storage of marine environmental and fisheries data from Norwegian waters. NMD maintains the largest collection of marine environmental and fisheries data in Norway.

NMD aims to provide access to all data objects archived for long-term periods and promotes re-use of data for research. With this in mind, the NMD is responsible for ensuring the authenticity and integrity of the archived data.

NMD data management, and thus the preservation policies, is based on the FAIR Principles for Data Management:

- Findable
- Accessible
- Interoperable
- Reusable

1.2 Scope of the NMD Preservation Plan

Good data management implies several activities which ensure that data are discoverable and accessible. This plan deals with all aspects of preservation when it comes to data archived at NMD. The plan does not consider internal documents, procedures and/or workflows.

The scope of the preservation plan is to point out and document the NMD procedures for long time archiving of data objects. It takes into account crucial aspects of preservation and is valid for all data objects archived at NMD.

2 NMD requirements

- The data that the Archive acquires are accompanied by adequate documentation to enable their use and re-use for analytical and research purposes
- The data is checked and validated according to the NMD Ingestion Policy
- The data is catalogued in internal metadata systems or with appropriate metadata standards when uploaded to NMDC.no
- The data, documentation, metadata and other representation information are preserved for the long term
- The authenticity, integrity and reliability of the data is preserved for the future

3 Rights of NMD regarding deposits in the repository

When data is deposited at NMD, all the rights to the data still belongs to the institution or depositor. This is also valid for data collected by the IMR. NMD does not take over the rights to the data.
If a depositor and/or institution provides the data, NMD is not liable for the contents of the data or for the documentation associated with the data.

Regarding data from IMR, NMD is responsible for archiving, documenting and maintaining the data in whatever means is necessary, but IMR still has all the rights to the data.

Regarding data deposited at NMD, NMD can modify the format and/or functionality of the data so it is ready for distribution or re-use. Also, if the data is to be provided to third parties, NMD can modify the format to be fit for purpose.

4 Transfer of assets

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. NMD coordinates publishing and sharing of data from IMR to international and national repositories such as SeaDataNet, CMEMS, GBIF, ICES, EMODNET and Geonorge.

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

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

Data archived at NMD usually has three originators:

- IMR

- Depositors, i.e.: Researcher who wants to archive her/his research data

- Third-parties, i.e.: Research institution other than IMR/university/company/research infrastructure

Data collected by IMR and archived at NMD follows implemented work procedures, documented in IMR quality system *Kvalitetsportalen*. Most of the data from IMR is licensed under Creative Commons Attribution 4.0 International (CC BY 4.0) and Norwegian License for Open Government Data, NLOD 2.0. NLOD 2.0 is compatible with CC BY 4.0.

A depositor or third-party must contact NMD at the general email address datahjelp@hi.no to deposit data. A data manager at NMD will assist the data owner to provide the data in a proper format, with associated metadata and documentation. This routine will change once a planned Data Deposit Portal at NMD is implemented and available.

Again, the data will be available under the licenses CC BY 4.0 and NLOD 2.0, if nothing else has been agreed upon with the depositor/third-party.

Regarding the rights to the data this is stated above, but in short NMD will not have the rights to the data, it will still belong to the originator. NMD's role, hence the responsibility of NMD is only to archive, document and maintain the data.

The archiving period of the data for all three originators is the same: NMD will archive the data for as long as possible, which is to be regarded as infinite.

6 Measures in place to handle technology change and file format obsolescence.

The data objects archived at NMD consist of many different types such as spreadsheets, texts, pictures, video, geographical information and more instrument specific data like echosounder and sonar data. Also, in-house developed data formats is used, i.e.: fishery catchments.

NMD has a list of preferred formats, documented here https://nmdc.no/resources/D3.1-Definition-of-data-formats-and-metadata-structure_V0.11.pdf. These formats has been selected by the community from the national infrastructure NMDC, where NMD is the coordinator. The formats has a high chance of remaining usable in the future, a measure taken to mitigate the risk of the data becoming obsolete in the future.

Depositors are encouraged to deliver data in the preferred formats corresponding to the type of data. In this way, the NMD offers the best long-term archiving of data objects in terms of usability, accessibility and interoperability. However, NMD does not reject any data that falls within the mission profile, i.e., marine data.

Hardware and software changes over time. Although software is not archived at NMD, data objects is archived with information on the applied software with brand, version and configuration

parameters. This information is valuable in the future if software has changed and the data objects is to be accessed and read.

Regarding the change of hardware over time, the IT section at IMR is responsible for maintaining the functionality of these and/or change to new hardware when needed. Procedures are in place to avoid loosing any data due to hardware failure.

7 Storage and backup

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.

The 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 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.

All data have an additional layer of data protection; GPFS Replication. GPFS Replication provide an active/active, synchronous Disaster Recovery (DR) and is independent of the storage hardware solution between the two sites.

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.

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.

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.

After storing cruise data on GPFS storage, the files are set as Immutable. These 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.

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.

In 2021, NMD stores more than 5 Petabyte data, a number that is expected to increase rapidly in the future.

8 Maintaining the preservation plan

Revision of the Preservation Plan will take place biannually and will be initiated by the data managers at NMD. The table below shows the planned scedule to keep up with preservation of the data objects archived at NMD.

Process	**Frequency**	**Responsible**
Monitor hardware failure	Daily	IT section
Review list of preferred formats	Twice a year	NMD
Any legal or regulatory changes	Biannually	Head of NMD
Revise the NMD Preservation Plan	Biannually	NMD

Kryssreferanser

Eksterne referanser