D7.1 – Identification of datasets

Year 3

Author: Ann Kristin Østrem
Contributing partners: Institute of Marine Research

Date: 2015-09-03

Contents

1 Introduction

The purpose of Deliverable 7.1 is to identify marine datasets from each partner to be made available through the NMDC infrastructure. This report is a status update of D7.1 for year 3 of the NMDC project.

2 Results

2.1   Metadata harvesting

In 2014 daily harvesting of metadata was set up for five of the partners:

- Norwegian Meteorological Institute

- Nansen Environmental and Remote Sensing Center

- Norwegian Polar Institute

- Norwegian Institute for Water Research (including Akvaplan-niva)

- Institute of Marine Research (pre-project metadatabase)

The metadata is now being harvested through the new architechture for the NMDC project. This was done when we found that the old system, geonetwork, did not meet the future requirements of the project.

IMR and other partners not in the list above have their metadata in the pre-project metadatabase. Only datasets that contains one or more URLs (links to data, maps, reports etc.) are harvested from the metadatabase.

Metadata harvesting will probably be set up for some of the other partners later.

2.2   Dataset inventory

In November 2014 an inventory was prepared with datasets that will be made available in NMDC from each partner, including delivery dates for most datasets. The inventory has now been updated by most of the partners, with new or altered datasets, new delivery dates for delayed data, and links to available datasets. The inventory is included as an attachment to this report.

Most partners will make data available in their own infrastructure. Datasets from the rest of the partners are available at ftp://ftp.nmdc.no/nmdc/ .

2.3   Pre-project Metadatabase

Metadata from IMR and several other partners is harvested from the metadatabase from the pre-project (http://talos.nodc.no/infrastruktur/).

Some new datasets have been added to the database this year, many datasets are updated and links to data are added.

There are a number of test datasets in the metadatabase, and duplicate datasets due to problems with the registration, that should be deleted.

3 Summary and conclusions

Daily harvesting of metadata for five of the partners has been moved from geonetwork to the new NMDC infrastructure.

A list of datasets with links to available data has been prepared, and is attached to this report.

Test and duplicate datasets has to be removed from the pre-project metadatabase.