Norwegian Marine Data Centre

Arnfinn Morvik, Institute of Marine Research (Norway), arnfinn.morvik@imr.no
Terry Hannant, Institute of Marine Research (Norway), terry.hannant@imr.no
Armin Pobitzer, Christian Michelsen Research (Norway), armin.pobitzer@cmr.no
Helge Sagen, Institute of Marine Research (Norway), helge.sagen@imr.no

Seamless access to marine data.

The main objective of NMDC is to serve the marine science community with seamless access to documented marine datasets covering waters of Norwegian interests. In order to fullfill this objective, a national infrastructure for marine datasets has been established.

NMDC is a distributed infrastructure for data with a central node holding a centralized metadata catalogue. The central node harvests metadata from the partners in the project and a web based data cataloge viewer gives access to the metadata.

The project has 16 national partners and is coordinated by IMR.

System design

Figure 1 shows the central node and the various data providers.

The central node harvests metadata from the data providers. The data providers are responsible for setting up repositories that expose structured metadata via the protocol OAI-PMH. The central node requests to harvest metadata once every day.

The central node accepts metadata in either ISO 19139 or GCMD DIF format.

After harvesting, the metadata are validated using standard xsd schema. The data providers are notified if their metadata fails to pass the validation.

The last part of the harvesting process is to transform the metadata to an internal NMDC metadata format.
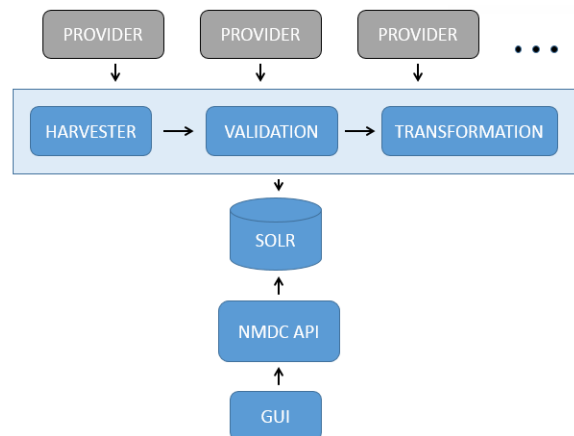

Figure 1

The transformation serves two purposes: All metadata are stored in the same format, and the format is designed to be well suited for geographical search.

Harvesting, validation, and transformation are standalone processes using AciveMQ as messaging server. Programs are written in Java and the JMS API are used to send and receive messages.

The central node uses Solr as its search platform. After transformation, the metadata are imported into Solr and the data are indexed due to a set of search criteria's defined by the central node. The current setup of the central node runs a Solr v5 installation. There is no immediate plan to move towards v6.

The NMDC API is a REST API available to anyone that wants to build applications on top of the central node. Solr is not directly accessible from the outside, and all searching in the indexed metadata are performed by calling methods through the NMDC API.

Solr can return various different formats, but JSON has been chosen as the only format the NMDC API will return.

Available methods are: getFacets, getMetadataDetails, getData, and search
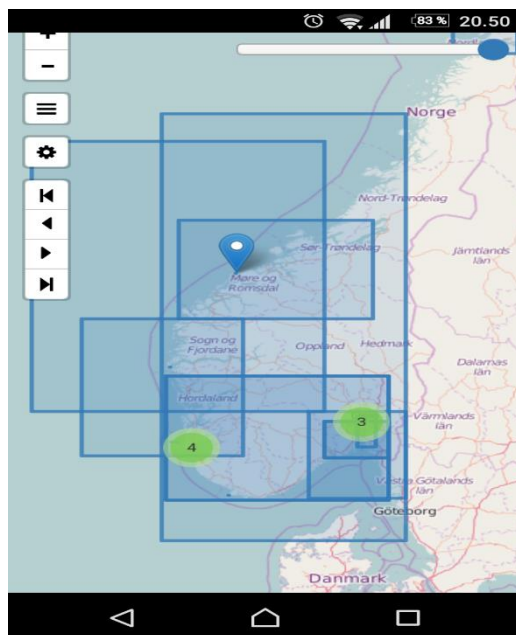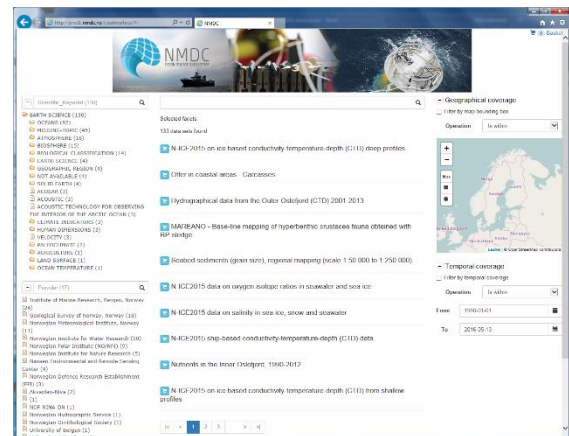
Applications

There are currently 2 applications using the NMDC API:

- The data catalogue as a web application on http://prod1.nmdc.no/UserInterface/
- An experimental map-based search interface available at Google Play: NMDC Mobile Portal

With the web application, the user can search for metadata on scientific keywords, data provider, free text, geographical coverage and temporal coverage. All criteria's can be combined in order to narrow the search.

It is possible to view the metadata for each dataset in more detail, and download the relevant dataset following the URL's given in the metadata files.

It will also be possible to download all the data from a selected group of metadata without using the individual download links from each dataset.

The map-based search uses the same NMDC API as the web application. The outline of all the available datasets are visible in a map as a rectangle for datasets with a geographical extent, or as a pin for time series from a point location. The user can select individual datasets to get more detailed information about the selected data.

The number of rectangles visible depends on the current geographical extent of the map. The default setup will only show datasets whose extent is completely within the current map extent. It is possible to change this to allow datasets with extent partially intersecting the map extent to be visible.

The same metadata details as in the web application are available.