



Norwegian Marine Data Centre

WP4

Data Storage

D4.3 – Implementation of the API's

Revision: 0.1 /AM
Date: 2015-12-17

Revision history

Version	Date	Comment	Responsible
0.1	2015-12-17	First version	Arnfinn Morvik

Contents

Introduction	4
Applicable documents.....	4
System overview	5
Harvester	6
Validator	6
Transformation.....	6
NMDC XML metadata extention	6
SOLR	8
NMDC API.....	8
GUI	17

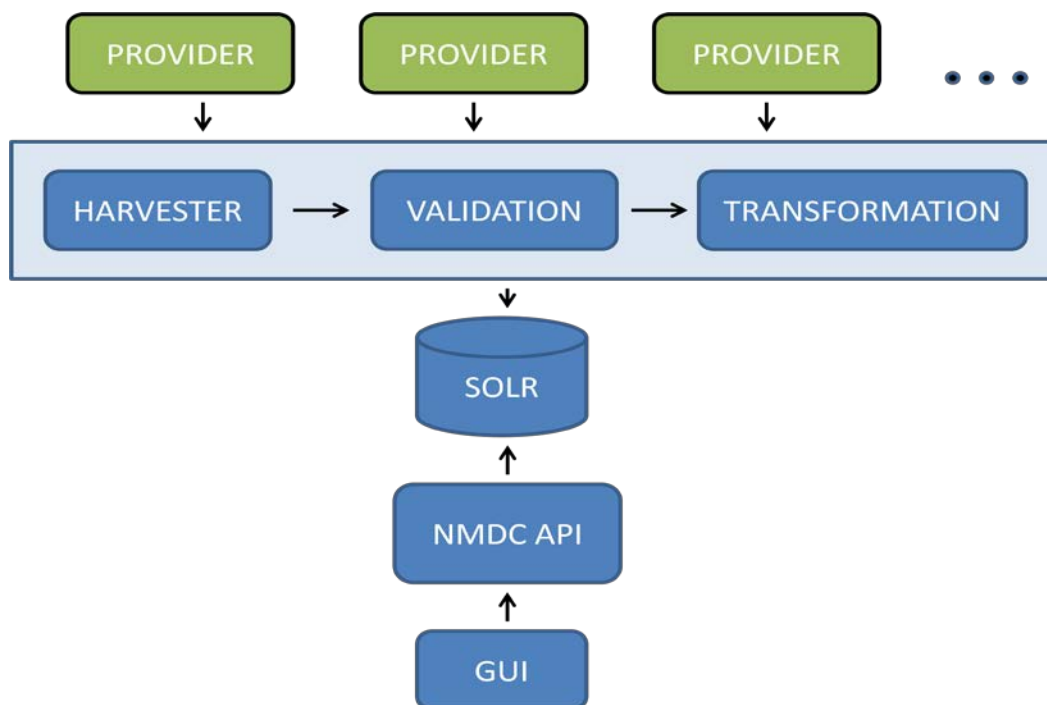
Introduction

This document describes the metadata harvesting and storage module and the indexing and search engine.

Applicable documents

- [1] NMDC System Requirements
- [2] NMDC Architecture Design Document
- [3] NMDC Use Cases
- [4] NMDC Data format and metadata structure
- [5] NMDC Data Catalogue Design
- [6] NMDC Online Data Catalogue

System overview



This document describes the implementation of the central node including the API to the search engine and the graphical user interface.

The figure above shows the components of the NMDc central node. The green parts are the responsibility of each metadata provider. Each metadata provider must set up an OAI-PMH provider for its own metadata. The central node contacts each provider at regular intervals and downloads new or updated metadata.

Harvesting, validation and transformation are separate processes. ActiveMQ is used as a messaging server.

Information about ActiveMQ is available here:

<http://activemq.apache.org/>

After transformation of metadata to NMDc XML, the metadata are indexed with Solr. An API is built on top of Solr, and all searching is performed through this API. A GUI is used for searching and viewing the results of the search.

This document describes the different components of the central node.

Harvester

The central node has implemented an OAI-PMH harvester.

More information about the OAI-PMH harvester can be found here:

<https://www.openarchives.org/OAI/openarchivesprotocol.html>

Validator

The central node validates harvested metadata against ISO 19139 and GCDM DIF xsd. Metadata gets rejected if the validation fails.

For more info about DIF and ISO 19139 see

<http://gcmd.gsfc.nasa.gov/add/difguide/index.html>

http://www.iso.org/iso/catalogue_detail.htm?csnumber=32557

Transformation

Validated metadata are transformed to NMDC xml. The format of the NMDC xml metadata is GCDM DIF with extra fields:

NMDC XML metadata extension

`<nmdc:parameters>`

Container for all the NMDC fields

Syntax

`<nmdc:parameters>`

`</nmdc:parameters>`

Usage

All other NMDC fields are contained within the start and end tag.

<nmdc:polygon>

Defines a polygon that represent the outer boundaries of the data. The polygon is computed from min/max latitude and longitude in the original metadata file. Note that the polygon needs to be closed. This means that first and last lat/long pair must be equal.

Syntax

```
<nmdc:polygon>POLYGON((decimalLatitude0 decimalLongitude0,decimalLatitude1 decimalLongitude1,  
..decimalLatituden decimalLongituden))</nmdc:polygon>
```

Example

```
<nmdc:polygon>POLYGON((11.0 59.0, 12.0 59.0, 12.0 60.0, 11.0 59.0, 11.0 59.0)) </nmdc:polygon>
```

<nmdc:point>

Defines a point that represents the position of the data. It is computed from the min/max latitude and longitude in the original metadata file. It is used if min/max values for latitude and min/max values for longitude in the original metadata file are equal.

Syntax

```
<nmdc:point>decimalLatitude decimalLongitude</nmdc:point>
```

Example

```
<nmdc:point>59.0 12.0</nmdc:point>
```

<nmdc:pDefs>

Container for parameters. Each parameter is a GCMC science keyword and is listed from Topic-Term-...

Syntax

```
<nmdc:pDefs>  
  <nmdc:pDef> .. </nmdc:pDef>  
  ...
```

</nmdc:pDefs>

<nmdc:pDef>

Parameter . Extracted from the original metadata science keywords.

Syntax

<nmdc:pDef>Category>Topic>Term>Variable_Level_1>.....</nmdc:pDef>

Example

<nmdc:pDef>EARTH SCIENCE>OCEANS>OCEAN TEMPERATURE>WATER
TEMPERATURE</nmdc:pDef>

SOLR

The central node uses Solr for indexing and search. Currently, the following parts of the metadata are indexed: Dataset title, Scientific keywords, Dates, Dataset provider, Related persons, Dataset summary, Dataset position/bounding box, Related URL's

More information about Solr can be found at:

<http://lucene.apache.org/solr/>

NMDC API

The API is implemented as a REST API. Currently, three methods are implemented: getFacets, search, and getMetadataDetail. The methods are described below. All the methods return JSON.

More information of the JSON format can be found at: <http://www.json.org/>

getFacets

get list of facets

Call method: <http://www.nmdc.no/rest/getfacets>

Parameters: none

search

Full Text- and Facet Search

Call method: <http://www.nmdc.no/rest/search>

Parameters:

- offset: record number to start list from
- hit: number of records to include in return set
- query: full-text search. Queries all facets except Date, depth and bounding box
- beginDate: optional – default 1800
- endDate: optional – default 2015
- fromDepth: optional – default 0
- toDepth: optional – default -13000
- queryFacet: provider=imr
- bbox: optional default -90, -180, 90, 180 defined in WGS:84
- location:optional

getMetadataDetail

Get metadata for a dataset

Call method: <http://www.nmdc.no/rest/getMetadataDetail?doi=<someOpenArchiveldentifier>>

Parameters:

- doi: the dataset id to get metadata

Returned JSON:

- **getFacets**

```
{
  "facets":{
    "provider":{
      "path":"provider",
      "children":[
        {
          "path":"provider=IMR",
          "label":"IMR",
          "matches":338
        }
      ]
    }
  }
}
```

```
]
},
"provider":{
  "path":"provider",
  "children":[
    {
      "path":"provider=MET",
      "label":"MET",
      "matches":338
    }
  ]
},
"provider":{
  "path":"provider",
  "children":[
    {
      "path":"provider=NPOLAR",
      "label":"NPOLAR",
      "matches":338
    }
  ],
  "label":"provider",
  "matches":1014
},

"author":{
  "path":"author",
  "children":[
    {
      "path":"author=John Doe",
      "label":"Johe Doe",
      "matches":26
    },
    {
      "path":"author=Mary Doe",
      "label":"MAry Doe",
      "matches":26
    }
  ]
}
```

```
    ],
    "label":"author",
    "matches":52
  },
  "parameter":{
    "path":"parameter",
    "children":[
      {
        "path":"parameter=Temperature, water",
        "label":"Temperature, water",
        "matches":62933
      },
      {
        "path":"parameter=Age",
        "label":"Age",
        "matches":52213
      }
    ],
    "label":"parameter",
    "matches":1065289
  },
  "project":{
    "path":"project",
    "children":[
      {
        "path":"project=Økosystem tok",
        "label":"Økosystem tok",
        "matches":48348
      },
      {
        "path":"project=ODP",
        "label":"ODP",
        "matches":37806
      }
    ],
    "label":"project",
    "matches":617976
  },
}
```

```
"title":{
  "path":"title",
  "children":[
    {
      "path":"title=from",
      "label":"from",
      "matches":153001
    },
    {
      "path":"title=station",
      "label":"station",
      "matches":83084
    }
  ],
  "label":"title",
  "matches":961814
},
"keywords":{
  "path":"device",
  "children":[
    {
      "path":"keyword=abc",
      "children":[ ],
      "label":"abc",
      "matches":58601
    },
    {
      "path":"keyword=123",
      "children":[ ],
      "label":"123",
      "matches":33752
    }
  ],
  "label":"device",
  "matches":320958
},
"location":{
```

```
"path":"device",
"children":[
  {
    "path":"keyword=Atlantic Ocean",
    "children":[ ],
    "label":"Atlantic Ocean",
    "matches":58601
  },
  {
    "path":"keyword=Baltic Sea",
    "children":[ ],
    "label":"Baltic Sea",
    "matches":33752
  }
],
"label":"device",
"matches":320958
},
"platform":{
  "path":"platform",
  "children":[
    {
      "path":"platform=G.O. Sars",
      "label":"G.O. Sars",
      "value":40112
    },
    {
      "path":"platform=Håkon Mosby",
      "label":"Håko Mosby",
      "value":28837
    }
  ],
  "label":"platform",
  "value":164592
}
}
```

- **search**

```
{
  "duration":578,
  "suggested":false,
  "offset":0,
  "data":[
    {
      "year":"2004",
      "provider":{
        "provider":"IMR",
        "url":"ftp://ftp.imr.no/nmdc/data/"
      },
      "identifiers":[
        "doi:10.1594/IMR.134595"
      ],
      "disseminations":[
        {
          "access":"open",
          "mime":"text/tab-separated-values",
          "type":"download",
          "title":"TestData",
          "fileType":"csv",
          "url":"ftp://ftp.imr.no/nmdc/data/134595?format=textfile"
        }
      ],
      "geometry":{
        "coordinates":[
          65.033,
          2.1433333
        ],
        "type":"Point"
      },
      "id":"",
      "title":"Test data",
      "expeditions":[
        "POS265"
      ],
      "authors":"Doe, Paul",

```

```
"platforms":[
  "G.O. Sars"
],
{
  "year":"2012",
  "provider":{
    "provider":"IMR",
    "url":"ftp://ftp.imr.no/nmdc/data/"
  },
  "identifiers":[
    "doi:10.1594/IMR.782995"
  ],
  "disseminations":[
    {
      "access":"open",
      "mime":"text/tab-separated-values",
      "type":"download",
      "title":"Test_2",
      "fileType":"csv",
      "url":"ftp://ftp.imr.no/nmdc/data/782995?format=textfile"
    }
  ],
  "geometry":{
    "coordinates":[
      130.891,
      -12.425
    ],
    "type":"Point"
  },
  "id":"",
  "title":"Testing 2",
  "expeditions":[
    "WCRP/GEWEX"
  ],
  "authors":"Doe, Danny",
  "platforms":[
  ]
}
],
"totalHits":338115,
"suggestion":null,
```

```
"maxHits":10,
"facets":{}
}
```

-getMetadataDetail

```
{
  "data":[
    {
      "year":"2004",
      "provider":{
        "provider":"IMR",
        "url":" ftp://ftp.imr.no/nmdc/data/"
      },
      "identifiers":[
        " doi:10.1594/IMR.134595"
      ],
      "disseminations":[
        {
          "access":"open access",
          "mime":"text/tab-separated-values",
          "type":"download",
          "title":"TestData",
          "fileType":"csv",
          "url":" ftp://ftp.imr.no/nmdc/data/134595?format=textfile "
        }
      ],
      "geometry":{
        "coordinates":[
          -20.235,
          49.0914
        ],
        "type":"Point"
      },
      "id":" doi:10.1594/IMR.134595",
      "title":"Nutrients measured on water bottle samples at station DI190_12034#1",
      "expeditions":[
        ""
      ],
      "authors":"author",
      "platforms":[
        "platform"
      ],
      "summary":"some summary",
    }
  ]
}
```



```

"personnel": {
  "Role": " aRole ",
  "First_name": " Jon ",
  "Middle_name": "",
  "Last_name": " Doe ",
  "Email": " rolle@domain.com "
},
"Project": "description",
"Keyword": "some keywords",
"bbox": "{-90,-180,90,180}",
"depth": "{-100, 0}"
}
],
"totalHits": 1,
"maxHits": 0
}

```

GUI

A graphical user interface are available at: <http://prod1.nmdc.no/UserInterface/#/>

The GUI is described in [6]

Below is a screenshot:

