

# eReefs Data Format and Access

Now that you've got a bit more of a handle on how the data are generated we're going to look at how you can access them.

To do this we are going to talk briefly about the data files and their structure, then give you an overview of how to access them.

**Key outcomes:** If you get nothing else out of this brief session i hope you will understand that:

- eReefs satellite data uses a netcdf file and what that looks like.
- eReefs satellite data can be accessed via two main ways - the dashboard webpage or thredds/opensap.

## Data File Structure

### NetCDF

**What do NetCDF's look like?** NetCDF is a “self-describing” binary file format that's used a lot for storing atmospheric and oceanographic data. “self-describing” means that each file contains all the metadata to describe itself - e.g. describing the spatial and temporal dimensions of variables, as well as information about units, and what missing values are set to etc. In the beginning it's a bit of an intimidating format to deal with, but you only need to know a small subset of things to get access to the data they hold.

NetCDF files contain dimensions, variables, and attributes:

- A dimension is just a name and a size that describes what the shape of the data inside the file is.
- A variable has a name, a type, a list of dimensions, some attributes, and some data (As well as specific attributes attached to specific variables NetCDF files can also have some overall global attributes).
- An attribute has a name, a type, and a value.

**Do a little drawing on the wall as we talk through the structure - this structure will be important later when we talk about more advanced data access techniques.**

**A real MWQ file example:**

```
netcdf A20150521.P1D.ANN_MIM_RMP {  
  dimensions:  
    time = 1 ;  
    lat = 1600 ;  
    lon = 1400 ;  
  variables:
```

```

int time(time) ;
    time:long_name = "time" ;
    time:standard_name = "time" ;
    time:units = "seconds since 1970-01-01 00:00:00" ;
    time:calendar = "gregorian" ;
    time:axis = "T" ;
    time:ioos_category = "Time" ;
float lat(lat) ;
    lat:long_name = "latitude" ;
    lat:standard_name = "latitude" ;
    lat:units = "degrees_north" ;
    lat:axis = "Y" ;
    lat:ioos_category = "Location" ;
float lon(lon) ;
    lon:long_name = "longitude" ;
    lon:standard_name = "longitude" ;
    lon:units = "degrees_east" ;
    lon:axis = "X" ;
    lon:ioos_category = "Location" ;
float Chl_MIM(time, lat, lon) ;
    Chl_MIM:_FillValue = 1.e+20f ;
    Chl_MIM:least_significant_digit = 3s ;
    Chl_MIM:comment = "chlorophyll-a concentration, MIM SVDC on Rrs" ;
    Chl_MIM:coordinates = "time lat lon" ;
    Chl_MIM:feature_id = "http://sweet.jpl.nasa.gov/2.2/realmOcean.owl#OceanRegion" ;
    Chl_MIM:ioos_category = "Ocean Color" ;
    Chl_MIM:long_name = "chlorophyll-a concentration, MIM SVDC on Rrs" ;
    Chl_MIM:matrix_id = "http://sweet.jpl.nasa.gov/2.2/matrWater.owl#SeaWater" ;
    Chl_MIM:procedure_id = "matrix_inversion_method_singular_value_decomposition" ;
    Chl_MIM:scaledQuantityKind_id = "http://environment.data.gov.au/water/quality/def/p
    Chl_MIM:standard_name = "mass_concentration_of_chlorophyll_in_sea_water" ;
    Chl_MIM:substanceOrTaxon_id = "http://environment.data.gov.au/water/quality/def/obj
    Chl_MIM:unit_id = "http://environment.data.gov.au/water/quality/def/unit/MicroGrams
    Chl_MIM:units = "mg m-3" ;
    Chl_MIM:valid_max = 64.f ;
    Chl_MIM:valid_min = 0.f ;

```

...more variables here...

// global attributes:

```

:Conventions = "CF-1.6" ;
:Metadata_Conventions = "Unidata Dataset Discovery v1.0" ;
:Metadata_Link = "Currently unavailable" ;
:keywords = "EARTH SCIENCE > OCEANS > OCEAN OPTICS" ;
:keywords_vocabulary = "NASA Global Change Master Directory (GCMD) Science
:standard_name_vocabulary = "CF-1.6" ;
:naming_authority = "au.gov.bom" ;

```

```

:netCDF_version = "4.1.3 of Aug 10 2011 04:46:23 $" ;
:id = "ABOM_EREEFS_MWQ_P1D_GRID_v20130813" ;
:title = "Marine water quality grids for the Great Barrier Reef region - da
:summary = "Parameters calculated using the MIM SVDC on Rrs approach as par
:references = "Brando, Dekker, Park and Schroder, 2012, Adaptive semianalyt
:comment = "NetCDF file contains water quality grids calculated using the M
:processing_level = "Automatically generated environmental data record with
:cdm_data_type = "grid" ;
:spatial_resolution = 0.01f ;
:geospatial_lat_units = "degrees_N" ;
:geospatial_lat_resolution = 0.01f ;
:geospatial_lat_min = -25.495f ;
:geospatial_lat_max = -9.505f ;
:geospatial_lon_units = "degrees_E" ;
:geospatial_lon_resolution = 0.01f ;
:geospatial_lon_min = 142.005f ;
:geospatial_lon_max = 155.995f ;
:start_time = "20150521T042008Z" ;
:stop_time = "20150521T042506Z" ;
:time_coverage_duration = "P1D" ;
:time_coverage_resolution = "P1D" ;
:institution = "ABOM" ;
:project = "eReefs - Marine Water Quality" ;
:creator_name = "National Plan for Environmental Information" ;
:creator_email = "ereefs_support@bom.gov.au" ;
:creator_url = "http://www.bom.gov.au/environment/" ;
:acknowledgment = "Any use of these data requires the following acknowledgm
:license = "Copyright for any data supplied by the Bureau of Meteorology is
:disclaimer = "You accept all risks and responsibility for losses, damages,
:history = "mwq_L3_processor using pyresample (EDR)" ;
:source = "A2015141023000.SEADAS.hdf, A2015141023000.ANN.hdf, A201514102300
:uuid = "def05096-0003-11e5-b7b0-001517d04ac5" ;
:date_created = "20150521T215352Z" ;
}

```

---

## Data Access

The Bureau stores the eReefs satellite data files on a centralised and backed up disk at head office in Melbourne. These data are accessible through a series of tools that allow external users to view, query, and retrieve the data via the internet and the world wide web.

## THREDDS

All of the data access methods for the eReefs satellite data are centralised around THREDDS (Thematic Real-time Environmental Distributed Data Services). THREDDS is middleware that bridges the gap between data providers and data users and consists of two main areas: the THREDDS Data Server (TDS) and the Common Data Model (CDM)/netCDF-Java library.

THREDDS is developed and supported by Unidata, a division of the University Corporation for Atmospheric Research ([UCAR](#)), and is sponsored by the [National Science Foundation](#).

eReefs THREDDS: <http://ereeftds.bom.gov.au/ereefs/tds/catalog.html>

eReefs ERDDAP: <http://ereferd.bom.gov.au/ereefs/erddap/index.html>

**The THREDDS Data Server** The THREDDS Data Server (TDS) provides catalog, metadata, and data access services for scientific data. The TDS publishes [THREDDS catalogs](#) that advertise the datasets and services it makes available. These THREDDS catalogs are XML documents that list the datasets and the data access services available for the eReefs system. Catalogs may contain metadata to document details about the datasets.

The TDS allows users to query the datasets listed in its catalogs using several standard data access protocols. The available remote data access protocols for eReefs include OPeNDAP, OGC WCS, OGC WMS, ERDDAP, and HTTP.

The TDS also supports several dataset collection services including some sophisticated dataset aggregation capabilities. This allows the TDS to aggregate a collection of datasets into a single virtual dataset, greatly simplifying user access to that data collection. For more information see the THREDDS Data Server page.

**Common Data Model** The Common Data Model (CDM) provides data access through the [netCDF-Java API](#) to a variety of data formats (e.g., netCDF, HDF, GRIB). Layered above the basic data access, the CDM uses the metadata contained in datasets to provide a higher-level interface to geoscience specific features of datasets, in particular, providing geolocation and data subsetting in coordinate space.

## OPeNDAP

[OPeNDAP](#) stands for “Open-source Project for a Network Data Access Protocol”. OPeNDAP is both the name of a non-profit organization and the commonly-used name of a data access protocol which the OPeNDAP organization developed. The OPeNDAP protocol provides a standardised means of requesting and providing data across the World Wide Web. The goal is to allow end users, whoever they may be, to access immediately whatever data they require in a form they can use, all while using applications they already possess and are familiar with. Ultimately, it is hoped, OPeNDAP will be a fundamental component of systems which provide machine-to-machine interoperability with semantic meaning in a highly distributed environment of heterogeneous datasets.

OPeNDAP allows you to access remote data over the internet using familiar data analysis and visualization packages/[clients](#) (i.e., applications able to access remote served data) like:

- GIS
- Python
- R
- Matlab
- Ferret
- IDL

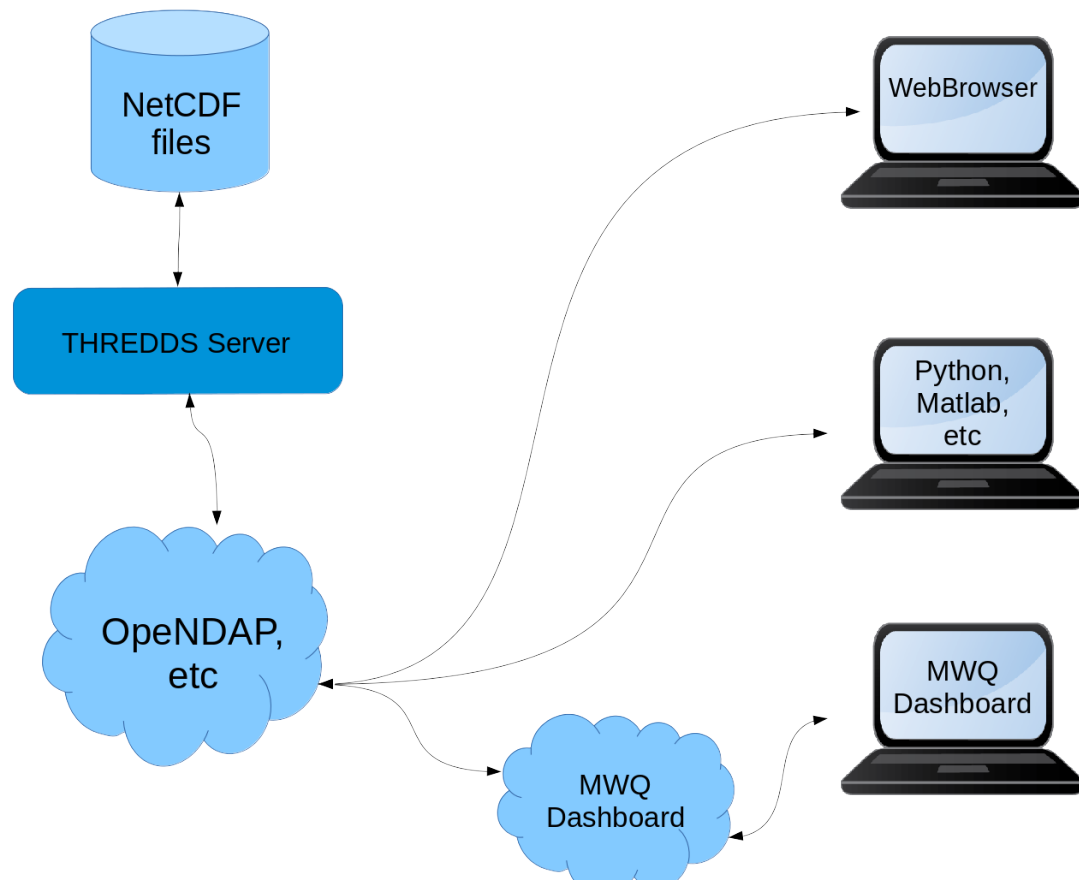


Figure 1: Simple diagram of opendap

## Dashboard

The eReefs Dashboard provides an world wide web interface to the eReefs satellite data files. It is a webpage that enables basic visualisation and acquisition of the core Marine Water Quality and ReefTemp Indices for each day of the year and for composite weeks, months, wet and dry seasons, and calendar years. The Dashboard is built on top of the THREDDS system and uses standard data access protocols through thredds to access the eReefs data files.

The eReefs Marine Water Quality Dashboard can be found at: <http://www.bom.gov.au/marinewaterquality/>