

ERDDAP service by NOAA

Easier access to scientific data

Claudia Fratianni, Paolo Oliveri
Sezione di Bologna

Outlines:

- Dissemination protocols in use at INGV OCEANO GROUP
- ERDDAP starting point: from user to provider
- ERDDAP general information and configuration
- ERDDAP interactive view
- Conclusions

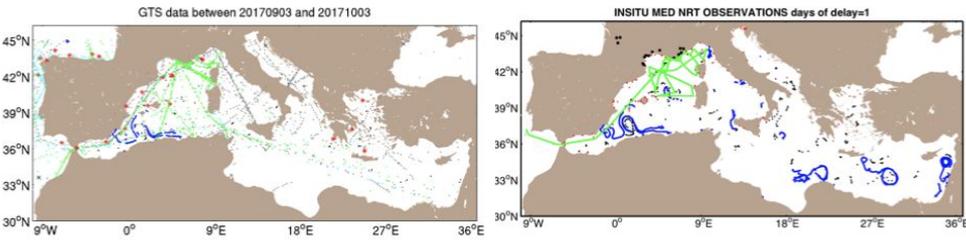
ERDDAP starting point: from user ...

Open Access to the GTS: A JCOMM Observations Coordination Group Pilot Project

The JCOMM Open Access to GTS project is taking steps to simplify the process of getting data onto and off of the GTS for the research community (Pinardi et al., 2019). NOAA's Observing System Monitoring Center (OSMC) provides access to near real-time ocean observations through the ERDDAP data platform.

Data Access test and comparison with Copernicus Marine Environment Monitoring service (CMEMS) and the Observing System Monitoring Center (OSMC) GTS feed for Mediterranean sea in month of September, 2017.

Simona Simoncelli and Alessandro Grandi (updated in 2019 Simoncelli-Fratianni)

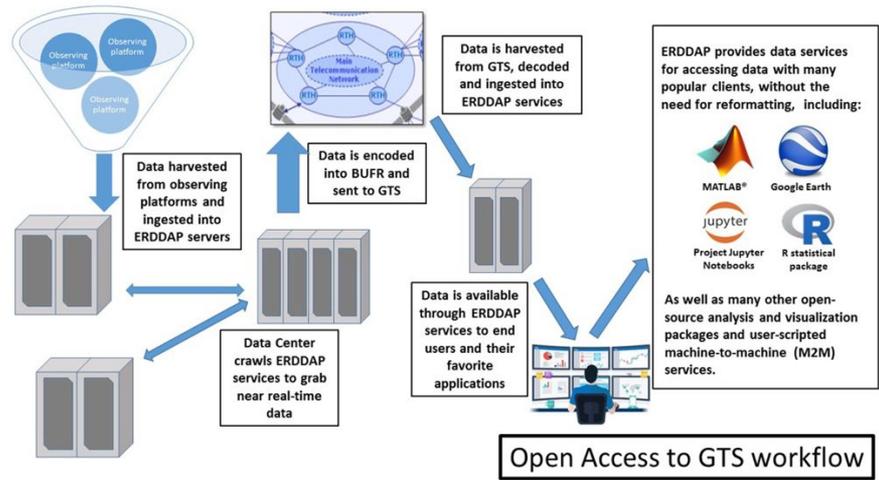


Exercise illustrated importance of GTS data for forecasting activities since:

- Most GTS data available within 12 hours
- 43% of CMEMS NRT data available after 24 hours

benefits as user

- Users can access data with clients of their choice without reformatting data
- Provides web accessible urls for data download and subsetting of data both temporally and spatially via other software packages



Open Access to GTS workflow

... to provider



ERDDAP > List of All Datasets

6 matching datasets, listed in alphabetical order.

Grid DAP Data	Sub-set	Table DAP Data	Make A Graph	W M S	Source Data Files	Title	Summary	FGDC, ISO, Metadata	Back-ground Info	RSS	E mail	Institution	Dataset ID
	set	data	graph			* The List of All Active Datasets in this ERDDAP *		M	background			Istituto Nazionale...	allDatasets
data			graph	M		North Adriatic Temperature and Salinity Climatology V1 monthly 1955-1984 and 1985-2016		F I M	background			Istituto Nazionale...	NADR_CLIM_TS_3KM_1_m_emt
data			graph	M		North Adriatic Temperature and Salinity Climatology V1 monthly 1955-2016		F I M	background			Istituto Nazionale...	NADR_CLIM_TS_3KM_1_m
data			graph	M		North Adriatic Temperature and Salinity Climatology V1 seasonal 1955-1984 and 1985-2016		F I M	background			Istituto Nazionale...	NADR_CLIM_TS_3KM_1_s_emt
data			graph	M		North Adriatic Temperature and Salinity Climatology V1 seasonal 1955-2016		F I M	background			Istituto Nazionale...	NADR_CLIM_TS_3KM_1_s
data			graph	M		North Adriatic Temperature and Salinity Climatology V1 seasonal decadal 1955-2016		F I M	background			Istituto Nazionale...	NADR_CLIM_TS_3KM_1_s_decades

The information in the table above is also available in other file formats (csv, htmlTable, .txt, .json, jsonCSV1, jsonCSV, jsonKVP, .mat, .nc, .ncsv, .tsv, .xhtml) via a RESTful web service.

benefits as provider

- provides data to users in many different formats and m2m service
- Can augment metadata without rewriting files
- generate ISO19115 and Schema.org compliant metadata on the fly
- A data brokerage service, reading from many different types of files, databases and services.



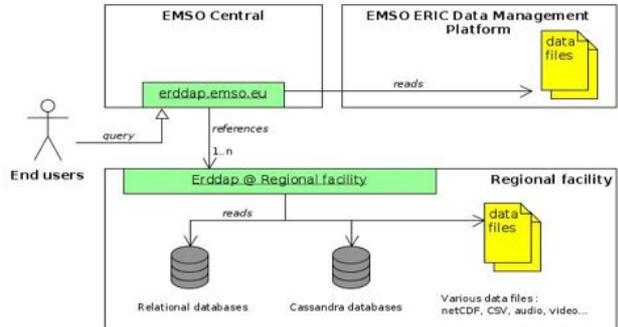
ERDDAP > List of All Datasets

4 matching datasets, listed in alphabetical order.

Grid DAP Data	Sub-set	Table DAP Data	Make A Graph	W M S	Source Data Files	Title	Summary	FGDC, ISO, Metadata	Back-ground Info	RSS	E mail	Institution	Dataset ID
	set	data	graph			* The List of All Active Datasets in this ERDDAP *		M	background			Western Ionian	allDatasets
		data	graph			2002-2003 NEMO-SN1 Observatory CTD data		F I M	background			Istituto Nazionale...	EMSO_Western_Ionian_Sea_CTD_2002_2003
		data	graph			2012-2013 NEMO-SN1 Observatory CTD data		F I M	background			Istituto Nazionale...	EMSO_Western_Ionian_Sea_CTD_2012_2013
		data	graph			2012-2013 NEMO-SN1 Observatory Pressure Gauge data		F I M	background			Istituto Nazionale...	EMSO_Western_Ionian_Sea_PRESSURE_GAUGE_2012_2013

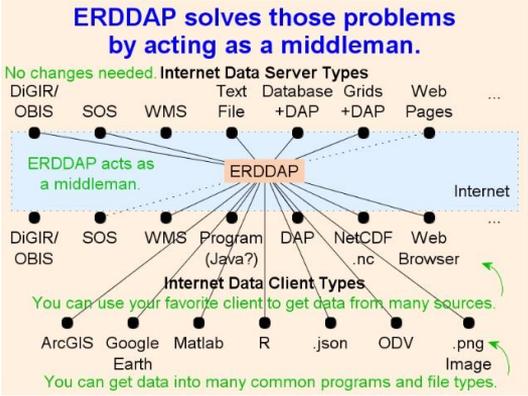
The information in the table above is also available in other file formats (csv, htmlTable, .txt, .json, jsonCSV1, jsonCSV, jsonKVP, .mat, .nc, .ncsv, .tsv, .xhtml) via a RESTful web service.

distributed/federated network of ERDDAP servers



ERDDAP general information

- ERDDAP (Environmental Research Division's Data Access Program) is a data server that gives a simple, consistent way to download subsets of scientific datasets in common file formats and make graphs and maps.
- ERDDAP is reusable, Free and Open Source software developed by the National Oceanographic and Atmospheric Administration (NOAA)



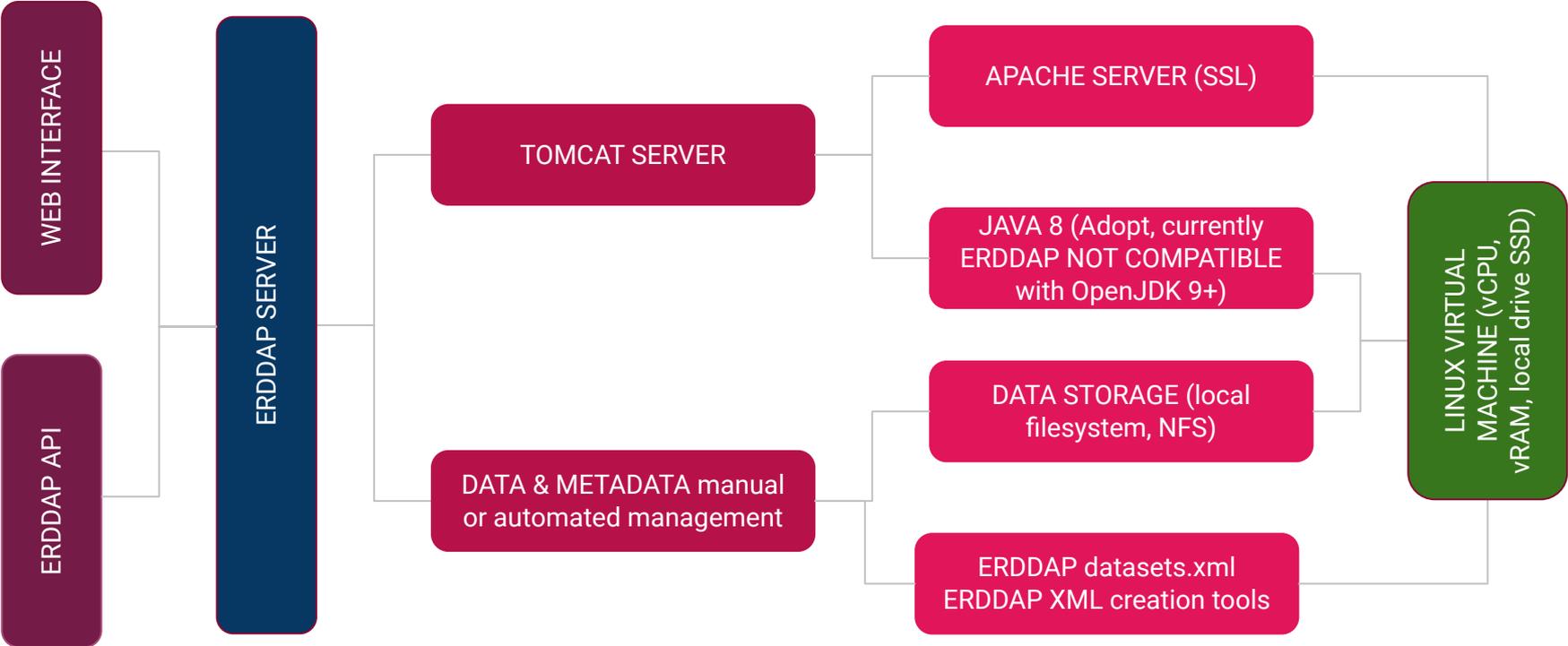
- ERDDAP takes a middleware approach: it can get data from lots of different types of remote data servers and it can provide that data to clients in lots of different file formats.
- ERDDAP is providing itself as an essential tool to improve data interoperability and integration among various data communities.

- Widely used for delivery of “FAIR” data in the geoscience community (many server deployments worldwide)

ERDDAP services and FAIR principles

- F**
 - ERDDAP can easily augment metadata with missing elements for compliance
 - ERDDAP can generate ISO metadata documents on demand (harvestable)
 - ERDDAP supports schema.org metadata
- A**
 - ERDDAP provides access to data in multiple formats
 - ERDDAP supports RESTful services for machine to machine access (harvestable)
 - ERDDAP provides a platform upon which data-specific applications can be built
- I**
 - All of these capabilities lead to improved interoperability
- R**
 - ERDDAP supports creation of BagIt packages for easy archival (DOI)
 - ERDDAP supports Climate and Forecast metadata, json-ld, schema.org

EXAMPLE ERDDAP CONFIGURATION



ERDDAP flexibility

- Can be installed on every modern machine with different O/S (**Windows, MacOS, Linux**) for **personal** use (small amount of data, single user or very small community);
- Can be provided as a standalone **Docker image** (for Docker users);
- To be ready to Big Data management (big amount of data, large metadata granularity, wide user community), we suggest to place the service in a **Linux VM** (virtual machine) with sufficient amount of vCPU, vRAM and SSD drive space to properly handle data and metadata management, without performance limitations or bottlenecks.

For further details, please follow this link:

<https://coastwatch.pfeg.noaa.gov/erddap/download/setup.html>

Example environment preparation

- **Build and deploy Linux VM.** Deeply inspect the vCPU and vRAM amounts to match the final configuration, based on data and intended use;
- **Install Apache** web server. SSL is a must for modern usage and security;
- **Install Java 8.** For example, Adopt OpenJDK. Note: Tomcat and other Tomcat bundled Java applications can use Open JDK 9+, but ERDDAP currently does not;
- **Install Tomcat.** Tomcat 7+ confirmed working with ERDDAP. Configure storage location, set Java executable path and Xmx and Xms RAM quantity in Tomcat service file using CATALINA_OPTS envi var;
- **Install ERDDAP.**

ERDDAP installation example

- **Download latest ERDDAP WAR** file, place it in the Tomcat web app directory and wait for the extraction and deployment;
- **Download erddapContent.zip** in a folder containing the needed content for the ERDDAP web app and properly set **-DerddapContentDirectory** option in CATALINA_OPTS envi var;
- **Configure** the ERDDAP service **setup.xml** file by editing at least:
 - **bigParentDirectory**: folder for ERDDAP **caching, flagging** and **data information retraining**. **Must be a folder with lot of space and sufficient I/O speed to ensure ERDDAP optimal performance**;
 - **http / https URL** for external server expose;
 - Admin email and work information for management.
- Properly set up **ProxyPass** in Apache to access the ERDDAP service;
- **Configure** the ERDDAP service **datasets.xml** that contains:
 - ERDDAP home page layout and options;
 - All metadata information and link to the physical data location.

Datasets.xml: the ERDDAP core

It's divided in two parts:

1. ERDDAP home page layout and options;
2. ERDDAP datasets metadata content: every content may be **Table** or **Grid**.

Datasets live into the tag **<dataset>** and there are two ways to add them:

1. **Manual compilation** of the dataset tag metadata hierarchy (for example, when the data that is going to add has no metadata directly associated (for ex. CSV tables));
2. **Automated compilation** using ERDDAP-provided data import tools in WEB-INF folder (for example, when using already well formatted data, such as netCDF).

For further details, please follow this link:

<https://coastwatch.pfeg.noaa.gov/erddap/download/setupDatasetsXml.html>

Automated import example: generateDatasetsXml

Semi automated script that creates a **<dataset>** tag entry starting from an already formatted data structure. Most important steps to run generator:

1. Select ERDDAP Dataset Type (**EDDType**):
 - a. GridFrom[something] (Dap, NcFiles, AsciiFiles, etc)
 - b. TableFrom[something] (Ascii, Database PostGreSQL driver, NcFiles, etc)
2. **Data directory or remote URL;**
3. **ReloadEveryMinutes:** time to reload the data from memory (for ex. in case of data update).

Other Useful information

1. There is a **status page** for ERDDAP server to show some basic statistics;
2. Other useful information can be acquired by the admin using **ERDDAP and Tomcat Log files**;
3. There is a basic IP based **Blacklist protection** that can be included in datasets.xml file (for ex. to forbid multiple simultaneous requests from bots);
4. ERDDAP default usage mode is **without login methods**. However, it can be added an **access restriction** to trace data access or to protect particular sensible data. Login can be provided using email, google credentials, ORCID, etc.;
5. While ERDDAP is online, some **datasets information is stored in json format** in order to speedup data access and navigation. This could result in a bottleneck if a huge ERDDAP server is executed in a weak performance configuration;
6. **Advanced statistics of ERDDAP** usage has not been found yet from us, but can be asked to the NOAA staff.

Interactive ERDDAP view



INGV OCEANO GROUP BO ERDDAP
Easier access to scientific data

Brought to you by INGV OCEANO GROUP

ERDDAP > List of All Datasets

6 matching datasets, listed in alphabetical order.

Grid DAP Data	Sub-set	Table DAP Data	Make A Graph	W M S	Source Data Files	Title	Summary	FGDC, ISO, Metadata	Back-ground Info	RSS	E mail	Institution
	set	data	graph			* The List of All Active Datasets in this ERDDAP *	?	M	background			Istituto Nazional... ? a
data			graph	M		North Adriatic Temperature and Salinity Climatology V1 monthly 1955-1984 and 1985-2016	?	F I M	background ?	RSS	✉	Istituto Nazional... ? N
data			graph	M		North Adriatic Temperature and Salinity Climatology V1 monthly 1955-2016	?	F I M	background ?	RSS	✉	Istituto Nazional... ? N
data			graph	M		North Adriatic Temperature and Salinity Climatology V1 seasonal 1955-1984 and 1985-2016	?	F I M	background ?	RSS	✉	Istituto Nazional... ? N
data			graph	M		North Adriatic Temperature and Salinity Climatology V1 seasonal 1955-2016	?	F I M	background ?	RSS	✉	Istituto Nazional... ? N
data			graph	M		North Adriatic Temperature and Salinity Climatology V1 seasonal decadal 1955-2016	?	F I M	background ?	RSS	✉	Istituto Nazional... ? N

The information in the table above is also available in other file formats (.csv, .htmlTable, .itx, .json, .jsonCSV1, .jsonCSV, .jsonKVP, .mat, .nc, .nccsv, .tsv, .xhtml) via a RESTful web service.

ERDDAP, Version 2.11
[Disclaimers](#) | [Privacy Policy](#) | [Contact](#)

http://oceano.bo.ingv.it/erddap/categorize/keywords/nadr_clim_ts_3km.1

From data dissemination to DOI

- To insert data into INGV registry it is necessary to have a valid dissemination point;
- ERDDAP is an easy instrument to be compliant with INGV data policy and to be inserted in the INGV registry → valorization of the research work;
- Steps to have DOI for your data:
 - Data page compilation in INGV data registry;
 - Interaction with INGV data office to create data DOI string;
 - DOI activation;
- Example for our data: https://doi.org/10.13127/nadr_clim_ts_3km.1

Discussion and questions