ERDDAP™ Updates and New Capabilities

Chris John

April 30

Governance summary

ERDDAP™ has transitioned from a Founder-Leader model to a governance model that involves the community.

NOAA Appointed Technical Director

● **Lead Developer**

Technical Board

● **Quarterly Meetings**
Prioritizes work, discusses large technical issues.

● **Coordinates efforts**

Strategic Insight Group

● **Supports ERDDAP™**
Makes sure there's long term support including funding.

● **Promotion of ERDDAP™**
Get the word out to the data community.

Documentation

Docusaurus

The main ERDDAP™ documentation has been migrated to Docusaurus. This is a widely used tool for documentation with a lot of great features.

The documentation is hosted on GitHub Pages: https://erddap.github.io

Markdown
Easy to edit pages

Search
Help users find content

Localization
Translated to 29 languages

Versioning
Support for different documentation by release.

Developer Focused Changes

Maven	Standardizing the developer environment. Used for dependency management, resources, builds, testing, running locally (Docker and Jetty).
Testing	Automated unit and integration tests. At about 60% code coverage and growing.
Static Analysis	Using ErrorProne and SpotBugs to find issues. We also apply standardized formatting to the code.
Refactoring	With the above in place we can make larger changes with more confidence.

Optimizations

Shared Watch Service

The SharedWatchService is a new approach for handling file updates. Previously every dataset that wanted to receive file updates used a thread to receive those updates. The SharedWatchServce has one thread that receives updates for all datasets and then notifies those datasets to handle the new files.

TouchThread
No duplicate pending entries

Buffered Random Access File
Speeds up Etopo reads and other similar grids

Many small changes
More efficient data types, avoid boxing/unboxing, avoid unnecessary work

Memory
Avoid memory leaks with try-with-resources, reduce memory usage

Updates for Admins

Sax XML Parser

● **Built by Ayush Singh**
This was a GSoC project

● **Supports XInclude**
This allows for reusable sections of dataset definitions and splitting dataset definitions across multiple files.

Prometheus Metrics

● **Available at erddap/metrics**
Includes a variety of data, including dataset usage.

● **Flexible Integrations**
Can be used for monitoring, alerting, and reporting

Improved File Support

● **Parquet**
Read and write with a new dataset type.

● **Zarr**
Read support for Zarr files added to *FromNcFiles dataset types.

User Visible Updates

displayInfo

● **Configurable**
New UI elements on dataset pages based on fields in datasets.xml

● **Dataset Citation**
One possible use is including how to cite datasets.
Built by Ayush Singh.

Change Notifications

● **RSS**
Updated when file changes are processed

● **Subscriptions**
Now notified on file changes, not waiting for the next reload
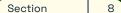
ISO19115.xml

● **New option**
Currently defaults to off, up to server admins to enable

● **Apache SIS**
Uses Apache SIS to aid in generating the ISO19115.xml files.

In Progress and Upcoming

Official Docker Image
There's been a lot of progress on this, mostly led by Shane St. Savage.

MQTT
Potential GSoC project to add support for MQTT.

Additional output file formats
Recently simplified how to add new file formats.

Localization of dataset metadata
Support providing localized versions of dataset metadata in datasets.xml.

HTML Update
Potential GSoC project, updating to use a modern templating framework to build html pages. Goals include supporting accessibility and responsive design.

Semantic Versioning
Changing ERDDAP™ to use a semantic version.