ERDDAP: SWFSC's Solution for Data Access

Satisfying the Presidential mandate to make data collected with federal funds accessible to the public: the Public Access to Research Results (PARR).

There are three aspects of data accessibility: *data discovery*, *data access*, and *data archiving*. To successfully satisfy the PARR requirements, all three components have to be fulfilled. NOAA has technical recommendations for data access and solutions for data discovery and data archiving.

Data Discovery:[1]

Data discovery is the process by which a user can discover what data sets exist; data discovery does not retrieve the data, only points to what exists. The process of data discovery involves repositories of metadata, the data that explains the data in a file. Metadata is used to describe the what, where, when, how and who about a particular data set. In order for metadata searches to be successful, there has to be accepted standards for reporting metadata; the previous U.S. Federal Geographic Data Commission (FGDC) standard has been deprecated. NOAA now follows the International Standards Organization (ISO) 19115 standard.

There exist a hierarchy of data discovery tools available to NMFS scientists. Data.gov is the federal metadata repository, data.noaa.gov is a NOAA implementation of data.gov and InPort is a metadata catalog developed within NMFS.

Data Archiving:

Data archiving ensures the preservation of federally-funded data for future usage. NOAA has recently reorganized the large data centers into the National Centers for Environmental Information (NCEI); this organization is responsible for determining NOAA archival responsibilities, policies and procedures.

Data Access:

The NOAA data access directive[2] is very broad in the requirements for public access. Submission to NCEI centers satisfies the requirement, but as a stand-alone solution this poses many problems regarding the curatorship of the data. The general requirements to ensure the data are accessible by the public include:

- approved submission of data to the NOAA NCEI,
- data shall be available in one or more machine-readable digital formats, suggested formats are via web service or application programming interface (API) that support machine-to-machine data access,

---

[1] NOAA Data Documentation Procedural Directive, Version 1.0, October 2011

[2] NOAA Data Access Procedural Directive

- web sites and pages may supplement the machine-readable access requirement,
- data shall be accessible over the internet without requiring a password,
- data shall be available free of charge.

These requirements, along with suggested software solutions are expanded upon in the NOAA Environmental Data Management Wiki.

From this brief listing, it is apparent that the data provider has to satisfy a number of requirements to meet the PARR mandates, many of which are above and beyond what they normally do. The goal for the data provider, and what will ensure for a lot more cooperation, "is one and done"; the data provider meets one mandate, and the rest "happen."

ERDDAP provides this capability. It provides data access, it will create the ISO metadata files and put them in a web accessible folder (WAF) for harvesting, it can provide the data to the archive in a standardized way, and, through subscriptions, it can provide notices of when a data set has been updated.

ERDDAP:

ERDDAP meets the metadata requirements with the added strength that it is designed around the user experience. When the information about the dataset is entered, ERDDAP produces both ISO and FGDC metadata files that are put in a Web Accessible Folder (WAF). These have been successfully harvested for use in data.noaa.gov. The NOAA Environmental Data Management Wiki's Data Access Technical Directive lists ERDDAP as the only NMFS-created or NOAA-created software that meets the access requirements.

A fundamental and basic design criterion for ERDDAP is to focus on the user experience and address the question "how do I get the data into my favorite application?[3]" To fulfill this design criterion, the software has to be simple, fast, interoperable and conforming to standards[4]. The elegant result is "middleware," a software data server designed to help the user get data from any data source into the user's client software.

ERDDAP's success is demonstrated by the over 50 organizations worldwide that have installed the software[5]. A useful presentation of ERDDAP's usability in terms of both data access and data visualization can be viewed online at http://coastwatch.pfel.noaa.gov/erddap/images/erddapTalk/erddapTalk4.html

ERDDAP can access data from types of data sources (including databases, many types of files, and remote services), and conforms to the accepted metadata standards. In that

---

[3] http://coastwatch.pfel.noaa.gov/erddap/images/erddapTalk/erdData.html

[4] http://coastwatch.pfeg.noaa.gov/erddap/information.html

[5] http://coastwatch.pfeg.noaa.gov/erddap/download/setup.html#organizations

regard, ERDDAP provides an excellent solution for both the data provider and the data user. Because the data are stored with the correct metadata, push transfers of metadata to the data discovery portals and of data files to the data archive sites are easily accomplished. The complete chain of data discovery and recovery is enabled.

ERDDAP provides discovery capabilities, as it has a powerful search capability as a service via its own search API, as well as support for the OpenSearch standard.

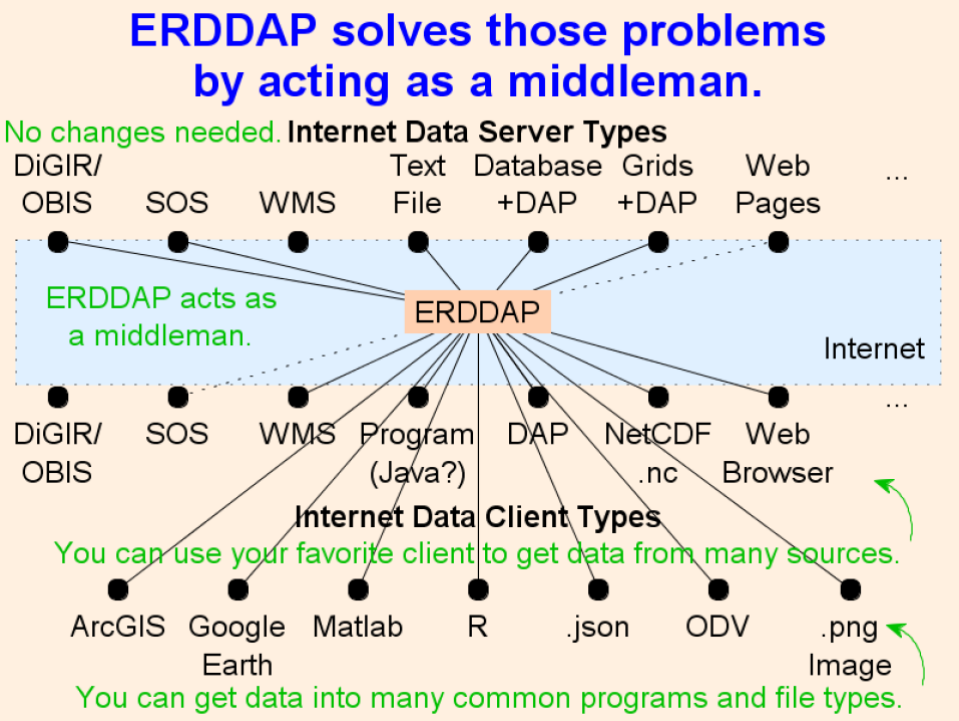

Figure 1: ERDDAP middleware solution.

ERDDAP can help with data archiving, as a large question is how to get the data into the archive in formats that the archive can accept. ERDDAP can translate the database tables into the standard netcdf files desired by NCEI. We are working with NCEI right now on the CalCOFI data as an example, where NCEI would update their archive (which is badly out-of-date and incomplete) by pulling from ERDDAP, and would subscribe to the dataset to be notified when an update occurs.

Note that once a dataset is put into ERDDAP, data access is satisfied, metadata requirements are satisfied, discovery is satisfied, and archiving is simplified. Instead of having to meet a host of data mandates, the data producer meets one, does not need to figure out how to reformat their data, and all the other requirements are met.

ERDDAP, combined with ERD and CoastWatch web services provides powerful tools for data access and utilization.

The ERDDAP homepage is http://coastwatch.pfeg.noaa.gov/erddap/index.html