Integrating Data and Information Across Observing Systems

Kevin O'Brien, University of Washington/JISAO, USA, kob@uw.edu
Steven Worely, University Corporation for Atmospheric Research, USA, worley@ucare.edu
Bob Simons, NOAA, USA, bob.simons@noaa.gov
Benjamin Pfeil, Bjerknes Climate Data Centre, Norway, Benjamin.Pfeil@uib.no
Eugene Burger, NOAA, USA, Eugene.Burger@noaa.gov

"Integrated data management" is best understood by contrasting it with what has historically been the norm for ocean data management. Ocean observations transmitted through the GTS have very high societal value, due to their potential use in ocean forecast models, climate state estimates, etc. Yet, for the average science data user, the ability to access and use data from the GTS has been such a prominent barrier that it was all but impossible. This creates a situation where much of this data is underutilized, despite its very high scientific value and the large investments made to collect it.

The potential for improved integration among delayed mode global ocean data assembly centers (GDACs) is another significant opportunity. Currently, each platform assembly center typically provides its own web site at which users (humans, but not machines) can select data of interest. Data can then be downloaded in the file format(s) available from that network. A user desiring interior ocean temperatures from both Argo and TAO, for example, must learn to navigate two independently designed (and very distinct) Web sites, each with its own unique technical language and each with its own style of selecting regions, times, platforms, and variables with various imposed groupings. When the use of these distinct Web sites have been mastered, a larger barrier remains – that of unstandardized data formats limiting the fitness of use of the data. Custom software is often needed to read the data from distinct networks. Variable names and metadata differ between providers. Non-experts would often struggle to pursue such a task to completion.

Rather than having data systems and offerings developed around individual networks, an integrated framework promises interoperable access that crosscuts the networks and offers a more intuitive research view for the users (Figure 1). In practical terms, this integrated access will allow users to focus on Essential Ocean Variables (EOVs), rather than on platform specific measurements. This framework will create interoperable access to data and metadata, and is not simply another data portal. Data providers and groups will be able to leverage the interoperable framework to build customized data portals, and with little advanced planning will be able to expand their impact with broader access support from the framework. As is, with the myriad of access services provided by the framework, users will already be able to access and use the data with clients they prefer (e.g. MatLab, iPython Notebook, etc.). They will not need to navigate through many different download paths or be subjected to having to reformat data specifically for their favorite client.

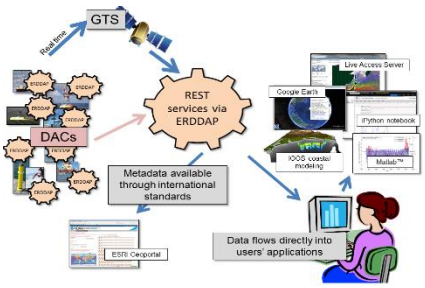

Figure 1 - Schematic of an integrated framework for serving ocean data

In this presentation, we will discuss how we have utilized such a framework to provide interoperable access to real time ocean observations from the GTS. We will also describe our efforts to integrate data

available from the heterogeneous platform networks providing ocean observations in support of the Tropical Pacific Observing System (TPOS) 2020 effort. We will discuss how this framework, utilizing a tool called ERDDAP (http://coastwatch.pfeg.noaa.gov/erddap), leverages the OPeNDAP protocol and supports RESTful web services, providing machine to machine access and easy client access. We will also highlight some complimentary efforts already underway within Europe to provide this integrated framework and interoperable data access.