

Model Data Interoperability for the United States Integrated Ocean Observing System (IOOS)

Richard P. Signell^{1,2}

Abstract

Model data interoperability for the United States Integrated Ocean Observing System (IOOS[®]) was initiated with a focused one year project. The problem was that there were many regional and national providers of oceanographic model data; each had unique file conventions, distribution techniques and analysis tools that made it difficult to compare model results and observational data. To solve this problem, a distributed system was built utilizing a customized middleware layer and a common data model. This allowed each model data provider to keep their existing model and data files unchanged, yet deliver model data via web services in a common form. With standards-based applications that used these web services, end users then had a common way to access data from any of the models. These applications included: (1) a 2D mapping and animation using a web browser application, (2) an advanced 3D visualization and animation using a desktop application, and (3) a toolkit for a common scientific analysis environment (Matlab[®]). Due to the flexibility and low impact of the approach on providers, rapid progress was made. The system was implemented in all eleven US IOOS regions and at the NOAA National Coastal Data Development Center, allowing common delivery of regional and national oceanographic model forecast and archived results that cover all US waters. The system, based heavily on software technology from the NSF-sponsored Unidata Program Center, is applicable to any structured gridded data, not just oceanographic model data. There is a clear pathway to expand the system to include unstructured grid (e.g. triangular grid) data.

Introduction

The IOOS model data interoperability project focused on a common problem in the ocean community. A number of ocean simulation models are in common use, each with strengths for certain types of problems. Each of these models typically has custom output file conventions, and a suite of analysis, visualization and web distribution software that is custom built for their particular format. With different conventions and access methods, it is difficult to perform common operations such as comparing results from different models running in the same region.

¹ Detail address (Dec 2008-Nov 2009): NOAA IOOS Program Office, 1100 Wayne Avenue, Silver Spring, MD 20910

² Permanent address: US Geological Survey, Woods Hole Coastal & Marine Science Center, 384 Woods Hole Road, Woods Hole, MA 02543; PH (508) 457-2229; email: rsignell@usgs.gov

As an example, a user in the Gulf of Maine would like to compare predicted sea surface temperature from two different forecast models in Massachusetts Bay. When the user visits the web sites, they have to navigate two different systems, and are eventually presented with maps that have different color scales and no clear way to access the data (Figure 1). Instead of each modeling group and each user developing custom solutions for each model, a standardized approach, if implemented with an eye to the existing community of practice and tested with real use cases, would enable innovation, encourage a wider community of participation, and make better use of resources through software reuse (Hankin et al, 2009).

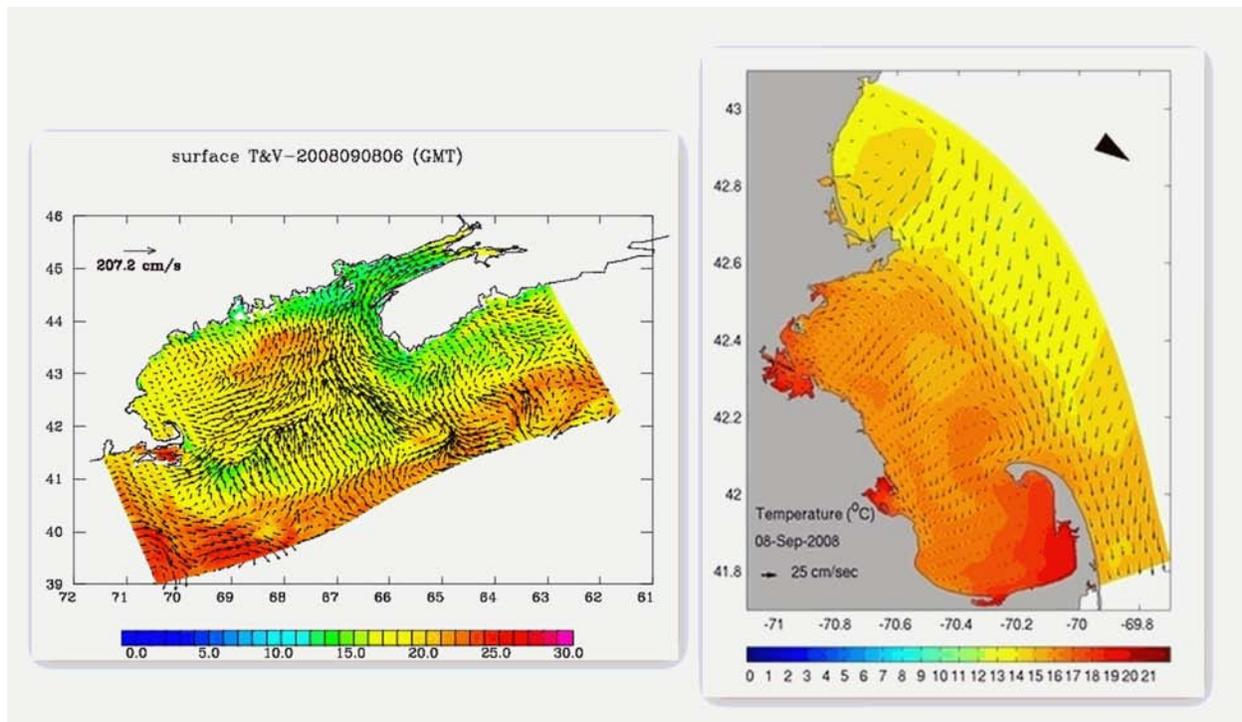


Figure 1. Comparison of sea surface temperature forecasts from the ocean modeling web sites of University of Maine (<http://www.umeoce.maine.edu/GoMPOM>) and University of Massachusetts, Boston (<http://www.harbor1.umb.edu/forecast/model.html>). The use of different color scales and map extents makes it difficult to qualitatively compare results, and difficulty in obtaining the actual forecast data makes it difficult to quantitatively compare results

The Integrated Ocean Observing System (IOOS) is a federal, regional, and private-sector partnership working to enhance the ability to collect, deliver, and use ocean information. A key part of this NOAA-led program is to supplement and integrate ocean measurements and modeling conducted by academic, government and private sectors in specific regions around the coastal US. Because each region has a unique collection of environmental, transportation and recreational issues of societal importance, each region has developed a customized observing and prediction system. IOOS is thus organized into eleven different regional associations and

observing systems (Figure 2). These regional systems both benefit from and feed information to the federally run observing and modeling systems (here called the National Backbone). IOOS seeks to improve the functionality of regional systems and the National Backbone by providing key funding for ocean observations, supporting workshops, developing standards and best practices, and establishing management structures that encourage and support collaboration.

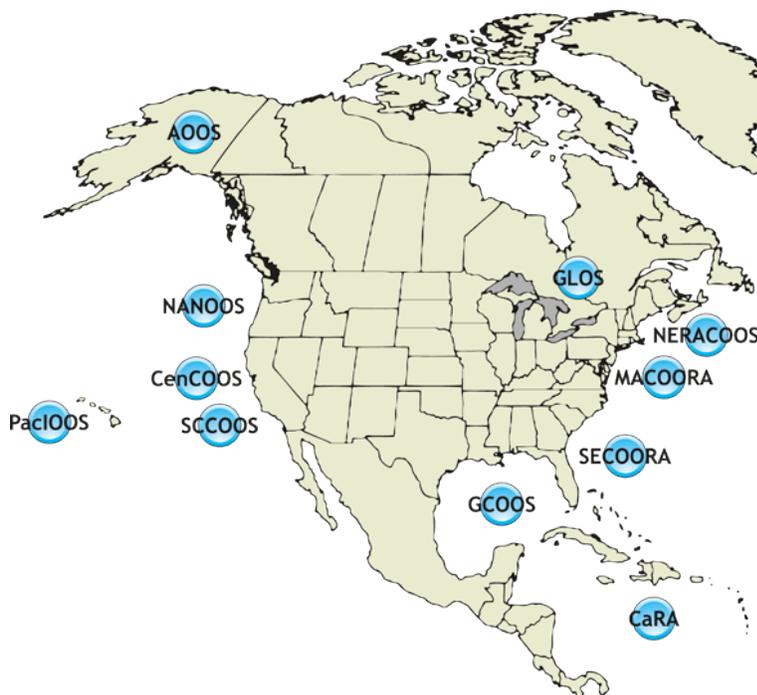


Figure 2. The eleven IOOS Regions across the United States. Figure from <http://www.usnfra.org>, used with permission of the National Federation of Regional Associations.

In the summer of 2008, IOOS held a community modeling workshop co-sponsored by the NSF-funded Ocean Observing Initiative (OOI). Attended by 57 members spanning federal, research and private sector, including modelers and stakeholders, the workshop produced a report with nine specific recommendations that would advance the state of ocean modeling in the US. One of recommendations was to “develop an implementation plan for a distributed, one-stop shopping national data portal and archive system for ocean prediction input and output data” (Ocean.US, 2008) which was to be a first-year task for a new yet-to-be-formed Ocean Prediction and Analysis Joint Working Group.

Realizing that it would take some time to assemble the Working Group and develop a master plan for all model inputs and outputs, the NOAA IOOS Program office proposed to implement a one year project to create interoperability and a single point of access for gridded model output data. The objective was to enable interoperability for a large amount of existing model output, but would provide a basis for future development. The USGS had been working on model data

interoperability for their collaborative projects on sediment transport modeling (Sherwood et al, 2004; Warner et al, 2008; Signell et al, 2008) and agreed to send one of their modelers to the NOAA IOOS office for a one year detail to lead the effort.

IOOS Data Standards

IOOS uses a distributed, federated approach to data management, with emphasis on standard access protocols via the web rather than on standard file formats. For gridded data, IOOS has approved the Open Geospatial Consortium (OGC) Web Coverage Service (WCS) and the OPeNDAP service with Climate and Forecast (CF) Conventions (de La Beaujardiere, et al, 2010)

The OGC WCS 1.1 works with only with gridded data that have rectilinear (1D) horizontal and vertical coordinates. If gridded data meets these requirements, WCS allows spatial and temporal subsetting of specific layers (e.g. variables output from a model), returning the data values along with geospatial coordinate information. Many ocean models do not meet these 1D coordinate requirements, however, having either curvilinear horizontal or stretched terrain-following vertical coordinates (Figure 3). Although complex native grid data could be interpolated to rectilinear coordinates to facilitate comparison and interoperability, this throws away potentially important scientific information. It is far better to preserve the data on the native grid, and build conventions and tools that can work with the more complex grid types. This is not only to preserve the scientific information for scientists, but for the developers who create custom end user applications for planners, decision makers and the general public. Even if they must create simple end user applications, they will have more options and the applications can be more powerful if the native grid data is preserved.

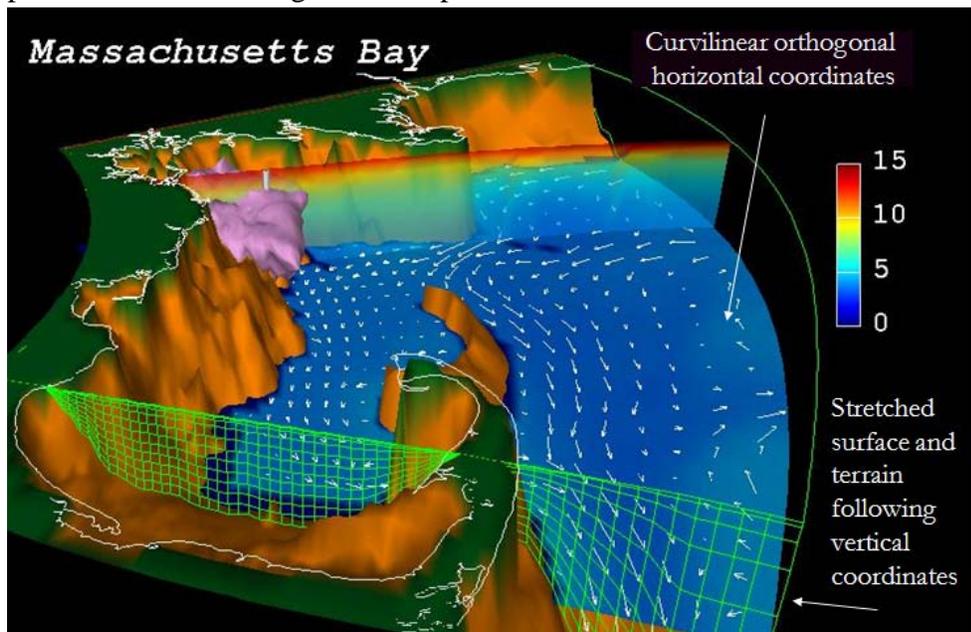


Figure 3. A typical ocean model grid (the UMASS Boston forecast model), with curvilinear horizontal coordinates and stretched surface and terrain and surface following vertical coordinates.

The CF Conventions were originally developed in 2003 by the climate community to create a more sophisticated set of conventions for NetCDF files (machine independent, self-describing binary files) to allow interoperability between the gridded data structures used by different models (<http://cf-pcmdi.llnl.gov>). The conventions were needed to facilitate access and analysis of data from the many different climate models used for the Intergovernmental Panel on Climate Change (IPCC) assessments, many of which had horizontal curvilinear coordinates and stretched vertical terrain-following coordinates. The CF group therefore developed attributes that allowed the horizontal coordinates to be explicitly specified (even if the coordinate variables are two-dimensional) and attributes to allow the functional representation of the vertical coordinate (e.g., the equation for the sigma coordinate) to be specified in terrain following models. Because ocean models use these types of coordinate systems as well, oceanographers contributed several specializations commonly used by ocean models (e.g., `ocean_sigma`, `ocean_s_coordinate`, `ocean_double_sigma`, `ocean_sigma_over_z`) to the initial CF 1.0 release and have continued to be engaged in the CF standards development process. The CF conventions have become a *de facto* standard in the atmospheric and meteorological communities and are now regarded as representing a “CF Data Model” that could be expressed in other formats than NetCDF.

Model Data Interoperability System Design

To encourage wide participation in the model data interoperability project with limited resources, an approach was needed that would be quick to implement, place minimal demands on the data providers, would not break their existing tool suites, and would not require additional software or hardware purchases. To convince providers that standardization had significant benefits, useful user tools had to be demonstrated, including toolkits and applications for standards-based access that simplified their current distribution methods and made their products accessible to more users.

The approach used relied on previously developed software components, many of them from the NSF-funded Unidata Program Center (Figure 4). At the heart of the system is the Unidata Common Data Model, which abstracts and unifies the NetCDF, GRIB, and HDF data models, and defines a GRID Feature Type which is the key to interoperability for structured grid model results. The CDM GRID Feature Type can be populated and accessed using high level methods in NetCDF-Java, a 100% Java library used on both the server and application side of the system. CF compliant NetCDF files can be read directly into the Common Data Model, but other GRIB and HDF files may also be read directly into the Common Data Model if their metadata is sufficient. If the metadata is not sufficient or incorrect, the NetCDF Markup Language (NcML) allows gridded data to be represented as a virtual dataset, where metadata can be modified or extended using XML without modifying the original data files. Once standardized in the Common Data Model, the THREDDS Data Server delivers data in a variety of web services, including OPeNDAP access and the Web Coverage Service. Finally, ncWMS, IDV and Matlab are applications that demonstrate several modes of access for the end user. These components will now be discussed in greater detail.

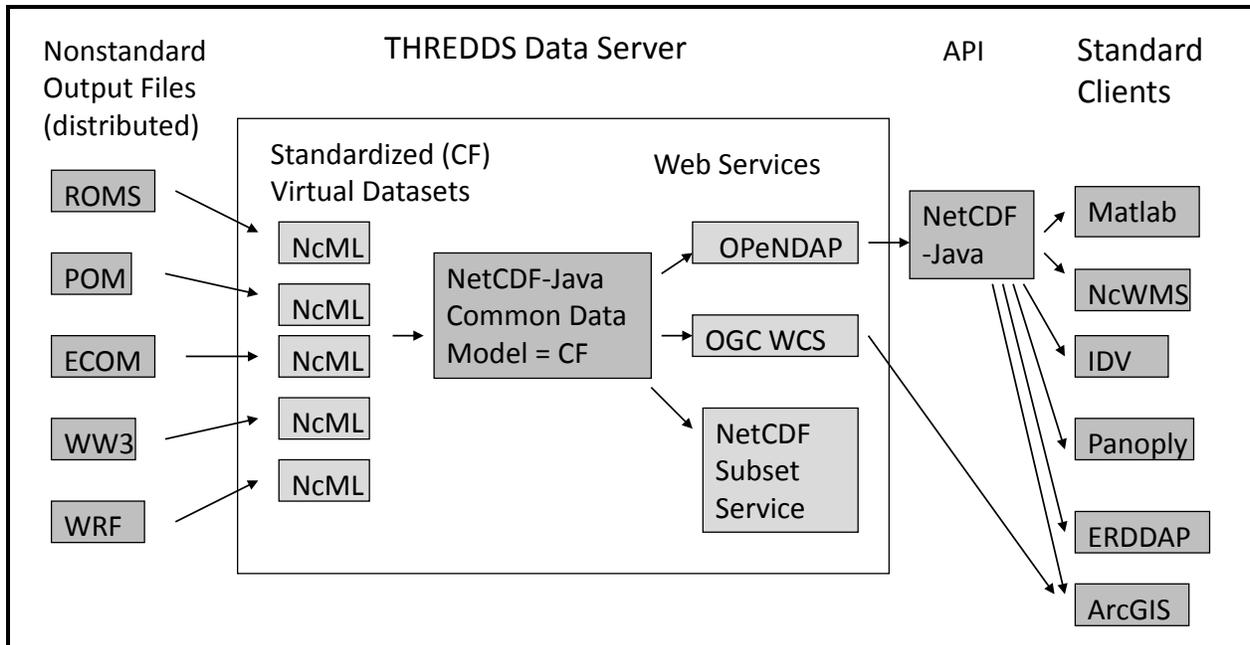


Figure 4. Schematic of the model data interoperability system design. Non-standard output files are converted into standardized and aggregated virtual datasets using the NetCDF Markup Language (NcML) middleware layer. Once the data has been standardized to the Unidata Common Data Model-Grid feature type, it is available via a number of different standards-based services and clients, providing model data interoperability for the user.

Components for Providers

The Unidata Common Data Model

The Unidata Common Data Model (<http://www.unidata.ucar.edu/software/netcdf-java/CDM/>) is a superset of the CF Data Model, and has three layers of increasing sophistication: (1) a data access component that handles data reading and writing; (2) a coordinate system component that identifies both georeferenced and non-georeferenced coordinate systems, and (3) a scientific feature layer, with feature types that represent commonly used data structures in the earth science community. The Common Data Model (CDM) Grid feature type handles structured grids, the CDM Point feature type handles scattered points, time series, profiles and trajectories, and the CDM Swath and CDM Radial and CDM Unstructured Grid feature types are under development. For model output, IOOS is using the CDM Grid feature type, which describes structured grids (uniform, rectilinear, curvilinear in the horizontal, uniform, rectilinear or terrain-following in the vertical). Data in NetCDF format with CF conventions can be read directly into the CDM without modification, but other gridded data formats and conventions can be handled as well, as described below.

NetCDF-Java

The Common Data Model is a conceptual model. Libraries of methods could be implemented in any language. Currently, however, the NetCDF-Java library is the most full-featured. (<http://www.unidata.ucar.edu/software/netcdf-java/>) . NetCDF-Java can populate the CDM Grid feature type with information from NetCDF3, NetCDF4, HDF4, GRIB, GRIB2, and NcML files, as well as from OPeNDAP servers and a dozen other formats and services. Gridded datasets supplied with the IOOS approved method of OPeNDAP with CF Conventions, therefore, are recognized as CDM Grids. An application written using NetCDF-Java to access data, utilizing the Common Data Model, can therefore enable interoperability across a wide range of input file types and conventions.

NcML

The NetCDF Markup Language (NcML) (<http://www.unidata.ucar.edu/software/netcdf/ncml/>) was originally created to allow additions and corrections to metadata in a NetCDF file using XML. For example, if the CF convention attributes for specification of the vertical or horizontal coordinates are missing or incorrect in the original NetCDF file, they can be supplied in an NcML file instead of rewriting the original file. The NcML file, which references the original NetCDF file, is then accessed instead of the original file. This technique allows the provider to generate NetCDF files without the need to understand CF conventions, leaving the job of creating the NcML to a CF expert. In addition, NcML can be used to create virtual datasets, presenting collections of NetCDF files or other gridded data files (e.g. GRIB, GRIB2, HDF4) as a single dataset to simplify access for the user. For example, a directory containing hundreds of GRIB2 files that each represents different variables at different time steps in a simulation can be represented as a single virtual NetCDF dataset using NcML.

An example of NcML that was actually used in the project, seen in Figure 5, shows: (1) addition of missing “coordinate” attributes to specify the curvilinear horizontal coordinates; (2) addition of missing dimensionless vertical coordinate attributes; (3) correction of the time units; (4) creation of a special aggregation dataset called a Forecast Model Run Collection (FMRC). The FMRC causes the data server to scan a directory to match files with a specified regular expression, and automatically generates a “best time series”, a continuous time series from a collection of three-day forecasts produced each day. The provider simply creates nonstandard forecast files and places them in a directory, and NcML allows the user access to the entire forecast archive, including the latest forecast, through a single CF-compliant virtual dataset.

```

<netcdf xmlns="http://www.unidata.ucar.edu/namespaces/netcdf/ncml-2.2">
  <aggregation dimName="runtime" type="forecastModelRunCollection" recheckEvery="10min"
    <scan location="/usr4/GoMOOS/cdfs/operational/"
      regexp=".*gomoos\.[0-9]{8}\.cdf$" dateFormatMark="gomoos.#yyyyMMdd"
      olderThan="5 min" subdirs="true" />
    <variable name="elev">
      <attribute name="coordinates" type="String" value="x y"/>
      <attribute name="_FillValue" type="short" value="16608"/>
    </variable>
    <variable name="ua">
      <attribute name="coordinates" type="String" value="x y"/>
      <attribute name="standard_name" type="String"
        value="barotropic_eastward_sea_water_velocity"/>
    </variable>
  </aggregation>
</netcdf>

```

Figure 5. NetCDF Markup Language (NcML) that aggregates existing forecast model output files into a single virtual dataset and also adds attributes for CF compliance, here adding the “coordinates” attribute for elevation and velocity to specify which variables contain the two-dimensional latitude and longitude arrays (here named “x” and “y”).

THREDDS Data Server

The THREDDS Data Server (TDS) is used to deliver the CF compliant gridded datasets created by NcML via OPeNDAP and several other standard web service protocols. The TDS uses NetCDF-Java on the server back end, and thus can read the common scientific formats mentioned previously. All files, whether standardized or not, can be served via OPeNDAP and an HTTP service which simply delivers a file in its original format. Gridded data that meet the CDM standards for the Grid scientific feature type and also have 1D coordinate axes (e.g. 1D lon and lat variables) can also be delivered through a subset service that allows users to specify a range of time, space, and variables to download to a NetCDF file. If the coordinate axes are uniformly spaced, the TDS can also serve data via the OGC Web Coverage Service. This allows users to inquire about the datasets, then select, subset and retrieve geospatial data as GeoTIFF or NetCDF files.

The TDS is a Java application that can be deployed on Tomcat or other Java application servers, making installation and configuration simple and platform-independent. As with other Unidata products, the software is free, with technical support via e-mail. Details of installation, configuration, security and more can be found on the THREDDS Data Server website (<http://www.unidata.ucar.edu/projects/THREDDS/tech/TDS.html>).

ncWMS

The final piece of provider software used in the IOOS data provider architecture is the NetCDF Web Map Service (ncWMS) developed by the University of Reading (Blower, 2009a). ncWMS is run in conjunction with TDS to deliver map visualizations of standards-conformant gridded data through an OGC Web Map Service, providing extensions for contain extra parameters such

as custom color range, and choice of linear or log scaling. Like the TDS, ncWMS is a Java application that can be deployed within Tomcat. The ncWMS service has been integrated into the TDS, but we still installed ncWMS because the TDS implementation was lagging behind the ncWMS, and did not support curvilinear coordinates at the time of deployment.

Components for Users

With standardized data access, interoperable applications for the end user become possible. Here we focus on three particular tools that use the Common Data Model via NetCDF-Java to enable standard access for (1) simple and rapid viewing of data maps in a browser (Godiva2 client for ncWMS); (2) more sophisticated 3D visualization and model comparison (the Integrated Data Viewer) and; (3) a toolbox for data access for a common scientific analysis package (the NetCDF-Java toolbox for Matlab).

Godiva2

Many people want to simply view data on a map in their browser. Godiva2 is a WMS client that comes bundled with the ncWMS distribution, and allows any CDM Grid data to be accessed in an identical fashion (Blower et al, 2009b). A variable is selected from a dataset list, a region is selected via zooming and panning on a world map with a variety of available base layers, a time range is selected using a calendar, depth is selected using a drop down box, and then color range, color pallet, and type of scale (log or linear) are selected (Figure 6). This defines the WMS URL that is sent to the ncWMS server, and the returned image is displayed in the browser. Godiva2 can allow for sequences of image requests to be made, allowing for simple animation, and ncWMS with a date range request can also deliver a KMZ file for animation in Google Earth or other applications. With the ncWMS service running on the machine where the data is resident, the maps are generated relatively quickly, so Godiva2 also becomes an important “first look” tool for scientists before data is downloaded.

Integrated Data Viewer

For users who are interested in more detailed three dimensional examinations of the data and are willing to invest some time in learning a new tool, the Unidata Integrated Data Viewer (IDV) (<http://www.unidata.ucar.edu/software/idv/>) has extensive visualization and data integration capabilities. Using the IDV, a user can perform construct isosurfaces of 3D data fields, interactive data slices across arbitrary horizontal and vertical sections, and combine data from multiple models (Figure 7). IDV contains many build-in functions, such as the ability to regrid data on the fly, allowing data on different grids to be compared easily. Users can also create their own functions easily using a Python language interface.

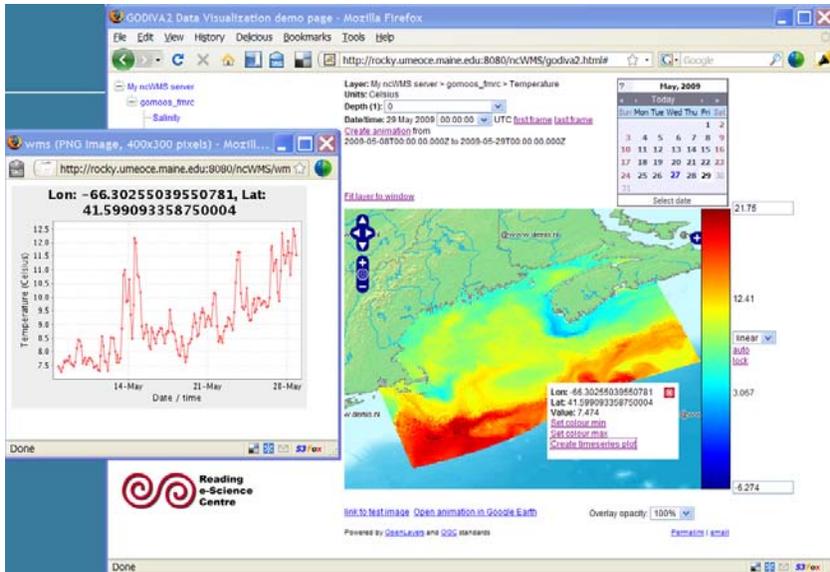


Figure 6. Using the GODIVA2 Web Map Service Client, users can use their web browser to choose ocean model variable, subset by longitude, latitude, depth and time, then create maps, animations and time series plots with custom color scales. Shown here is the GODIVA2 interface where a map of surface temperature has been selected along with a time series of temperature over a specified three week period at a fixed location.

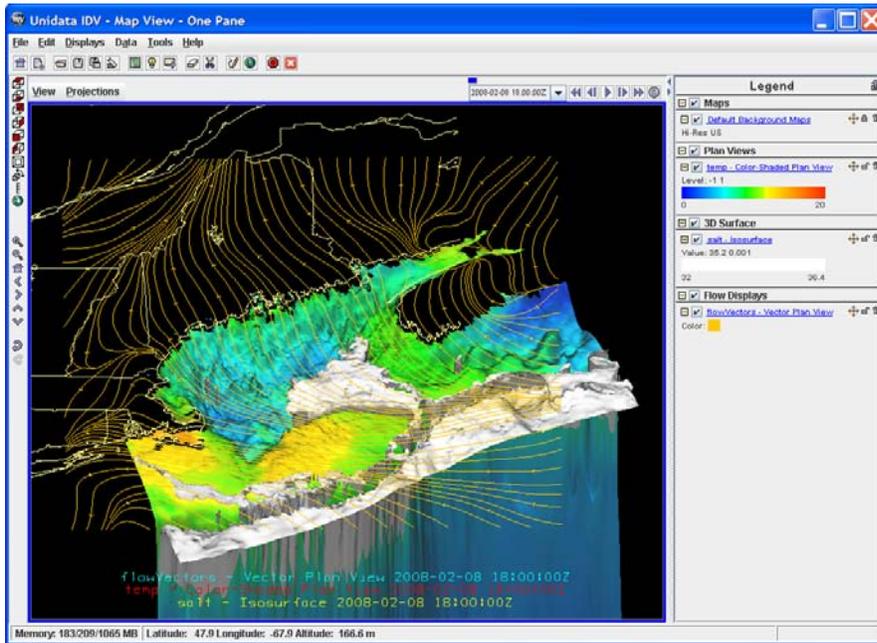


Figure 7. Using the Unidata Integrated Data Viewer (IDV), users can download data and perform sophisticated 3D visualization and analysis. Here a color-shaded map of bottom temperature and a white isosurface of salinity from the UMAINE Gulf of Maine ocean model are combined with surface wind streamlines from the UNH regional weather model.

NetCDF Java Toolbox for Matlab

Many oceanographers use Matlab for analysis and visualization. The NetCDF Java Toolbox for Matlab (<http://njtbx.sourceforge.net>) enables the power of the NetCDF-Java library for Matlab users so that they can perform both (1) low-level commands that simply read data from a NetCDF, GRIB, or HDF file, or (2) high-level routines that use the Common Data Model to obtain and manipulate georeferenced gridded data. Matlab users can access georeferenced data from any CF compliant model output or other gridded data using the same commands, removing the need for model-specific routines. Thus Matlab users can easily compare results between modeled and observed data, whether performing sophisticated analysis, or simply plotting model results and data in the same geographic range using the same color scale (Figure 8).

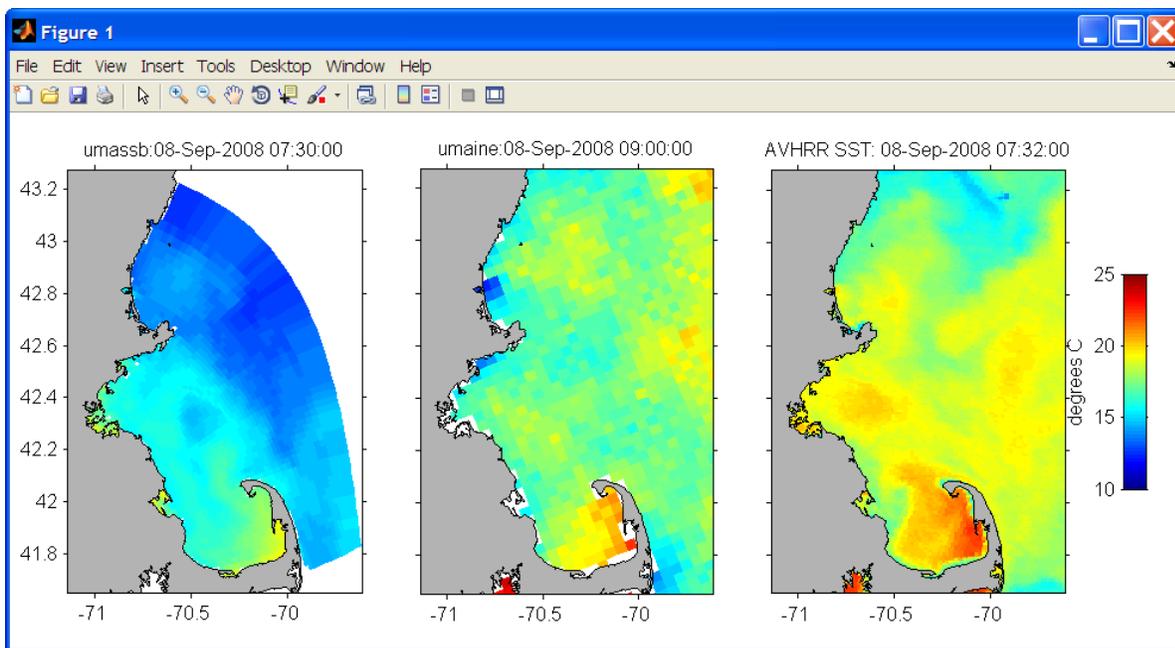


Figure 8. Using the NetCDF Java Toolbox for Matlab, users can use the power of both NetCDF-Java and Matlab to access numerous high-level routines that perform common tasks such as extracting model data at a fixed horizontal level or vertical section data along a specified transect. Here the forecast SST data from the two different models shown in Figure 1 are brought into Matlab along with SST remote sensing data and compared using the same color scale. Data from the two different models and from the observations is obtained using the exact same command, demonstrating interoperability.

System Implementation

When the IOOS project began, the basic system had already been prototyped by the Gulf of Maine Ocean Data Partnership Modeling Committee, which was using it to deliver CF compliant data from five different models: the ocean models ROMS, POM, and ECOM, the atmosphere model WRF and the wave model WaveWatch 3 (Signell, 2008). To kick off the project, the Gulf of Maine system was demonstrated to the National Federation of Regional Associations

Modeling Committee and the DMAC leads in the eleven IOOS regions. The objective was to identify at least one group in each IOOS region who was interested in implementing the system in their region, and then to schedule one-on-one consultations with these groups.

Over the course of the year, each group received consultations either in person or via WebEx to enable the system. First the group was interviewed and sample model output examined to identify problems or metadata gaps to be corrected using NcML. If there were collections of files that could be aggregated, the appropriate NcML was then added. Once CF-compliant aggregated virtual datasets that could be understood as CDM Grids were created, the TDS and ncWMS were installed. This step often first required installing the latest version of Java and installing Tomcat. The time to deploy the system varied substantially from site to site, but generally decreased from a few days to a few hours as the year went on, since problems tackled early were often met again with only slight differences later in the year. Examples of the NcML solutions for various types of dataset problems encountered on this project may be found on the NOAA Global Earth Observation - Integrated Data Environment (GEO-IDE) Guidelines and Best Practices Wiki (<http://www.nosc.noaa.gov/dmc/swg/wiki>).

By November 2009 the project had helped enable THREDDS Data Servers in 17 locations in all 11 IOOS regions to serve aggregated, CF-compliant model data. In addition, a THREDDS catalog that aggregated results of the NAVO NCOM model, part of the National Backbone of operational ocean models had also been developed. Because THREDDS catalogs can be accessed remotely by other THREDDS catalogs, a master catalog of IOOS-Compliant Modeling Data was created, providing a single point of access to all the regional THREDDS Data Server datasets. See the GEO-IDE wiki (<http://www.nosc.noaa.gov/dmc/swg/wiki>) for the location of this catalog, as a permanent location has not yet been determined. This catalog may be browsed and OPeNDAP Data URLs copy and pasted into OPeNDAP applications, or the catalog can be used directly with tools like IDV to explore and combine IOOS Compliant data from any model provider. The catalog provides access to model data covering much of the US waters in a completely standardized way (Figure 9).

Lessons Learned

The success of this project was largely due to: (1) a suite of flexible, powerful, easy-to-install software from Unidata; (2) a significant allotment of man hours for configuration, training and support to install the system throughout the community and; (3) enthusiasm of model providers in both the regions and in the federal operational centers to participate.

Unidata software is unusual in that it is both free and supported, and development by Unidata is largely set by two external committees: a Users Committee and a Policy Committee, both drawn largely from the academic geoscience community (mostly atmospheric). By relying on these committees, Unidata is virtually guaranteed that they are developing tools useful for the community they wish to serve. By providing training and support, and continuous software

improvement in response to user needs, they build healthy connections with the community as well. Their budget is largely from NSF Division of Atmospheric Sciences through a block-funded proposal every 5 years.

The Unidata Common Data Model exemplifies the Unidata approach of community-driven software. When the CF Conventions for NetCDF became the *de facto* community standard, Unidata built a data model for geospatial and other gridded data around it. In addition to the CDM Grid feature type, Unidata implemented a prototype CDM Point feature type for profile, time series and trajectory data, and are currently proposing conventions for CF that would enable this feature type. Because there is an alpha version software implementation, members of the CF Standards Committee and the community can test out how well the proposed conventions will work for their data, and give feedback that will help evolve the standard.

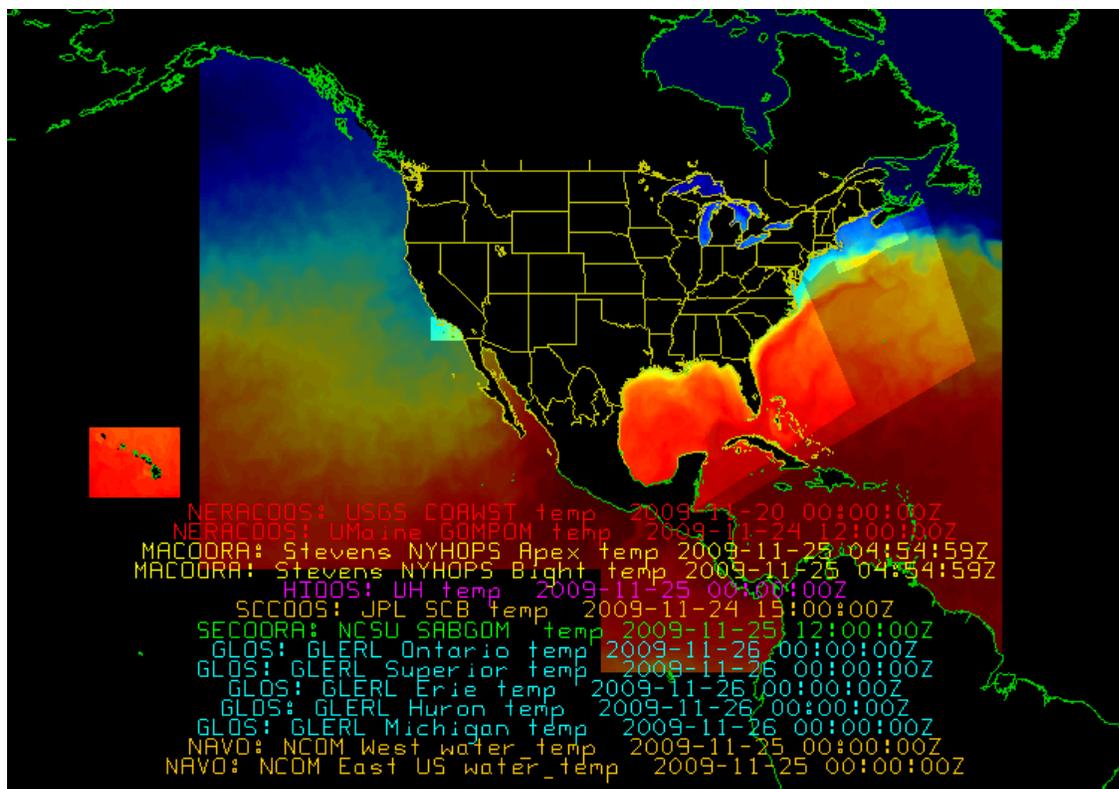


Figure 9. Most recent Sea Surface Temperature Forecasts from 14 different ocean models obtained on November 21, 2009, visualized with the IDV. The different models, variable names and time stamps of the latest forecast fields are recorded on the figure. This not only illustrates the coverage of available interoperable ocean model data, but serves as a useful “dashboard” application to visually check for system-wide forecast model problems.

If the Common Data Model provided the foundation of interoperability, NcML was the enabler. Before NcML, the way to achieve CF-compliance was to tell providers to rewrite their existing output files. Even if the model data providers could have been convinced to change their existing routines, they would have needed to understand the CF conventions, find time to write

and test the new procedures, possibly breaking their existing software tool suites in the process. They also would have had to manage collections of duplicate files, wasting effort and disk space. With the NcML approach, providers could allow an expert to create a middleware layer that provided a standardized interface to their non-standard data files.

Although the TDS is easy to deploy, developing the NcML to aggregate and standardize existing files can be daunting due to the breadth of approaches possible. In addition, technical people are typically overloaded and it makes little sense to force them to learn the details of a sophisticated standard (CF) and toolset (NcML) that they may need only one time to perform the custom-model-to-standardized-model mapping. Thus face-to-face meetings and one-on-one support proved to be a very effective way to get started, and follow-up support could be provided via e-mail or the IOOS Model Data Interoperability Google Group (http://groups.google.com/group/ioos_model_data_interop).

Conclusions

The one year project to develop model data interoperability for IOOS was successful. Our system design used existing software components that allowed modelers to serve their data in a standardized manner via IOOS-approved web services without modifying their original data files or their models. Users were then able access these standardized data streams using a variety of tools, from simple map-based browsing, to more sophisticated 3D visualization, to full scientific exploration on the scientific desktop. Servers were installed in all eleven IOOS regions, hopefully forming a nucleus of expertise in each region which can support continued expansion of regional gridded data products. The same techniques were used to aggregate and standardize data from the national operational center, providing nationwide access to ocean model products.

The system design is not only applicable to ocean model data, but any gridded geospatial data. NOAA is currently working on a project called a Unified Access Framework to Gridded Data across the agency that uses this design, and the USGS Council for Data Integration is using the approach for one of its three focus projects in FY2010. The approach should be extremely useful for projects that require large and diverse gridded data resources such as marine spatial planning.

Recommendations for Future Work

Although the interoperability system works well for structured grid data, there is a critical need to extend the system to work with unstructured (e.g. triangle based) grids. Unstructured grid modeling is growing in popularity for the coastal zone due to the ability to arbitrarily refine the grid around complex geometry and in regions of particular interest. One particular difficulty with unstructured grids is that instead of just accessing a certain range of data along the horizontal coordinate dimensions (a feature supported natively by NetCDF and OPeNDAP), one has to compute which elements or nodes fall within the bounding box and deliver topology data along with the scientific data and geospatial coordinate data. This function should take place on the server to avoid downloading a large amount of unnecessary data, which in turn requires a

standard syntax to communicate the subsetting operation to the server. Since the subsetting operation may take significant processing power, it would be desirable to handle these operations on a scalable computing environment such as the Amazon and Google computing clouds. Indeed, this is exactly the approach that the NSF-supported Oceans Observing Initiative Cyber Infrastructure project is taking with their Data Exchange effort (Meisinger et al, 2009). These ideas are being discussed on the (<http://groups.google.com/group/ugrid-interoperability>) Google Group, with a plan to develop CF Conventions for unstructured grids and implement in the Unidata Common Data Model as a new UGRID class in NetCDF-Java.

Other recommendations related to the overall system infrastructure are:

- Continue to work with additional gridded data providers to implement the system.
- For structured grids, develop CF conventions for staggered grids, masked regions, and velocity component relationships
- Continue to extend the high-level functionality in the NetCDF-Java toolkit for Matlab to reduce programming time for scientists and developers.
- Document lessons learned and solutions on the GEO-IDE wiki and the IOOS Model Interoperability Google Group
- Develop CF toolkits are needed for other common languages such as Python and IDL.

By focusing on improvements to the existing system, the community will see continuous improvement in their ability to access and work with ocean modeling data. This in turn should lead to more effective use of model data, more rapid feedback to the modelers, and more rapid improvement of models.

Acknowledgements

Thanks to the NOAA IOOS Office for making this a rewarding and productive detail: Zdenka Willis, Charly Alexander, Jeff de La Beaujardière, Marcia Weeks, Rob Ragsdale, Carmel Ortiz, Alex Birger and Carl Gouldman. Thanks also to the many modelers and technical folks from the federal centers and IOOS regions who made this work a success: Frank Bub, Scott Cross, Charles Carleton, Avichal Mehra, Arun Chalwa, Jordan Alpert, Hendrik Tolman, Huijie Xue, Steve Cousins, Mingshun Jiang, Doug Vandemark, Dennis McGillicuddy, Ruoying He, Avijit Gangopadhyay, Andre Schmidt, Changsheng Chen, Geoff Cowles, Alan Blumberg, Nickitas Georgas, Don Chesley, Will Perry, Bash Toulany, John Warner, Brandy Armstrong, John Wilkin, Dave Robertson, Ruoying He, Rob Hetland, Steve Baum, Damian Ruiz, Jorge Capella, Brian Powell, Jim Potemra, John Maurer, Rob Cermak, Yi Chao, Peggy Li, Quoc Vu, Antonio Baptista, Paul Turner, Joseph Zhang, Alexander Kurapov, Craig Risien, Dave Schwab, Guan Wang, Greg Lang, and Stuart Eddy. Any use of trade, product, or firm names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

References

- Blower, J.D, Mak, P., Caron, J., David, E., and A. Santokhee , 2009a, “Visualizing Earth Science data using a Web Map Service interface to the THREDDS Data Server”, *Geophysical Research Abstracts*, v11, EGU2009-5879.
- Blower, J.D., Haines, K., Santokhee, A., and C.L. Liu, 2009b, “GODIVA2: interactive visualization of environmental data on the Web”, *Phil. Trans. R. Soc. A*, 367(1890), 1035-1039. doi: 10.1098/rsta.2008.0180
- de La Beaujardiere, J., Beegle-Krause, C.J, Bermudez, L., Hankin, S., Hazard, L., Howlett, E., Le, S., Proctor, R., Signell, R.P., Snowden, D., and J. Thomas, 2010, “Ocean and Coastal Data Management”, In *Proc. OceanObs’09: Sustained Ocean Observations and Information for Society (Vol 2)*, Venice Italy, 21-25 September 2009, Hall, J., Harrison., D.E. and Stammer, D., Eds., ESA Publication WPP-306.
- Hankin, S., Blower, J.D., Carval, T., Casey, K., Donlon, C., Lauret, O., Loubieu, T., de La Villeon, L.P., Srinivasan A., Trinanes, J., Godøy, Ø., Mendelssohn, R., Signell, R., de La Beaujardiere, J., Cornillon, P., Blanc, F., Rew, R., 2009, “NetCDF-CF-OPeNDAP: Standards for Ocean Data Interoperability and Object Lessons for Community Data Standards Processes”, In *Proc. OceanObs’09: Sustained Ocean Observations and Information for Society (Vol 2)*, Venice Italy, 21-25 September 2009, Hall, J., Harrison., D.E. and Stammer, D., Eds., ESA Publication WPP-306.
- Meisinger, M., Farcas, C., Farcas, E., Alexander, C., Arrott, M., de La Beaujardière, Paul Hubbard, P., Mendelssohn, R., and R. Signell, 2009. “Serving Ocean Model Data on the Cloud”. In *Proceedings of the Oceans 2009 IEEE/MTS Meeting*, Biloxi, MS.
- Ocean.US., 2008. “The Integrated Ocean Observing System (IOOS) Modeling and Analysis Workshop Report”, Arlington VA, June 22-28, 2008, Ocean.US Publication No. 18, 21 pp. (http://www.ocean.us/files/MAST_Report_2008.pdf)
- Sherwood C. R., J. W. Book, S. Carniel, L. Cavaleri, J. Chiggiato, H. Das, J. D. Doyle, C. K. Harris, A. W. Nedoroda, H. Perkins, P.-M. Poulain, J. Pullen, C. W. Reed, A. Russo, M. Sclavo, R. P. Signell, P. Traykovski and J. C. Warner, 2004. “Sediment dynamics in the Adriatic Sea investigated with coupled models”. *Oceanography*, 17(4), 58-69.
- Signell, R.P., 2008. “The model interoperability experiment in the Gulf of Maine: A success story made possible by NetCDF, CF-1.0, NcML, NetCDF-Java, THREDDS, OPeNDAP and MATLAB”, *American Geophysical Union*, Fall Meeting 2008, abstract #IN22A-07.

Signell, R.P., Carniel, S., Chiggiato, J., Janekovic, I., Pullen, J.D. and C.R. Sherwood, 2008. "Collaboration tools and techniques for large model datasets", *Journal of Marine Systems*, v69, 154-161.

Warner, J.C., Sherwood, C.R., Signell, R.P., Harris, C.K., and Arango, H.G., 2008. "Development of a three-dimensional, regional, coupled wave, current, and sediment-transport model", *Computers and Geosciences*, v34, 1284-1306