13.1 TOOLS FOR ACCESSING DISTRIBUTED COLLECTIONS OF OBSERVED IN SITU DATA

Nancy N. Soreide[1]*, Donald W. Denbo[2], Jason E. Fabritz[2], Bernard Kilonsky[3],
John R. Osborne[4], L. Charles Sun[5], and Willa H. Zhu[2]
[1]NOAA/OAR/PMEL, 7600 Sand Point Way NE, Seattle, Washington
[2]Joint Institute for the Study of the Atmosphere and Ocean, University of Washington, Seattle
[3]University of Hawaii, Honolulu
[4]OceanAtlas Software, Vashon Island, Washington
[5]NOAA/NODC, Silver Spring, Maryland

1. INTRODUCTION

The process of locating and selecting individual observations from numerous large collections (millions of individual profiles, time series, etc.) of in situ data observations is complex. This paper describes a suite of tools that have been developed to facilitate the process of locating, subsetting, viewing, intercomparing, and downloading both data and metadata from geographically distributed collections of in situ data. Developed in Java and utilizing OpenSource components, development and maintenance for the tools is low, and they operate smoothly across a variety of hardware platforms and operating systems.

2. CLIMATE DATA PORTAL

2.1 *Data Available*

The Climate Data Portal (http://www.epic.noaa.gov/cdp) provides unified, networked access to geographically distributed data from ocean observing systems and historical data archives of in situ observed ocean data (Soreide *et al*., 2002). Data presently available through the portal includes:

- TAO El Niño buoy data (realtime)—PMEL
- Global sea level data (realtime)—University of Hawaii
- Global Temperature-Salinity Profile Program (GTSPP)—NODC
- International. Pacific Research Center in situ collection—University of Hawaii
- PMEL hydrographic data collection—PMEL
- WOCE Acoustic Doppler Current Profiler data—University of Hawaii

Data collections in progress or planned include (a) WOCE hydrographic data—Scripps Institution of Oceanography, and (b) Marine Environmental Data Service (MEDS) Canada, (c) ARGO profiling float data, and (d) WOCE/CLIVAR data. The Climate Data Portal may be extended in the future to allow access to other types of data, such as gridded data and model results and satellite data.

2.2 *Features*

Features of the Climate Data Portal include uniform network access to distributed observational datasets with sophisticated Java tools for subsetting large, irregular, in situ data collections, and data fusion and data download from a desktop Java application or the web.

2.3 *Technology*

The software is modern, scalable, and extensible. Built from off-the-shelf OpenSource components (Java, CORBA, LDAP, XML), the Climate Data Portal is poised to utilize new networking technology such as network Quality of Service and collaboration tools (Zhu and Denbo, 2002; Denbo, 2002; Fabritz and Denbo, 2002). Pending proposals will provide interoperability with the OPeNDAP (formerly DODS) networking protocols. Significant technology spin-offs from the Climate Data Portal have already been realized, and are described below.

3. OCEAN-SHARE

3.1 *Collaboration features*

Close collaborations amongst scientific colleagues involve frequent discussions of specific data sets. Given the geographic distribution of the investigators and the time and financial costs of frequent face-to-face meetings, a collaboration tool that utilizes the network for exploration and annotation of data, enabling colleagues in other locations to see those annotations in real time, is vital. OceanShare (http://www.epic.noaa.gov/collab/) has been developed to meet this need (Denbo, 2002).

The ability to interactively graph data from each of the collaborating sites, or from other remote network sites, point to a specific feature in the data while conducting a conference call, and have collaborators immediately see the item of interest and/or point out other features in the data, is extremely powerful and beneficial for these collaborations. Another useful aspect of this tool is the ability to establish a longer-term topic of discussion on-line, where individual participants may participate in a "networked conference session" to discuss a topic of

* *Corresponding author address:* Nancy N. Soreide, NOAA/PMEL/OD, Route: R/PMEL, 7600 Sand Point Way NE, Seattle, WA 98115-6349 (nns@pmel.noaa.gov)

interest, save the session, work off-line to investigate various aspects of this topic, and then rejoin the session later to post results for comment. It should be noted that so-called "screen scrapers," such as Microsoft NetMeeting, do not provide this functionality.

3.2 *Data Available*

All datasets available through the Climate Data Portal are available within the OceanShare environment. Local data can be added to the data fusion graphics by any collaborating colleague. Although OceanShare has been initially focused on oceanographic in situ observations and the needs of the Fisheries Oceanography community, pending proposals will extend its capabilities from oceanographic in situ observed data to include gridded data sets and model results.

3.3 *Technology*

OceanShare was developed by starting with the Climate Data Portal, and uses Jini and JavaSpaces, a simple unified mechanism for dynamic communication, coordination, and sharing of objects between Java™ technology-based network resources, to provide the basic collaborative framework and the Scientific Graphics Toolkit (SGT) to provide the interactive Java graphics.

4. JAVA OCEAN ATLAS

4.1 *Features*

Java OceanAtlas (http://www.epic.noaa.gov/epic/software/JavaJOA.htm and http://odf.ucsd.edu/joa/jsindex.html) (or simply JOA) is a software application for viewing and manipulating oceanographic profile data. JOA was designed primarily for oceanographic sections but is also useful for looking at data also in the latitude-longitude domain (Osborne, 2002). Java OceanAtlas provides import capabilities for a wide variety of profile data formats. Features include:

- Open a wide variety of standard oceanographic profile data file types.
  Including EPIC netCDF, WOCE bottle and CTD "EXCHANGE" files, WOCE netCDF bottle and CTD files, spreadsheet or tab-separated value, NODC SD2, and Mac OceanAtlas binary files.

- Powerful data collection filtering and selection via NdEdit.

- Property-property plots, profile (waterfall) plots, station value plots, contour plots, residual contour plots, and maps in a variety of projections with coastlines and bathymetry.

- Linked browsing between all views.
  Clicking on a point in any plot will identify the same point in all other views. A central data window shows the values of parameters at the selected point.

- Wide variety of built-in calculated parameters.
  Including many popular parameters, such as potential temperature, density (built-in and custom reference pressures), heat storage, specific volume anomaly, spiciness, sound velocity, etc. A custom calculator can be used to create new parameters by arithmetic operations and derivatives on existing parameters.

- Station calculations.
  JOA can calculate the mixed-layer depth (using a variety of techniques), or the integral (or weighted average) of any parameter between surfaces defined by any other parameter.

- Plots can be filtered by station or observation criteria.
  Station filters include geographic region, missing parameters, and individual station selection (include or exclude selected stations). Observation filters consists of up to four criteria that can be grouped using and/or logic. A criterion can test whether a particular parameter is inside (or outside) a given range or whether its quality code (if present) matches a certain value. Plots can show or highlight the observations that match the filter criteria.

- Extensive customization for your data or area of interest.
  JOA has tools for creating color bars (used for coloring plots as well as contouring), interpolation surfaces (used for contour plots), and color palettes. JOA also can save custom map settings, observation filter settings, and custom CTD decimation schemes. Existing sections can be sorted by latitude, longitude, date, and station number.

- Plot output can be printed, saved to GIF files or "printed" to Postscript files.
  MAC OS-X can print graphics to PDF files.

Written in Java, Java OceanAtlas is supported for Unix, Linux, Windows, and MAC OS.

5. NcBROWSE

NcBrowse (http://www.epic.noaa.gov/java/ncBrowse) (Denbo, 2002) provides flexible, interactive graphical displays of data and metadata (attributes) from a wide range of netCDF data file conventions, including:

- 1- and 2-dimensional cuts through gridded data
- Handles files with character variables
- Handles dimensions without associated variables
- Properly decodes several time formats
- Parametric plots (e.g., temp-salinity)
- Flexible mapping of independent variables to axes.

- Vector plots, Animations, 3D visualizations
- Graphics appearance is easily modified
- Publication quality graphics, printing support
- Remote netCDF file access with simple HTTP or DODS/OPeNDAP
- Access to remote OPeNDAP datasets (beta version)
- Supported on Unix, Linux, Windows, and Mac OS-X

NcBrowse is registered with Unidata as a netCDF client, and is increasingly referenced as an OPeNDAP Java client. It is popular with users, with over 5497 downloads by 2524 unique sites in 56 countries, and an active email discussion group.

6. JAVA TOOLS

A number of useful tools have been utilized in development of the software applications that have been described in this paper, and these too are freely available for download.

6.1 *Scientific Graphics Toolkit (SGT)*

SGT (http://www.epic.noaa.gov/java/sgt/) facilitates easy development of platform independent Java applications to produce highly interactive, flexible, publication quality, object oriented graphics of scientific data (Denbo, 2000; Denbo, 2001). Features include user settable or automatically scaled axes, sophisticated, automatically self-scaling time axes, labels as movable, customizable objects, automatic generation of legends to explain the data being displayed, and many more.

SGT is the graphics engine for the Climate Data Portal, OceanShare, and ncBrowse. SGT conforms to the JavaBeans component architecture guidelines and is presently being made more accessible to a wider community of non-expert users by the creation of easy-to-use SGT JavaBeans.

SGT is freely available, and the download site includes many demonstration programs as well as a streaming video tutorial and an active email discussion group. It has been very popular, with 11042 downloads from 3955 unique sites in 65 countries as of September 2002.

6.2 *NdEdit*

NdEdit (http://www.epic.noaa.gov/epic/software/JavaNdedit.htm) is an interactive Java tool for selecting and subsetting large collections of in situ observations (Osborne and Denbo, 2002). This is a challenging task that has not been addressed successfully by any other tool that we are aware of. The user can graphically view the locations of selected data sets in any 2D view of space and time (e.g., latitude-longitude map view or latitude-time view) and use graphical tools to select a desired subset of the displayed data (Figure 1). Actions such as selecting or zooming in one 2D view are duplicated in the other 2D views. NdEdit also features coastline and bathymetry display in the longitude-latitude view, user-settable horizontal and vertical axes, and zooming. Data sets in a selected box, polygon, or multi-section line can be deleted or retained.

NdEdit is written in Java, for platform independence, and is utilized by the Climate Data Portal, OceanShare, the Java Ocean-Atlas, and as a stand alone tool for the PMEL EPIC (http://www.epic.noaa.gov/epic/) system for management, display, and analysis of in situ ocean observational data.

6.3 *The EPIC Java library*

The EPIC Java Library (http://www.epic.noaa.gov/java/epic/) is a library of useful Java utility classes for the developer of Java applets and applications. These classes primarily support the development of scientific applications that deal with geophysical data sets. It includes Swing classes that provide a graphical user interface (GUI) for setting dates, latitude, and longitude, Formatter classes for handling text, dates, numbers and messages, and Utility classes with date and time facilities and miscellaneous utility classes. All are fully documented with Javadoc pages at http://www.epic.noaa.gov/java/doc/.

7. WEB TOOLS

In contrast to the Java tools described in this paper, several web tools have also been developed. Both feature data selection based on user input time, space and keyword criteria, and are freely available from http://www.pmel.noaa.gov/epic/.

The Web access to the Climate Data Portal (http://www.epic.noaa.gov/cdp/cdpweb.htm) provides access to the distributed data collections as the Java Climate Data Portal application, and includes the ability to locate, subset, and co-plot data from the different in situ data sources (Zhu and Denbo, 2002). An interactive Java map applet provides geographic range selection features. Graphics are created with the Java Scientific Graphics toolkit and presented to the user on the web in GIF format.

The EPIC web browser (http://www.epic.noaa.gov/epic/ewb/) provides customized access to local collections of time series and profile ocean observations (Zhu *et al.*, 2001). A Java map applet allows for data selection based on geographic range. The user can select from many special customized plots, such as a histogram plot of the time series values and stacked plots of all selected time series, or depth vs. temperature, salinity, oxygen, sigma-t plots of ocean profile data. Although developed for selecting, graphing, and downloading data from oceanographic data collections, it has been found that these tools apply well to atmospheric time series and profile data collections as well.

A collection of EPIC tools for data input/output and web presentation of data are available for download from the web. The EPIC I/O library, epslib, consists of a set of

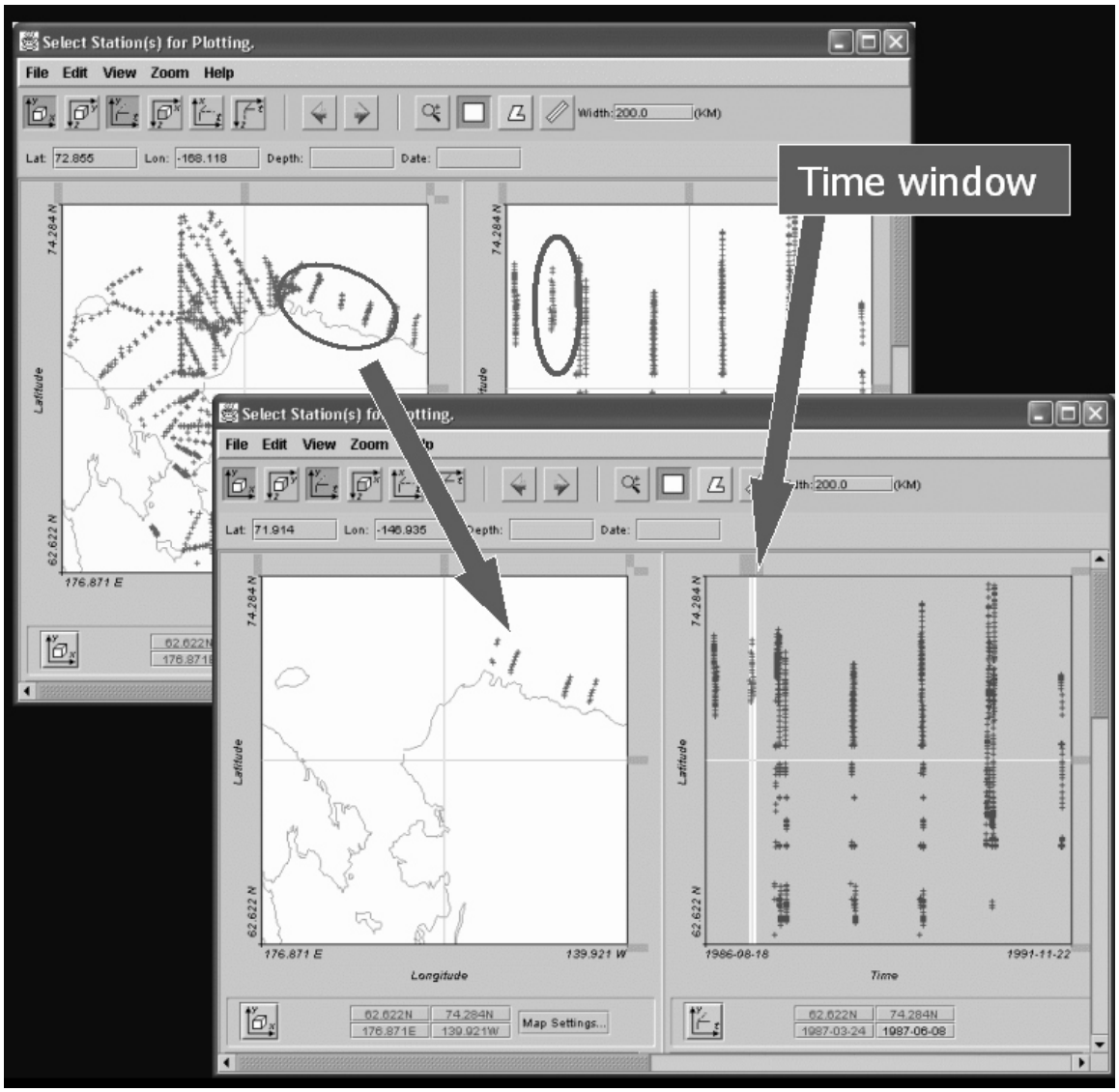

Figure 1. Screen snapshot of "ndEdit," the Java data selection tool utilized in the Climate Data Portal, OceanShare and Java OceanAtlas. Linked windows showing two views of in situ data locations (latitude-longitude map view and latitude-time view). The user can filter in the time window to subset the data or view the data locations.

routines which provides data file Input/Output for the EPIC system for management, display and analysis of oceanographic or meteorological data (http://www.pmel.noaa.gov/epic/software/epslib.htm). Epslib supports multiple data file formats, including netCDF format, transparently (Figure 2).

A collection of useful utilities which simplify creation of web pages to obtain input from a user and point to a static URL or to pass the input to a Unix script for dynamic generation of graphics or listings on the web are available at http://www.pmel.noaa.gov/epic/software/webutil.htm.

8. CONCLUSIONS

A significant collection of Java and web applications and tools has been developed and made freely available for use in the community. Please see http://www.pmel.noaa.gov/epic/software/ep_java.htm for more information, or to download these tools.

9. REFERENCES

Denbo, D.W., 2000: The Scientific Graphics Toolkit. Proceedings of Oceans '99 MTS/IEEE Conference, 13–16 September, Seattle, WA (on CD).

Denbo, D.W., 2001: Interactive graphics toolkit for Java applications and web applets. 17th Conference on

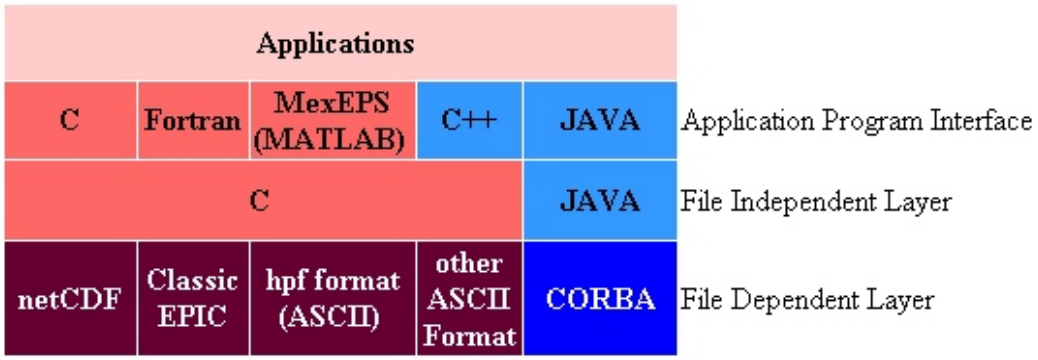

Figure 2. Application programs call a "file independent layer," which is a set of generic input/output routines for data file I/O. A lower level "file dependent layer" of routines actually performs the I/O. Application programs are independent of the data file formats, and additional formats can be supported by the addition of a single set of file format dependent routines.

Interactive Information and Processing Systems (IIPS) for Meteorology, Oceanography, and Hydrology, AMS, 14–19 January 2001, Albuquerque, NM, 372–375.

Denbo, D.W., 2001: ncBrowse: A graphical netCDF file browser. 17th Conference on Interactive Information and Processing Systems (IIPS) for Meteorology, Oceanography, and Hydrology, AMS, 14–19 January 2001, Albuquerque, NM, 473–476.

Denbo, D.W., 2002: Fisheries oceanography collaboration software. 18th Conference on Interactive Information and Processing Systems (IIPS) for Meteorology, Oceanography, and Hydrology, AMS, 13–17 January 2002, Orlando, FL, 245–246.

Denbo, D.W., 2002: Using a LDAP directory server for environmental data discovery. 18th Conference on Interactive Information and Processing Systems (IIPS) for Meteorology, Oceanography, and Hydrology, AMS, 13–17 January 2002, Orlando, FL, 58–59.

Fabritz, J.E., and D.W. Denbo, 2002: Automated scanning of observational datasets for the generation of formal metadata. 18th Conference on Interactive Information and Processing Systems (IIPS) for Meteorology, Oceanography, and Hydrology, AMS, 13–17 January 2002, Orlando, FL, 60–61.

Osborne, J.R, 2002: Java OceanAtlas: A Cross-Platform Application for Visualization and Selection of Oceanographic Profile Data, Presentation for NVODS/DODS Technical Working Conference, Boulder, CO, January 9–11, 2002. (http://www.epic.noaa.gov/talks/oz/dods).

Osborne, J.R., and D.W. Denbo, 2002: NdEdit: Interactive data selection tool. 18th Conference on Interactive Information and Processing Systems (IIPS) for Meteorology, Oceanography, and Hydrology, AMS, 13–17 January 2002, Orlando, FL, 68–69.

Soreide, N.N., C.L. Sun, B.J. Kilonsky, D.W. Denbo, W.H. Zhu, and J.R. Osborne, 2002: A climate data portal. 18th Conference on Interactive Information and Processing Systems (IIPS) for Meteorology, Oceanography, and Hydrology, AMS, 13–17 January 2002, Orlando, FL, 66–67.

Zhu, W.H., and D.W. Denbo, 2002: Distributed data servers and client interfaces for the climate dataportal. 18th Conference on Interactive Information and Processing Systems (IIPS) for Meteorology, Oceanography, and Hydrology, AMS, 13–17 January 2002, Orlando, FL, 64–65.

Zhu, W. H., D. W. Denbo, and N. N. Soreide, 2001: EPIC web browser: On-line access to large oceanographic and meteorological in-situ data sets. 17th Conference on Interactive Information and Processing Systems (IIPS) for Meteorology, Oceanography, and Hydrology, AMS, 14–19 January 2001, Albuquerque, NM, 376–379.