

IM2: A Web-based Dissemination System for Now and Forecast Data Products

Michael Piasecki¹, Luis Bermudez², Saiful Islam³, Frank Sellerhoff⁴

Abstract

This paper describes the various steps to develop a web-based information system for now- and forecast oceanographic and meteorological data products. An initial user survey delineates the user needs and requirements and determines to what extent and level the web-system would need to be developed. In order to have a globally relocatable system it needs to be operating system independent (achieved through the use of JAVA) and have a underlying data structure that separates data content from data description (achieved through the use of metadata concepts) and at the same time permit data display of various kinds, for example time series for stations, contour plots for spatially continuous data variables, animations over time. Also, the system needs to be flexible to accept various raw data sets in “pull” and “push” mode, be easily adaptable to different data products, and at the same time permit speedy access and response times for the user. The latter involved the use of data compression techniques, and also the use of “thin” client configurations to accommodate less powerful client-end machines.

Introduction

Web-based information systems have become more and more state-of-the-art in all fields of science and engineering. Funded through the National Oceanographic Partnership Program (NOPP, 2003), and under the leadership of Drexel/URI, a group of researchers have developed a globally relocatable information system including a Web-based dissemination sub-system (IM2) for now- and forecast data products. The test bed for this is the Narragansett Bay and adjacent RI coastal waters, which serve as a demonstration of the practical use of the system to support environmental monitoring, marine pollutant transport and fate, marine transportation, and search and rescue operations.

The development objectives for the IM2 system are twofold. First, it is supposed to cater to a wide user audience with varying demands on data analysis, hardware setup, and variety of sought data products. Secondly, it needs to be a system that is fast (small wait times for client and server communications), is operating system independent and flexible in its setup requirements (therefore globally relocatable), and that is user friendly by providing an intuitive user interface that does not require a steep learning curve. Also, the system was to be designed that no additional software requirements would arise through licensing fees of proprietary components.

¹ Assist. Prof., Dept. of Civil, Architectural & Environ. Engrg, Drexel University, Philadelphia, PA 19104

² Research Assist., Dept. of Civil, Architectural & Environ. Engrg, Drexel University, Philadelphia, PA 19104

³ Research Assist., Dept. of Civil, Architectural & Environ. Engrg, Drexel University, Philadelphia, PA 19104

⁴ CEO smile consult GmbH, Vahrenwalder Str. 7, D-30165 Hannover, Germany

This paper describes the development of IM2 using a number of approaches to address the above objectives, more specifically i) the use of user survey to identify the needs and priorities of the community, ii) the use of state-of-the-art technology like JAVA and XML to develop an OS independent information system, iii) the use of object oriented paradigms to facilitate the development of the software iii) the use of a Metadata standard to develop the system specific subset of data descriptors, iv) the use of a flat file system for storage, compression techniques, and a multi-server configuration to increase the performance of client- server communications, v) and the utility of a workshop and demonstration phase to receive feedback from the user community to improve the system functionality.

User Survey

Due to the large and different amount of data that is either collected or produced, the decision was made to investigate, via a survey, what exactly the data needs in the Narragansett Bay user community are. After collection and compilation of the survey returns, the results of this survey were used to aid the Drexel research team to focus on specific issues that relate to the design, content, and most of all functionality of the proposed data dissemination web-site. To this end a small survey was developed that sought

- to identify the user group at large and in detail
- to identify the data variable that is being sought
- to identify the time mode, i.e. hind- now- forecasting
- to identify the display frequency
- to identify the preferred display mode, i.e. time-series plot, contours, etc

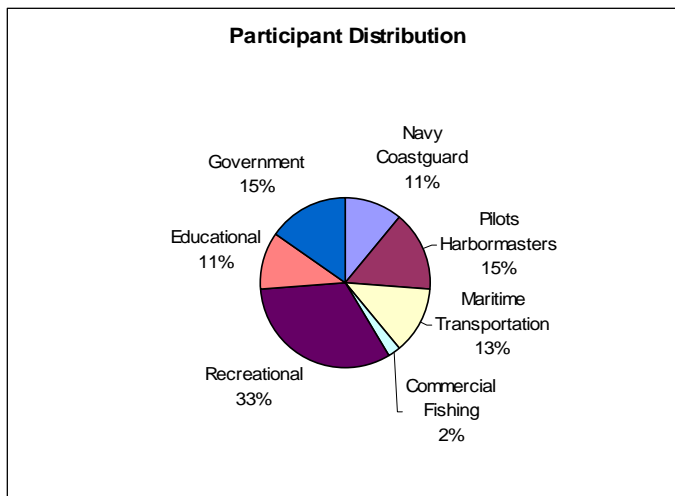


Figure 1 Survey Participant Distribution

The survey was send out to a number of institutions, groups, individuals, and governmental departments in order to capture a large and most encompassing cross-section of potential users in the Narragansett Bay area. Because of the relatively fine-grained initial identification of user community groups, participant numbers from group to group varied substantially (numbers from the recreational community were quite high, while

participants from Navy and Pilot organization remained limited). However, it was felt

that as many identified groups, even when low in numbers should be polled separately in order to have a more detailed description as far as the need for specific data products are concerned.

The survey results are based on a response rate of about 50% resulting in 47 surveys that were returned. While this number may not seem to be very high and therefore representative, it was felt that the number of returned surveys provided a sufficient basis for a good assessment of the survey effort, particularly in light of the large diversity among the participants, i.e. the institutional affiliation, which is shown in Figure 1.

In order to obtain a better understanding what state variables the user community is most interested in seeing disseminated or displayed, the survey asked to identify and rank the available variable information on a scale from 1 (least) to 5 (most). The ranks were added up and then divided by the number of participants to compute an average ranking. From Figure 2 it becomes clear that wind, waves, currents, visibility, water, and air temperature are the most desired states (in that order) by the community.

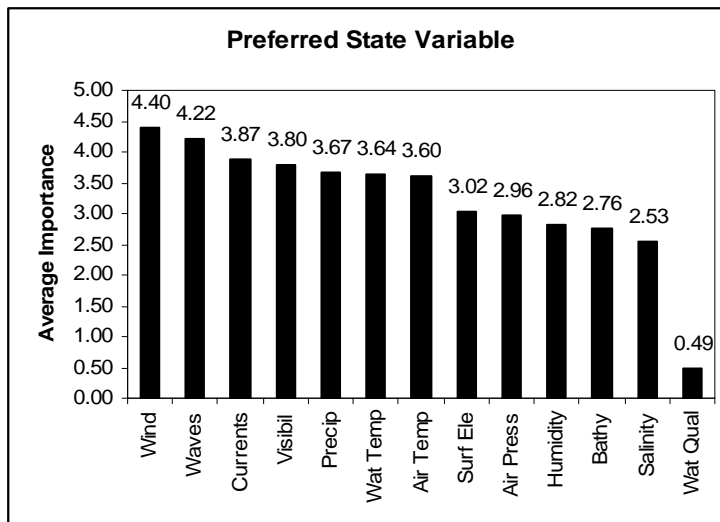


Figure 2 Data Preference

While the overall preference is an important indicator for what needs to be made available, it is also important to notice that preferences change from user group to user group. For instance, surveys submitted by members of the Naval Station (Submarine Warfare Center) indicated that bathymetry, salinity, and water temperature are of significant importance, which can be attributed to the mission of the

Warfare Center. Additionally, while water quality received only a small score, members of the scientific community (both University and governmental) placed high marks on the importance of knowing about the water quality situation even though the list of potential pollutants is long and would certainly need to be more specific. Another important target variable was the reporting of Wave Conditions, coming in second, which was not included in the initial list of available products.

The survey also tried to identify what casting mode is preferred by the user community. The choice was fore-casting, real-time observations, and access to historical records. This part was particularly important for identifying what data

products needed to be generated for the user community, and also whether a history record was needed mandating the provision of a large data storage facility. The results showed a clear overall average preference for real-time and forecasting information. As for the data products (variables) preferences in the previous section, this result is dominated by the structure of the user community that in general, preferences real-time and fore-casting data products, while historical records are only important for the scientific and regulatory bodies of the Narragansett Bay community. These results clearly indicated that the proposed development of fore-casting models and the resulting fore-casting data products was prudent and reflected a data product line that met the need of potential users. As a result, the development of the WEB-portal, IM2, needed to include options for both real-time and fore-casting data display.

The last item addressed in the survey was aimed at finding out how frequent the data should be made available. The strong desire for real-time data (see above) clearly indicated that continuous data display should be a high priority. Figure 4 shows the results in percent of participants that ranked a certain frequency as highest priority, where the 33% vote for continuous display reflects the desire for real-time observations.

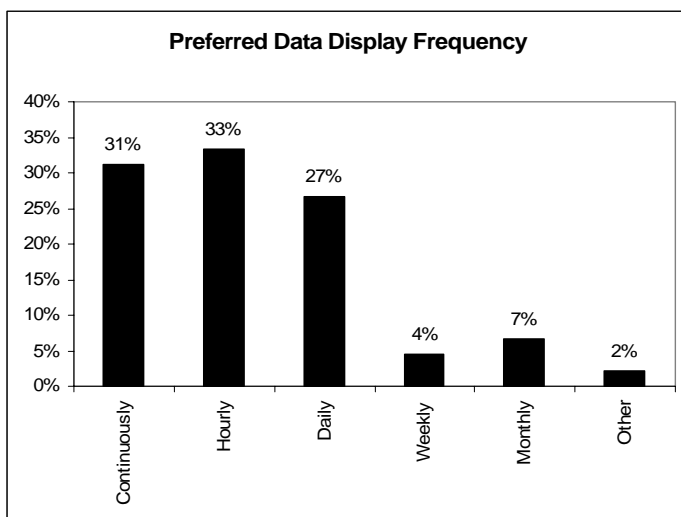


Figure 3 Preferred Data Display Frequency

Clearly, most of the participants thought that data dissemination on an hourly basis would satisfy their needs best, followed by daily updates. Since the daily frequency is a subset of an hourly display, it was concluded that data should be made available on an hourly basis. This was an important decision because some of the data is collected or computed with time-steps that are much shorter hence creating an overhead that would not be used. The only exception was

the real-time mode where data was displayed as often as possible, pending the frequency with which it was sent to the data server and with which it could be processed prior to display.

Metadata and Data Storage

The concept of metadata is to describe data in a standardized form, hence, “data to describe data” (Gilliland-Swetland, 2000). Instead of searching through piles of collected data (often insufficiently described), data availability is searched based on the metadata description of the raw data set. The objective for this project was to

identify a minimum metadata set that would permit the exchange of data sets among the collectors/creators and the web-based system and at the same time permit the end-user an easy and intuitive access to the data holdings of the IM2 system. In addition, the metadata set had to be simple and short to reduce capacity demands on the data transfer bandwidth when pushing or pulling the data sets through the internet. On the other side, the set had to be extensive enough to sufficiently describe the given data sets and flexible enough to be expanded or altered when used for a different task elsewhere.

Currently, there are a number of metadata standards in circulation, all of which are in a state of continued expansion and development. The Federal Geospatial Data Committee (FGDC, 2002) has been leading the effort to develop metadata standards for the US-based scientific and public community. The FGDC coordinates the development of the National Spatial Data Infrastructure (NSDI, 2002), and hosts about a dozen sub-committees and workgroups that attempt to develop metadata standards for a variety of disciplines that span the entire spectrum of the scientific community, e.g., Ground Transportation, Marine and Coastal Spatial Data, Soils, Cadastral, Geologic, and Remote Sensing, to name just a few. Perhaps the most intensive and globally reaching metadata initiative is carried out by the International Standards Organization (ISO) through its Technical Committee 211 (TC211). The ISO TC211 combines over 50 participating and observing member states and their respective Standard Organizations. The scope of the TC211 is the general “Standardization in the field of digital geographic information”. Findings and recommendations of the TC211 are disseminated through the publication of ISO norms, of which the 19115 norm deals with metadata. This metadata standard defines a much larger set of descriptive elements that are grouped in mandatory, optional, and voluntary sets, and which has also become the Draft International Standard (DIS) (ISO, 2002).

While above standards suggest a comprehensive coverage and therefore appear to be an obvious choice, these standards are designed to be quite generic and to cover a large spectrum of geo-science communities. As a result, their strengths are also their weaknesses, as the given metadata element set is too extensive to work well for a system framework that requires only a bare minimum set. In fact, the objectives of the system were to merely identify a metadata set such the data products could be easily found in the IM2 system. A more suitable standard for this task is the Dublin Core Metadata Standard (Dublin Core Metadata Initiative DCMI, 2002) that emerged in 1995 and has since seen a wide acceptance in the library community. Its advantage is the small number of required elements and the ease with which it can be extended to include descriptions that are not part of the standard set. A list of the DCMI elements used for the development of the IM2 system in addition to the extensions is given in Table 1. All data sets use this boiler plate template to fully describe the received datasets. All internal querying, parsing, retrieval, and subsequent task execution is based on the metadata information provided in the table.

Table 1 Dublin Core Metadata Elements used in IM2

Metadata Element	Purpose	Type
Title	General Title for Data-set	character string
Subject	Variable identifier	character string
Description	What data variable is collected?	character string
Publisher	Who posts it on the WEB?	character string
Contributor	Who collected/prepared it?	character string
Date	When the data collected?	real
Type	What type of data?	character string
Format	Is the data text/ASCII?	character string
Identifier	Pointer to the data file	URL
Source	Where did the data originate from?	character string
Language	What language is being used?	character string
Coverage	Temporal/spatial coverage of the data set?	
Box	What is the Geo-spatial extent?	
Northlimit	Northernmost limit	real
Southlimit	Southernmost limit	real
Westlimit	Westernmost limit	real
Eastlimit	Easternmost limit	real
Period	What is the temporal extent?	
Start	Time for first data digit [millisecs]	integer
End	Time for last data digit [millisecs]	integer
Frequency	Δt between one data item and the next	Real
Rights	Copyright identifier	character string

The cells highlighted in green represent standard DCMI elements, while the ones highlighted in turquoise are DCMI qualifiers, i.e. extensions to the given set generated by following the extension rules of DCMI.

Data Storage

In view of the objective to develop a globally-relocatable web-based information system, much of the initial planning centered on the need how to facilitate this objective. Important aspects to be considered were: costs (license fees), maintenance, performance (fast access, querying & retrieval), and the need to accommodate a variety of platforms, like WINDOWS, MAC, LINUX, and several other UNIX derivatives (Lehfeldt and Piasecki, 2002; Bermudez and Piasecki, 2002). While several public domain Rational Database Management Systems (RDBMS) are available to the user community (like MySQL, 2002; PostgreSQL, 2003) and as such would not demand a license fee (unlike systems as ORACLE, Dbase, ACCESS), the decision to install and implement a RDBMS must take all other aspects as well. To

aid in the decision making process, we set up a decision matrix to weigh the different components against each other, as shown in Table 2.

Table 2 Decision Matrix for Choice of Storage Model

Database Type	Pros	Cons
Proprietary Database	Support, functionality	Costly, a lot of overhead
Pub Domain Database	Free, less overhead (lean)	No support, vulnerable
File System Database	Very lean, no costs, easy	Limitation on structure

Since the complexity of the data structure is quite manageable, that is the number of different data files (data variables) is limited and there is little to none cross-referencing necessary among the different data items, the choice was an easy one to make. Because of its leanness, ease of installation, no additional training requirements, and no-cost allure, the File System DataBase (FSDB) was selected and implemented in the server system, as shown schematically in Figure 4.

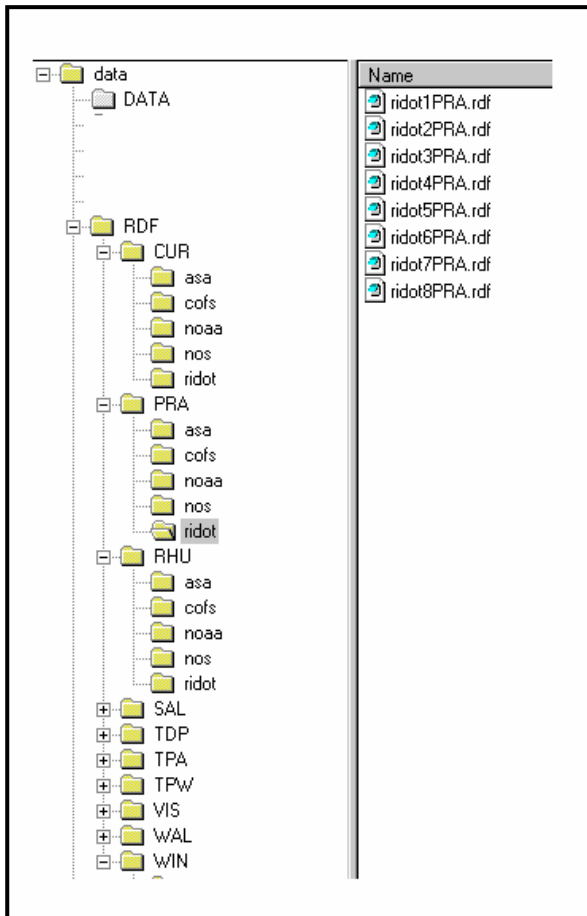


Figure 4 Schematic of the File System Database

The data-root (**data**) is essentially divided up into two branches: one for the actual data files called **DATA**, and one for the data file descriptions called **RDF**. These two directories are mirror images of each other as far as the subdirectory structure in each of these branches is concerned. They also contain the exact same number of files, i.e. each file containing “real” data has a complementary one located in the RDF branch.

Each branch is then subdivided into the 11 data variables (Wind, Salinity, Air Pressure, Relative Humidity, Water Surface Elevation, Bathymetry, Wind Gusts, Visibility, Water Temperature, Wave Height, and Dew Point Temperature) accounted for, which are then divided into each of the data providers. Some of these directories are empty since a certain variable may not be supplied by a specific data provider, but it was reasoned that it is better to maintain a symmetrical structure both for

maintenance reasons and future expansion potential. At any given time during the operation of the system, several thousand files are stored in the directory tree, of which the RDF part is parsed based on the queries sent by the client machine (then processed by SERVLETS residing on the TOMCAT server). Despite this large number of files, the parsing process is very fast taking only several thousand milliseconds to parse through 10,000 RDF files.

Encoding

While the elements of the DCMI metadata set define the semantic requirements of description framework, the syntactic base, or encoding approach, remains to be defined so the metadata descriptions can be machine read or “parsed”. The syntactic framework can be provided by using the Extensible Markup Language (XML, 2003). It is considered a meta-language (similar to HTML but much more powerful) that can be used to describe any kind of information and that is platform independent. It is purely syntactic and therefore constitutes a “grammar” framework that a user can utilize to embed his/her specific semantic (e.g., a metadata catalogue) requirements to form a new markup language suitable for a specific purpose. The language is easy to read and ideally suited for structured documents, i.e., repetitive occurrences of certain information pieces, as can be found in data sets (e.g., seismic array data).

Its flexible structure makes it an ideal vehicle for the metadata descriptions. Once the metadata elements have been identified, they can be coded in XML, which in turn allows these data descriptions to be disseminated across the intra- or internet. Dissemination or querying is facilitated by placing the XML-encoded in a separate file, which in turn is then stored on the server’s local directory structure for querying purposes. An example **RDF**-file for the salinity filed of the ASA model is shown in Figure 5.

The use of the XML language to encode metadata descriptions is favorable for a number of reasons. First, the data exchange between the various modules can be accomplished using XML documents, i.e., the modules communicate through these documents. Second, XML allows sending data files (or descriptions of these files) across the internet allowing for a standardized data exchange among a variety of users through the internet. Third, XML can easily be extended in case future applications demand an extension of the functionality of the IM2 system.

System Design and Functionality

Performance

The fact that much of the data is computed and measured at time levels that are not synchronized nor are collected exactly at the top of the hour, mandated the establishment of a common clock or time reference. In order to unify the time levels at which data will be displayed, we developed an automated and server-based interpolation routine that maps any data delivered at any time to a time-frame that

uses the full hour mark as reference point. Hence, data is displayed at 1.00am, 2.00am, 3.00am, ..., 10.00pm, 11.00pm, 12.00pm, for a total of 24 data points per day

```

<?xml version="1.0" encoding="UTF-8" ?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
xmlns:dc="http://purl.org/dc/elements/1.1/"
xmlns:eor="http://dublincore.org/2000/03/13/eor#"
xmlns:dcq="http://purl.org/dc/terms/extent/">
<rdf:Description rdf:about="DATA/SAL/asa/01010600noaa1SAL.dat">
<dc:title>Salinity values of ASA - Model from January 1, 2001 at 03:03-
05:00 to January 5, 2001 at 04:03-05:00</dc:title>
<dc:subject>Salinity</dc:subject>
<dc:description>Salinity values, in PSU, collected every 1hr </dc:description>
<dc:publisher>Drexel University</dc:publisher>
<dc:contributor>University of Rhode Island, NOAA</dc:contributor>
<dc:date>2001-01-06T00:00-05:00</dc:date>
<dc:type>Dataset</dc:type>
<dc:format>text/ascii</dc:format>
<dc:identifier>DATA/SAL/asa/01010600asa1SAL.dat</dc:identifier>
<dc:source>http://www.appsci.com/</dc:source>
<dc:language>en</dc:language>
<dc:coverage>
<dcq:box>
<northlimit units="degrees">42.0</northlimit>
<eastlimit units="degrees">-72.0</eastlimit>
<southlimit units="degrees">41.0</southlimit>
<westlimit units="degrees">-70.0</westlimit>
</dcq:box>
<dcq:period>
<start>2001010103030500</start>
<end>"2001010504030500</end>
<frequency units="hr">1</frequency>
</dcq:period>
</dc:coverage>
<dc:rights>Coyright 2001, ASA</dc:rights>
</rdf:Description>
</rdf:RDF>

```

Figure 5 Example XML-encoded RDF file for Salinity Array

for any grid or measurement point in the domain. An exception to this rule were the time series data received from the University of Rhode Island's COASTMAP server which collected and passed data from the Narragansett Bay PORTS program on to the collection bin on the Drexel server whenever it became available, typically in 6 minute intervals. Because the data storage requirements were relatively small it was decided that an interpolation onto full and half hour marks would not serve any higher purpose and the time intervals as well as the delivery times were kept as is. The IM2-

server was configured such that it would routinely check in the server's delivery bin for newly arrived data, move it into a temporary storage area, execute the recognition and interpolation routines, and then pass on the processed data to the respective directories in the FSDB.

In order to further improve the performance of the IM2 system, it was decided that the density of the results of the hydrodynamic simulations, executed by ASA on a curvilinear grid comprised of approximately 15,000 nodes, could be reduced without losing a significant amount of information content. This was accomplished by "thinning" out the ASA grid, mostly in the regions of high resolution, i.e. inside the bay. This reduction resulted in a "Drexel-display-grid" with only approximately 5000 grid points, i.e. a reduction to one third of the original grid. The server routines (servelets executed by the TOMCAT server engine) would take the ASA files and interpolate the results based on the 15K node grid onto the 5K node reduced grid (shown in Figure 6) using a bi-variate interpolation. The interpolation of a complete 24hrs file (for one variable) just took a few seconds, hence a complete 24 hour forecast dataset would be ready for display in less than 5 minutes after it had been delivered from the ASA server. The grid used for the atmospheric simulations (by NOAA) is comprised of a regular 4*4 kilometer mesh size, which did not lend itself for further reduction hence the results of all atmospheric simulations were retained for that grid.

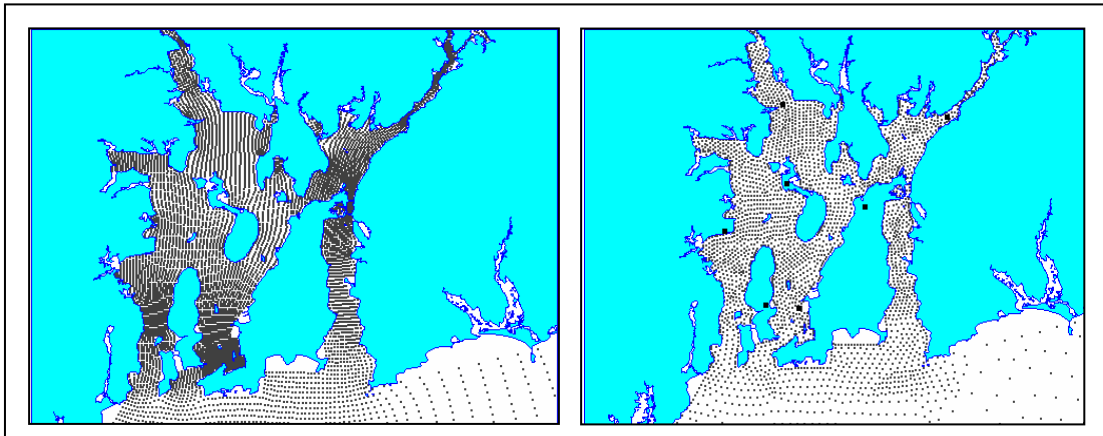


Figure 6 Original ASA Grid (left panel) and reduced Drexel Grid (right panel)

Two additional aspects needed attention during the development phase of the IM2 portal. First, a decision had to be made regarding the uncertainty of what type of browser the end user might be using. Since the targeted user community was very diverse and a prediction therefore quite difficult, the development team had to assume a mixture of the most commonly used browsers (NetScape, Windows Explorer, and Mozilla). This posed a problem, because of the inconsistency of JAVA versions used and supported by any of these browsers. While the Explorer supported JAVA version 1, JAVA2 was not supported but supplanted by a slightly different Microsoft version of JAVA. This made the use of JAVA2 API like SWING difficult, because Explorer users would have to disable the embedded Java Virtual Machine, and download the

JVM from SUN Microsystems instead. While this is a fairly simple action to execute, it provides a considerable level of inconvenience to the end user. Nevertheless, the advantages of using JAVA2 were apparent, so development of the portal proceeded using the JAVA2 system. Fortunately in early 2003, a court ruling appeared that prohibited Microsoft to continue using its own JAVA derivative and as a result cleared the way for all browsers now using JAVA2 and a compatible JVM to run the applet applications.

Second, while much of the data transfer occurred over the internet and then inside high performance networks typically employed at university campuses and governmental institutions, it remained unclear how the end user might be connected to the WWW, i.e. how the “last mile” would be covered. Considering that many of the potential end users maybe re-recreational boaters, the likelihood of 56KB modem connections was high, which prompted the development team to seriously consider the choice of a light- or thin- versus a heavy-weight solution. More specifically, the design of the portal had to balance the advantage of a system that would receive the raw data and then execute all calculations on the client end (heavy weight) versus the advantage of having all calculations executed on the (powerful) server side, leaving the portal rather lean, but then requiring a considerable amount of client-server communications. In view of a commonly accepted “patience threshold”, i.e. the time that should at most elapse between a click of the mouse button and the results of this action showing on the screen, of 6 seconds it was decided to adopt a middle-of-the-road approach. This approach focused on leaving the most computationally intensive computations on the server, but giving the portal (client machine) some computations to carry out, which would reduce the frequency of communication. Key to this approach was to try to reduce the size of the data objects sent between the client and the server. The adoption of a short integer notation, a binary format, and the use of JAVA’s JAR classes (a compression system like ZIP) accomplished this goal in a very satisfactory manner. Some of the data files sent during a session were reduced by over 90%, as shown in Table 3.

Table 3 Example Data Files and their Reduction

Data File	Original Size	Compressed Size
Coastline Polygon	292KB (float, ascii)	114KB (sint, bin, jar)
ASA single WSE data set	91KB (float, ascii)	4KB (sint, bin, jar)
ASA grid	1.2MB (float, ascii)	148KB (sint, bin, jar)
Drexel grid	794KB (float, ascii)	60KB (sint, bin, jar)

The portal was designed to permit background uploading, i.e. while some of the graphics appear in the viewer pane, other data continues to be transferred. Also, once the initial data was uploaded it stayed active throughout the duration of the client

session, reducing the data upload needs to just the result files, which were small (for example 4KB).

IM2-System Configuration

The initial setup of the system included a single server running the LINUX operating system on Pentium 3 processor and having 512MB of RAM. While this configuration was sufficient during the initial development phase, it became clear in the course of the system expansion that the LINUX machine reached its limits handling the raw data streams, processing of those, hosting the web-site, delivering the portal and maintaining sufficient speed during a client-server session. An upgrade to 1GB solved the problem only partially as the main performance problem was encountered during client sessions. In order to mitigate the problem the development team decided to pursue upgrade options for the existing hardware and turned to SUN Microsystems by applying for an Academic Equipment Grant (AEG). This grant, totaling \$128K, was installed in phases. In phase 1 the development team received a 280R server with 2 SPARCIII processors, which greatly improved the computational speed available to the IM2 system. The second phase saw the installation of 0.5 TerraByte of hard-disk and tape-backup in addition to 3 SUNBLADE 2000 workstations that were used for continued development and testing. Also, it permitted the IM2 system to incorporate 2 servers, one server for collecting the data, hosting the website (APACHE Web server, v. 2.0.40), and delivering the IM2 portal to the user, while the other server would handle the number crunching intensive part, i.e. the interpolations, reformatting, and running the servlet engine (Jarkarta TOMCAT server, v. 5.0.1), as shown in Figure 7.

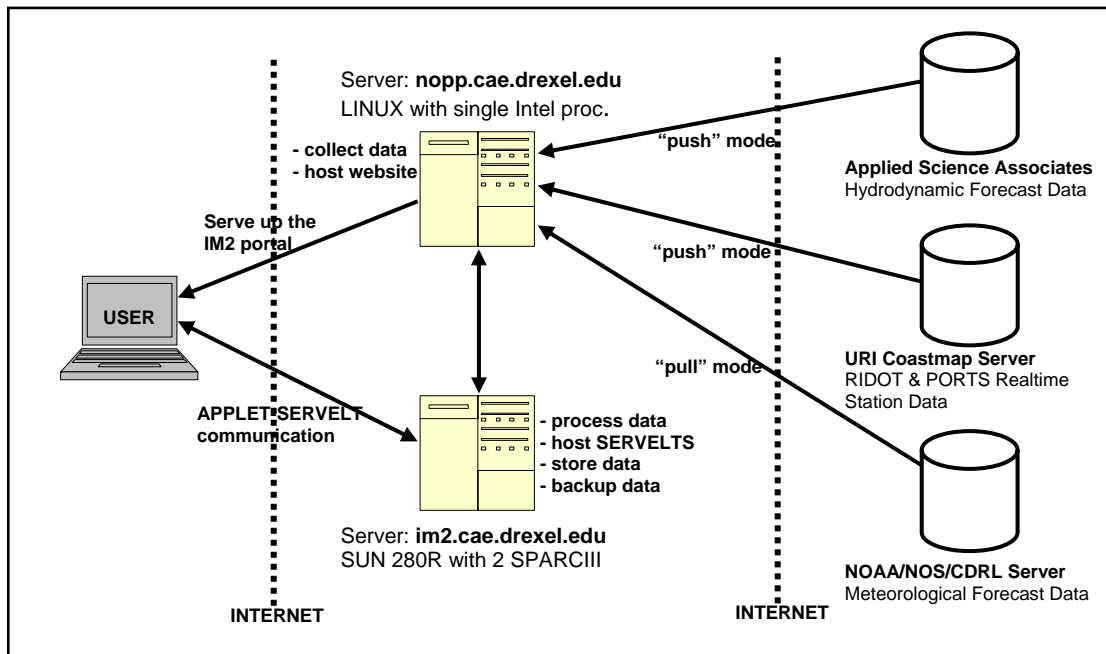


Figure 7 Components of the IM2 System

The IM2 system is configured such that it receives both *.DAT and *.RDF files from Applied Science Associates, ASA, (hydrodynamic forecast data) as well as from the COASTMAP server at the University of Rhode Island, URI, (real-time station data collected from the PORTS and RIDOT systems). The frequency for the ASA data sets is once per 24 hours received during the early morning hours, while the station data from URI arrives in the collecting bin whenever it becomes available, i.e. about every 6 minutes. In contrast, the meteorological data sets are “pulled” from the NOAA server and delivered in netCDF encoding. These data sets become available around 5.00am every day and contain 24 hours forecast information presented at the full hour. The IM2 system is fully automated, i.e. all data collection procedures, recognition, interpolation, and subsequent storage is carried out either using SOLARIS scripts or executing servlets on the TOMCAT server. Despite occasional interventions and subsequent necessary re-starts of the system, the configuration proved and proves to be quite stable with only few disruptions during the course of its operation.

Portal Design and Functionality

One of the main objectives of the IM2 development was to make sure that the system is user friendly and could be used with only a small period of learning. This meant to keep the portal well organized and structured and also mandated to minimize the number of available options and buttons that could be used to alter, modify, and select display settings as well as the variable selection, as shown in Figure 8.

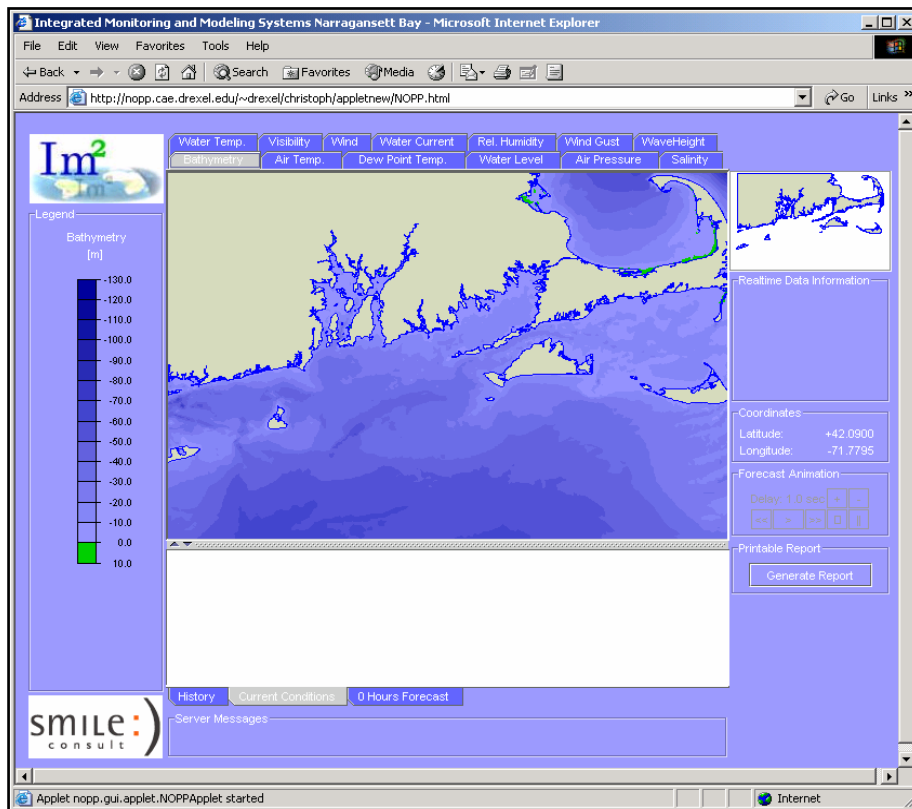


Figure 8 IM2 Portal Layout (Bathymetry is shown as contour)

In addition, the system was designed such that the appearance as well as the functionality would be controlled through configuration files. In other words, each change or expansion of the IM2 portal is controlled in a central configuration module which makes sure that relevant changes trickle through and modify the system accordingly. Also, much of the IM2 portal design was influenced by the outcome of the user survey, particularly the selection of desired state variables, display frequency, and display options. It was clear that both real-time and forecast data were foremost on the mind of the user group, displayed every hour, and presented as either time-series, contours, or vector plots (for current and wind velocities) the latter two preferably animated, i.e. cycling through the forecast hours.

The view-pane dimensions (about 480*300pixels) was aligned with a physical view box bounded by 70 to 72 degrees West and 40.75 to 42 degrees North. All geographical references with respect to grid points and data stations are given in LON/LAT coordinates and are transformed to the horizontal view plane using a Universal Transverse Mercator, UTM, transformation (region 17). During zoom operations, the screen coordinates are transferred back to LON/LAT coordinates applying a reverse UTM transformation (there is also a small viewer pane tracking the mouse movements) so the FSDB can be parsed using LON/LAT coordinates.

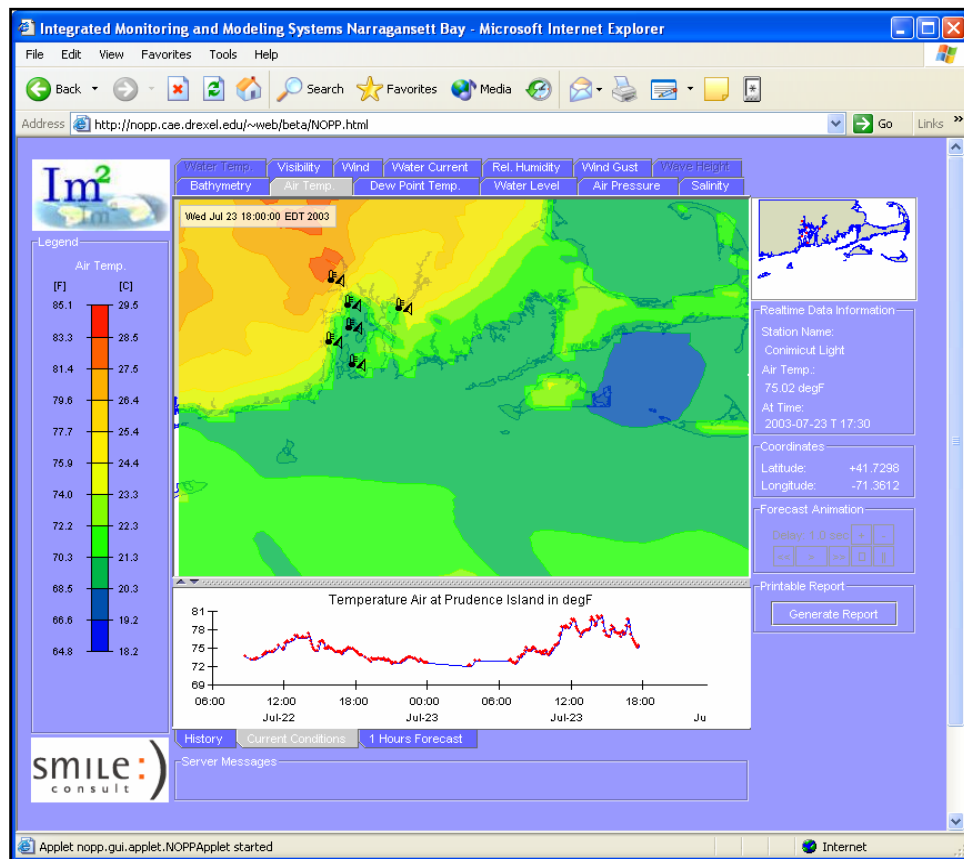


Figure 9 Area Contour and Station History Data for current Air Temperature Distribution

The IM2 portal furthermore contains a pane for the legend (left side), a view pane for time histories (below main pane), and an action pane (on the right) where the user can control animations by stepping through the time levels one by one, or automatically animating them (there is also a button that can be used for controlling the display speed). The variables can be selected from a set of tabs on the top pane, while real-time, forecast or historic data can be selected from tabs at the bottom of the main view pane. Also there is an option to select a specific station (real time data) and obtain a data summary of all data that is currently being selected (the latest available data bit will be displayed). Also, the summary report plots a history of the last 60 hours of data that has been selected at that station.

The system was designed such, that it would automatically display icons for the measurement stations if the selected variable was present both in the forecast simulations and real-time collections (for example air-temperature), as shown in Figure 9. In addition, a sub-pane in the right action pane automatically displays what the current value is (for the selected variable) on a mouse-over of one of the displayed station icons. It should be noted that real-time (or current) data measures are not necessarily synchronized between the various sources. Since forecast data is available only at the hour and station data becomes available about every 6 minutes, the system uses the convention that “real-time” is defined as either the last available data set (station data) or the contour time-step closest to the actual local time. In other words, the real-time display originating from the forecast calculations may be off by 30 minutes while the one from the stations may just encounter a 6 minute delay.

Demonstration Periods and Updates

In order to test and improve the IM2 system (also the COASTMAP system), it was decided by the research group that 2 demonstration periods should be identified interspersed by a workshop to which the expected user-community would be invited. IM2 went online for the first demonstration in March of 2003, which was concluded with a training and user workshop on April 22, 2003 held at the University of Rhode Island’s Narragansett Bay Campus. The purpose of this workshop was to give the potential user community an opportunity to sit down and use the system hands on, while providing feedback to the development team on how the system could be improved to better suit the needs of the various user groups.

The workshop proved to be a success as the users took to the IM2 with ease and positively commended on the ease-of-use, layout of the portal, and information content of the data displayed. However, some suggestions were made to improve the system, which were largely related to the fact that different users or user communities have different expectations as far as the data display conventions are concerned. Not surprisingly, the comment made most concerned the unit system available in the legend pane as well as the vector display of wind speed and direction. The former addressed the fact that, depending on what home institution the individual would come from, SI or English units were preferred while others (Navy, Pilots, USCG)

preferred current and wind speeds in knots rather than feet/s or m/s. Also, it was noted that the magnitude of the original vector display (vectors scaled to a reference vector) was very hard to discern, which prompted the team to change the vector plot to include a color fill that would be matched in a color coded legend, as shown in Figure 10.

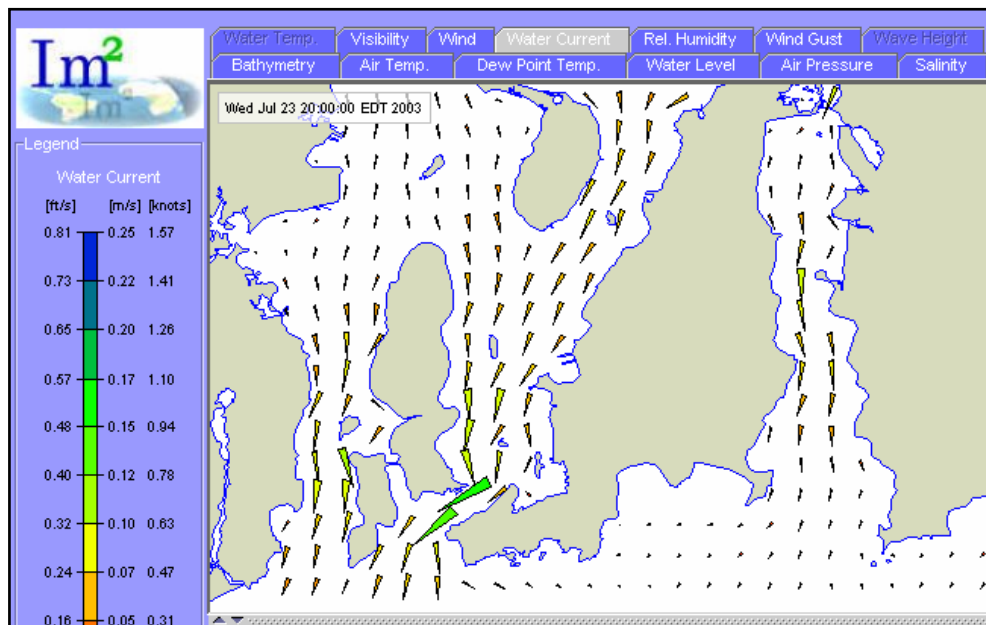


Figure 10 Vector Plot for Currents in the NB Area

Also notice that as a result of the suggestions made the legend was changed to display all unit systems. It should be noted that the vector density, i.e. vectors plotted per screen segment, is fixed to one vector per 20*20 pixel box. This feature was implemented, because of the dense grid distribution inside the bay. If a vector would be plotted for each node no information would be discernable because i) the vector heads would have to be extremely small, or ii) the vectors would overlap to a degree that they are not distinguishable. The result is a fixed 20*20 pixel grid that is associated with information from the underlying grid points irrespective of what the zoom level is. This effectively means that low zoom levels data is being interpolated from a fine grid (model grid with high density of nodes per 20*20 square) to a coarse pixel grid. On the other hand, during high zoom levels the pixel grid might become denser than the underlying model grid and consequently data is interpolated from a coarse to a fine grid. If the fine-to-coarse and coarse-to-fine grid ratios are too high (or too small) interpolation tends to act as filter and information detail is either lost or a detail is suggested that does not really exist. Yet, the development team felt that was and is more important to discern data trends that, even though slightly inaccurate at times, will fulfill the data needs of the user.

One other concern voiced during the workshop was how the wind field is displayed. The meteorological community is used to indicate wind direction with a straight line and to associate magnitude with a set of perpendicular line segments attached at the

end of the base (direction) line segment. The number of these perpendicular line segments then depicts the wind speed bracket. An example of the wind field plot is shown in Figure 11, together with a legend showing all unit systems.

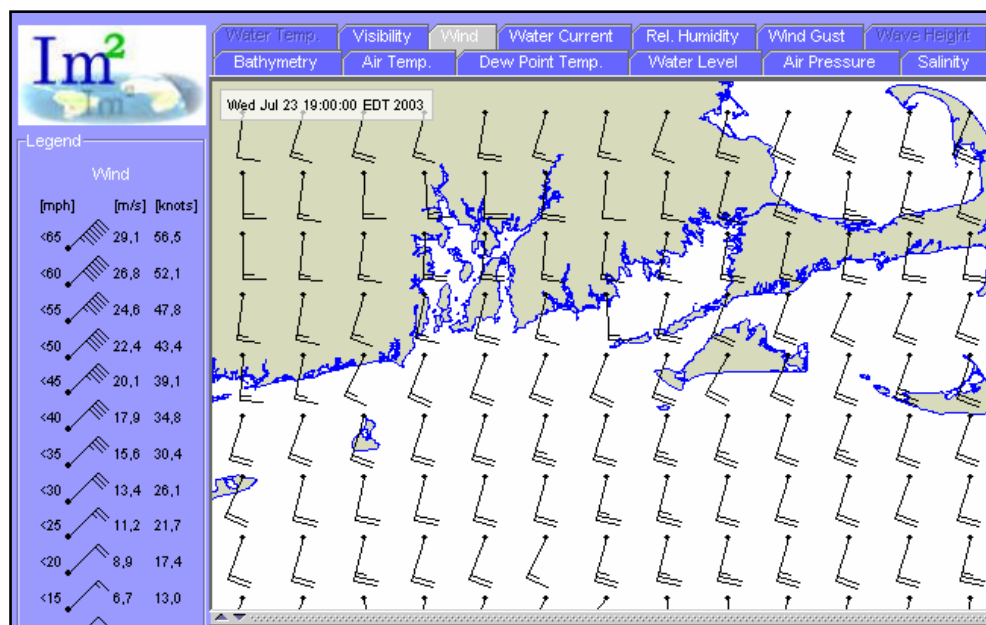


Figure 11 Wind Field Plot

The system is fully operational at and continues to receive data from the ASA, NOAA, and URI servers. Some of the data items, however, are not consistently being delivered, like Air Pressure and Visibility, as well as Currents and Wave Height. These data products depend on the continued execution of the forecast models which, without funding can only be maintained for a limited period of time. Efforts are under to make this system an integral part of other modeling domains, i.e to transfer the basic concepts and apply it, for example, to environmental information systems.

The IM2 portal can be accessed at <http://nopp.cae.drexel.edu/~web/beta/NOPP.html> using “beta” and “beta” as loginID and passwd when prompted. The site also contains an online comment-submission feature so users can continue to submit their comments for further improvement of the system.

Acknowledgements

This research was made possible through a National Ocean Partnership Program (NOPP) grant, administered by NASA under grant number NAG13-00040. The research team would also like to acknowledge the generous Academic Equipment Grant (AEG) received from SUN Microsystems, who provided an essential building blocks of the server system (280R, tape backup system, additional hard-disk storage, free-of-charge software licenses, 3 SunBlade 2000 workstations for system development).

References

Bermudez, L. and Piasecki, M. (2002), "Development of a Web-based Hydrologic Information System", American Water Research Association Annual Conference, Philadelphia, Nov. 3-7, 2002.

DCMI (2002), Dublin Core Metadata Initiative, Retrieved Sep. 2002 from <http://dublincore.org/>.

FGDC (2002), Federal Geographic Data Committee. Retrieved Nov. , 2002 from <http://www.fgdc.gov/>

Gilliland-Swetland, A.J.(2002), "Introduction to Metadata: Setting the Stage", Pathways to Digital Information, <http://www.getty.edu/research/institute/standards/>, accessed 2002.

ISO, (2002). International Organization for Standardization. Retrieved Dec. , 2002 from <http://www.iso.ch/>

MySQL v.4.0, (2002) <http://www.mysql.com/>, accessed Sep. 2002.

National Spatial Data Infrastructure (NSDI) (2002); <http://www.fgdc.gov/nsdi/nsdi.html>, accessed 2002.

National Ocean Partnership Program, NOPP (2000), accessed Sep. 2003 <http://www.nopp.org/>

PostgreSQL v.7.3 (2003) <http://www.postgresql.org/>, accessed Sep. 2003

World Wide Web Consortium (W3C), (2003), "Extensible Markup Language", <http://www.w3.org/XML/>, accessed Sep. 2003.

R. Lehfeldt, and M. Piasecki, (2002), "Components of Web Portals in Coastal Engineering", 5th International Conference on Hydroinformatics, July 1 - 5, 2002, Cardiff and Bristol, Wales.