U.S. JGOFS Data Management

a retrospective

Cyndy Chandler

U.S. JGOFS Data Management Office

25 January 2005

NACP Data Management Planning Workshop

New Orleans, LA
26 January 2005

blizzicane of ‘05
Topics in today’s presentation:

- **Introduction**
  - What is U.S. JGOFS?
  - What is the DMO?

- **Lessons Learned**
  - U.S. JGOFS data server
    - JGOFS d-DBMS and user interface
    - Live Access Server
What is U.S. JGOFS?

U.S. Joint Global Ocean Flux Study

- part of multinational JGOFS
- U.S. Global Change Research Program (US GCRP), Scientific Committee on Oceanic Research (SCOR), and International Geosphere-Biosphere Programme (IGBP)
- long term (U.S. 1989-2005)
- multidisciplinary (bio, chem, PO, geology)
- process studies (U.S. 1989-1998), time-series, global surveys, synthesis and modeling, data management
- investigate ocean carbon flux
U.S. JGOFS
Data Management Office (DMO)

- formed in 1988 specifically to meet needs of U.S. JGOFS
- assist PIs to submit their data to DMO
- ongoing quality control of data
- develop and maintain simple, reliable interface to program data
- provide timely, easy access to project results
- collaborate with other program DMO
- publish U.S. JGOFS data reports
- plan final archive of U.S. JGOFS information
Basic Principles of Data Management
from 1988 JGOFS Working Group on Data Management

- scientists will generate data in a format useful for their needs
- oceanographic data sets are best organized in terms of metadata (temporal and geographical)
- data managers should avoid use of coded data values
- users should be able to obtain all the data they require from one source and in a consistent format
- data interchange formats should be designed for the convenience of scientific users
Additional Guidelines

- metadata is critical and therefore mandatory
- data managers should maintain awareness of emerging standards and strive for compliance
- whatever interface is used to provide access to the data, it’s still important to provide subset and download capability
- data management systems must be dynamic - balancing tension between existing and new technologies
Basic Components of a Data Management System

First four components are active throughout the program:

- data acquisition
  - from variety of sources
- quality assurance
- data publication
- synthesis and modeling
- archive
lesson defined . . .

1: a passage from sacred writings read in a service of worship

2 a: a piece of instruction b: a reading or exercise to be studied by a pupil c: a division of a course of instruction

3 a: something learned by study or experience b: an instructive example

4: an edifying example or experience

5: a reprimand

Lessons Learned . . .
The first lesson learned . . .

is a meta lesson . . .

All the lessons learned take on enhanced meaning when applied to science programs of increasing size and complexity.

Lessons Learned . . .
Lessons Learned . . .

- Technology is good; people are more important
  - Diverse range of expertise and personalities
    - Designers, programmers, data managers
  - Someone with authority to set overall vision
- Qualified staff to make effective use of technology
- Guidance from advisory committee which includes active investigators
U.S. JGOFS DMO personnel

- David Glover (director)
- Cyndy Chandler (manager)
- Jeff Dusenberry (data specialist)

Previous staff members

- Christine Hammond (manager)
- George Heimerdinger (data specialist)
- David Schneider (data specialist)
Lessons Learned...

- develop a data policy, publicize and follow it
  - guidance from steering committee
  - consent from participating investigators
  - reiterate at conferences and workshops
  - encourage compliance

- agree on method of enforcement (used as last resort)
Lessons Learned . . .

- establish protocols at program start with mechanism for adaptation when necessary
  - sampling methodologies
  - naming conventions
    - parameter dictionary, controlled vocabulary
    - XML schema, thesauri, ontologies
  - units of measurement
Lessons Learned . . .

- facilitate contribution of data to collection
  - compile an inventory of expected results
  - publish and maintain the inventory
  - remind investigators of opportunities to contribute data and results to the growing inventory
  - review procedure at conferences and workshops
  - accept all formats of data
  - work with investigators to complete metadata records
JGOFS distributed database management system

JGOFS object = method + data

U.S. JGOFS DMO accepted any format data from the field study investigators and used methods to locate and translate the data objects.
Lessons Learned . . .

- **metadata is of critical importance**
  - accurate, complete, available with data
  - monitor emerging standards
  - define minimal metadata requirements
  - standards-compliant solutions where possible
  - complete metadata record enables reuse of data

metadata assembly is time consuming, but is the key to enabling secondary reuse of data
**bottle**

**PI:** Lou Codispoti  
**dataset:** Temp, salinity, nutrients from Niskin bottles  
**project/cruise:** Arabian Sea/TTN045 - Process Cruise 2  
**ship:** Thomas Thompson  

**Final Recalibration changes (10/28/96)**  
**PI Notes and Methodology**  
**DNO note on calculated depth**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>event</td>
<td>a unique number assigned to each sampling operation consisting of month MM, day DD, hour HH and minute mm</td>
<td></td>
</tr>
<tr>
<td>sta</td>
<td>station number</td>
<td></td>
</tr>
<tr>
<td>sta_std</td>
<td>Arabian Sea standard station identifier</td>
<td></td>
</tr>
<tr>
<td>cast</td>
<td>CTD cast number</td>
<td></td>
</tr>
<tr>
<td>date</td>
<td>date (YYYYMMDD) decoded as follows YYYY = year, MM = month, DD = day Date converted to UTC(GMT).</td>
<td></td>
</tr>
<tr>
<td>time_begin</td>
<td>start time of cast in UTC (GMT)</td>
<td>decimal hours</td>
</tr>
<tr>
<td>time_end</td>
<td>end time of cast in UTC (GMT)</td>
<td>decimal hours</td>
</tr>
<tr>
<td>lat_begin</td>
<td>start latitude of cast (negative = South)</td>
<td>decimal degrees</td>
</tr>
<tr>
<td>lon_begin</td>
<td>start longitude of cast (negative = West)</td>
<td>decimal degrees</td>
</tr>
<tr>
<td>lat_end</td>
<td>end latitude of cast</td>
<td>decimal degrees</td>
</tr>
<tr>
<td>lon_end</td>
<td>end longitude of cast</td>
<td>decimal degrees</td>
</tr>
<tr>
<td>bot</td>
<td>CTD rosette bottle number</td>
<td></td>
</tr>
<tr>
<td>press</td>
<td>sample depth reported as pressure</td>
<td>decibars</td>
</tr>
</tbody>
</table>
Lessons Learned . . .

- quality assurance is an ongoing process
  - intense QA process during initial acquisition and ingestion into data system
  - problems discovered as data are utilized by others
    - insufficient or inaccurate metadata
  - process of data product synthesis becomes a valuable diagnostic tool for improving data quality
Lessons Learned . . .

- begin synthesis early
  - do not wait until all the data has been collected
  - synthesized products greatly enhance the data collection
Lessons Learned . . .

- provide timely, easy access to project results
- develop and maintain simple, reliable interface to data collection
  - single interface to entire data collection
  - balance tension between new innovative technologies and existing stable implementations
  - if an interface is broken, it doesn’t matter how great the original concept was
U.S. JGOFS Data Server

- JGOFS distributed database management system (d-DBMS) used for field data
- Live Access Server (LAS) used for gridded, synthesis and model results
US JGOFS Data System

Data Acknowledgement Policy

The data available here are intended solely for scholarly use by the academic and scientific community, with the express understanding that any such use will properly acknowledge the originating investigator. Anyone wishing to use U.S. JGOFS data in a presentation, report, thesis or publication should contact the originating PI. It is expected that all customary courtesies and privileges attached to data use will be strictly honored. Use or reproduction of any material herein for any commercial purpose is prohibited without prior written permission from the U.S. JGOFS Data Management Office. The complete copyright information is available [here](http://usjgofs.whoi.edu/dm/)

The merged data products are available via the [Live Access Server](http://usjgofs.whoi.edu/dm/)

US JGOFS Data Categories

Go to the indicated category of data by clicking on its name

- **jgofs**
- **skp results**
- **arabian**
- **ctt-039**
- **ctt-042**
- **ctt-044**
- **ctt-045**
- **ctt-046**
- **ctt-050**
- **ctt-053**
- **ctt-054**
- **beta**
- **espar**
- **ctl07**
- **ctl08**
- **ctl12**
JGOFS Distributed Database Management System (d-DBMS)

- distributed, object-oriented system
- originally developed by Glenn Flierl, James Bishop, Satish Paranjpe, David Glover
- supports multidisciplinary, multi-institutional data acquisition project
- multiple data storage formats and locations
- data interpreted by ‘methods’
<table>
<thead>
<tr>
<th>event</th>
<th>sta</th>
<th>sta_std</th>
<th>cast</th>
<th>date</th>
<th>time_begin</th>
<th>time_end</th>
<th>lat_begin</th>
<th>lon_begin</th>
<th>lat_end</th>
<th>lon_end</th>
</tr>
</thead>
<tbody>
<tr>
<td>03142214</td>
<td>l</td>
<td>NI</td>
<td>l</td>
<td>15950314</td>
<td>22.2461</td>
<td>22.9030</td>
<td>22.38423</td>
<td>59.88413</td>
<td>23.38506</td>
<td>58.87998</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>hot_press</th>
<th>depth</th>
<th>temp</th>
<th>sal_std</th>
<th>sal_bot</th>
<th>02_ml_L</th>
<th>02_ml_0 kg</th>
<th>02_ml_O_L</th>
<th>NO3</th>
<th>PO4</th>
<th>SiO4</th>
<th>N02</th>
</tr>
</thead>
<tbody>
<tr>
<td>24</td>
<td>1.9</td>
<td>1.9</td>
<td>23.752</td>
<td>36.594</td>
<td>36.593</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
<td>5.07</td>
<td>0.75</td>
<td>3.4</td>
</tr>
<tr>
<td>23</td>
<td>1.9</td>
<td>1.9</td>
<td>23.738</td>
<td>36.594</td>
<td>36.590</td>
<td>4.254</td>
<td>183.33</td>
<td>189.98</td>
<td>5.12</td>
<td>0.71</td>
<td>3.4</td>
</tr>
<tr>
<td>22</td>
<td>12.6</td>
<td>11.9</td>
<td>23.753</td>
<td>36.594</td>
<td>36.593</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
<td>5.25</td>
<td>0.71</td>
<td>3.4</td>
</tr>
<tr>
<td>21</td>
<td>12.6</td>
<td>11.9</td>
<td>23.762</td>
<td>36.594</td>
<td>36.590</td>
<td>4.214</td>
<td>183.58</td>
<td>188.19</td>
<td>5.30</td>
<td>0.71</td>
<td>3.4</td>
</tr>
<tr>
<td>20</td>
<td>22.0</td>
<td>21.9</td>
<td>23.791</td>
<td>36.594</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
</tr>
<tr>
<td>19</td>
<td>22.3</td>
<td>22.2</td>
<td>23.796</td>
<td>36.594</td>
<td>36.593</td>
<td>4.084</td>
<td>177.92</td>
<td>182.39</td>
<td>6.12</td>
<td>0.71</td>
<td>3.3</td>
</tr>
<tr>
<td>18</td>
<td>32.1</td>
<td>31.9</td>
<td>23.755</td>
<td>36.606</td>
<td>36.602</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
<td>5.89</td>
<td>0.71</td>
<td>3.2</td>
</tr>
<tr>
<td>17</td>
<td>32.6</td>
<td>31.0</td>
<td>23.755</td>
<td>36.606</td>
<td>36.601</td>
<td>4.157</td>
<td>161.10</td>
<td>165.65</td>
<td>5.69</td>
<td>0.71</td>
<td>3.2</td>
</tr>
<tr>
<td>16</td>
<td>42.1</td>
<td>41.0</td>
<td>23.659</td>
<td>36.592</td>
<td>36.590</td>
<td>4.076</td>
<td>177.66</td>
<td>182.12</td>
<td>6.69</td>
<td>0.71</td>
<td>3.3</td>
</tr>
<tr>
<td>15</td>
<td>42.2</td>
<td>41.9</td>
<td>23.660</td>
<td>36.592</td>
<td>36.593</td>
<td>4.083</td>
<td>177.68</td>
<td>182.34</td>
<td>6.76</td>
<td>0.71</td>
<td>3.3</td>
</tr>
<tr>
<td>14</td>
<td>52.1</td>
<td>51.8</td>
<td>23.203</td>
<td>36.590</td>
<td>36.594</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
<td>9.81</td>
<td>1.05</td>
<td>4.8</td>
</tr>
<tr>
<td>13</td>
<td>52.9</td>
<td>52.0</td>
<td>23.302</td>
<td>36.522</td>
<td>36.594</td>
<td>3.421</td>
<td>159.04</td>
<td>152.78</td>
<td>9.92</td>
<td>1.05</td>
<td>4.8</td>
</tr>
<tr>
<td>12</td>
<td>62.4</td>
<td>62.0</td>
<td>22.733</td>
<td>36.450</td>
<td>36.453</td>
<td>2.660</td>
<td>118.28</td>
<td>118.79</td>
<td>13.05</td>
<td>1.35</td>
<td>7.8</td>
</tr>
<tr>
<td>11</td>
<td>62.4</td>
<td>62.0</td>
<td>22.738</td>
<td>36.450</td>
<td>36.454</td>
<td>2.639</td>
<td>114.96</td>
<td>117.86</td>
<td>13.15</td>
<td>1.35</td>
<td>7.8</td>
</tr>
<tr>
<td>10</td>
<td>72.3</td>
<td>71.8</td>
<td>21.968</td>
<td>36.336</td>
<td>36.332</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
<td>18.49</td>
<td>1.77</td>
<td>11.4</td>
</tr>
<tr>
<td>9</td>
<td>75.1</td>
<td>75.0</td>
<td>21.967</td>
<td>36.336</td>
<td>36.320</td>
<td>1.435</td>
<td>62.51</td>
<td>64.05</td>
<td>18.64</td>
<td>1.77</td>
<td>11.4</td>
</tr>
<tr>
<td>8</td>
<td>102.4</td>
<td>101.7</td>
<td>20.761</td>
<td>36.136</td>
<td>36.135</td>
<td>0.531</td>
<td>23.13</td>
<td>23.71</td>
<td>22.43</td>
<td>2.15</td>
<td>15.4</td>
</tr>
<tr>
<td>7</td>
<td>102.3</td>
<td>101.6</td>
<td>20.760</td>
<td>36.136</td>
<td>36.135</td>
<td>0.535</td>
<td>23.17</td>
<td>23.40</td>
<td>22.47</td>
<td>2.15</td>
<td>15.6</td>
</tr>
<tr>
<td>6</td>
<td>192.4</td>
<td>191.5</td>
<td>19.935</td>
<td>36.134</td>
<td>36.134</td>
<td>0.220</td>
<td>9.58</td>
<td>9.83</td>
<td>23.40</td>
<td>2.38</td>
<td>18.3</td>
</tr>
<tr>
<td>5</td>
<td>192.4</td>
<td>191.5</td>
<td>19.934</td>
<td>36.134</td>
<td>36.134</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
</tr>
<tr>
<td>4</td>
<td>162.6</td>
<td>161.4</td>
<td>18.967</td>
<td>36.220</td>
<td>36.223</td>
<td>0.127</td>
<td>5.53</td>
<td>5.67</td>
<td>22.93</td>
<td>2.39</td>
<td>21.1</td>
</tr>
<tr>
<td>3</td>
<td>162.4</td>
<td>161.3</td>
<td>18.965</td>
<td>36.218</td>
<td>36.222</td>
<td>0.130</td>
<td>5.66</td>
<td>5.81</td>
<td>22.92</td>
<td>2.25</td>
<td>20.9</td>
</tr>
<tr>
<td>2</td>
<td>202.6</td>
<td>201.4</td>
<td>18.153</td>
<td>36.356</td>
<td>36.356</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
</tr>
<tr>
<td>1</td>
<td>232.2</td>
<td>231.5</td>
<td>17.420</td>
<td>36.404</td>
<td>36.414</td>
<td>0.190</td>
<td>7.40</td>
<td>7.59</td>
<td>23.32</td>
<td>2.35</td>
<td>24.6</td>
</tr>
</tbody>
</table>
subset, plot, download data

Plotting and Other Operations Menu

Current object is: //usjgofs.whoi.edu/jgofs/arabian/tn-045/bottle

- Listing and downloading data
  - List at this level
  - Other data listing formats
  - Matlab file format of all data at this level and further in.
  - Download utility

- Manipulating data
  - Math operations for calculating values from existing parameters.
  - Join 2 objects having at least 1 parameter in common.
  - Statistics

- Plotting data
  - Simple X-Y plot
  - Mapping of data locations (JGOFS software)

- Subsetting data

Version: April 1, 2004
Live Access Server Interface provides access to synthesis and model results.
Live Access Server ~ LAS

- LAS Development Team (original)
- Steve Hankin
- Jon Callahan
- Joe Sirott

- located at UW JISAO/NOAA-PMEL

University of Washington’s Joint Institute for the Study of the Atmosphere and Ocean and NOAA Pacific Marine Environmental Laboratory
Live Access Server Interface

- configurable Web server
- data and metadata interface
- provides access to geo-referenced scientific data
- presents distributed data sets as a unified virtual data base (DODS/OPeNDAP)
- uses Ferret as the default visualization application
- visualize data with on-the-fly graphics
- request custom subsets of variables in a variety of file formats
- access background reference material (metadata)
- compare variables from distributed sources
LAS enables a data server to ...

- unify access to multiple types of data in a single interface
- create thematic data collections from distributed data sources
- offer derived products on the fly
- offer variety of visualization styles
  - customized for the data
U.S. JGOFS LAS

- MySQL database of netCDF and JGOFS format data objects
- interface to project data and metadata
- data sub-selection (selections, projections)
- multi-variable support
- gridded vs. in-situ data differencing
- multiple views
  - (property-property, depth horizon, cruise tracks, overplots)
- multiple products (ps, gif, text, NetCDF)
select dataset and variables
LAS v6 constraints

- select dataset
- select variable
- set constraints

select output type

- lat/lon
- time
- depth range

26 January 2005
select dataset
select variable
set constraints
output
U.S. JGOFS Arabian Sea Niskin Bottle Data

Longitude: 34°E to 71°E
Latitude: 3°N to 18°N
Date: 01-JAN-1995 to 31-DEC-1995

4492 points

Depth (meters)

Nitrate (micromoles/liter)

from merged product generated by DMO from *in situ* Niskin bottle data
Comparison Overlay Plot

Live Access to U.S. JGOFS SMP Data (Test Server)

1: Datasets > Global SeaWiFS chlorophyll 1998-2002
1: Variable(s): Chlorophyll
2: Datasets > Najjar monthly nutrients
2: Variable(s): nitrate

Select your desired view (geometry of output) and output (type of product). Then set the 4-D region (lon-lat-depth-time) and any additional constraints.

Select view: xy (lat/lon) slice
Select output: Overlay plot (GIF)
Select region: Full Region

Select time for first variable: 17-May-1998
Select time for second variable: 16-Apr
Select depth for second variable: 10
chlorophyll (May) shaded
SeaWiFS data contributed by: Yoder and Kennelly
surface nitrate (April) contours
Nutrient Climatologies contributed by: Ray Najjar
U.S. JGOFS Data System Summary

- supports a variety of data formats
- OPeNDAP used to access data collection
- coupled metadata and data
- supports data subselection
- offers variety of products for download
Lessons Learned . . .

- encourage data managers to collaborate
  - other data managers within program
    - JGOFS DMTT
  - data managers from other programs
    - GLOBEC, LTER
  - program investigators and participants
    - attend conferences and workshops
    - offer data system tutorials
Lessons Learned . . .

- publish data reports
  - archive data in one place
  - easy access to project results
  - most complete and accurate form of database
United States Joint Global Ocean Flux Study
Final Data Report, Volume 1
Process Study Data

Funded primarily by the U.S. National Science Foundation
with additional support from NOAA, NASA, DOE and ONR
Lessons Learned . . .

- plan early for final archive of program results
  - digital records
  - don’t forget the boxes of stuff!

26 January 2005
Lessons Learned . . .

- develop a data policy, publicize and follow it
- establish protocols at program start with mechanism for adaptation when necessary
- facilitate contribution of data to collection
- **metadata is of critical importance**
  - accurate, complete, available with data
- quality assurance is an ongoing process
- begin synthesis early
Lessons Learned . . .

- provide timely, easy access to project results
- develop and maintain simple, reliable interface to data collection
- encourage data managers to collaborate
- publish data reports
- plan early for final archive of program results
Challenges

- functioning amid the chaos
  - maintaining a healthy data management system amid the chaos of rapidly changing information technology
  - distinguishing between enabling and disruptive technologies

- data and results – what to preserve?
  - raw and processed data, synthesized products, model code, inputs, results

- increasing volume and diversity

- long-term preservation of data
The United States Joint Global Ocean Flux Study is a national component of International JGOFS and an integral part of global climate change research.

A Sea of Change: JGOFS Accomplishments and the Future of Ocean Biogeochemistry
Final Open Science Conference 5-8 May 2002 • Washington, DC USA
Click here for more information

Click on Field Studies and Offices for more information

The U.S. launched the Joint Global Ocean Flux Study (JGOFS) in the late 1980s to study the ocean carbon cycle. An ambitious goal was set to understand the controls on the concentrations and fluxes of carbon and associated nutrients in the ocean. A new field of ocean biogeochemistry emerged with an emphasis on making measurements of carbon system parameters and interdisciplinary field studies of the biological, chemical, and physical processes which control the ocean carbon cycle. As we studied ocean biogeochemistry, we learned that our simple views of carbon uptake and transport were severely limited, and a new "w卖al" of ocean science was born. U.S. JGOFS has been supported primarily by the U.S. National Science Foundation in collaboration with the National Oceanic and Atmospheric Administration, the National Aeronautics and Space Administration, the Department of Energy and the Office of Naval Research. The current and final phase of U.S. JGOFS is the Anthropogenic and Natural Ecosystems (ANDE). Additional information about the history and mission of U.S. JGOFS is available.

U.S. JGOFS web site
http://usjgofs.whoi.edu