1. What is BCO-DMO?
2. How does BCO-DMO relate to US GEOTRACES?
3. Important concepts in data management...
4. US GEOTRACES data management
5. Lessons learned from 2010 & 2011 North Atlantic Transect cruises
bco-dmo.org for NSF OCE project data

current projects, and legacy data from large coordinated research programs
(e.g. US GLOBEC and US JGOFS)
BCO-DMO staff members provide data management support for investigators funded by the Biological or Chemical Oceanography sections of NSF Ocean Sciences or OPP ANT—*at no additional cost to the project*. This represents a new model for data management; a long-term commitment to community-wide data management as opposed to a project specific data management office.
US GEOTRACES NAT DATA
NSF OCE RESEARCH PROJECT DATA

BCO-DMO MapServer Geospatial Interface

Map

[Map Interface]

Contact | Help | NSF Acknowledgment
BCO-DMO: what we do …

- Provide data management support
  - [http://bcodmo.org/resources](http://bcodmo.org/resources)
    - Best Practice Guidelines Manual
    - Data Management Plan
    - How to contribute data

- Make data and metadata available
  - Restricted or public access as appropriate

- Ensure final archive of data in appropriate National Data Center (NODC)
U.S. GEOTRACES DATA MANAGEMENT

based on recommendations from GEOTRACES DMC and SSC

Pacific Section cruise:

- Chief Scientist
  - point of contact for BCO-DMO
  - Cruise report
    - Cruise metadata including participant list
  - Navigation data (X,Y,T cruise track data)
  - Scientific sampling event log
  - final data inventory (list of expected data sets)

- phased reporting of data sets from cruise
  - cruise track (nav data) event log and basic hydrography (CTD and Rosette base data); sampling ID log
  - bathymetry, ADCP, underway and meteorological
  - measurements from on board investigator teams

http://www.bodc.ac.uk/geotraces/data/policy/
ODF personnel are an excellent resource for hydrographic cruise planning

Recommendation:
include representatives from ODF in pre-cruise planning workshops

(Chandler & German, 2011)
GEOTRACES DATA TIMELINE

Data/Metadata Submission (timeline):

- As soon as a cruise is organized: submit pre-cruise metadata to GEOTRACES IPO and BCO-DMO. (complete form)

- Within one week of cruise completion (Chief Scientist):
  - Submit Post-cruise metadata forms (update pre-cruise)
  - Submit electronic versions (scanned or original) of event log and sample log sheets, [and copies of the bridge log]
  - Submit copy of ROSCOP/CSR form or equivalent cruise report
GEOTRACES DATA TIMELINE

Data/Metadata Submission (timeline):

- Within 6 months of end of cruise:
  - Chief scientist submits final cruise report
  - Chief Scientist does cruise status review with BCO-DMO
  - Data and metadata for shared ancillary parameters (e.g., nutrients) submitted to BCO-DMO
  - Submit CTD and underway data (both raw and processed files; sensor information and calibration) to BCO-DMO
Data/Metadata Submission (timeline):

As soon as possible, within 2 years of data generation:

- Submit all data sets and accompanying metadata to GDAC
- GDAC: In most cases, data will be submitted initially to a national data centre (DAC or GEOTRACES Data Assembly Center, e.g., BCO-DMO). BCO-DMO is responsible for submitting US GEOTRACES data to BODC.
BCO-DMO contributes US GEOTRACES data to the Data Portal at BODC

GEOTRACES International Programme data are managed by BODC/NERC

http://www.bodc.ac.uk/geotraces/
GEOTRACES International Data Assembly Centre

GEOTRACES (www.geotraces.org) is an international programme which aims to improve our understanding of biogeochemical cycles and large-scale distribution of trace elements and their isotopes (TEIs) in the marine environment. The global field programme will run for at least a decade and will involve cruises in all ocean basins run by a variety of nations.

Planning has involved scientists from around 30 countries. GEOTRACES is expected to become the largest programme to focus on the chemistry of the oceans and will improve our understanding of past, present and future distributions of TEIs and their relationships to important global processes.

GEOTRACES mission is:

*To identify processes and quantify fluxes that control the distribution of key trace elements and isotopes in the ocean, and to establish the sensitivity of these distributions to changing environmental conditions.*

Our aim as the GEOTRACES International Data Assembly Centre (GDAC) is to provide the data management to promote data sharing and collaboration between research groups and to ensure data are made widely accessible for long-term use.

To find out more follow these links

- **Introduction** — A non-technical insight into the main goals and themes of the GEOTRACES programme.
- **Benefits** — A brief description of the long term benefits of the programme.
- **Role** — The role of the International Data Management Office.

http://www.bodc.ac.uk/geotraces/
IMPORTANT LINKS

GEOTRACES data portal at BODC
http://www.bodc.ac.uk/geotraces/

BCO-DMO is the US GDAC
(US GEOTRACES Data Assembly Center)
http://bco-dmo.org/

US GEOTRACES data managers at BCO-DMO
info@bco-dmo.org
Important to understand:

- station identification system (e.g., super, deep and shallow stations)
- event identification system; event log entries
- robust metadata records to support shared use of data, and future unanticipated use
  - sampling and analytical protocols
  - quality assurance and control procedures
LESSONS LEARNED

US GEOTRACES North Atlantic Transect
lessons learned, based on feedback from NAT
Chief Scientists, PIs and BCO-DMO staff

Bob Groman  Nancy Copley
Dicky Allison  Terry McKee
Danie Kinkade  Steve Gegg
Shannon Rauch
Total of 80 data sets so far
bco-dmo.org
Metadata:
Event number
Station
Cast
GEOTRACES SAMPLE NUMBER
Methods and reference samples

Metadata are essential for accurate and efficient data interpretation and use.
LESSONS LEARNED (NAT)

Courtesy of everyone from US GEOTRACES North Atlantic Transect cruises …

• Every instrument deployment is an event and must be entered in the event log

• Report event numbers and GEOTRACES sample numbers with the data
<table>
<thead>
<tr>
<th>STNNBR</th>
<th>CASTNO</th>
<th>GEOTRC_EVENTNO</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>2008</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>GEOTRC_SAMPNO</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>5523</td>
<td></td>
</tr>
<tr>
<td>5522</td>
<td></td>
</tr>
<tr>
<td>5521</td>
<td></td>
</tr>
<tr>
<td>5520</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>GEOTRC_SAMPNO</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>5519</td>
<td></td>
</tr>
</tbody>
</table>
The unique identifier that enables you to connect data from different investigators.

For example …
Iron speciation from Kristen Buck (BIOS)
Nanomolar Nutrients from Greg Cutter (ODU)
Cobalt data from Mak Saito (WHOI)

… event numbers indicate that all were sampled from Go-Flo bottles.

Data can be merged easily by matching the GEOTRCE_SAMPNO.
LESSONS LEARNED (NAT)

The event number and GEOTRACES sample number enable you to answer this important question, all by yourself …

“when was my sample taken?”
BCO-DMO data managers added the GEOTRACES SAMPLE NUMBERS (and other key fields) to the NAT data sets.

It is highly desirable to include event, station, cast, and GEOTRACES sample numbers in every data set contributed to BCO-DMO.
Importance of event log and base bottle file:

- the event log and the base bottle data file (Niskin and GoFlo) enable merging of different datasets;
- it is imperative that the event log and bottle files be accurate;
- QC review is required;
- access should not be restricted.
Consistent naming of common fields is highly desirable; BCO-DMO chose these for NAT data:

- station_GEOTRC
- cast_GEOTRC
- event_GEOTRC
- sample_GEOTRC (GEOTRACES sample #)
- sample_bottle_GEOTRC (sample bottle #)
- bottle_GEOTRC (Niskin or GoFlo bottle)
- depth_GEOTRC_CTD (depth from CTD P)
Depth and/or Pressure

• Know the difference!
• If a column is labeled depth, be certain that is what those numbers are.
• Perhaps reporting both would help?
Units: 

- Report data as “per unit analyzed”
- e.g., if analysis results are pmol/kg then data should be reported in pmol/kg
- Conversion to per liter can be done using salinity from the bottle data file and lab temperature recorded during analysis (linked by the event number and GEOTRACES sample number).
LESSONS LEARNED (NAT)

Quality flags:

- Many investigators reported these with the NAT data
- But there was a lot of variation
- Flag definitions were not specified

Recommendation: adopt a common flag scheme for use in GEOTRACES

Ocean Data Standards volume 3
http://www.iode.org/mg54_3
The secondary level flags are optional, but can be used to represent details of quality assessment and control or data processing history.
Reference Materials: (Ken Bruland)

- Comparison with GEOTRACES Reference Station Samples (e.g. SAFe) and/or internal laboratory references must be reported with the metadata for each dataset
- Trace element reference concentrations are essential for inter-comparison

Reference:
http://www.geotraces.org/science/intercalibration
THANK YOU

Questions?

Woods Hole, Massachusetts, USA

bco-dmo.org & info@bco-dmo.org