Better Practices for Shipboard Data Management

Cyndy Chandler
Biological and Chemical Oceanography Data Management Office

12 November 2009
Ocean Acidification Short Course
Woods Hole, MA USA
Discussion Topics

- Research Cruise
  - Allocation of sample (wire) time
  - Allocation of sample water
  - Cruise report
  - Data inventory
  - Cruise Sampling Event Log

- Data and Metadata Reporting
  - Data Quality (review)
  - Metadata and Standards
  - Data Centers and National Archives
What about experiments?

- presentation will be specific to cruise activities, but the concepts apply to lab experiments, perturbation or mesocosm experiments.
European Project on OCEan Acidification (EPOCA)

"Guide for Best Practices on Ocean Acidification Research and Data Reporting"


Editors in chief:
Ulf Riebesell, Victoria J. Fabry, Jean-Pierre Gattuso
Pre-cruise Planning

- Station plan
- Allocation of sample (wire) time
- Allocation of sample water

These arrangements should be made prior to the cruise and then reviewed at the first science briefing on board.

Have a plan, write it down, communicate it . . . early and often.
# Cast Plan ~ Sampling Time and WaterAllocation

<table>
<thead>
<tr>
<th>Bottle #</th>
<th>Depth</th>
<th>SF6</th>
<th>SF8</th>
<th>O2</th>
<th>NO2</th>
<th>N2</th>
<th>pCO2</th>
<th>DIC</th>
<th>Talk</th>
<th>DMS</th>
<th>DOC/GDOM</th>
<th>Productivity</th>
<th>POC</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>1</td>
<td>1000</td>
<td>1000</td>
<td>1000</td>
<td>1000</td>
<td>1000</td>
<td>1000</td>
<td>600</td>
<td>600</td>
<td>600</td>
<td>600</td>
<td>600</td>
<td>600</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>1100</td>
<td>1000</td>
<td>1000</td>
<td>1000</td>
<td>1000</td>
<td>1000</td>
<td>600</td>
<td>600</td>
<td>600</td>
<td>600</td>
<td>600</td>
<td>600</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>700</td>
<td>1000</td>
<td>1000</td>
<td>1000</td>
<td>1000</td>
<td>1000</td>
<td>600</td>
<td>600</td>
<td>600</td>
<td>600</td>
<td>600</td>
<td>600</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
<td>600</td>
<td>1000</td>
<td>1000</td>
<td>1000</td>
<td>1000</td>
<td>1000</td>
<td>600</td>
<td>600</td>
<td>600</td>
<td>600</td>
<td>600</td>
<td>1000</td>
</tr>
<tr>
<td>6</td>
<td>5</td>
<td>500</td>
<td>1000</td>
<td>1000</td>
<td>1000</td>
<td>1000</td>
<td>1000</td>
<td>600</td>
<td>600</td>
<td>600</td>
<td>600</td>
<td>600</td>
<td>1000</td>
</tr>
<tr>
<td>7</td>
<td>6</td>
<td>400</td>
<td>1000</td>
<td>1000</td>
<td>1000</td>
<td>1000</td>
<td>1000</td>
<td>600</td>
<td>600</td>
<td>600</td>
<td>600</td>
<td>600</td>
<td>600</td>
</tr>
<tr>
<td>8</td>
<td>7</td>
<td>300</td>
<td>1000</td>
<td>1000</td>
<td>1000</td>
<td>1000</td>
<td>1000</td>
<td>600</td>
<td>600</td>
<td>600</td>
<td>600</td>
<td>600</td>
<td>600</td>
</tr>
<tr>
<td>9</td>
<td>8</td>
<td>250</td>
<td>1000</td>
<td>1000</td>
<td>1000</td>
<td>1000</td>
<td>1000</td>
<td>600</td>
<td>600</td>
<td>600</td>
<td>600</td>
<td>600</td>
<td>600</td>
</tr>
<tr>
<td>10</td>
<td>9</td>
<td>200</td>
<td>1000</td>
<td>1000</td>
<td>1000</td>
<td>1000</td>
<td>1000</td>
<td>600</td>
<td>600</td>
<td>600</td>
<td>600</td>
<td>600</td>
<td>600</td>
</tr>
<tr>
<td>11</td>
<td>10</td>
<td>175</td>
<td>1000</td>
<td>1000</td>
<td>1000</td>
<td>1000</td>
<td>1000</td>
<td>600</td>
<td>600</td>
<td>600</td>
<td>600</td>
<td>600</td>
<td>600</td>
</tr>
<tr>
<td>12</td>
<td>11</td>
<td>150</td>
<td>1000</td>
<td>1000</td>
<td>1000</td>
<td>1000</td>
<td>1000</td>
<td>600</td>
<td>600</td>
<td>600</td>
<td>600</td>
<td>600</td>
<td>600</td>
</tr>
<tr>
<td>13</td>
<td>12</td>
<td>125</td>
<td>1000</td>
<td>1000</td>
<td>1000</td>
<td>1000</td>
<td>1000</td>
<td>600</td>
<td>600</td>
<td>600</td>
<td>600</td>
<td>600</td>
<td>1000</td>
</tr>
<tr>
<td>14</td>
<td>13</td>
<td>100</td>
<td>500</td>
<td>1000</td>
<td>1000</td>
<td>1000</td>
<td>1000</td>
<td>600</td>
<td>600</td>
<td>600</td>
<td>600</td>
<td>600</td>
<td>1000</td>
</tr>
<tr>
<td>15</td>
<td>14</td>
<td>75</td>
<td>500</td>
<td>1000</td>
<td>1000</td>
<td>1000</td>
<td>1000</td>
<td>600</td>
<td>600</td>
<td>600</td>
<td>600</td>
<td>600</td>
<td>1000</td>
</tr>
</tbody>
</table>
Keeping records (recording metadata)

- Log Sheets (formal way to record metadata)
  - station logs
  - sample logs

- Cruise report (cruise metadata)
- Data inventory (dataset metadata)
- Event log (device deployment metadata)
## Log Sheet per Sampling Device

**CTD/Rosette cast**

<table>
<thead>
<tr>
<th>N</th>
<th>Depth</th>
<th>Niskin</th>
<th>Helium</th>
<th>Oxygens</th>
<th>DIC/Alk</th>
<th>TOC/TON</th>
<th>Salts</th>
<th>Nuts</th>
<th>Bact</th>
<th>FRRF</th>
<th>POC/PON</th>
<th>HPLC</th>
<th>Chla</th>
<th>Flow Cyt</th>
<th>DO</th>
<th>pH</th>
<th>15</th>
<th>10</th>
<th>1</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>23.8</td>
<td>1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>0</td>
<td>23.8</td>
<td>2</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>0</td>
<td>23.8</td>
<td>3</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>0</td>
<td>23.8</td>
<td>4</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>0</td>
<td>23.8</td>
<td>5</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>6</td>
<td>6</td>
<td>0</td>
<td>23.8</td>
<td>6</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>7</td>
<td>7</td>
<td>0</td>
<td>23.8</td>
<td>7</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>8</td>
<td>8</td>
<td>0</td>
<td>23.8</td>
<td>8</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>9</td>
<td>9</td>
<td>0</td>
<td>23.8</td>
<td>9</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>10</td>
<td>10</td>
<td>0</td>
<td>23.8</td>
<td>10</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>9</td>
<td>9</td>
<td>9</td>
<td>9</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>11</td>
<td>11</td>
<td>0</td>
<td>23.8</td>
<td>11</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>12</td>
<td>12</td>
<td>0</td>
<td>23.8</td>
<td>12</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>11</td>
<td>11</td>
<td>11</td>
<td>11</td>
<td>11</td>
<td>11</td>
</tr>
<tr>
<td>13</td>
<td>13</td>
<td>0</td>
<td>23.8</td>
<td>13</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>12</td>
<td>12</td>
<td>12</td>
<td>12</td>
<td>12</td>
<td>12</td>
</tr>
<tr>
<td>14</td>
<td>14</td>
<td>0</td>
<td>23.8</td>
<td>14</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>13</td>
</tr>
<tr>
<td>15</td>
<td>15</td>
<td>0</td>
<td>23.8</td>
<td>15</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>14</td>
<td>14</td>
<td>14</td>
<td>14</td>
<td>14</td>
<td>14</td>
</tr>
<tr>
<td>16</td>
<td>16</td>
<td>0</td>
<td>23.8</td>
<td>16</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>17</td>
<td>17</td>
<td>0</td>
<td>23.8</td>
<td>17</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>16</td>
<td>16</td>
<td>16</td>
<td>16</td>
<td>16</td>
<td>16</td>
</tr>
<tr>
<td>18</td>
<td>18</td>
<td>0</td>
<td>23.8</td>
<td>18</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>17</td>
<td>17</td>
<td>17</td>
<td>17</td>
<td>17</td>
<td>17</td>
</tr>
<tr>
<td>19</td>
<td>19</td>
<td>0</td>
<td>23.8</td>
<td>19</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>18</td>
<td>18</td>
<td>18</td>
<td>18</td>
<td>18</td>
<td>18</td>
</tr>
<tr>
<td>20</td>
<td>20</td>
<td>0</td>
<td>23.8</td>
<td>20</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>19</td>
<td>19</td>
<td>19</td>
<td>19</td>
<td>19</td>
<td>19</td>
</tr>
<tr>
<td>21</td>
<td>21</td>
<td>0</td>
<td>23.8</td>
<td>21</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td>22</td>
<td>22</td>
<td>0</td>
<td>23.8</td>
<td>22</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>21</td>
<td>21</td>
<td>21</td>
<td>21</td>
<td>21</td>
<td>21</td>
</tr>
<tr>
<td>23</td>
<td>23</td>
<td>0</td>
<td>23.8</td>
<td>23</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>22</td>
<td>22</td>
<td>22</td>
<td>22</td>
<td>22</td>
<td>22</td>
</tr>
<tr>
<td>24</td>
<td>24</td>
<td>0</td>
<td>23.8</td>
<td>24</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>23</td>
<td>23</td>
<td>23</td>
<td>23</td>
<td>23</td>
<td>23</td>
</tr>
</tbody>
</table>
Cruise Report

- basic cruise metadata
  - Cruise ID - a way to identify the cruise
    - KN195-08 (ship, voyage and leg)
    - KM0908 (ship, 2 digit year and sequential voyage for year)
  - dates and ports

- personnel manifest
  - list of everyone on board and contact information
  - their role during the cruise

- data inventory
  - list of who is expecting to collect what data during cruise

- event log
  - list of every device deployment
Better Practices for Shipboard Data

- Data Management Best Practices Guide compiled by BCO-DMO based on experience from US GLOBEC and US JGOFS

- a collection of better practice recommendations for management of data from research cruises

- available as a PDF download from: http://bco-dmo.org/resources
## Data Inventory  (list of expected measurements)

<table>
<thead>
<tr>
<th>Instrument</th>
<th>Measurement</th>
<th>PI_name</th>
<th>co-PI_name</th>
</tr>
</thead>
<tbody>
<tr>
<td>TMR</td>
<td>Bottle O2</td>
<td>Casciotti</td>
<td>Frame; Sieracki</td>
</tr>
<tr>
<td>TMR</td>
<td>Nitrate isotope</td>
<td>Casciotti</td>
<td>nd</td>
</tr>
<tr>
<td>TMR</td>
<td>Uptake Expts-Fe Cd Zn Hg Ni</td>
<td>Cox</td>
<td>Saito</td>
</tr>
<tr>
<td>CTD</td>
<td>Productivities; selected stations</td>
<td>DiTullio</td>
<td>nd</td>
</tr>
<tr>
<td>CTD</td>
<td>Pigments</td>
<td>DiTullio</td>
<td>nd</td>
</tr>
<tr>
<td>CTD</td>
<td>Uptake Expts-carbon C14</td>
<td>Ditullio</td>
<td>Riseman</td>
</tr>
<tr>
<td>ON_DECK_PUMP</td>
<td>Incubation Expts-Iron; DMSP effects</td>
<td>DiTullio</td>
<td>nd</td>
</tr>
<tr>
<td>TMR</td>
<td>N2O</td>
<td>Frame</td>
<td>Casciotti</td>
</tr>
<tr>
<td>TMR</td>
<td>Methyl Mercury</td>
<td>Hammerschmidt</td>
<td>nd</td>
</tr>
<tr>
<td>CTD</td>
<td>nifH gene expression</td>
<td>Hilton</td>
<td>Zehr; Webb</td>
</tr>
<tr>
<td>TMR</td>
<td>FeL</td>
<td>Lam</td>
<td>Buck</td>
</tr>
<tr>
<td>MCLANE</td>
<td>Fe-Metal Particulates</td>
<td>Lam</td>
<td>nd</td>
</tr>
<tr>
<td>MCLANE</td>
<td>POC</td>
<td>Lam</td>
<td>nd</td>
</tr>
<tr>
<td>nd</td>
<td>Aerosol metals</td>
<td>Lamborg</td>
<td>nd</td>
</tr>
<tr>
<td>nd</td>
<td>Sediment trap fluxes including metals</td>
<td>Lamborg</td>
<td>nd</td>
</tr>
<tr>
<td>TMR</td>
<td>Total Dissolved Mercury</td>
<td>Lamborg</td>
<td>nd</td>
</tr>
<tr>
<td>TMR</td>
<td>DOC</td>
<td>Morris</td>
<td>Carlson</td>
</tr>
<tr>
<td>CTD</td>
<td>Heterotrophic bacterial counts-act</td>
<td>Morris</td>
<td>nd</td>
</tr>
<tr>
<td>CTD</td>
<td>Proteomics</td>
<td>Morris</td>
<td>Rocap</td>
</tr>
<tr>
<td>CTD</td>
<td>Pro and Syn phylgeny-ecotype</td>
<td>Rocap</td>
<td>Webb</td>
</tr>
<tr>
<td>ON_DECK_PUMP</td>
<td>Incubation Expts-Phosphate</td>
<td>Rocap</td>
<td>nd</td>
</tr>
<tr>
<td>LAB</td>
<td>Sampling Event Log</td>
<td>Saito</td>
<td>nd</td>
</tr>
</tbody>
</table>
What is a ‘Cruise Sampling Event Log’?

- a chronological record of all scientific sampling events that happened during a cruise, wherein each sampling event is assigned a unique identifier

Why is an event log important?

- event logs with unique sampling event identifiers help to …
  - integrate observations from the plethora of sampling devices deployed during a cruise
  - understand relative timing between events
### July 9th final summary of cruise activities

<table>
<thead>
<tr>
<th>file VERTIGO final cruise activities.xls</th>
<th>172_173</th>
<th>173_174</th>
<th>174_175</th>
<th>175_176</th>
<th>176_177</th>
<th>177_178</th>
<th>178_179</th>
<th>179_180</th>
<th>180_181</th>
</tr>
</thead>
<tbody>
<tr>
<td>Julian Day</td>
<td>172_173</td>
<td>173_174</td>
<td>174_175</td>
<td>175_176</td>
<td>176_177</td>
<td>177_178</td>
<td>178_179</td>
<td>179_180</td>
<td>180_181</td>
</tr>
<tr>
<td>hours</td>
<td>day 1</td>
<td>day 2</td>
<td>day 3</td>
<td>day 4</td>
<td>day 5</td>
<td>day 6</td>
<td>day 7</td>
<td>day 8</td>
<td>day 9</td>
</tr>
<tr>
<td>0</td>
<td>SS#1 CTD 2-18</td>
<td>SS#1 CTD 2-15</td>
<td>Trull trap in-300</td>
<td>Survey CTD #24 &amp; drifters</td>
<td>MCCNESS</td>
<td>MULVFS</td>
<td>NBST 300 out</td>
<td>MOCNESS</td>
<td>MULVFS</td>
</tr>
<tr>
<td>2</td>
<td>SS#1 CTD 2-18</td>
<td>SS#1 CTD 2-18</td>
<td>Launch Clap 150</td>
<td>Survey CTD #25</td>
<td>MULVFS</td>
<td>NBST 300 out</td>
<td>MULVFS</td>
<td>NBST 300 out</td>
<td>MCCNESS</td>
</tr>
<tr>
<td>4</td>
<td>SS#1 CTD 2-18</td>
<td>SS#1 CTD 2-18</td>
<td>Std Tid CTD #10</td>
<td>Std Tid CTD #15</td>
<td>MULVFS</td>
<td>NBST 300 out</td>
<td>MULVFS</td>
<td>NBST 300 out</td>
<td>MCCNESS</td>
</tr>
<tr>
<td>8</td>
<td>SS#1 CTD 2-18</td>
<td>SS#1 CTD 2-15</td>
<td>Launch Clap 500</td>
<td>Launch-optical trap</td>
<td>MULVFS</td>
<td>MOCNESS</td>
<td>MULVFS</td>
<td>MOCNESS</td>
<td>MULVFS</td>
</tr>
<tr>
<td>10</td>
<td>SS#1 CTD 2-18</td>
<td>SS#1 CTD 2-15</td>
<td>Launch-optical trap</td>
<td>MOCNESS</td>
<td>MULVFS</td>
<td>MOCNESS</td>
<td>MULVFS</td>
<td>MOCNESS</td>
<td>MULVFS</td>
</tr>
<tr>
<td>12</td>
<td>SS#1 CTD 2-15</td>
<td>MOCNESS test</td>
<td>Launch NBST 300</td>
<td>MOCNESS</td>
<td>MULVFS</td>
<td>MOCNESS</td>
<td>MULVFS</td>
<td>MOCNESS</td>
<td>MULVFS</td>
</tr>
<tr>
<td>14</td>
<td>SS#1 CTD 2-15</td>
<td>MOCNESS test</td>
<td>Launch NBST 300</td>
<td>MOCNESS</td>
<td>MULVFS</td>
<td>MOCNESS</td>
<td>MULVFS</td>
<td>MOCNESS</td>
<td>MULVFS</td>
</tr>
<tr>
<td>16</td>
<td>SS#1 CTD 2-15</td>
<td>MOCNESS test</td>
<td>Launch NBST 300</td>
<td>MOCNESS</td>
<td>MULVFS</td>
<td>MOCNESS</td>
<td>MULVFS</td>
<td>MOCNESS</td>
<td>MULVFS</td>
</tr>
<tr>
<td>18</td>
<td>SS#1 CTD 2-15</td>
<td>MOCNESS test</td>
<td>Launch NBST 300</td>
<td>MOCNESS</td>
<td>MULVFS</td>
<td>MOCNESS</td>
<td>MULVFS</td>
<td>MOCNESS</td>
<td>MULVFS</td>
</tr>
<tr>
<td>20</td>
<td>SS#1 CTD 2-15</td>
<td>MOCNESS test</td>
<td>Launch NBST 300</td>
<td>MOCNESS</td>
<td>MULVFS</td>
<td>MOCNESS</td>
<td>MULVFS</td>
<td>MOCNESS</td>
<td>MULVFS</td>
</tr>
<tr>
<td>22</td>
<td>SS#1 CTD 2-15</td>
<td>SS#1 CTD 2-15</td>
<td>Trull trap in-300</td>
<td>Survey CTD #21 &amp; drifters</td>
<td>MCCNESS</td>
<td>MULVFS</td>
<td>optical trap out</td>
<td>MCCNESS</td>
<td>MULVFS</td>
</tr>
</tbody>
</table>

**VERTIGO project KM0414 ALOHA cruise**

**Sampling event matrix**

**R/V Kilo Moana** (University of Hawaii Marine Center)
Why is an event log important?

- The unique sampling event identifier helps to integrate observations from discrete data sets.

- Example: CTD station 4 cast 2 is assigned event number 20080904.1342 … the Niskin bottle nutrient data and pigment data from that cast can be integrated using that event number.
What types of information are important to record in the event log?

- the unique sampling event identifier
  - Example: YYYYMMDD.hhmm
  - YYYYMMDD + time/2400
  - SCYYYYMMDD.hhmm (SC=ship code)
- station and cast number
- date and time (UTC)
- position (latitude and longitude)
other important fields in the event log . . .

- instrument [package] type
  CTD, TM, drifter net, penguin

- name of person
  responsible for sampling event

- activity descriptor
  e.g. deployment, recovery
  start, max depth, end, abort
additional possible fields . . .

- description of activity and/or comments (free text)
  e.g. first cast after retermination

- local time (important for biology cruise)
- timezone

- cruise notebook or subsample log page

- position relative to a feature (eddy center or treatment patch)
### shipboard sampling event log

generated automatically using some algorithm
date, time and position from shipboard system

<table>
<thead>
<tr>
<th>event</th>
<th>date</th>
<th>time</th>
<th>time_L</th>
<th>lon</th>
<th>lat</th>
<th>ev_type</th>
<th>person</th>
<th>activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>0212208</td>
<td>20020121</td>
<td>2208</td>
<td>1108</td>
<td>-175.220</td>
<td>-53.572</td>
<td>CTD001</td>
<td>nd</td>
<td>CTD001</td>
</tr>
<tr>
<td>0230442</td>
<td>20020123</td>
<td>0442</td>
<td>1742</td>
<td>-171.480</td>
<td>-55.398</td>
<td>CTD002</td>
<td>Wang</td>
<td>CTD002</td>
</tr>
<tr>
<td>0231556</td>
<td>20020123</td>
<td>1556</td>
<td>0456</td>
<td>-171.583</td>
<td>-55.407</td>
<td>ZooTow</td>
<td>Landry</td>
<td>ZooplankTow</td>
</tr>
<tr>
<td>0240153</td>
<td>20020124</td>
<td>0153</td>
<td>1453</td>
<td>-171.490</td>
<td>-55.329</td>
<td>TM001</td>
<td>Wang</td>
<td>TM001</td>
</tr>
<tr>
<td>0240356</td>
<td>20020124</td>
<td>0356</td>
<td>1656</td>
<td>-171.336</td>
<td>-55.314</td>
<td>CTD003</td>
<td>Bailey</td>
<td>CTD004</td>
</tr>
<tr>
<td>0240745</td>
<td>20020124</td>
<td>0745</td>
<td>2045</td>
<td>-171.408</td>
<td>-55.335</td>
<td>Pump_Cast</td>
<td>Andrews</td>
<td>PumpCast01</td>
</tr>
<tr>
<td>0241133</td>
<td>20020124</td>
<td>1133</td>
<td>0033</td>
<td>-171.405</td>
<td>-55.324</td>
<td>TM002</td>
<td>Wang</td>
<td>TM002</td>
</tr>
<tr>
<td>0241319</td>
<td>20020124</td>
<td>1319</td>
<td>0219</td>
<td>-171.384</td>
<td>-55.333</td>
<td>CTD005</td>
<td>Timothy</td>
<td>CTD005</td>
</tr>
<tr>
<td>0241435</td>
<td>20020124</td>
<td>1435</td>
<td>0335</td>
<td>-171.385</td>
<td>-56.333</td>
<td>HPT</td>
<td>Tanner</td>
<td>HandPlankTow</td>
</tr>
<tr>
<td>0241520</td>
<td>20020124</td>
<td>1520</td>
<td>0420</td>
<td>-171.383</td>
<td>-55.337</td>
<td>TM003</td>
<td>Landry</td>
<td>TM003</td>
</tr>
</tbody>
</table>
Event Log Data Sources

- arrangements were made, agreed upon and reviewed at the first science briefing on board . . .

- and everyone agreed on the common data source for:

  - date and time
  - shipboard network and UTC *(not your wristwatch)*
  - position information
    - decimal degrees lat/lon *(agree on required precision)*
Final Event Log

- should be an electronic file in plain text (TSV or CSV)

- many researchers record events on paper logs in the main lab, and then enter the records into Excel

- if the original event log entries were made on paper log sheets, scan the originals and convert to PDF

- some research vessels support event logging applications, and NSF is funding the development of an event logger for use on UNOLS vessels (R2R project, rvdata.us)
Constructing the cruise report . . .

Cruise ID  manifest  data inventory  sampling event log
Discussion Topics

- Research Cruise
  - Allocation of sample (wire) time
  - Allocation of sample water
  - Cruise report
  - Data inventory
  - Cruise Sampling Event Log

- Data and Metadata Reporting
  - Data Quality (review)
  - Metadata and Standards
  - Data Centers and National Archives
Data Quality

- Quality Assessment
  It is important to have a system in place to assure outside users that the analytical results produced are of proven and known quality. (Andrew Dickson, 2009)

- to make an assessment of quality we must also keep in mind the time [the data were] collected, the methodology – the capability of the time (A.K. Sinha, Virginia Tech, 2006)

- Data Quality involves
  quality assurance – done prior to measurement
  quality control – done after measurement
Data Quality

- It’s important to understand that reporting on the ‘quality’ of a dataset (a set of measurements) is not a statement about its value to the community.

- Just because a dataset might be ranked lower on some scale used to assess quality, does not mean it is of less value to the community.

- A dataset of known quality is more valuable than one that lacks the quality assessment metadata.
Data Quality (of your data)

Questions asked during sampling and analysis (related to accuracy and precision*):
- How good do I need the measurement? (QA)
- How good did I get the measurement? (QC)

Data Quality Metadata:
- report the questions above and answers with the data
- much of this information still fits in the methods section of the peer-reviewed publication – but the problem is that all the data no longer fit in that same publication
- important to document the data quality assessment with the published dataset (reported as metadata)
Data Quality (of colleague’s data)

Thank you to Andrew Dickson (previous lecture)

In order to assess ‘fitness for purpose’ it helps to know why the measurement was made
ones ability to ascertain ‘good enough’ is related to the uncertainty associated with the measurement
uncertainty relative to your needs as defined by the research topic
Metadata

- Needed to automate the process of data discovery
  - like using a library catalog to locate a resource
- Needed to determine the fitness of a data set for use
  - particularly regarding quality (“fitness for purpose”)
- Needed to facilitate accurate data interpretation
  - e.g. units of measurement, data format

- Metadata records are expensive to generate
  - and may require additional expertise to define
- But the benefits are substantial
  - metadata make it possible to find data sets, and use them effectively
  - they allow the benefits of investments in data to be realized
Metadata

- in the US, 1995 was the year that state government agencies started devoting resources to metadata capture

- motivated by:
  - 1995 Paperwork Reduction Act (104th US Congress)
  - Internet and HTML = World Wide Web
  - desire to automate public access to government documents following the requirement that agencies establish ‘locator services’ for federal information

- Dublin Core Metadata Initiative (2001)
Work in Progress . . .

There is no cookbook of instructions – or at least the book isn’t finished – and establishing best practices will continue to be iterative.

- research themes are becoming more complex
- cost of doing research will continue to increase
- *in situ* data can not be collected ‘again’

The water samples collected in March 2009 from 22° 45'N, 158° 00'W can not be collected again.
Metadata ~ the goal

- document the quality assurance and control measures applied to the measurements during sampling and analysis
  - what protocols were followed (include reference)
  - were replicates done, include results of control chart
  - were inter-comparisons done (analytical techniques, different labs?)
  - were reference materials used (which ones)
  - what was done to account for T and P dependencies
  - were data adjusted based on results of quality control procedures

- objective: to report sufficient metadata to support
  - determination: are these data 'fit for purpose'
  - accurate re-use of the data

- metadata reporting is especially important when the protocols are still being developed (e.g. OA sampling and analytical techniques)

- remember local v global (in space and time); resultant data will be used and re-used by colleagues
European Project on OCEan Acidification (EPOCA)

"Guide for Best Practices on Ocean Acidification Research and Data Reporting"


Editors in chief:
Ulf Riebesell, Victoria J. Fabry, Jean-Pierre Gattuso
recording metadata at sea... is problematic

Who’s recording the metadata?

Think I’ll go record some metadata.

... this is the office!
Metadata matter

- Ocean acidification research is and will continue to be...
  - Expensive (research cruises are resource intensive)
    - Fuel costs
    - Equipment allocation
    - People time (highly trained people at sea)
  - Collaborative
    - Team projects are more complicated than individual research
  - Important – answers are needed to enable science-based decision support for legislative policies

- Means the metadata matter more
Data management partners: BCO-DMO

- metadata forms ([http://bcodmo.org/resources](http://bcodmo.org/resources))
  - Program
  - Project
  - Deployment (e.g. cruise)
  - Dataset metadata contributed with the data

- data contributed in any format (often as Excel files)

- researchers work in partnership with BCO-DMO staff members to manage data through all phases of a project
Standards

- community adopted standards for . . .
  - sampling and analytical protocols
    - assessment of quality assurance and control
  - metadata content standards
    - FGDC or ISO may be required by some Data Centers

Use of standards can facilitate data integration.

At the moment, most of the effort relating to standards is being handled by data centers.
Data Centers and National Archives

- BCO-DMO
  - Biological and Chemical Oceanography Data Management Office
  - for researchers funded by US NSF OCE

- CDIAC (Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory)

- NODC
  - National Oceanographic Data Center
  - permanent data archive for US researchers funded by NOAA or NSF

- analogous *ODC or WDC in other nations
Questions?

conclusion part 2 of 2

thank you
end of day 1
part 2 of
total 3 part data management section

Part 1: Monday, ODV Introduction, Reiner Schlitzer

Part 2: Thursday, Data Management: Introduction and Shipboard Data, Cyndy Chandler

Part 3: Friday, Contributing Data to Data Centers, Alex Kozyr