

Topic 4 – Data Formats

Adam Leadbetter

Roy Lowry

Justin Buck

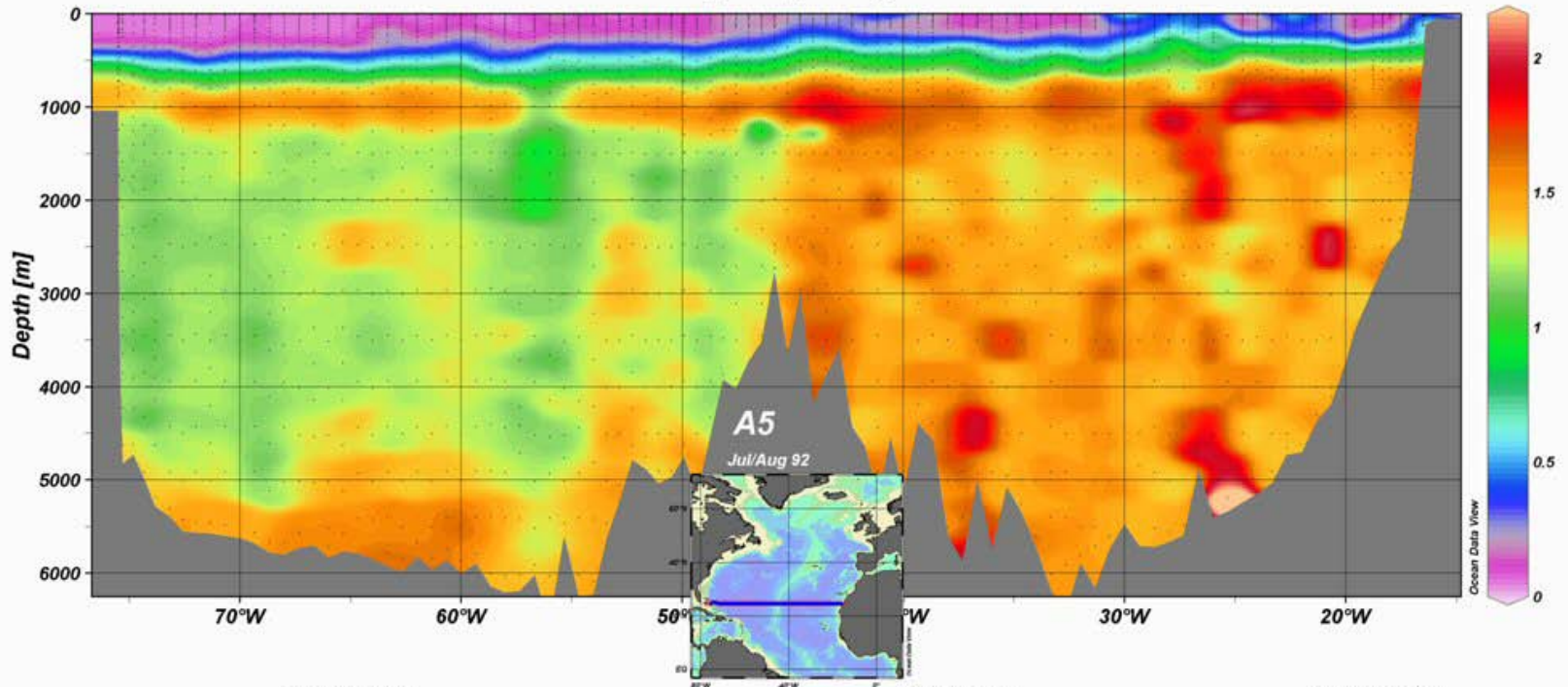
British Oceanographic Data Centre

Introduction

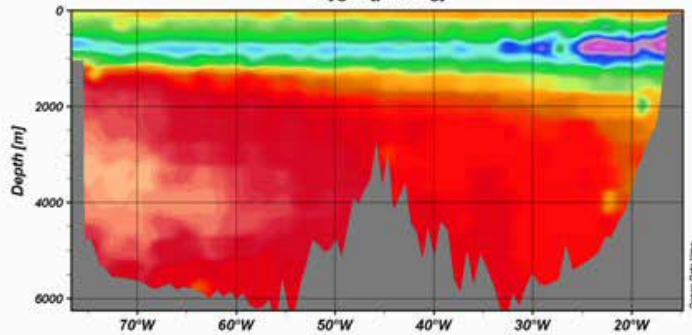
- Why have common / interoperable data file formats?
 - Aids in data assimilation
 - Gives options of building common tools
- Which:
 - Makes collaborative science projects possible with fewer “startup costs”

eWOCE

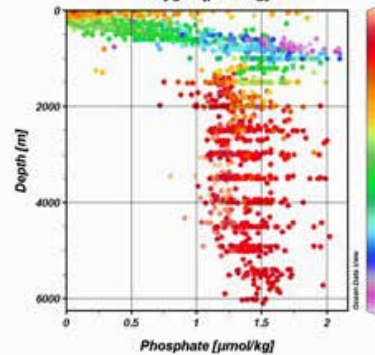
Phosphate [$\mu\text{mol/kg}$]



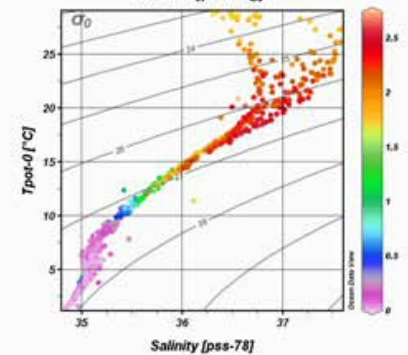
Oxygen [$\mu\text{mol/kg}$]



Oxygen [$\mu\text{mol/kg}$]



CFC-11 [pmol/kg]



SeaDataNet

- “Delayed mode” data files
- Three formats
 - Two ASCII based
 - ODV and Medatlas
 - One NetCDF based
 - Using the CF Conventions
- All carry data quality flags

SeaDataNet

- ASCII formats carry the “semantic header”
 - Linking local variable names to common vocabulary terms

```
<subject>SDN:LOCAL:{VarName}</subject>
```

```
<object>SDN:P01::{ParamKey}</object>
```

```
<units>SDN:P06::{UnitsKey}</units>
```

SeaDataNet

- “SeaDataNet-CF”
 - Conforms to CF 1.6
 - Profile, timeseries or trajectory
 - Extensions
 - Link the metadata to the data
 - Using a SDN namespace
 - CDI
 - EDMO

SeaDataNet

- Provide links to standardised concept description
 - Again using SDN namespace

```
PSAL:sdn_parameter_urn = "SDN:Po1::PSLTZZo1" ;  
PSAL:sdn_parameter_name = "Practical salinity of the  
water body" ;  
PSAL:sdn_uom_urn = "SDN:Po6::UUUU" ;  
PSAL:sdn_uom_name = "Dimensionless"
```

- SDN attributes easily added to other “flavours”

OceanSites

- NetCDF based
 - Real-time and delayed mode data
 - Global attributes
 - What, Where, When, Who, How
 - Data variables use
 - CF Standard Names
 - Parameter names based on SDN Po2
 - Documentation notes not strictly standardised
 - Uncertainty information and quality flags
 - Platform and sensor information

Argo

- NetCDF based
 - Profiles and trajectories
 - Global attributes from CF conventions
 - Who, What
 - Data quality flag scale
 - History of each action performed on the measurement
 - Parameters currently encoded as per OceanSites

Argo

- NetCDF based
 - Parameter encoding is currently evolving
 - Driven by new sensor developments (biogeochemical)
 - Under consideration is assigning SDN parameters & units to each variable

Everyone's Gliding Observatories

- NetCDF format
 - Based on OceanSites
 - Uses CF conventions
 - Global attributes
 - What, Where, When, Who, How
 - Glider characteristics, sensor & deployment info
 - Parameters encoded as per OceanSites
 - Uses SDN parameter names and units – Interoperability!

Common Ground

- NetCDF
 - Particularly the CF conventions
- Formats include data quality flags
- See also:
 - Glider format intercomparison spreadsheet

Challenges

- Namespaces for attributes
 - How do I know whose conventions the file uses?
 - Need to persuade others to declare their NS's
- Make sure your CF really is CF!
- Encoding data quality flags