Extensions to Better Support the Needs of \textit{in situ} Data and Observational Communities

\textbf{Focus:}
\begin{itemize}
  \item uncertainty in geolocation data
  \item domain/community-specific metadata
  \item support for “summary”, non-scalar data
\end{itemize}
OIIP Goals

- **Extend available (higher TRL) technologies** to address key interoperability and data challenges associated with oceanographic *in situ* datasets, focusing on marine animal electronic tagging data as a representative (but also more challenging) use case.

  *Components Leveraged:* NCEI.nc templates, ROSETTA, THREDDS, CMC, DMAS, Tagbase

- **Engage Instrument manufacturers** (Wildlife Computers)

- **Develop improved PO.DAAC capacity to support NASA field campaign** data (SPURS, OMG) via the integration of these technology components within system workflows with a view to operational DAAC infusion.
eTag Sensors & Data

- Biological “Gliders”
- Horizontally & vertically resolved physical data
  minimally: light level, pressure/Z, temperature
- Movement patterns, habitat utilization, stock structure

Mako Shark – N. Atlantic, 6 months migration

Albacore Tuna– E. Trop. Pacific, 2 years of Archival Data
I. Support for Geolocational Uncertainty

- Errors in positional data are ubiquitous and important but rarely represented

- CF standard provides the \textbf{Cell Bounds} construct for defining grid cell extents, but unclear whether this is applicable as a framework for representing uncertainty in geolocations of point, profile, trajectory series data

Two eTAG use cases

\textit{Data from Implantable Archival Tags}
- series of Lat/Lon positional estimates
- with associated estimation error values Lat\_err, Lon\_err

\textit{Argos positions from SPOT and PAT tags}
- series of Lat/Lon positional estimates
- with associated accuracy Class codes

<table>
<thead>
<tr>
<th>Class</th>
<th>Type</th>
<th>Estimated error*</th>
<th>Number of messages received per satellite pass</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Least Squares</td>
<td>Kalman Filter</td>
</tr>
<tr>
<td>G</td>
<td>GPS</td>
<td>&lt; 100m</td>
<td>1 message or more</td>
</tr>
<tr>
<td>3</td>
<td>Argos</td>
<td>&lt; 250m</td>
<td>4 messages or more</td>
</tr>
<tr>
<td>2</td>
<td>Argos</td>
<td>250m &lt; &lt; 500m</td>
<td>4 messages or more</td>
</tr>
<tr>
<td>1</td>
<td>Argos</td>
<td>500m &lt; &lt; 1500m</td>
<td>4 messages or more</td>
</tr>
<tr>
<td>0*</td>
<td>Argos</td>
<td>&gt; 1500m</td>
<td>4 messages or more</td>
</tr>
<tr>
<td>A</td>
<td>Argos</td>
<td>No accuracy estimation</td>
<td>Unbounded accuracy estimation</td>
</tr>
<tr>
<td>B</td>
<td>Argos</td>
<td>No accuracy estimation</td>
<td>Unbounded accuracy estimation</td>
</tr>
<tr>
<td>Z</td>
<td>Argos</td>
<td>Invalid location (available only for Service Plus/Auxiliary Location Processing)</td>
<td></td>
</tr>
</tbody>
</table>
II. Improved Support for Community Metadata

- CF/ACDD provides comprehensive standards for geospatial attributes ...

.. but what about support for domain-specific metadata that may critical to preservation, discoverability and interpretation of *in situ* data?

Towards a framework: eTag metadata use case

- Developed a community vetted inventory & specification of eTag metadata attributes (130) categorized by:
  - Thematic type (10):
    - Animal, Device, Attachment, Deployment, Recovery ...
  - Disposition:
    - Required, Recommended, Optional

- Solicited/included comments from tagging community & external collaborators (IATTC, SWFSC, WC)
II Support for Community Metadata (2/2)

- Developing a Framework for packaging such rich metadata attribute sets in .nc4 files

- Utilization of **Group** structures to **organize** metadata thematically/hierarchically

- Approach to **encoding** attributes
  - Currently: simple key-value pairs
  - Future:
    - Explore ISO, RDF?
    - Attribute vocabulary standardization/reconciliation (mappings to EML, SensorML, etc)
III. Support for “Summary” data

- Best practices/standards for representing “summary” /non-scalar data?

**PAT tag data use case**

- Detailed time series available on if the tag is physically retrieved
- Only position and daily summary data transmitted to satellite upon surfacing (depth/temperature bin-frequencies, PDT min/max)

**Satellite-in situ data Matchup use Case**

- .nc format specification for matchup output file (NASA/AIST-DOMS project)
- Use of Groups to separately package matched satellite and in-situ records (variables: Lat, Lon, Z, Time, Measurements)
- Reconciliation of matched records between satellite and in situ groups via Matchup lookup array of matching record IDs (supports many-to many relationships)
<table>
<thead>
<tr>
<th>Group</th>
<th>Satellite Group Variables</th>
<th>In-situ Group Variables</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(lat, lon, time, Measurements)</td>
<td>(lat, lon, Z, time, Measurements)</td>
</tr>
</tbody>
</table>

**Satellite Group Variables**

- **Variables**: (lat, lon, time, Measurements)
- **Definition**: Variables associated with satellite data, including latitude, longitude, time, and various measurements such as temperature, humidity, or precipitation.

**In-situ Group Variables**

- **Variables**: (lat, lon, Z, time, Measurements)
- **Definition**: Variables associated with in-situ data, including latitude, longitude, depth (Z), time, and various measurements similar to satellite data.

**Matchup ID Array**

- **Definition**: Array containing matchup IDs for satellite and in-situ data pairs for comparison.

**Global Attributes**

- **CF/ACCD**
- **+matchup query URL**

**Descriptions**

- **CF/ACCD**: Clouds and Fire Atlas of the Continental United States Data Format
- **+matchup query URL**: URL for accessing matchup data queries