ACCESS15-0017

Oceanographic In-situ Interoperability Project (OIIP)

V. Tsontos, S. Arms, C. Lam, N. Quach, C. & Thompson
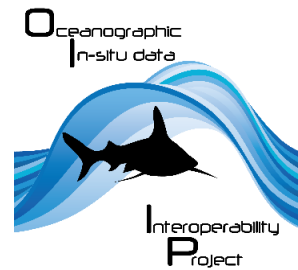
JPL   UCAR   UMASS-Boston

Extensions to Better Support the Needs of *in situ* Data and Observational Communities

Focus:
- uncertainty in geolocation data
- domain/community-specific metadata
- support for "summary", non-scalar data

CL#17-2286

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
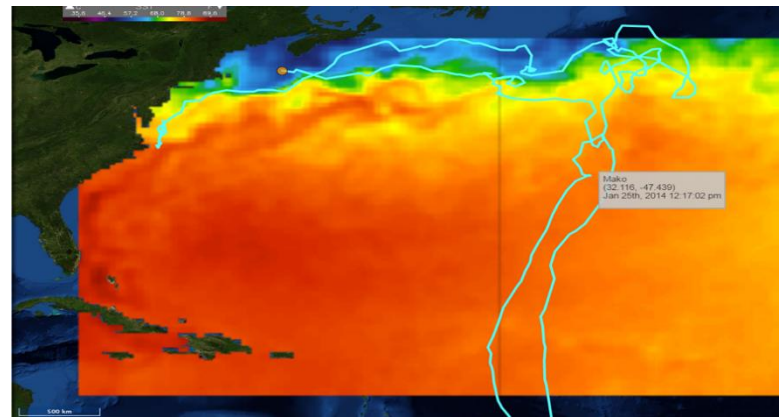
eTag Sensors & Data
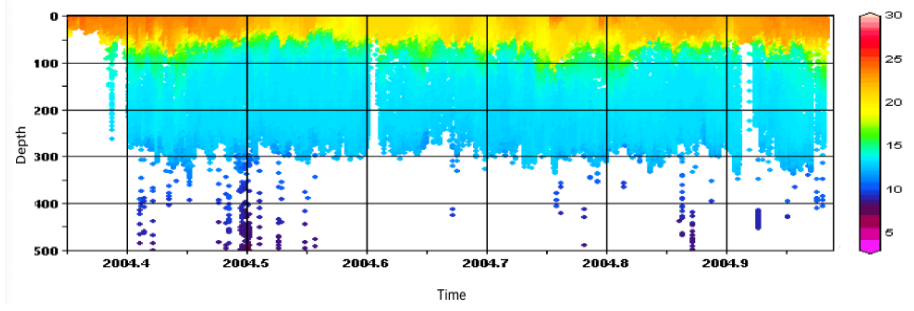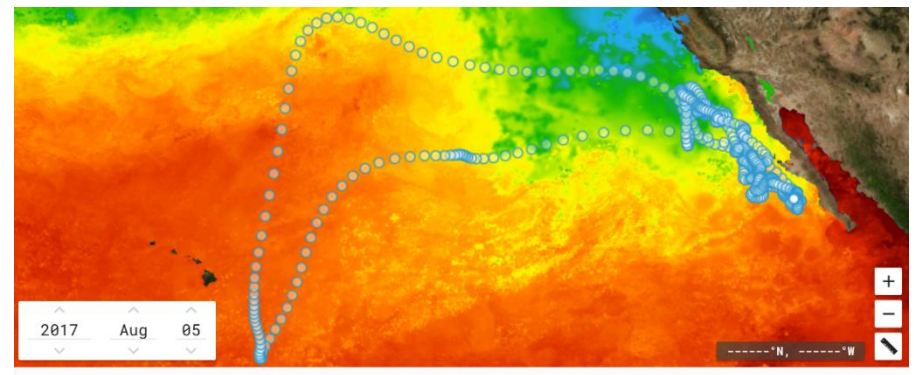

SPOT tag

PAT tag on Bluefin Tuna

Implantable Archival tag

- 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, <u>2 years</u> of Archival Data

I. Support for Geolocational Uncertainty

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

- CF standard provides the *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

Data from Implantable Archival Tags

- series of Lat/Lon positional estimates
- with associated estimation error values Lat_err, Lon_err

Argos positions from SPOT and PAT tags

- series of Lat/Lon positional estimates
- with associated accuracy Class codes
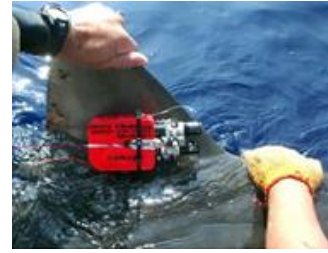
Class	Type	Estimated error*		Number of messages received per satellite pass	
		Least Squares	Kalman Filter	Least Squares	Kalman Filter
G	GPS	< 100m		1 message or more	
3	Argos	< 250m		4 messages or more	
2	Argos	250m < < 500m		4 messages or more	
1	Argos	500m < < 1500m		4 messages or more	
0*	Argos	> 1500m		4 messages or more	
A	Argos	No accuracy estimation	Unbounded accuracy estimation	3 messages	
B	Argos	No accuracy estimation	Unbounded accuracy estimation	messages	1 or 2 messages
Z	Argos	Invalid location (available only for Service Plus/Auxiliary Location Processing)			

SPOT

Pop-up Archival (PAT)

II. Improved Support for Community Metadata (1/2)

- 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)

Category/ Group	Attribute name	Description	Example	Comments	Necessity	fish shark	turtle	mam mal	bird	Notes
device	manufacturer	Name of manufacturer	Wildlife Computers, Microwave Telemetry, Lotek Wireless, Desert Star Systems, CEFAS, StarOddi, Sea Mammal Research Unit, Vemco, Loggerhead Instruments, Biologging Solutions, Little Leonardo, Teleonics etc.		required	1	1	1	1	Some edits to the exam by IATTC
device	model	Model name	MiniPAT		required	1	1	1	1	
device	serial_number	Serial number	16P0100		required	1	1	1	1	

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)
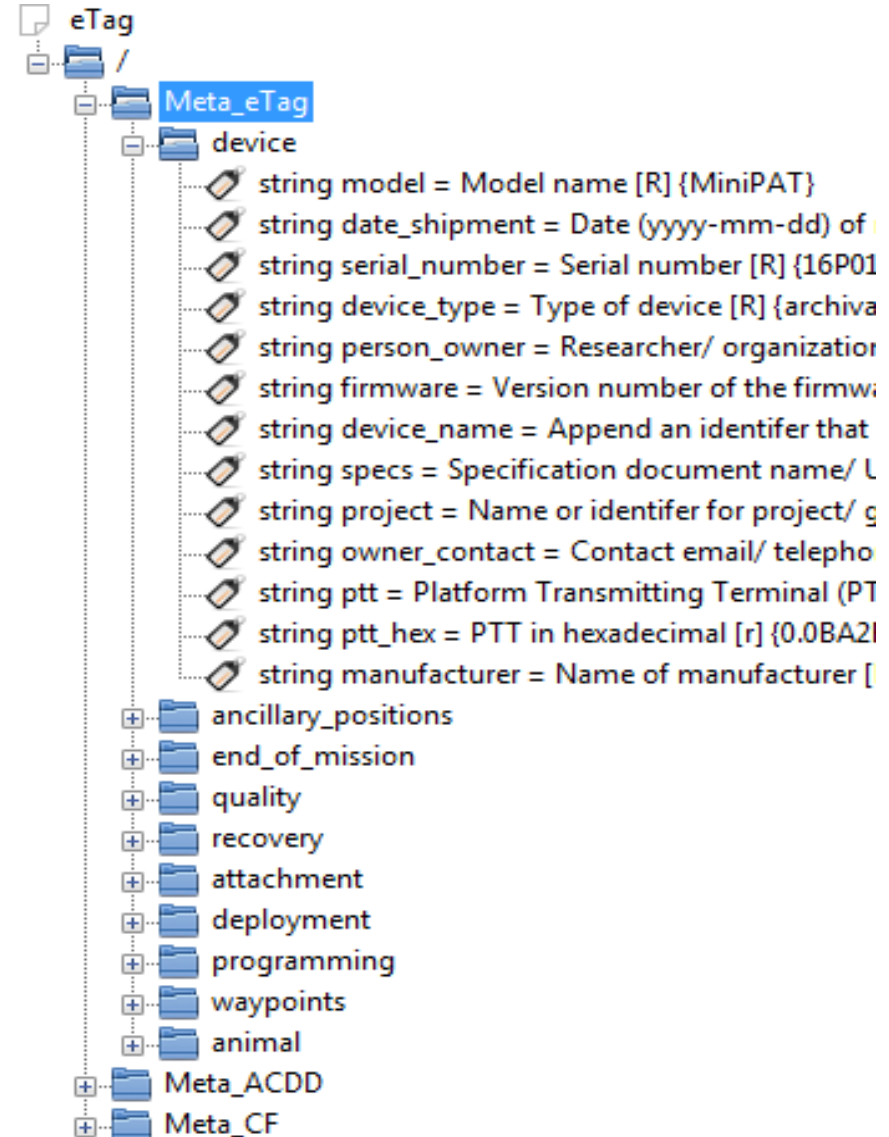
III. Support for "Summary" data

● Best practices/standards for representing "summary" /non-scalar data?

<u>PAT tag data use case</u>

• 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)
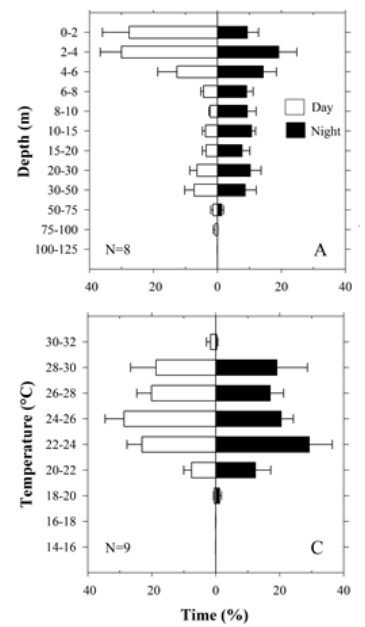
<u>Satellite-in situ data Matchup use Case</u>

• .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)

Left-margin annotations:
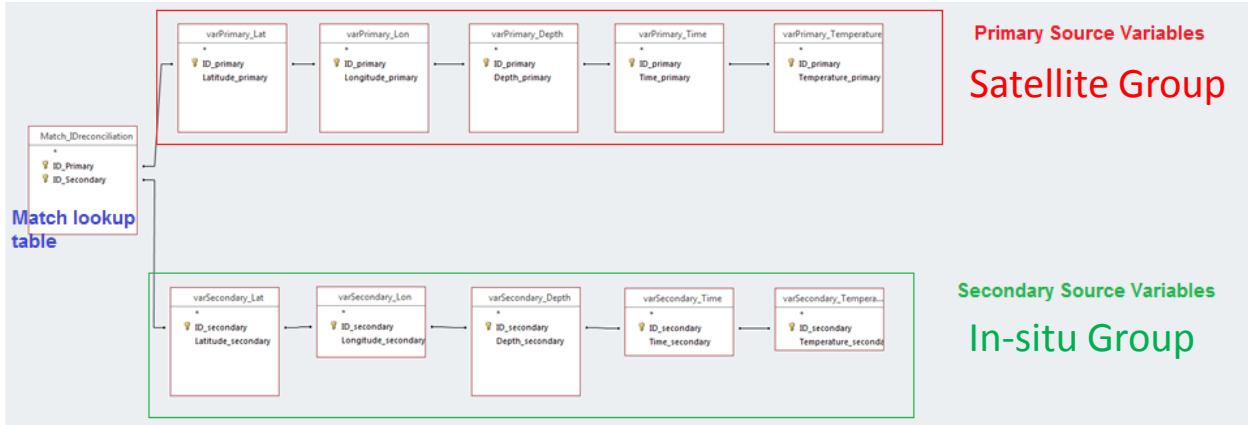
Global Attributes CF/ACCD +matchup query URL

Satellite Group Variables (lat, lon, time, Measurements)

In-situ Group Variables (lat, lon, Z, time, Measurements)

Matchup ID Array

Block (element)	ID	Inclusion Flag	Attribute (with example value)	Notes/Comments	Definitions (see https://www.nodc.noaa.gov/data/formats/netcdf/v2.0/)
DIM	1	TRUE	dimensions:		
DIM	2	TRUE	satelliterecordID = unlimited;		
DIM	3	TRUE	insiturecordID = unlimited;		
Glob - CF	5	TRUE	Conventions = "CF-1.6, ACDD-1.3";		A comma-separated list of the conventions that are followed by the dataset
Glob - CF	6	TRUE	title = "DOMS satellite-insitu matchup output file";		A short phrase or sentence describing the dataset. In many discovery systems, the title will be displayed in the results list from a search, and therefore should be human readable and reasonable to display in a list of such names. This attribute is also recom...
Glob - CF	7	TRUE	history = "Processing_Version = V1.0, Software_Name = DOMS, Software_Version = 1.03, Processing_Time = 2012071626320000	Include key DOMS processing/provenance info captured here	Provides an audit trail for modifications to the original data. This attribute is also in the NetCDF Users Guide: 'This is a character array with a line for each invocation of a program that has modified the dataset. Well-behaved generic netCDF applications should append a line containing: date, time of day, user name, program name and command arguments.' To include a more complete description you can append a reference to an ISO Lineage entity; see NOAA EDM ISO Lineage guidance.
Glob - CF	8	TRUE	institution = "JPL, FSU, NCAR";		The name of the institution principally responsible for originating this data. This attribute is recommended by the CF convention.
Glob - CF	9	TRUE	source = "doms.jpl.nasa.gov";		The method of production of the original data. If it was model-generated, source should name the model and its version. If it is observational, source should characterize it. This attribute is defined in the CF Conventions. Examples: 'temperature from CTD...
Glob - ACDD	10	TRUE	standard_name_vocabulary = "CF Standard Name Table v27", "BODC controlled vocabulary";		The name and version of the controlled vocabulary from which variable standard names are taken. (Values for any standard_name attribute must come from the CF Standard Names vocabulary for the data file or product to comply with CF.)
Glob - ACDD	11	TRUE	cdm_data_type = "Point/Profile, Swath/Grid";		The data type, as derived from Unidata's Common Data Model Scientific Data types and understood by THREDDS. The current choices are: Grid, Image, Point, Radial, Station, Swath, and Trajectory.
Glob - ACDD	12	TRUE	processing_level = "4";	Assigned L4 since this is a value added dataset	A textual description of the processing (or quality control) level of the data. This attribute should also identify the authority who defined the processing level itself. The processing levels defined by NOAA for satellite data...
Glob - ACDD	13	TRUE	date_modified = "20140402";		The date on which the data was last modified. Note that this applies just to the data, not the metadata. The ISO 8601:2004 extended date format is recommended, as described in the Attributes Content Guidance section.
Glob - ACDD	14	TRUE	date_created = "20130508";	Potentially use Equivalent attribute Processing_Time attribute value	The date on which this version of the data was created. (Modification of values implies a new version, hence this would be assigned the date of the most recent values modification.) Metadata changes are not considered when assigning the date_created.
Glob - ACDD	15	TRUE	time_coverage_start = "20130318";		Describes the temporal coverage of the data as a time range. The temporal coverage of the data can be described with any of the following pairs of values: start/end, start/duration, or end/duration. Use ISO 8601:2004 for date and time; preferably the exte...
Glob - ACDD	16	TRUE	time_coverage_end = "20130406";		Describes the temporal coverage of the data as a time range. The temporal coverage of the data can be described with any of the following pairs of values: start/end, start/duration, or end/duration. Use ISO 8601:2004 for date and time; preferably the exte...
Glob - ACDD	18	TRUE	geospatial_lat_max = 34.0f; // float		These attributes define the latitude and longitude coordinates of the bounding box of the data set; may be part of a 2- or 3-dimensional bounding region. Report the latitude coordinates in decimal degrees north (south is negative). These values should b...
Glob - ACDD	19	TRUE	geospatial_lat_min = 23.0f; // float		These attributes define the latitude and longitude coordinates of the bounding box of the data set; may be part of a 2- or 3-dimensional bounding region. Report the latitude coordinates in decimal degrees north (south is negative). These values should b...
Glob - ACDD	20	TRUE	geospatial_lon_max = -37.0f; // float		These attributes define the latitude and longitude coordinates of the bounding box of the data set; may be part of a 2- or 3-dimensional bounding region. Report the latitude coordinates in decimal degrees north (south is negative). These values should b...
Glob - ACDD	21	TRUE	geospatial_lon_min = -63.0f; // float		These attributes define the latitude and longitude coordinates of the bounding box of the data set; may be part of a 2- or 3-dimensional bounding region. Report the latitude coordinates in decimal degrees north (south is negative). These values should b...
Glob - ACDD	22	TRUE	geospatial_lat_resolution = "point";	point or Lat resolution interval if known	These attributes define the latitude and longitude coordinates of the bounding box of the data set; may be part of a 2- or 3-dimensional bounding region. Report the latitude coordinates in decimal degrees north (south is negative). These values should b...
Glob - ACDD	23	TRUE	geospatial_lon_resolution = "point";	point or Lon resolution interval if known	These attributes define the latitude and longitude coordinates of the bounding box of the data set; may be part of a 2- or 3-dimensional bounding region. Report the latitude coordinates in decimal degrees north (south is negative). These values should b...
Glob - ACDD	24	TRUE	geospatial_lat_units = "degrees_north";		This attribute defines the units applied to the geospatial_lat_min and geospatial_lat_max attributes. CF recommends the units should be "degrees_north". Regardless of what is used, it must conform to udunits (http://www.unidata.ucar.edu/software/u...
Glob - ACDD	25	TRUE	geospatial_lon_units = "degrees_east";		This attribute defines the units applied to the geospatial_lon_min and geospatial_lon_max attributes. CF recommends the units should be "degrees_east". Regardless of what is used, it must conform to udunits (http://www.unidata.ucar.edu/software/ud...
Glob - ACDD	26	TRUE	geospatial_vertical_min = 0.0f; // float		These attributes define the numerically smallest and largest highest vertical measurements of the data set's bounding box; may be part of a 2- or 3-dimensional bounding region.
Glob - ACDD	27	TRUE	geospatial_vertical_max = 5.0f; // float		These attributes define the numerically smallest and largest highest vertical measurements of the data set's bounding box; may be part of a 2- or 3-dimensional bounding region.
Glob - ACDD	28	TRUE	geospatial_vertical_units = "m";		This defines the units applied to the geospatial_vertical_min and geospatial_vertical_max attributes.
Glob - ACDD	29	TRUE	geospatial_vertical_resolution = "point";	point or depth resolution interval if known	Information about the targeted regular vertical spacing of points. Example: '25 meters'
Glob - ACDD	30	TRUE	geospatial_vertical_positive = "down";		One of 'up' or 'down'. If up, vertical values are interpreted as 'altitude', with negative values corresponding to below the reference datum (e.g., under water). If down, vertical values are interpreted as 'depth', positive values correspond to below the reference datum.
Glob - ACDD	31	TRUE	platform = "Endeavor";	Comma separated list of platforms used to derive matchup outputs?	If using as a global attribute: Name of the platform(s) that supported the sensor data used to create this data set or product. Platforms can be of any type, including satellite, ship, station, aircraft or other. Indicate controlled vocabulary used in platform_vo...
Glob - ACDD	32	TRUE	instrument = "Endeavor on-board sea-bird SBE 9/11 CTD";	Comma seperated list of instruments/sensors where known used to derive matchup outputs? Optional? Frank: We don't currently store	If using as a global attribute: Name of the contributing instrument(s) or sensor(s) used to create this data set or product. Indicate controlled vocabulary used in instrument_vocabulary.
Glob - ACDD	33	TRUE	project = "Distributed Oceanographic Matchup System (DOMS)";		The name of the project(s) principally responsible for originating this data. Multiple projects can be separated by commas, as described under Attribute Content Guidelines. Examples: 'PATMOS-X', 'Extended Continental Shelf Project'.
Glob - ACDD	34	TRUE	keywords_vocabulary = "NASA Global Change Master Directory (GCMD) Science Keywords";		Identifies the controlled list of keywords from which the values in the "keywords" attribute are taken. Good choices include: GCMD Science Keywords Version 8.1, CF Standard Names, ISO Topic Categories, NCEI Data Types, NCEI Observation Types.
Glob - ACDD	35	TRUE	keywords = "Salinity, Upper Ocean, SPURS, CTD, Endeavor, Atlantic Ocean";	example keywords provided here. Modify as necessary. Frank: What Could be PODAAC?	A comma separated list of keywords and phrases.
Glob - ACDD	36	TRUE	creator_name = "NASA PO.DAAC";		The name of the person (or other creator type specified by the creator_type attribute) principally responsible for creating this data.
Glob - ACDD	37	TRUE	creator_email = "podaac@podaac.jpl.nasa.gov";	add a contact email here. Could be podaac helpdesk/user services	The email address of the person (or other creator type specified by the creator_type attribute) principally responsible for creating this data.
Glob - ACDD	38	TRUE	creator_url = "https://podaac.jpl.nasa.gov";	Could be podaac portal URL or DOMS website?	The URL of the person (or other creator type specified by the creator_type attribute) principally responsible for creating this data.
Glob - ACDD	39	TRUE	publisher_name = "NASA PO.DAAC";		The name of the person (or other entity specified by the publisher_type attribute) responsible for publishing the data file or product to users, with its current metadata and format.
Glob - ACDD	40	TRUE	publisher_email = "podaac@podaac.jpl.nasa.gov";		The email address of the person (or other entity specified by the publisher_type attribute) responsible for publishing the data file or product to users, with its current metadata and format.
Glob - ACDD	41	TRUE	publisher_url = "https://podaac.jpl.nasa.gov";		The URL of the person (or other entity specified by the publisher_type attribute) responsible for publishing the data file or product to users, with its current metadata and format.
Glob - ACDD	42	TRUE	acknowledgment = "DOMS is a NASA/AIST-funded project. NRA NNH14ZDA001N.";		
Glob - DOMS	71	TRUE	Matchup_TimeWindow = 24.0f;//float	potentially prefix attribute name with "DOMS_" if a DOMS metadata	DOMS time window tolerance for matchup specified by user in units given below
Glob - DOMS	72	TRUE	Matchup_TimeWindow_Units = "hours";	potentially prefix attribute name with "DOMS_" if a DOMS metadata	Units for matchup time window
Glob - DOMS	73	TRUE	Matchup_SearchRadius = 1000.0f;//float	potentially prefix attribute name with "DOMS_" if a DOMS metadata	DOMS horizontal search window/tolerance for matchup specified by user in units given below
Glob - DOMS	74	TRUE	Matchup_SearchRadius_Units = "m";	potentially prefix attribute name with "DOMS_" if a DOMS metadata	DOMS horizontal search window/tolerance for matchup specified by user in units given below
Glob - DOMS	75	TRUE	URI_Subset = "http://webservice subsetting query request";	potentially prefix attribute name with "DOMS_" if a DOMS metadata	DOMS subsetting URI used to subset the data prior to matchup
Glob - DOMS	76	TRUE	URI_Matchup = "http://webservice matchup query request";	potentially prefix attribute name with "DOMS_" if a DOMS metadata	DOMS matchup URI used to collocate the data after subsetting
Glob - DOMS	77	TRUE	Matchup_ParameterPrimary = "Temperature";	potentially prefix attribute name with "DOMS_" if a DOMS metadata	DOMS primary matchup parameter (temperature, salinity etc) cited using the CF standard name for that parameter
Group-SatelliteData	100	TRUE	\\ Group : SatelliteData	**Group encapsulating all SatelliteData-related variables**	Group encapsulating all SatelliteData-related variables
Group-SatelliteData	101	TRUE	variables:		
Group-SatelliteData: Var-Lon	102	TRUE	float lon(satelliterecordID);		
Group-SatelliteData: Var-Lon	103	TRUE	long_name = "Longitude";		
Group-SatelliteData: Var-Lon	104	TRUE	standard_name = "longitude";		
Group-SatelliteData: Var-Lon	105	TRUE	axis = "X";		
Group-SatelliteData: Var-Lon	106	TRUE	units = "degrees_east";		
Group-SatelliteData: Var-Lon	107	TRUE	valid_min = -63.0f; // float		
Group-SatelliteData: Var-Lon	108	TRUE	valid_max = -37.0f; // float		
Group-SatelliteData: Var-Lon	109	TRUE	_FillValue = -32767.0f; // float	Apply the designated product fill-value as necessary	
Group-SatelliteData: Var-Lat	114	TRUE	float lat(satelliterecordID);		
Group-SatelliteData: Var-Lat	115	TRUE	long_name = "Latitude";		
Group-SatelliteData: Var-Lat	116	TRUE	standard_name = "latitude";		
Group-SatelliteData: Var-Lat	117	TRUE	axis = "Y";		
Group-SatelliteData: Var-Lat	118	TRUE	units = "degrees_north";		
Group-SatelliteData: Var-Lat	119	TRUE	valid_min = 23.0f; // float		
Group-SatelliteData: Var-Lat	120	TRUE	valid_max = 34.0f; // float		
Group-SatelliteData: Var-Lat	121	TRUE	_FillValue = -32767.0f; // float	Apply the designated product fill-value as necessary	
Group-SatelliteData: Var-Time	127	TRUE	int time(satelliterecordID);		
Group-SatelliteData: Var-Time	128	TRUE	long_name = "Time";		
Group-SatelliteData: Var-Time	129	TRUE	standard_name = "time";		
Group-SatelliteData: Var-Time	130	TRUE	axis = "T";		
Group-SatelliteData: Var-Time	131	TRUE	units = "seconds since 1970-01-01 00:00:00 0:00";	suggest ISO 8601 date-time value here to be a global reference time	
Group-SatelliteData: Var-Time	132	TRUE	calendar = "standard"		
Group-SatelliteData: Var-Platform	137	TRUE	string PlatformType(satelliterecordID);	Astring variable holding standard platform designation - possibly add additional variable attributes if necessary	
Group-SatelliteData: Var-Measurement	140	TRUE	float SatelliteMeasurements(satelliterecordID);	This varible will hold the satellite data and attribute values for the primary matchup parameter. (eg. Salinity, WindSpeed, Temperature)	
Group-SatelliteData: Var-Measureme	141	TRUE	long_name = "sea surface salinity";	For Wind speed, Density, Spiciness, Salinity-Uncertainty, substitute an	
Group-SatelliteData: Var-Measureme	142	TRUE	standard_name = "sea_surface_salinity";	substitute "sea_surface_density" or "wind_speed". No CF standard name available for Spiciness or SSS-uncertainty so just apply a sensible	
Group-SatelliteData: Var-Measureme	143	TRUE	units = "1e-3";	this is CF-canonical salinity unit rather than PSU; For Wind speed product substitute 'm s-1'. Density CF canonical units are given as 'kg m-3' (see http://cf-pcmdi.llnl.gov/documents/cf-standard-names cf-	
Group-SatelliteData: Var-Measureme	144	TRUE	_FillValue = -32767.0f;	Apply the designated product fill-value as necessary	
Group-SatelliteData: Var-Measureme	145	TRUE	valid_min = 30.0f	Apply the appropriate min value range for the variable field	
Group-SatelliteData: Var-Measureme	146	TRUE	valid_max = 40.0f	Apply the appropriate max value range for the variable field	
Group-SatelliteData: Var-Measureme	149	TRUE	coordinates = "lon lat time"	Eliminate unless measurment data values are "packed"	
Group-InsituData	100	TRUE	\\ Group : InsituData	**Group encapsulating all insitu data-related variables**	Group encapsulating all insitu data-related variables
Group-InsituData: Var-Lon	101	TRUE	variables:		
Group-InsituData: Var-Lon	102	TRUE	float lon(insiturecordID);		
Group-InsituData: Var-Lon	103	TRUE	long_name = "Longitude";		
Group-InsituData: Var-Lon	104	TRUE	standard_name = "longitude";		
Group-InsituData: Var-Lon	105	TRUE	axis = "X";		
Group-InsituData: Var-Lon	106	TRUE	units = "degrees_east";		
Group-InsituData: Var-Lon	107	TRUE	valid_min = -63.0f; // float		
Group-InsituData: Var-Lon	108	TRUE	valid_max = -37.0f; // float		
Group-InsituData: Var-Lon	109	TRUE	_FillValue = -32767.0f; // float	Apply the designated product fill-value as necessary	
Group-InsituData: Var-Lat	114	TRUE	float lat(insiturecordID);		
Group-InsituData: Var-Lat	115	TRUE	long_name = "Latitude";		
Group-InsituData: Var-Lat	116	TRUE	standard_name = "latitude";		
Group-InsituData: Var-Lat	117	TRUE	axis = "Y";		
Group-InsituData: Var-Lat	118	TRUE	units = "degrees_north";		
Group-InsituData: Var-Lat	119	TRUE	valid_min = 23.0f; // float		
Group-InsituData: Var-Lat	120	TRUE	valid_max = 34.0f; // float		
Group-InsituData: Var-Lat	121	TRUE	_FillValue = -32767.0f; // float	Apply the designated product fill-value as necessary	
Group-InsituData: Var-Lat	122	TRUE			
Group-InsituData: Var-Time	127	TRUE	int time(insiturecordID);		
Group-InsituData: Var-Time	128	TRUE	long_name = "Time";		
Group-InsituData: Var-Time	129	TRUE	standard_name = "time";		
Group-InsituData: Var-Time	130	TRUE	axis = "T";		
Group-InsituData: Var-Time	131	TRUE	units = "seconds since 1970-01-01 00:00:00 0:00";	suggest ISO 8601 date-time value here to be a global reference time such as start of mission with appropriate time units for required level of precision (eg.milliseconds, seconds) given also the defined variable data type (integer, double). Times are UTC by default unless designated in the timestamp terminus eg. 2012-08-25 00:00:00 -6:00 .	
Group-InsituData: Var-Time	132	TRUE	calendar = "standard"		
Group-InsituData: Var-Platform	137	TRUE	string PlatformType(insiturecordID);	Astring variable holding standard platform designation - possibly add additional variable attributes if necessary	
Group-InsituData: Var-Measurement	139				
Group-InsituData: Var-Measurement	140	TRUE	float InSituMeasurements(insiturecordID);	This varible will hold the insitu data and attribute values for the primary matchup parameter. (eg. Salinity,WindSpeed, Temperature)	
Group-InsituData: Var-Measurement	141	TRUE	long_name = "sea surface salinity";	For Wind speed, Density, Spiciness, Salinity-Uncertainty, substitute an	
Group-InsituData: Var-Measurement	142	TRUE	standard_name = "sea_surface_salinity";	substitute "sea_surface_density" or "wind_speed". No CF standard name available for Spiciness or SSS-uncertainty so just apply a sensible	
Group-InsituData: Var-Measurement	143	TRUE	units = "1e-3";	this is CF-canonical salinity unit rather than PSU; For Wind speed product substitute 'm s-1'. Density CF canonical units are given as 'kg m-3' (see http://cf-pcmdi.llnl.gov/documents/cf-standard-names cf-	
Group-InsituData: Var-Measurement	144	TRUE	_FillValue = -32767.0f;	Apply the designated product fill-value as necessary	
Group-InsituData: Var-Measurement	145	TRUE	valid_min = 30.0f	Apply the appropriate min value range for the variable field	
Group-InsituData: Var-Measurement	146	TRUE	valid_max = 40.0f	Apply the appropriate max value range for the variable field	
Group-InsituData: Var-Measurement	149	TRUE	coordinates = "lon lat time"	Eliminate unless measurment data values are "packed"	
Group-InsituData: Var-MetaString	154	TRUE	string MetaString(insiturecordID);	Astring variable holding ancillary metastring information - possibly add additional variable attributes if necessary	