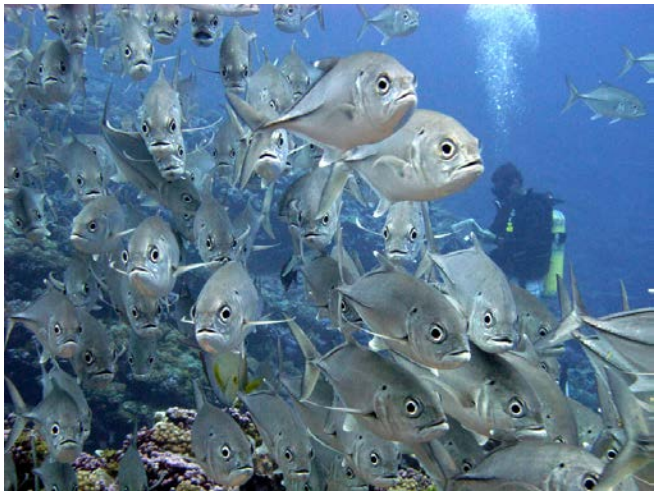
U.S. Integrated Ocean Observing System (U.S.IOOS)

IOOS Biological Data Services Enrollment Procedures

Hassan Moustahfid[1] and Philip Goldstein[2]

[1] NOAA/ U.S. Integrated Ocean Observing System, Silver Spring, MD
hassan.moustahfid@noaa.gov
[2] University of Colorado/Ocean Biogeographic Information System-USA, Boulder, CO



NOAA photo by Ben Ruttenburg.

Version 1.0
March 2014

DOCUMENT VALIDATION

U.S. IOOS Program Office Validation

Zdenka S. Willis 04/23/14

Zdenka S. Willis, Director, U.S. IOOS Program Office — Date

TABLE OF CONTENTS

REVISION HISTORY

Date	Author	Version	Notes
15 August 2013	Hassan Moustahfid	0.9	Created
4-6 September 2013	Philip Goldstein	0.9	Added information to "Enrollment Procedures" section added text, pp 9, 12, 13, 14; appended enrollment journal, pp 15-49
6 September 2013	Hassan Moustahfid	0.9	Edits and revision to the drafts
20 December 2013	Hassan Moustahfid	0.9	Final changes and edits
10 January 2014	Hassan Moustahfid	0.9	Revised in response to Charly Alexander, Rob Ragsdale and Derrick Snowden (US IOOS) comments
14 February 2014	Hassan Moustahfid	0.9	Revised in response to Jim Potemra (PACIOOS), Vembu Subramanian (SECOORA), Matt Howard (GCOOS-RA) and Bob Simons (NOAA/SWFSC/ER) comments/edits.
28 March 2014	Hassan Moustahfid	1.0	

ENDORSEMENT DISCLAIMER

Mention of a commercial company or product does not constitute an endorsement by NOAA. Use of information from this publication for publicity or advertising purposes concerning proprietary products or the tests of such products is not authorized

ACKNOWLEDGEMENTS

Special thanks go to Jim Potemra (PacIOOS), Vembu Subramanian (SECOORA), Matt Howard (GCOOS), Bob Simons (NOAA/SWFSC/ERD) and Charles Alexander, Rob Ragsdale and Derrick Snowden (U.S. IOOS) for comments and suggestions that considerably improved this document content.

I. INTRODUCTION

"This IOOS Biological data solutions and linking to users applications will allow people to do research that simply could not be done before due to access and format challenges. It will save weeks to months of time acquiring, reformatting, understanding the collection techniques and conducting the analyses"

Mitch Roffer (ROFFS ™ Inc.)

"The obtainable consistency of the biological data due to standards, as well as the richness and standard approach to detailed circumstances of observation, are what enable IOOS Biological Data for applications"

Philip Goldstein (OBIS-USA)

This document describes the data enrollment procedures proposed by IOOS Data Management and Communication (DMAC) to support sharing and integration of aquatic biological data (i.e. IOOS core variables: fish species, fish abundance, zooplankton species, zooplankton abundance, phytoplankton species). However these data services are applicable to any aquatic taxa that have presence, absence and abundance information.

The document has the following major sections:

IOOS Basic Principle
Describes the IOOS data principle for sharing ocean observing data.

IOOS Biological Data Flow
Describes the biological data flow from IOOS Data providers through IOOS Regional Association to US IOOS DMAC system.

Enrollment Procedures
Defines the types of enrollment procedures and the steps of enrolling a data provider.

II. BASIC PRINCIPLE

IOOS was established with the core principle that observation data should be made freely available to all. This has implications: 1) Data providers must be made aware that their data will be offered freely through IOOS. 2) Data providers will be responsible for restricting access to any data which may be too sensitive for general release (e.g. location data for endangered species). Providers are free to manage such control in any way that makes sense in their context.

III. IOOS BIOLOGICAL DATA FLOW

Data flow is represented in Figure 1. IOOS Regional Associations (RA) can enroll biological data providers (i.e. IOOS Biological Core variables: fish species, fish abundance, zooplankton species, zooplankton abundance and phytoplankton species) following the IOOS enrollment procedures outlined and register their web services to the IOOS Registry (http://www.ioos.noaa.gov/catalog/register.html). For broader dissemination, the IOOS registry will then connect the data to other data portals such as Data.gov (https://www.data.gov/), OBIS (http://www.usgs.gov/obis-usa/), NMFS InPort (https://inport.nmfs.noaa.gov/inport/) etc. and national Archiving Center (NODC, http://www.nodc.noaa.gov/) and make the data accessible to data users.

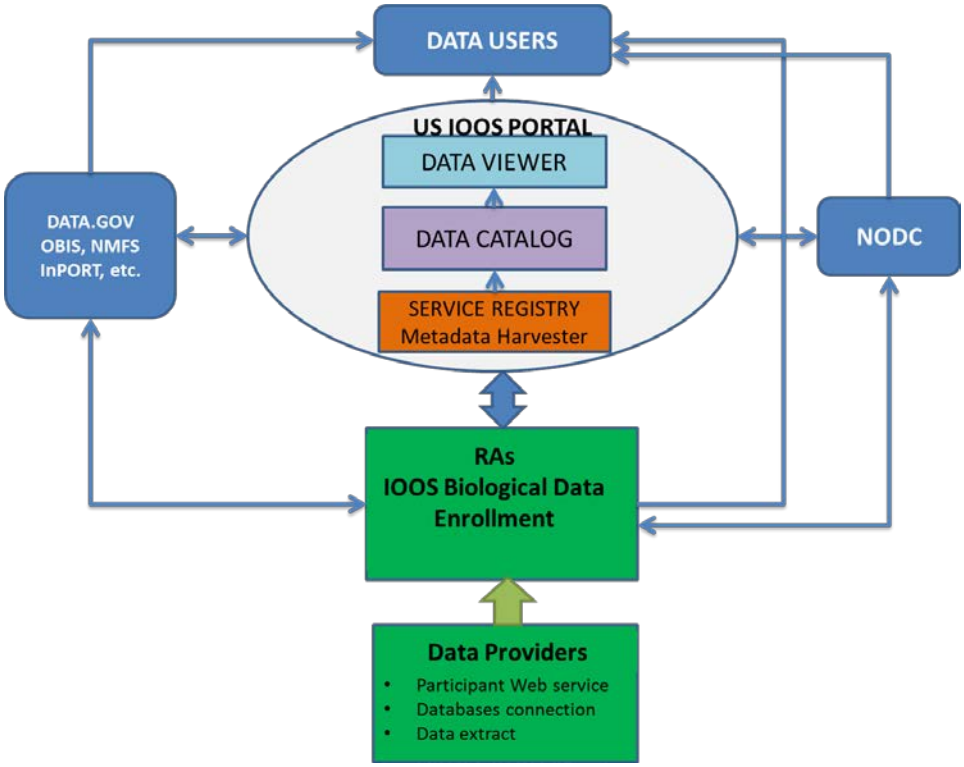


Figure 1. IOOS Biological data flow data providers through RAs to IOOS Registry/Catalog/Viewer (i.e. US IOOS Portal) and other National portals (e.g. OBIS-USA, , Data.gov, NMFS InPORT) and archiving center (NODC). Data providers can provide their biological data to RAs through accepted RAs webservice, database connection or data extract. RAs will then enroll the data using IOOS Biological Data Services (IOOS BDS) enrollment procedures (see below) and register the data to US IOOS service registry. The data users (e.g. scientists, managers) can access the data from US IOOS portal, RA portals or from other national portal and archiving center (NODC) in a standard form.

In order for an RA to enroll biological data in the IOOS DMAC system, the RA must first agree to the following:

- The RA will work with their region's data providers to enroll their data through IOOS DMAC system. This will include assisting/training the data providers in formatting their original data into IOOS Biological Data Services terms (IOOS BDS) to enhance standardization and integration with IOOS system, configuring the access service and helping data to serve data and how to help the data providers to improve discovery metadata.

- The RA will deploy standard-based data access services such as ERDDAP (http://coastwatch.pfeg.noaa.gov/erddap/index.html) and /or other accepted IOOS data access service (see details here: http://www.ioos.noaa.gov/data/contribute_data.html).

- The RA will ensure that the data gets registered into the IOOS service registry according to the IOOS DMAC guidance that can be found here: https://github.com/ioos/registry.

IV. ENROLLMENT PROCEDURES

Data providers can provide their data set in electronic form to the RA in their region. Then, the RA will publish the data. The RA will then work with the provider to determine the necessary steps to ensure the data are compliant with IOOS BDS. Data providers, in turn, will work with the RA to complete the necessary data transformation according to IOOS BDS guidelines available at:
http://www.ioos.noaa.gov/biological_observations/welcome.html and biological data services standard vocabulary:
http://mmisw.org/orr/#http://mmisw.org/ont/ioos/biological.

Documentation and reference examples for the aspects of the enrollment process will be continually updated with community engagement. Further, the RA will provide resources

(DMAC Coordinator and DMAC team) to assist the data provider with the enrollment process while continuing the essential involvement of data providers who know the content-specific details. Then, the RA will then publish the data via web services.

The optimal model for sharing data for a prospective provider depends on whether or not the group or individual is interested in maintaining their own data server and whether or not they expect to be making regular updates to the data set.

The IOOS model is one of a network of distributed data services. One benefit of such a model is data providers maintain and control distribution of their own data while at the same time these data are exposed through the wider network. However, it is recognized that many providers may not have the resources or interest in maintaining local data services. In these cases the RA can act as the initial data access point.

The following section of this document describe the steps (Figure 2) by which an RA would enroll biological data providers in their region.

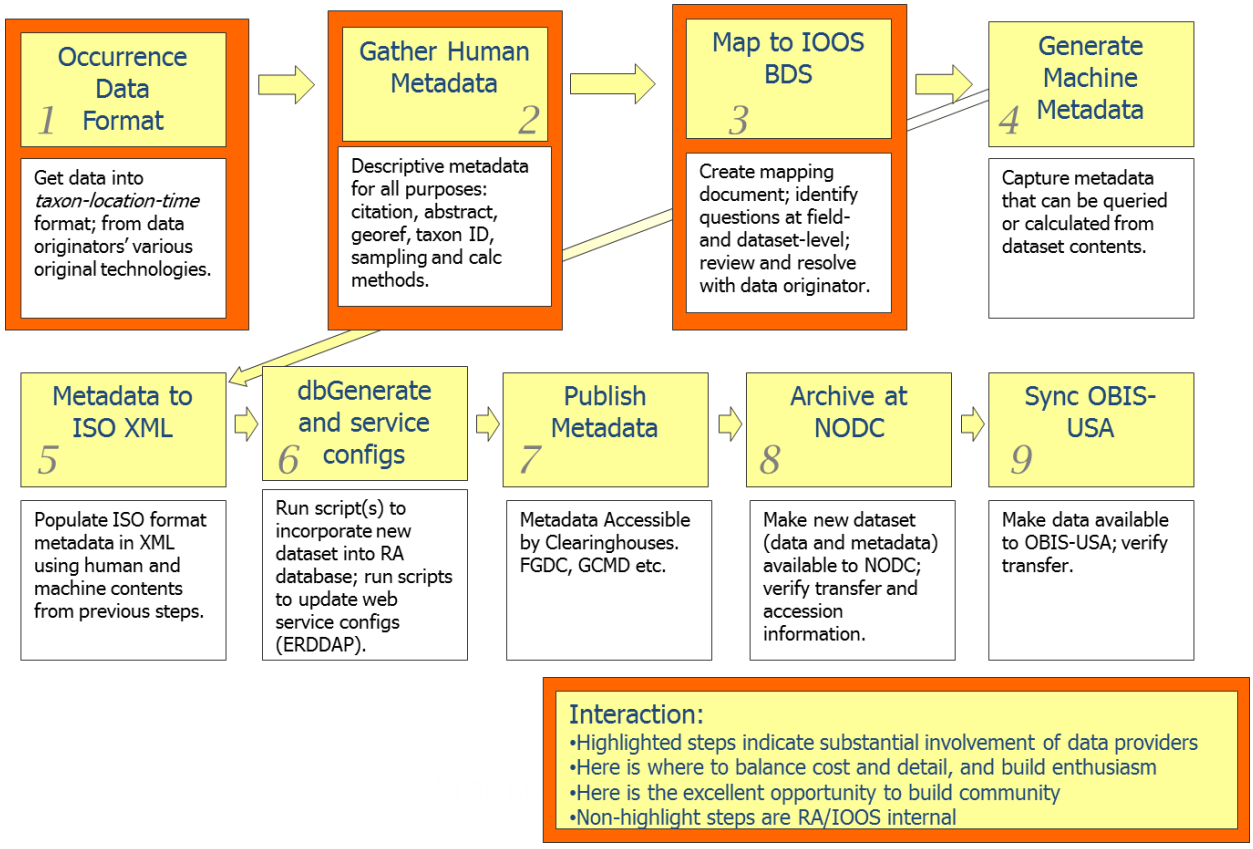


Figure 2. Biological data provider's enrollment procedure steps.

Data files are the most common way for data providers to send their data to the RA. Data in relational databases can be dumped to data files are the next format that data providers use to transfer the data.

The core component of IOOS BDS is an occurrence record. An occurrence is an observer's, or an observing system's record of a named taxon at a specific location and time along with detailed information about the circumstances of the observation. A collection of occurrence records forms a database-like table of data, with columns (each with a specific type of information) and rows (each with an occurrence record). Though occurrence information can exist in other data structures, a table with columns and rows is a helpful model.

Occurrence data can be made more consistent by using standard observation methods and by following standards, such as IOOS BDS, for formatting and presenting the data. Data that is more consistent is more useful.

Step #1 - Format Occurrence Data to enroll data in IOOS BDS

The goal is to obtain data in taxon-location-time format. The RA will work with the provider in two steps: (1a) verify that the contents of the dataset do in fact contain occurrence information (observed taxon-location-date-details), and (1b) perform data manipulation, such as relational joins, to gather the occurrence data in a single table with columns and rows with each row representing a taxon-specific occurrence. A row may represent a group of occurences, after binning by sex, size, or other characteristics of biology or the observation method, or they may be aggregated as totals without division into such bins. Each column from left to right across the row represents detail about the occurrence. For more on the use and meaning of each column, see Step #3, Format Data to IOOS BDS.

Step # 2. Gather Human Metadata

The RA will work with the data providers to gather descriptive metadata for the datasets, including: title, summary/abstract, the data provider's name and email address, the recommended way to cite the dataset, keywords, georeference method, coordinate uncertainty, taxonomic identification details, the sampling protocol / effort / conditions, derived data calculation methods (e.g. for fish abundance estimate), history, and related resources.
Machine metadata is metadata can be generated from the data. See step #5. This will be done by RA data staff.
Human metadata refers to essential descriptive details, vital to effective use of the data, which must be gathered, written, edited by humans. Human metadata cannot be generated automatically from the contents of a dataset. Human metadata is the principle method used to capture expertise from the data providers; the RA data staff will assist in every way feasible.

Step # 3. Format Data to IOOS BDS

This step involves matching columns in the provider's data table with IOOS BDS terms (http://www.ioos.noaa.gov/biological_observations/welcome.html or in MMI http://mmisw.org/ont/ioos/biological) and documenting exceptions, questions, and issues that need any kind of follow-up. This step requires a thorough understanding of how each term is used in the original data table as well as a thorough understanding of the IOOS BDS terms. In many cases, terms in the original data table map directly to IOOS BDS terms, and a rich, standardized dataset can be created. In other cases, columns of data from the original data table need to be modified or reformatted to match the IOOS BDS specification. Precise attention to data formats and term definitions is required. IOOS Biology terms are specific: they are not used as "catch-all" terms that have general or approximate meaning. An "enrollment journal" approach is recommended (See example in Appendix). An enrollment journal is a document with a table that contains columns for: the IOOS Biology term, the definition of the IOOS Biology term, the name of the corresponding column in the original data, and documentation/comments regarding format changes or conversion requirements, related issues, processing details, and other issues that need further investigation. The enrollment journal is intended to be a multi-party document where enrollers and RA staff can communicate such issues and questions among those whose expertise is required to resolve the issues and complete the enrollment. The enrollment journal encourages an analytical, quality-based approach to gaining consensus of multiple parties within the data provider's organization and the RA as to the best way to depict the original data in the IOOS BDS standard. The enrollment journal also provides a rich method to gather and refine information that will go into the human metadata that describes the dataset and how it was made available via IOOS BDS.

Step # 4. Generate Machine Metadata

Machine metadata is metadata that can be generated from data.
Examples: number of records, number of taxa and locations, date range, ranges and enumeration of other values, and geographic extent.
Other machine metadata comes from the IOOS BDS terminology definitions.

Step # 5. Metadata to ISO 19115 XML

IOOS recommends using ISO 19115 XML for metadata standards. In this step, the RA data staff can assist data providers with creating an ISO 19115 metadata document that follows the ISO 19115 guidance from NGDC and NCDDC. See https://geo-ide.noaa.gov/wiki/index.php?title=Category:ISO_19115 . Or, if an RA is using ERDDAP, then ERDDAP will automatically generate the ISO and FGDC XML documents (seehttp://coastwatch.pfeg.noaa.gov/erddap/download/setupDatasetsXml.html#globalAttributes).

Step #6. Service Configurations

If the dataset will be published via ERDDAP, an XML description of the dataset needs to be created so that ERDDAP knows how to access the data and how to present the data to users. See http://coastwatch.pfeg.noaa.gov/erddap/download/setupDatasetsXml.html.

Step # 7. Publish Metadata

The dataset's metadata, or a link to it, must be distributed to IOOS Service Registry (https://github.com/ioos/registry). IOOS Registry then will make it available to desired locations, such as repositories (e.g., NASA GCMD etc.), portals (e.g., Data.gov), search engines, or other systems that can utilize the metadata. Requirements for this step include assuring that the master copy of metadata is under control, and that there are procedures to assure that all secondary copies will get updated whenever the master copy is changed.

Step # 8. Archive with NOAA National Oceanographic Data Center (NODC)

Send the new dataset (data and metadata) to NODC. This should be done with the process that is already in place between NOAA IOOS Program Office, RAs and NODC to archive observing data. Two important aspects of this coordination with NODC are (1) make sure that NODC has the same version of the data (i.e., the same snapshot in the lifecycle of the data contents) that IOOS is serving, and (2) as much as possible, the same metadata should be used by IOOS BDS and NODC.

Step #9. Synchronizing data and metadata with OBIS-USA and IOOS System

This step requires OBIS-USA to establish a relationship with IOOS to query the IOOS Service Registry/Catalog and vise-versa. Since OBIS-USA is a national resource for discovery, attribution, access, and use of data from any and all US marine biogeographical data sources, each RAs data providers dataset's participation in OBIS-USA via IOOS can be an additional means of discovery, referral and attribution for IOOS data. In most cases, OBIS-USA will carry a more limited "biogeographic" form of data, with a reduced level of detail, than the IOOS nodes will be serving. If users find data at OBIS-USA and want more detailed data, they will be referred to the IOOS service.

The practical steps of integrating with OBIS-USA will largely be accomplished by IOOS's enrollment/publication process, due to the shared definitions within IOOS BDS and OBIS-USA's MBG (Marine BioGeography) standard, as well as common metadata practices.

REFERENCES

IOOS DMAC Guidance. http://www.ioos.noaa.gov/data/contribute_data.html

Guidance for Implementation of the Integrated Ocean Observing System (IOOS) Data Management and Communications (DMAC) Subsystem Version July 2013. http://www.ioos.noaa.gov/data/contribute_data.html

IOOS Biological Data Terminology version 1.1.
Or http://mmisw.org/ont/ioos/biological

ISO 19115 Metadata Standard for Geographic Data NOAA NGDC. https://geo-ide.noaa.gov/wiki/index.php?title=Category:ISO_19115

IOOS Service Registry. https://github.com/ioos/registry

APPENDIX: An example of enrollment journal to develop a data resource for IOOS BDS.

The example used here is the Comparative Assessment of Gulf Estuarine Systems (CAGES) Louisiana data from NOAA Southeast Fisheries Science Center Galveston Lab (More details about this data can be found at http://www.galvestonlab.sefsc.noaa.gov/stories/2013/ecology/).

Key to shaded table cells for IOOS Biology terminology v1.1; changes since v1.0:

New term	
Term name changed, same definition and usage as an earlier term	
Term depricated from IOOS Biology	

IOOS Biological Data Terminology (http://mmisw.org/orr/#http://mmisw.org/ont/ioos/biological)- aligned with CAGES Louisiana contents

Biological Data Term	**Term Definition**	**Alignment with terms in source data** CAGES Louisiana	**Comments**
modified	The most recent date the data originator updated or verified the record, expressed in the standard ISO 8601:2004(E). Can be the record creation date or date of subsequent update or verification. If the data originator does not record or provide this information, IOOS will populate this date with the date the record became available to IOOS.	'2013-08-14'	As given by data holder.

Biological Data Term	**Term Definition**	**Alignment with terms in source data** CAGES Louisiana	**Comments**
verbatimModified	The date and time value of the modified term, recorded in the exact format that the data originator provided it, for purposes of audit trail, since IOOS will convert this to ISO 8601:2004(E) for operational purposes.	NULL	Not required.
datasetID	An abbreviated name for the dataset that contains the occurrence record. This may also be the technical name for the data resource in a database or web service. The datasetID occurs on each row to preserve record identity in case records served by the resource become distributed to other resources or applications.	'CAGES_Louisiana_Lengths_IOOS_Standard'	
datasetName	This is a long-form name for the dataset, more explanatory than the datasetID. The datasetName will read as a recognizable title that reflects information about the dataset. Suggested information to include in the datasetName can refer to originator, content, purpose, method, geography, or other characteristics of the dataset. This datasetName term in the data record will match the dataset name in metadata.	'CAGES Louisiana Length Data with CPUE'	

Biological Data Term	**Term Definition**	**Alignment with terms in source data** CAGES Louisiana	**Comments**
higherInstitutionCode	Institution codes, in the form of abbreviated institution names, separated by semicolons, that define the hierarchy of institutions within which the data originator operates. Includes the originating institution as the lowest level (rightmost) code. Abbreviations for institutions in all levels of the hierarchy will be explained in dataset metadata.	'DOC;NOAA;NMFS;SEFSC;Fishery Ecology Branch'	
institutionCode	An abbreviation for the institution that is the originator of this data resource; the institution involved in research, data collection and/or data management that most directly produced this dataset. This institution also appears as the lowest level (rightmost) code in the higherInstitutionCode term.	'NMFS'	
ownerInstitutionCode	An abbreviation that identifies the institution, within the higherInstitutionCode hierarchy, that is considered the owner or controller of the data.	'NMFS'	
collectionCode	An identifier for a subset(s) of data within the dataset, partitioned by methods or parameters meaningful to the data originators. The system and purpose for defining and partitioning by collectionCode within a dataset will be explained in metadata.	'CAGES Louisiana'	

Biological Data Term	**Term Definition**	**Alignment with terms in source data** CAGES Louisiana	**Comments**
catalogNumber	A record identifier provided by the data originator. The identifier is controlled by the data originator and reflects the originator's practices of data administration. Special concerns about the identifier, such as whether it is unique, how it is controlled, if it is persistent, or if it contains embedded information, can be addressed in metadata if applicable.	NULL	Ok to leave NULL.
eventDate (term name revised in v1.1; formerly observationEventDateTime)	The date and time of observation expressed in local time using the standard ISO 8601:2004(E). Where date and time are imprecise, for example, time not recorded or day not recorded, ISO 8601 practice is to omit the components of the date and time string for those unspecified details. Where date and time imprecision are more complex, for example, "observed between 10am and noon" or "observed during the first half of July", supplement the use of ISO 8601 in this term with the use of verbatimEventDate, time, timeUncertainty, and eventDateRemarks.	"YYYY" || '-' || lpad("MM",2,'0') || '-' || lpad("DD",2,'0') || 'T' || substr(lpad("Time",4,'0'),1,2) || ':' || substr(lpad("Time",4,'0'),3,2) (pseudocode above is in postgreSQL syntax)	Use the separate year, month, day, and time data elements in the original data to construct eventDate in ISO 8601 format.

Biological Data Term	**Term Definition**	**Alignment with terms in source data** CAGES Louisiana	**Comments**
eventDateTimeZone (term name revised in v1.1; formerly observationDateTimeZone)	The time zone of the observation event, expressed as +/- hh:mm offset from UTC. While time zone information can also be included in the full ISO 8601 expression in eventDate, time zone is stored here for reliable use in time and timeUncertainty calculations. The term time must be expressed in UTC.	"TimeZone" from reference table "CdtCstReferenceTable"	Made a cross reference table for platform-independent handling of daylight saving time.
eventDateRemarks (term name revised in v1.1; formerly observationDateTimeAnnotation)	eventDateRemarks explains information about date and time that may not be evident in verbatim, ISO 8601, or time form of the event date and time. For example, date and time may include uncertainty, such as time of day left null because time was not recorded. However, data originators may be able to say that although time was not recorded, the observation is known to have been made between 8am and 5pm local time. eventDateRemarks can provide a concise explanation of this condition. eventDateRemarks can also provide explanation for how timeUncertainty was determined.	NULL	Not required for CAGES Louisiana.

Biological Data Term	**Term Definition**	**Alignment with terms in source data** CAGES Louisiana	**Comments**
verbatimEventDate (term name revised in v1.1; formerly verbatimObservationDateTime)	The expression of the date and time of the observation in language identical to the originator's description of the date and time of observation. If the originator used a non-standard format or descriptive language for date and time, verbatimEventDate contains this information.	NULL	Not required for CAGES Louisiana.

Biological Data Term	**Term Definition**	**Alignment with terms in source data** CAGES Louisiana	**Comments**
time (term name revised in v1.1; formerly nominalObservationDateTime)	This term identifies a specific instant chosen to represent the date and time of the observation, in Coordinated Universal Time (UTC). Both date and time are specified in this term in order to facilitate some searches and services that require resolution of events to this level of precision in time. Some observation events occur with uncertainty at the level of minutes, hours, or even days or more. A nominal event date and time is determined, even for observations with date and time uncertainty, for IOOS services to best represent the instant of the event. The companion term, nominalObservationDateTimeUncertainty, defines an interval of uncertainty that quantifies precision to correspond to the precision the original data recorded. Methods for determining nominal time and uncertainty will be documented in metadata, and they can be explained (if concise) in the "observationDateTimeAnnotation" term.	replace((("YYYY" || '-' || lpad("MM",2,'0') || '-' || lpad("DD",2,'0') || ' ' || substr(lpad("Time",4,'0'),1,2) || ':' || substr(lpad("Time",4,'0'),3,2))::timestamp + "TimeZone"::interval)::character varying,' ','T') || 'Z'	Use the separate year, month, day, and time data elements in the original data and extend to seconds and add Z to indicate UTC. This requires offset for converting CAGES Louisiana time zone to UTC.

Biological Data Term	**Term Definition**	**Alignment with terms in source data** CAGES Louisiana	**Comments**
timeUncertainty (term name revised in v1.1; formerly nominalObservationDateTimeUncertainty)	This is the uncertainty in seconds before and after the nominalObservationDateTime, as a result of both precision and accuracy factors in determining observation time. This use of uncertainty establishes mid-point in the range of uncertainty equal to twice the value of this term.	360	Calculated as the sum of two uncertainty factors: (1) Duration: 10 minutes divided by 2 = 5 minutes, and (2) measurement error = 1 minute. Total: 6 minutes. In seconds: 6 minutes times 60 seconds/minute, equals 360 seconds.

Biological Data Term	**Term Definition**	**Alignment with terms in source data** CAGES Louisiana	**Comments**
latitude	The position of the observation north or south of the equator in decimal degrees. Latitudes north of the equator are positive and range to +90 degrees. Positions south of the equator are negative, and range to -90 degrees. The data type for latitude and longitude terms is defined as "string" for xsd purposes. These data elements may be converted upon use to platform-specific data types for GIS, web services and other applications. The use of "xsd:string" in the term definition is to preserve the exact information provided by the data originator, because exact features, particularly the number of decimal places, must be retained to remain true to the original information.	"Latitude"	Are these available in five-decimal-place form in the original data? Some working data may be truncated to two decimal places? If truncated, explain the number of decimal places in metadata. Explain georeferencing method or somehow explain the background of the georeference information. Explain (estimated) uncertainty.

Biological Data Term	**Term Definition**	**Alignment with terms in source data** CAGES Louisiana	**Comments**
longitude	The position of the observation east or west of the prime meridian in decimal degrees. Positions west of the prime meridian are negative, and range to -180. Positions east of the prime meridian are positive, and range to +180. The data type for latitude and longitude terms is defined as "string" for xsd purposes. These data elements may be converted upon use to platform-specific data types for GIS, web services and other applications. The use of "xsd:string" in the term definition is to preserve the exact information provided by the data originator, because exact features such as the number of decimal places must be retained to remain true to the original information.	"Longitude"	(see comments for "latitude")

Biological Data Term	**Term Definition**	**Alignment with terms in source data** CAGES Louisiana	**Comments**
footprintWKT	A Well-Known Text (WKT) representation of geometry associated with the observation. For example, this can be a line segment or polyline representing a cruise path or transect. The location for an occurrence record may have both a point-radius representation (see latitude, longitude) and a footprint representation. This would be a case known to occur in marine data, where the record contains both a point along a transect, as well as a representaiton of the transect itself. The footprintWKT can also represent other geometry such as a polygon.	NULL	Not required for CAGES Louisiana. If start and end trawl positions were known, footprintWKT could be applicable.
coordinateUncertaintyInMeters	An expression in meters of the overall uncertainty of the georeference provided by the latitude and longitude data. The uncertainty can combine issues of both accuracy and precision, and depends on the method the data originator used to determine coordinates. Methods and considerations that go into determinations of uncertainty will be described in metadata.	1250	Estimated based upon a starting point and a distance calculation based on trawl duration.
verbatimCoordinates	Record original coordinate information here (both latitude and longitude) for audit trail purposes where necessary.	NULL	Not required for CAGES Louisiana.
verbatimCoordinateSystem	The spatial coordinate system for the VerbatimCoordinates.	NULL	Not required for CAGES Louisiana.

Biological Data Term	**Term Definition**	**Alignment with terms in source data** CAGES Louisiana	**Comments**
verbatimSRS	The ellipsoid, geodetic datum, or spatial reference system (SRS) upon which coordinates given in VerbatimCoordinates are based.	NULL	Not required for CAGES Louisiana.
geodeticDatum	The ellipsoid, geodetic datum, or spatial reference system used in latitude and longitude. IOOS preference is to provide all coordinates using EPSG:4326, also known as WGS 84. This element will identify the actual datum used, confirming WGS 84 or identifying what other datum applies.	'ESPG:4326 WGS 84'	
georeferencedBy	The name or other identifier of individual(s) or institution(s) that determined the georeference. Can be a list delimited by semicolon.	'LDWF'	'SEFSC'? same as 'recordedBy'? Explain in metadata that LDWF is 'Louisiana Department of Wildlife and Fisheries'
georeferenceProtocol	A description or reference to the methods used to determine the spatial footprint, coordinates, and uncertainties. Georeference methods, including determionation of coordinates and uncertainty, should be explained in dataset metadata.	'Coordinate uncertainty is based on starting point plus travel during trawl.'	
depth (new term in v1.1)	The depth of the observed biological occurrence in meters.	NULL	Not available in CAGES Louisiana.

Biological Data Term	**Term Definition**	**Alignment with terms in source data** CAGES Louisiana	**Comments**
minimumDepthInMeters (no longer used in v1.1; use of min and max replaced by single term depth)	The minimumDepthInMeters and maximumDepthInMeters express the depth range in which the observation was made. If the data originator provides a single depth measurement, the minimumDepthInMeters and maximumDepthInMeters show that measurement and will be equal. If no depth information is provided, both the min and max terms will be NULL.		These min and max columns for depth will not be used in IOOS Biology at this time.
maximumDepthInMeters (no longer used in v1.1; use of min and max replaced by single term depth)	The minimumDepthInMeters and maximumDepthInMeters express the depth range in which the observation was made. If the data originator provides a single depth measurement, the minimumDepthInMeters and maximumDepthInMeters show that measurement and will be equal. If no depth information is provided, both the min and max terms will be NULL.		These min and max columns for depth will not be used in IOOS Biology at this time.
basisOfRecord	Identifies the source of information or observation that generated the biological occurrence record.	'Human Observation'	
recordedBy	The name or other identifier of individual(s) or institution(s) responsible for making the observation and recording it in electronic form. Can be a list delimited by semicolon.	'LDWF'	Explain in metadata that LDWF is 'Louisiana Department of Wildlife and Fisheries'

Biological Data Term	**Term Definition**	**Alignment with terms in source data** CAGES Louisiana	**Comments**
vernacularName	A common or vernacular name for the taxon observed. A vernacularName is not required. Use of this term is at the discretion of the data originator. If this term is used, then authorities, references and procedures for making identifications and translating vernacular name to scientific name should be documented in metadata. The recommended practice is that vernacular names be used consistently within a dataset.	initcap(split_part("Common Name",' ',1)) || ' ' || lower(split_part("Common Name",' ',2)) || ' ' || lower(split_part("Common Name",' ',3))	Convert from all-uppercase to mixed case.
scientificName	The taxonomic identification of the observation as either 1) Genus and species (and subspecies if provided) in Latin binomial nomenclature form, or 2) the lowest-level taxonomic name to which the observation is identified, expressed in Latin form. scientificName is a required term. Authorities, references and procedures for making identifications should be documented in metadata.	initcap(split_part("Scientific Name",' ',1)) || ' ' || lower(split_part("Scientific Name",' ',2)) || ' ' || lower(split_part("Scientific Name",' ',3))	Convert from all-uppercase to mixed case.
taxonRank	The taxonRank term is a companion to the scientificName term. taxonRank identifies the taxonomic level of the lowest-level name in the scientificName term, if the ScientificName refers to a level above Genus.	formula below in postgreSQL pseudocode.	Determine rank from Harmon's spreadsheets .

Biological Data Term	Term Definition	Alignment with terms in source data CAGES Louisiana	Comments
	CASE -- no taxon rank if no sci name WHEN "Scientific Name" IS NULL THEN NULL -- no taxon rank for the following list WHEN "Species Code" IN ('999','1120','2000','9999') THEN NULL -- taxon rank = Genus for the following list WHEN "Species Code" IN ('993','1115','1116','1210','1213','1409','1506','2057','2127','2142','2184','2185','2186','2188','2190','2191','2197','2199','2200','2201','2202','2211','2221','2222','2227','2230','2233','2240','2252','2258','2313','2322','2366','2385','2392','2429','2434','2446','2454','2459') THEN 'Genus' -- taxon rank = Family for the following list WHEN "Species Code" IN ('994','1521','2194','2300','2321','2387','2425','2427','2428','2441','2457','2458','2460') THEN 'Family' -- taxon rank = Order for the following list WHEN "Species Code" IN ('992','2380','2383','2384') THEN 'Order' -- taxon rank = Subclass for the following list WHEN "Species Code" IN ('2386') THEN 'Subclass' -- taxon rank = Class for the following list WHEN "Species Code" IN ('989','990','991','995','996','997') THEN 'Class' -- taxon rank = Subphylum for the following list WHEN "Species Code" IN ('998') THEN 'Subphylum' ELSE 'Species' END		
aphiaID	A unique taxon identifier obtained by validation of the taxon name with the World Register of Marine Species (WoRMS), www.marinespecies.org.		To be determined.
tsn	A unique taxon identifier obtained by validation of the taxon name with the Integrated Taxonomic Information System (ITIS), www.itis.gov.		To be determined.

Biological Data Term	**Term Definition**	**Alignment with terms in source data** CAGES Louisiana	**Comments**
genus	The full scientific name of the genus in which the taxon is classified.	CASE WHEN "Species Code" IN ('989','990','991','992','994','995','996','997','998','999','1120','1521','2000','2194','2300','2321','2380','2383','2384','2386','2387','2425','2427','2428','2441','2457','2458','2460','9999') THEN NULL ELSE initcap(split_part("Scientific Name",' ',1)) END	Parse from "Scientific Name". Handle exceptions by referring to Harmon's taxonomic spreadsheets. For anything not identified to genus, don't populate genus.
subgenus	The full scientific name of the subgenus in which the taxon is classified.	NULL	Not used in CAGES Louisiana.

Biological Data Term	**Term Definition**	**Alignment with terms in source data** CAGES Louisiana	**Comments**
species	The full scientific name of the species in which the taxon is classified. If a subspecies identification is provided, add the subspecies name after the species name separated by a space.	CASE WHEN "Species Code" IN ('989','990','991','992','993','994','995','996','997','998','999','1115','1116','1120','1210','1213','1409','1506','1521','2000','2057','2127','2142','2184','2185','2186','2188','2190','2191','2194','2197','2199','2200','2201','2202','2211','2221','2222','2227','2230','2233','2240','2252','2258','2300','2313','2321','2322','2366','2380','2383','2384','2385','2386','2387','2392','2425','2427','2428','2429','2434','2441','2446','2454','2457','2458','2459','2460','9999') THEN NULL ELSE lower(split_part("Scientific Name",' ',2)) END	Parse from "Scientific Name". Handle exceptions by referring to Harmon's taxonomic spreadsheets. For anything not identified to species, don't populate species.
infraspecificEpithet	The full scientific name of the subspecies, variant or other finer level identification.	NULL	Not available in CAGES Louisiana?
scientificNameAuthorship	The authorship information for the scientificName.	NULL	Not used in CAGES Louisiana.

Biological Data Term	**Term Definition**	**Alignment with terms in source data** CAGES Louisiana	**Comments**
identifiedBy	The name or other identifier of individual(s) or institution(s) that assigned the taxonomic name. Can be a list delimited by semicolon.	'LDWF'	Explain in metadata that LDWF is 'Louisiana Department of Wildlife and Fisheries'
identificationDate	The date the identification on this record was made, expressed in ISO 8601 date format.	"YYYY" || '-' || lpad("MM",2,'0') || '-' || lpad("DD",2,'0')	Same as "eventDate" (but only to day; disregard time of day).
identificationQualifier	A brief phrase or a standard term ("cf.", "aff.") to express the determiner's uncertainty about the Identification.	NULL	Not used in CAGES Louisiana.
identificationRemarks	Comments or notes about the identification, including any applicable reference to process, authorities, keys, uncertainties, or special conventions used in making the identification and assigning the name and/or formatting the name.	NULL	Not used in CAGES Louisiana.

Biological Data Term	**Term Definition**	**Alignment with terms in source data** CAGES Louisiana	**Comments**
individualCount	The number of individuals represented in the observation record. Valid values include positive integers, zero, and null. Positive integers represent presence. If the observation record and metadata also contain other required information such as details about the sampling activity, individualCount can contribute to the abundance calculations. Null value for individualCount represents a record of presence, with quantity unspecified, and null values do not support abundance calculations. Zero value represents absence. Zero values for absence are only valid if methodology for establishing absence is recorded in metadata.	1	All records represent individual measured animals.
sex	The sex of biological individuals represented in the observation record. Vocabulary will be consistent within a dataset and will be explained in metadata. Methods of determination (where applicable) will be explained in metadata.	NULL	Not available in CAGES Louisiana.
weightInKg (new term in v1.1)	The weight in kilograms of the biological individual(s) in this record. Contrast to totalInSample, that can report the total in the original sample in cases of subsampling. In cases of subsampling, weightInKg will be the weight of the subsample.	NULL	Only poplulate for CAGES Louisiana if weight of individual fish (not available).

Biological Data Term	**Term Definition**	**Alignment with terms in source data** CAGES Louisiana	**Comments**
lifeStage	An expression or description of age or lifestage of biological individual(s) in the observation record. Vocabulary will be consistent within a dataset and will be explained in metadata. Methods of determination (where applicable) will be explained in metadata.	NULL	Not available in CAGES Louisiana.
dynamicProperties (new term in v1.1)	A list of observed measurements and facts about a biological occurrence record that are not captured by other terms. Recorded in text format according to the following example: "weight=8.4kg". Multiple measurements can be recorded, separated by semi-colons. Facts about the sampling activity, in contrast to the biological entity, will be recorded in the terms for sampling protocol, effort, and conditions, using the same convention. Development of data using "dynamicProperties" may be preparatory for future use of the Darwin Core MeasurementOrFact extension.	NULL	Nothing applicable at this time for CAGES Louisiana.
observedIndividualLengthInCm	If a single individual is observed and measured, record length in centimeters here.	to_char("Length"::numeric/10, 'FM99.9')::numeric	Source data "Length" is in mm; convert to cm; retain significant digits.
observedMeanLengthInCm	If measuring more than one individual for aggregate length values, record mean length in centimeters here.	NULL	Not used in CAGES Louisiana.
observedMaxLengthInCm	If measuring more than one individual for aggregate length values, record maximum length in centimeters here.	NULL	Not used in CAGES Louisiana.

Biological Data Term	**Term Definition**	**Alignment with terms in source data** CAGES Louisiana	**Comments**
observedMinLengthInCm	If measuring more than one individual for aggregate length values, record minimum length in centimeters here.	NULL	Not used in CAGES Louisiana.
lengthType	The type or method of length measurement used in this observation record, such as TL for total length, FL for fork length, SL for standard length.	(see pseudocode below)	'Carapace Width' for crabs (determined from list) or else default to 'Total Length' for fish.
	CASE WHEN "Species Code" IN ('988','2003','2072','2118','2159','2160','2161','2162','2163','2164','2165','2183','2184','2185','2196','2230','2240','2241','2242','2243','2244','2245','2250','2257','2259','2268','2273','2275','2276','2279','2281','2283','2284','2285','2299','2302','2304','2312','2314','2318','2328','2337','2377','2378','2389','2402','2418','2422','2424','2425','2426','2428','2432','2442','2443','2445','2446','2449','2451','2458','2462') THEN 'Carapace Width' ELSE 'Total Length' END		
OccurrenceRemarks	Information, that is not addressed elsewhere in IOOS Biological Terminology, about the observation event and biological or other context details about what was observed. Can be general remarks directly from a corresponding column in the original data. Can also include miscellaneous but applicable information carried over from conversion to Darwin Core / IOOS Biological Terminology and placed here for future reference.	NULL	

Biological Data Term	**Term Definition**	**Alignment with terms in source data** CAGES Louisiana	**Comments**
surveyEventID	The SurveyEventID identifies a tow, trawl, cast, cruise, transect or other event or unit that survey operators, data gatherers and data managers want to treat as a single consistent event or unit. A survey event can contain one or many sample events or SampleIDs.	NULL	There is no cruise ID.
sampleID	A unique identifier for a unit of material, or other medium of observation, gathered at a discrete point in time, that serves as a sample for a survey of biological occurrence (presence, quantification, absence, or derived value). The definition of a sample and assignment of IDs is controlled by the data originator. If applicable, details about how sample are are defined and IDs assigned will be provided in metadata.	"Sample Code"	
subsampleID	If a sample is further divided by the data originator for purposes of organization, handling, or analysis, and if the data originator wants to maintain the identity of the subsample, capture the ID of the subsample here. Explain in metadata if applicable.	"Sample Code" || '-' || "Species Code"	Concatenate "Sample Code" and "Species Code"; Harmon verifies that this is an effective identifier.

Biological Data Term	**Term Definition**	**Alignment with terms in source data** CAGES Louisiana	**Comments**
samplingProtocol	Information about the intended sampling method as defined by the data originator. Contents in this term (either verbatim or by vocabulary) will be used consistently in a dataset and will be explained in metadata. If there is more than one aspect of sampling protocol worth capturing, information can be concatenated into this term, delimited by semicolon. Contents in this term (either verbatim or by vocabulary) will be used consistently in a dataset. More extensive discussion of sampling protocol, effort and conditions may be provided in metadata as needed.	'Each sample was subsampled by species and individual lengths recorded.'	

Biological Data Term	**Term Definition**	**Alignment with terms in source data** CAGES Louisiana	**Comments**
samplingEffort	Information about the extent, duration or other variable aspects of the sampling activity, as conducted by the data originator, that occurred at the time of the sampling event. In contrast to the samplingProtocol, which describes the intended method of sampling prior to the event, samplingEffort describes actual phenomena as they occur at the time of sampling. More than one type of effort information can be concatenated into this term, delimited by semicolon. Other terms are available in this terminology for some specific details of sampling effort, such as sample dimensions, that are expected to be frequently used in IOOS services. Use the samplingEffort term for additional effort information that is not addressed by other terms. Contents in this term (either verbatim or by vocabulary) will be used consistently in a dataset. More extensive discussion of sampling protocol, effort and conditions may be provided in metadata as needed.	'Duration = ' || "Duration" || ' minutes'	

Biological Data Term	**Term Definition**	**Alignment with terms in source data** CAGES Louisiana	**Comments**
samplingConditions	Information about the state of the sampling location, determined by external factors such as weather, water conditions, lighting and visibility, etc., at the time of the sampling event. More than one type of information about conditions can be concatenated into this term, delimited by semicolon. Other terms are available in this terminology for some specific details of sampling conditions, such as habitat, bottomType, visibilty and temperature, that are expected to be frequently used in IOOS services. Use the samplingConditions term for additional information about conditions that is not addressed by other terms. Like samplingEffort, this term describes phenomena other than the method intended by the data collector in the samplingProtocol. The difference between samplingEffort and samplingConditions is that samplingConditions describes the state of the environment, and samplingEffort describes the actions of the sampler. More than one type of information about conditions can be concatenated into this term, delimited by semicolon. Contents in this term (either verbatim or by vocabulary) will be used consistently in a dataset and will be explained in metadata. More extensive discussion of sampling protocol, effort and conditions may be provided in metadata as needed.	NULL	This term is where hydrological data can reside, but the decision at this time is to defer putting extensive hydrological data in IOOS Biolgy for the CAGES dataset.

Biological Data Term	**Term Definition**	**Alignment with terms in source data** CAGES Louisiana	**Comments**
totalInSample (new term in v1.1)	In cases where subsampling has been done and the data record represents a subsample, the totalInSample term contains information about the total quantity in the original sample, usually expressed as total count or wieght, for example "Count=85" or "Weight=4.4kg". Specific details about how subsampling was done in relation to the total sample must be explained in samplingProtocol, and/or in metadata.	'Total Number = ' || "Total Number"	
sampleShape	The shape of the sample space in which the observation was made, such as R for rectangular, C for cylindrical.	'R'	
sampleLengthInMeters	The length in meters of the sample space in which the observation was made.	NULL	
sampleWidthInMeters	The width in meters of the sample space in which the observation was made.	NULL	
sampleHeightInMeters	The height in meters of the sample space in which the observation was made.	NULL	
sampleRadiusInMeters	The radius in meters of a cylindrical sample space.	NULL	
sampleAreaInSquareMeters	The area in square meters, if projected vertically to a plane representing the ocean surface, of the sample space in which the observation was made.	NULL	

Biological Data Term	**Term Definition**	**Alignment with terms in source data** CAGES Louisiana	**Comments**
sampleVolumeInCubicMeters	The volume in cubic meters of the sample space in which the observation was made.	NULL	
visibilityInMeters	Visibility through the water column in meters.	NULL	
visibilityType	The direction of estimated visibility, horizontal or vertical.	NULL	
waterTemperatureInCelsius	The temperature of the water at the time of the observation, expressed in Celsius.	NULL	
habitat	A description or identifier of general information about the habitat in which the observation was made. Vocabulary will be consistent within a dataset and will be explained in metadata. Methods of determination (where applicable) will be explained in metadata.	NULL	
bottomType	A description or identifier of general information about the sea floor at the coordinates or locality where the observation was made. Vocabulary will be consistent within a dataset and will be explained in metadata. Methods of determination (where applicable) will be explained in metadata.	NULL	

Biological Data Term	**Term Definition**	**Alignment with terms in source data** CAGES Louisiana	**Comments**
quantificationType	This is a one-character string with enumerated values indicating what type of quantification this record represents. "P" for presence (and no quantification), "Q" for quantified or "A" for absence.	'P'	Each CAGES Louisiana record will actually include both presence ('P') and quantification ('Q'); choice couild be either of these, but it seems primary purpose of the record is presence for calculating quantification. Harmon agrees, record is primarily for presence.

Biological Data Term	**Term Definition**	**Alignment with terms in source data** CAGES Louisiana	**Comments**
quantificationVocabulary	The names for derived values, referred to as "quantification", can be drawn from a controlled vocabulary or vocabularies of quantification names. Or they can be reported using the verbatim name given by the data originator. Either the name of the controlled vocabulary or the word "verbatim" should be used in this term to record the level of control over names. "Verbatim" means the names are not using a controlled vocabulary other than the names assigned just as they are by the data originator. Either way, within a dataset, identifiers for quantification names must be used consistently, and explained in metadata.	'Verbatim'	

Biological Data Term	**Term Definition**	**Alignment with terms in source data** CAGES Louisiana	**Comments**
quantificationName	A name for a derived value or processed data about the occurrence, for example, "Biomass". The word used in this terminology for such information is "quantification", rather than "abundance", because the word "abundance" may already be associated with specific, rather than general, practices of quantification in biology. In this example, biomass would not be obtained at survey time, rather biomass would be calculated from input data that was obtained at survey time, combined with a biomass calculation methodology. This term is used along with companion terms for quantification value, units, uncertainty, method, and determination details. It is suggested that names for derived values be more specific than "Biomass", because many institutions will develop biomass information using diverse methods. Suggested practice is to include some more specific reference(s) in the quantification name, such as institution, survey, date or other specifics. For example, "Biomass-PIFSC-2011", and explain the name and method of calculation in metadata.	'CPUE'	

Biological Data Term	**Term Definition**	**Alignment with terms in source data** CAGES Louisiana	**Comments**
quantificationValue	The value of the derived information product, such as the numerical value for biomass. This term does not include units. Units are contained in another term.	"CPUE"	
quantificationUnit	The units in which the value expresses the quantification, for example, "kg/m^2 surface" or "kilograms per square meter of ocean surface".	hectare^-1	UDUnits standard for "per hectare"
quantificationUncertainty	An estimate of the uncertainty, positive and negative (+/-) offset of the stated value, that expresses the uncertainty of the value as calculated and reported. This can be reported in the units of the quantification, or as a percentage. Note that the expression in this term can combine both accuracy and precision considerations to a more general estimate, uncertainty, hence the slightly different name for this term from compared to the Darwin Core (dwc) reference term.	NULL	
quantificationDetermined Date	The date the quantification was determined. Can be different from the observation date.	NULL	
quantificationDetermined By	The name or ID of person(s) or institution(s) that determined the quantification.	'SEFSC'	

Biological Data Term	**Term Definition**	**Alignment with terms in source data** CAGES Louisiana	**Comments**
quantificationMethod	A reference to the method used to determine the quantification. This will be different from the method used to make the observation of biological occurrence; it will be the companion calucation method that followed the observation method. Quantification information is expected to be useful only if method is sufficiently documented, so this term is vitally important for use of quantification terms for derived values. This term can contain a code or other abbreviated reference to a method. In all cases quantificationMethod should be explained in metadata and supporting publications or related resources identified.	'refer to metadata'	

Biological Data Term	**Term Definition**	**Alignment with terms in source data** CAGES Louisiana	**Comments**
waterBody	The name of the water body in which the observation was made. For marine data, this is the ocean name. If there is a more specific named region, such as a bay or reef, or sector of the ocean (for example, northwest Pacific), add the regional name after the ocean name, separated by a semicolon. Note that the locality term will contain the most specific name, such as a pier or other named feature, so that lowest level of detail need not be included in the waterbody term. The waterbody term is for the ocean and any named intermediate geography such as a region.	"Water Body"	
islandGroup	The name of an island group associated with the location where the observation was made.	NULL	
island	The name of an island associated with the location where the observation was made.	NULL	

Biological Data Term	**Term Definition**	**Alignment with terms in source data** CAGES Louisiana	**Comments**
locality	The locality is the most specific named place to identify the location of the observation. Many marine data rely strictly on geographic coordinates, so the locality term may be null. If a locality is used, it may be a specific location or feature name, such as "Pier 39" or "Atafu Atoll", or it may include a descriptive phrase, such as "a tidal pool 1500 meters north of Pier 39". Note that in marine observations, the latitude and longitude may be more precise than any named place would suggest, because the coordinates often come from navigation of GPS systems. Geographic names can be secondary locality identifiers and they can be less precise than coordinates.	"Site Name"	
country	The name of the country in which the observation location occurs. Such geographic named locations may not be available or applicable for marine data, and acordingly the use of this term is optional. In some cases such jurisdiction identification may be crucial for data attribution, management and/or use, so this term is retained in the terminology for cases where it may be essential.	'USA'	

Biological Data Term	**Term Definition**	**Alignment with terms in source data** CAGES Louisiana	**Comments**
stateProvince	The name of the state or province in which the observation location occurs. Such geographic named locations may not be available or applicable for marine data, and acordingly the use of this term is optional. In some cases such jurisdiction identification may be crucial for data attribution, management and/or use, so this term is retained in the terminology for cases where it may be essential.	'Louisiana'	
county	The name of the county or other comparable geographic unit in which the observation location occurs. Such geographic named locations may not be available or applicable for marine data, and acordingly the use of this term is optional. In some cases such jurisdiction identification may be crucial for data attribution, management and/or use, so this term is retained in the terminology for cases where it may be essential.	NULL	

Biological Data Term	**Term Definition**	**Alignment with terms in source data** CAGES Louisiana	**Comments**
municipality	The name of the municipality in which the observation location occurs. Such geographic named locations may not be available or applicable for marine data, and acordingly the use of this term is optional. In some cases such jurisdiction identification may be crucial for data attribution, management and/or use, so this term is retained in the terminology for cases where it may be essential.	NULL	
kingdom	The full scientific name of the kingdom in which the taxon is classified.	NULL	
phylum	The full scientific name of the phylum in which the taxon is classified.	NULL	
class	The full scientific name of the class in which the taxon is classified.	NULL	
order	The full scientific name of the order in which the taxon is classified.	NULL	
family	The full scientific name of the family in which the taxon is classified.	NULL	