Guide to the Marine Salmon Diet Database (MSDD)

Caroline Graham, Evgeny Pakhomov, and Brian Hunt

April 9, 2020

TABLE OF CONTENTS

INTRODUCTION

The Marine Salmon Diet Database (MSDD) was designed to be a collaborative, open-access database system for salmon researchers to store and share salmon diet data, specifically from the marine environment. This tool can be used by researchers across the North Pacific to collaboratively work on critical questions related to salmon marine survival and ocean conditions. The database is built to house all types of diet information ranging from direct measurements (e.g., stomach content analyses) to indirect measurements (e.g., fatty acids, stable isotopes). In addition to diet data, the database is designed to include related information on salmon biological parameters, prey biological parameters, and environmental data. The MSDD was built using MySQL, an open-source relational database management system. While this database was specifically designed to house salmon diet data, the core structure was designed in a manner that could allow it to be applied to other types of predator and prey data.

The initial data input was compiled through a systematic literature review of published and gray literature, followed by data extraction from these sources. Data were extracted in the same format and at the same data resolution as reported in the source in almost all cases. Due to the different analytical approaches used by various studies, the data have varying resolutions, with different metrics and units. It is up to the database user to collate the data as they see fit for their particular interest. As a general rule, if data are available for a certain attribute, then it is entered in the database, if not, it is assigned an “NA” value.

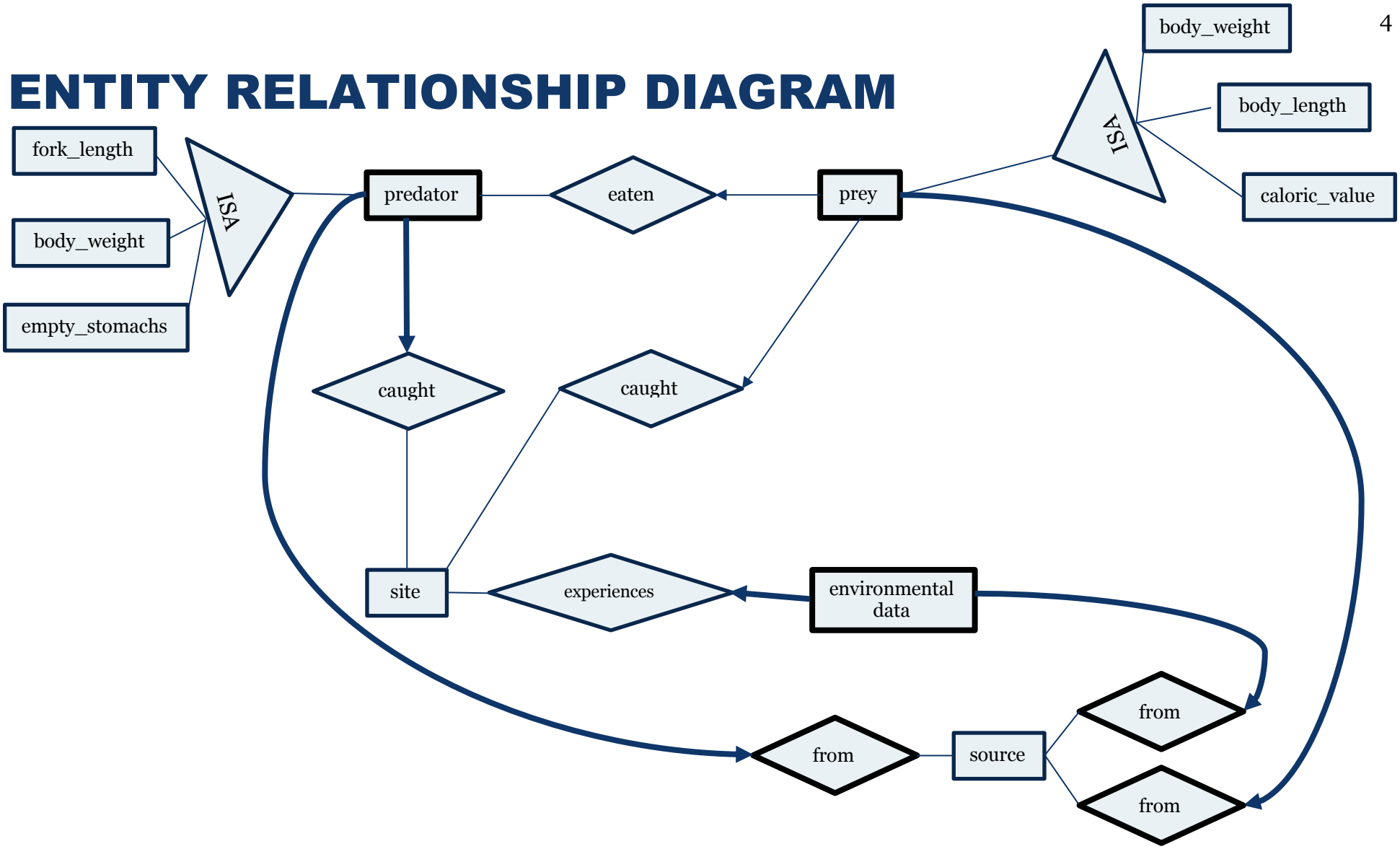

Figure 1. The basic entity relationship diagram that underlies the structure of the database. This figure does not include all of the attributes (columns) associated with the various database entities (tables). Not all predator (salmon) and prey biological parameter entities are displayed in the diagram

RELATIONAL MODEL

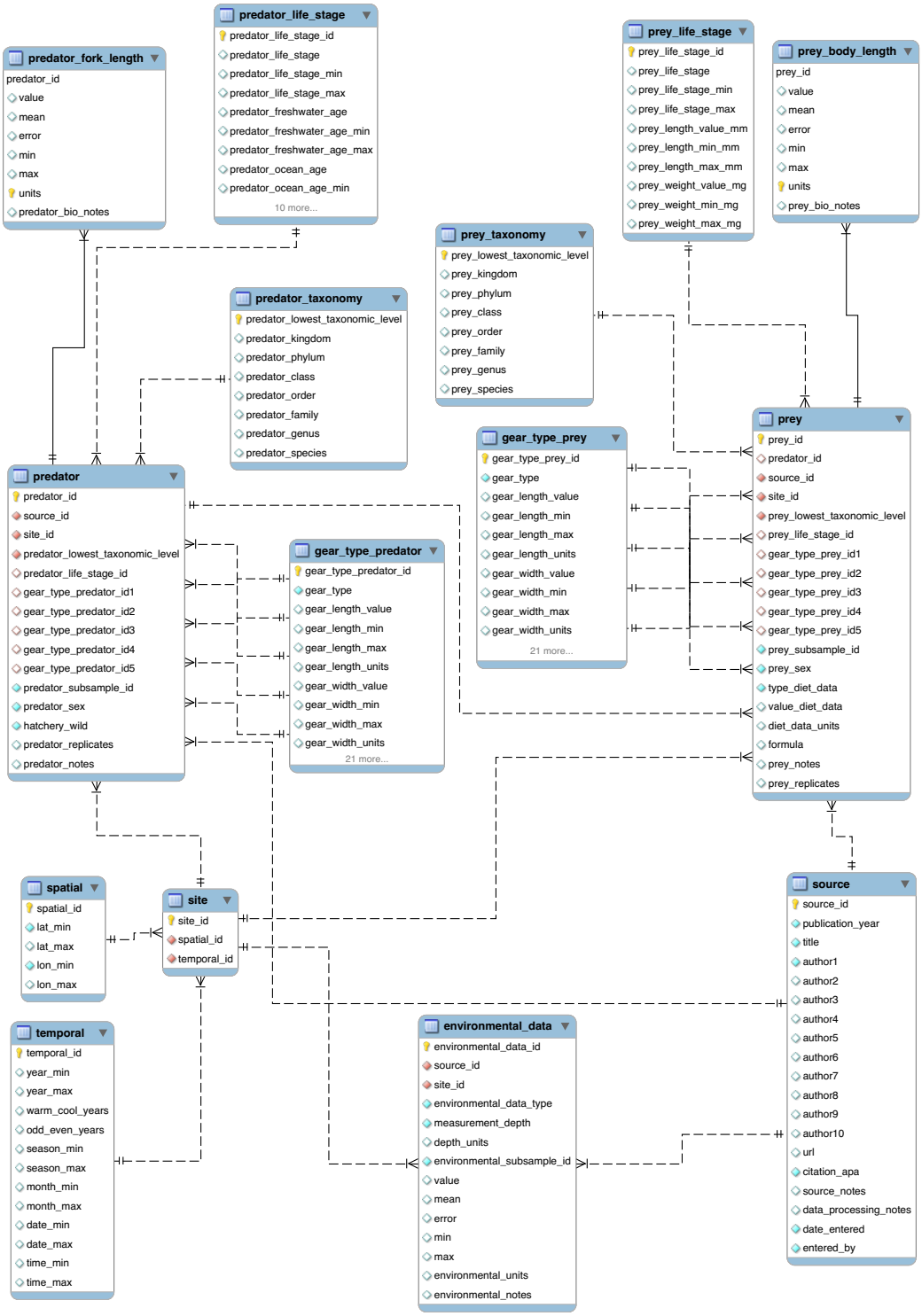

Figure 2. The relational model for the database. The yellow symbol represents the primary key, red represents foreign keys, blue represents not NULL attributes. Not all predator and prey biological parameter relations (tables) are displayed in the diagram. Just one example is given for one predator and one prey biological parameter.

DATABASE RELATIONS AND ATTRIBUTE DESCRIPTIONS

1) 'source' relation

This relation describes the reference sources for the data in the database. All diet, predator biological, prey biological, and environmental data is associated with a source so that the data can be traced back to its origin. This relation also includes the date that the source was entered into the database and the person who entered the data.

Attribute	Key	Data Type	Explanation
source_id	PK*	INT(10)	A unique number that is generated and assigned to each source
publication_year		YEAR(4)	Publication year in the format: YYYY; not NULL
title		VARCHAR(500)	Title of the source; not NULL
author1		VARCHAR(45)	First author of the source in the following format: M. C. Graham or M. Graham (first and middle names are abbreviated by initials and a period while last names are completely spelled out) ; not NULL
author2		VARCHAR(45)	Second author of the source in the following format: M. C. Graham or M. Graham (first and middle names are abbreviated by initials and a period while last names are completely spelled out)
author3		VARCHAR(45)	Third author of the source in the following format: M. C. Graham or M. Graham (first and middle names are abbreviated by initials and a period while last names are completely spelled out)
author4		VARCHAR(45)	Fourth author of the source in the following format: M. C. Graham or M. Graham (first and middle names are abbreviated by initials and a period while last names are completely spelled out)
author5		VARCHAR(45)	Fifth author of the source in the following format: M. C. Graham or M. Graham (first and middle names are abbreviated by initials and a period while last names are completely spelled out)
author6		VARCHAR(45)	Sixth author of the source in the following format: M. C. Graham or M. Graham (first and middle names are abbreviated by initials and a period while last names are completely spelled out)

author7		VARCHAR(45)	Seventh author of the source in the following format: M. C. Graham or M. Graham (first and middle names are abbreviated by initials and a period while last names are completely spelled out)
author8		VARCHAR(45)	Eighth author of the source in the following format: M. C. Graham or M. Graham (first and middle names are abbreviated by initials and a period while last names are completely spelled out)
author9		VARCHAR(45)	Ninth author of the source in the following format: M. C. Graham or M. Graham (first and middle names are abbreviated by initials and a period while last names are completely spelled out)
author10		VARCHAR(45)	Tenth author of the source in the following format: M. C. Graham or M. Graham (first and middle names are abbreviated by initials and a period while last names are completely spelled out)
url		VARCHAR(200)	URL associated with source
citation		VARCHAR(700)	The full citation for the source
source_notes		VARCHAR(3000)	Any additional notes about the source; for example, if the data might be overlapping with another source, this should be indicated in this attribute
data_processing_notes		VARCHAR(3000)	Notes about how the data were processed – in the lab or field, were quantities measured using scales or visually estimated, etc.
date_entered		DATE	The date the source was added to the database in the format: YYYY-MM-DD; not NULL
entered_by		VARCHAR(45)	The full name of the person who entered the data (e.g., Caroline Graham) ; not NULL

*PK indicates a primary key, while FK indicates a foreign key.

2) 'predator_life_stage' relation

This relation describes the life stage of the salmon, hereafter referred to as the predator, using a variety of metrics. Data are entered for whichever metric(s) is used in the source.
The first metric is the most general life stage metric and predators are either defined as juvenile or adult. The second metric is ocean age, which is the number of years a predator has spent living in the ocean. The third metric is maturity, where predators are defined as one of the following: juvenile, immature, maturing, or kelt. In some cases, predators are not identified as any of these life stage categories but are instead identified by a length or weight. Lengths and weights fall both into the 'predator_life_stage' relation and also their own relations ('predator_body_length', 'predator_fork_length', 'predator_weight'), as part of the associated predator biological parameters. To avoid repetitive data, length or weight information is entered into the 'predator_life_stage' relation only if there is no other way to determine life stage, or if length or weight categories are the only way to uniquely identify samples of diet data from a source. Otherwise length and weight

information are entered into their respective predator biological parameters relations. Information on length and weight is not duplicated in the predator biological parameter relations unless it is a value that needs to be further specified as a mean in the predator biological parameters relation. Another instance in which data are reported in both the predator_life_stage and the predator biological parameters relations (for length/weight) is if a mean is given in addition to a range of length of weight values, the mean is reported in the predator biological parameters relation while the ranges are reported in the predator_life_stage relation.

Attribute	Key	Data Type	Explanation
predator_life_stage_id	PK	INT(10)	A unique number that is generated and assigned to each predator life stage
predator_life_stage		VARCHAR(9)	Either 'juvenile', 'adult' or NA
predator_life_stage_min		VARCHAR(9)	If there are a mixture of juveniles and adults, then 'juvenile' is entered here
predator_life_stage_max		VARCHAR(9)	If there are a mixture of juveniles and adults, then 'adult' is entered here
predator_freshwater_age		INT(1)	An integer to indicate the number of years spent living in freshwater
predator_freshwater_age_min		INT(1)	If there are a mixture of freshwater ages, then this attribute represents the minimum age; an integer to indicate the number of years spent living in freshwater
predator_freshwater_age_max		INT(1)	If there are a mixture of freshwater ages, then this attribute represents the maximum age; an integer to indicate the number of years spent living in freshwater
predator_ocean_age		INT(1)	An integer to indicate the number of years spent living in the ocean
predator_ocean_age_min		INT(1)	If there are a mixture of ocean ages, then this attribute represents the minimum age; an integer to indicate the number of years spent living in the ocean
predator_ocean_age_max		INT(1)	If there are a mixture of ocean ages, then this attribute represents the maximum age; an integer to indicate the number of years spent living in the ocean
predator_maturity		VARCHAR(9)	Either 'juvenile', 'immature', 'maturing' 'mature', 'kelt' (for steelhead) or NA
predator_maturity_min		VARCHAR(9)	If there are a mixture of maturity levels, then the minimum maturity level is found here

predator_maturity_max		VARCHAR(9)	If there are a mixture of maturity levels, then the maximum maturity level is found here
predator_length_value_cm		DECIMAL(6,2)	The length of a predator in centimeters (could be either fork length or total length); only reported if there is no other way to determine life stage, or if length or weight categories are the only way to uniquely identify samples of diet data from a source
predator_length_min_cm		DECIMAL(6,2)	The minimum length of a predator in centimeters (could be either fork length or total length) if there are a range of sizes; only reported if there is no other way to determine life stage, or if length or weight categories are the only way to uniquely identify samples of diet data from a source
predator_length_max_cm		DECIMAL(6,2)	The maximum length of a predator in centimeters (could be either fork length or total length) if there are a range of sizes; only reported if there is no other way to determine life stage, or if length or weight categories are the only way to uniquely identify samples of diet data from a source
predator_weight_value_g		DECIMAL(6,2)	The weight of a predator in grams; only reported if there is no other way to determine life stage, or if length or weight categories are the only way to uniquely identify samples of diet data from a source
predator_weight_min_g		DECIMAL(6,2)	The minimum weight of a predator in grams if there are a range of sizes; only reported if there is no other way to determine life stage, or if length or weight categories are the only way to uniquely identify samples of diet data from a source
predator_weight_max_g		DECIMAL(6,2)	The maximum weight of a predator in grams if there are a range of sizes; only reported if there is no other way to determine life stage, or if length or weight categories are the only way to uniquely identify samples of diet data from a source

3) 'predator_taxonomy' relation

This relation describes the taxonomy of the predators. Since sources will sometimes report to different taxonomic levels (not in the case of Pacific salmon, which are all reported to the species level), a taxonomy table allows the database user to specify at what resolution they want to examine the data. The taxonomy relation is filled in based on the lowest taxonomic level reported in the source. Taxonomies are verified and updated using the World Register of Marine Species (WoRMS - http://www.marinespecies.org).

Attribute	Key	Data Type	Explanation
predator_lowest_taxonomic_level	PK	VARCHAR(45)	The lowest taxonomic level reported in the source; if a source reports to the species level then this attribute should include both the genus and species names (e.g., Oncorhynchus nerka); only scientific names are reported, not common names; first letter is capitalized
predator _kingdom		VARCHAR(45)	The kingdom based on the lowest taxonomic level; first letter is capitalized
predator _phylum		VARCHAR(45)	The phylum based on the lowest taxonomic level; first letter is capitalized
predator _class		VARCHAR(45)	The class based on the lowest taxonomic level; first letter is capitalized
predator _order		VARCHAR(45)	The order based on the lowest taxonomic level; first letter is capitalized
predator _family		VARCHAR(45)	The family based on the lowest taxonomic level; first letter is capitalized
predator _genus		VARCHAR(45)	The genus based on the lowest taxonomic level; first letter is capitalized
predator_species		VARCHAR(45)	The species based on the lowest taxonomic level; all lowercase

4) 'predator' relation

This relation describes the predator sample information related to both the diet data in the prey table (through combination of predator_id and prey_id) and the predator biological parameters. (This is the parent entity in the ISA hierarchy for predator biological parameters.) Each predator sample is uniquely identified, first through the source, and then through other attributes – site_id,

predator_lowest_taxonomic_level, predator_life_stage_overall_id, predator_subsample_id, predator_sex, and hatchery_wild. Predator samples can be just one predator, or a number of predators for which diet data have been collated and reported.

Attribute	Key	Data Type	Explanation
predator_id	PK	INT(10)	A unique number that is generated and assigned to each predator sample
source_id	FK	INT(10)	This id number corresponds to the 'source relation; not NULL
site_id	FK	INT(10)	This id number corresponds to the 'site' relation; not NULL
predator_lowest_taxonomic_level	FK	VARCHAR(45)	This corresponds to the 'predator_taxonomy' relation; not NULL
predator_life_stage _id	FK	INT(10)	This id number corresponds to the 'predator_life_stage' relation
gear_type_predator_id1	FK	INT(10)	This id number corresponds to the 'gear_type_predator' relation
gear_type_predator_id2	FK	INT(10)	This id number corresponds to the 'gear_type_predator' relation; this attribute is required if there are at least 2 types of gear used to sample predators
gear_type_predator_id3	FK	INT(10)	This id number corresponds to the 'gear_type_predator' relation; this attribute is required if there are at least 3 types of gear used to sample predators
gear_type_predator_id4	FK	INT(10)	This id number corresponds to the 'gear_type_predator' relation; this attribute is required if there are at least 4 types of gear used to sample predators
gear_type_predator_id5	FK	INT(10)	This id number corresponds to the 'gear_type_predator' relation; this attribute is required if there are at least 5 types of gear used to sample predators
predator_subsample_id		INT(10)	A unique number that is generated and assigned to predator samples if a source reports the diets of individual predators with no unique

			identifiers; values are assigned for each source starting from 1 and increasing by a value of 1 each time (e.g., 1,2,3…); not NULL; default value is 0
predator_sex		VARCHAR(11)	Either 'male' or 'female'; not NULL; default value is 'unspecified'
hatchery_wild		VARCHAR(11)	Either 'hatchery' or 'wild'; not NULL; default value is 'unspecified'
predator_replicates		INT(10)	The total number of predator replicates per sample
predator_notes		VARCHAR(300)	Any additional comments on the predator

5) 'prey_taxonomy' relation

This relation describes the taxonomy of the prey. Since sources report to different taxonomic levels a taxonomy table allows the database user to specify at what resolution they want to examine the data. The taxonomy relation is filled in based on the lowest taxonomic level reported in the source. Taxonomies are verified and updated using the World Register of Marine Species (WoRMS - http://www.marinespecies.org). If a taxonomic record is not able to be found, then the prey item is assigned to the lowest taxonomic level that could be identified or to 'Miscellaneous'. For some prey descriptions that were challenging to define taxonomically, the prey_lowest_taxonomic_level was determined based on a description of a group of prey, not on taxonomy. The following prey types had specially designated names that were not related to taxonomy: 'Gelatinous collective', 'Zooplankton collective', 'Nekton collective', 'Arthropoda (terrestrial)', and 'Invertebrate collective'.

Attribute	Key	Data Type	Explanation
prey_lowest_taxonomic_level	PK	VARCHAR(45)	The lowest taxonomic level reported in the source; if a source reports to the species level then this attribute includes both the genus and species names (e.g., Calanus pacificus); first letter is capitalized
prey_kingdom		VARCHAR(45)	The kingdom based on the lowest taxonomic level; first letter is capitalized
prey_phylum		VARCHAR(45)	The phylum based on the lowest taxonomic level; first letter is capitalized
prey_class		VARCHAR(45)	The class based on the lowest taxonomic level; first letter is capitalized
prey_order		VARCHAR(45)	The order based on the lowest taxonomic level; first letter is capitalized

prey_family		VARCHAR(45)	The family based on the lowest taxonomic level; first letter is capitalized
prey_genus		VARCHAR(45)	The genus based on the lowest taxonomic level; first letter is capitalized
prey_species		VARCHAR(45)	The species based on the lowest taxonomic level; all lowercase

6) 'prey_life_stage' relation

This relation provides data on the prey life stage. In some cases the different prey items were just differentiated based on whether they were small, medium or large. In this case, these descriptors were entered into the 'life_stage' attribute. To avoid repetitive data, length or weight information is entered into the 'prey_life_stage' relation only if there is no other way to determine life stage, or if length or weight categories are the only way to uniquely identify samples of diet data from a source. Otherwise length and weight information can be found in their respective prey biological parameters relations. Information on length and weight is not duplicated in the prey biological parameter relations unless it is a value that needs to be further specified as a mean in the prey biological parameters relation. Another instance in which data are reported in both the prey_life_stage and the prey biological parameters relations (for length/weight) is if a mean is given in addition to a range of length of weight values. The mean is reported in the prey biological parameters relation while the ranges are reported in the prey_life_stage relation.

Attribute	Key	Data Type	Explanation
prey_life_stage_id	PK	INT(10)	A unique number that is generated and assigned to each predator life stage
prey_life_stage		VARCHAR(45)	The prey life stage
prey_life_stage_min		VARCHAR(45)	If there are a mixture of life stages, then this attribute represents the minimum life stage
prey_life_stage_max		VARCHAR(45)	If there are a mixture of life stages, then this attribute represents the maximum life stage
prey_length_value_mm		DECIMAL(6,2)	The length of prey in millimeters; only reported if there is no other way to determine life stage, or if length or weight categories are the only way to uniquely identify samples of prey data from a source
prey_length_min_mm		DECIMAL(6,2)	The minimum length of prey in millimeters if there are a range of sizes; only reported if there is no other way to determine life stage, or if length or weight categories are the only way to uniquely identify samples of prey data from a source
prey_length_max_mm		DECIMAL(6,2)	The maximum length of prey in millimeters if there are a range of sizes;

			only reported if there is no other way to determine life stage, or if length or weight categories are the only way to uniquely identify samples of prey data from a source
prey_weight_value_mg		DECIMAL(6,2)	The weight of prey in milligrams; only reported if there is no other way to determine life stage, or if length or weight categories are the only way to uniquely identify samples of prey data from a source
prey_weight_min_mg		DECIMAL(6,2)	The minimum weight of prey in milligrams if there are a range of sizes; only reported if there is no other way to determine life stage, or if length or weight categories are the only way to uniquely identify samples of prey data from a source
prey_weight_max_mg		DECIMAL(6,2)	The maximum weight of prey in milligrams if there are a range of sizes; only reported if there is no other way to determine life stage, or if length or weight categories are the only way to uniquely identify samples of prey data from a source

7) 'prey' relation

This relation describes the diet information related to the 'predator' relation and the prey sample information related to prey biological parameters. (This is the parent entity in the ISA hierarchy for prey biological parameters.) The prey information presented in this relation is not necessarily connected to the predator data as diet information. This relation can also store data that are collected on "potential" prey items from the environment. For example, if a zooplankton tow was conducted in Pacific salmon habitat and "potential" prey size was recorded, this information could be entered here. This data can be connected to the 'predator' relation via the 'site' relation. The types of diet data currently in the database include: percent weight of prey, absolute weight of prey, average weight of prey, percent volume of prey, percent number of prey, stomach content index, absolute number of prey, index of fullness, average number of prey, frequency of occurrence (numerical and percent), and index of relative importance.

Attribute	Key	Data Type	Explanation
prey_id	PK	INT(10)	A unique number that is generated and assigned to each prey sample based on the following attributes: source_id, site_id, prey_lowest_taxonomic_level, prey_life_stage_overall_id,

			prey_subsample_id, prey_sex, type_diet_data, units
predator_id	FK	INT(10)	This id number corresponds to the 'predator' relation
source_id	FK	INT(10)	This id number corresponds to the 'source' relation; not NULL
site_id	FK	INT(10)	This id number corresponds to the 'site' relation; not NULL
prey_lowest_taxonomic_level	FK	VARCHAR(45)	This id number corresponds to the 'prey_taxonomy' relation; not NULL
prey_life_stage_overall_id	FK	INT(10)	This id number corresponds to the 'prey_life_stage' relation
gear_type_prey_id1	FK	INT(10)	This id number corresponds to the 'gear_type_prey' relation; if the prey is part of a diet data sample then the id should be 1 which corresponds to the 'predator' gear type (i.e. sample came from a predator stomach)
gear_type_prey_id2	FK	INT(10)	This id number corresponds to the 'gear_type_prey' relation; this attribute is entered if there are at least 2 types of gear used to sample prey; gear information is only entered if potential prey items are collected from the environment
gear_type_prey_id3	FK	INT(10)	This id number corresponds to the 'gear_type_prey' relation; this attribute is entered if there are at least 3 types of gear used to sample prey; gear information is only entered if potential prey items are collected from the environment
gear_type_prey_id4	FK	INT(10)	This id number corresponds to the 'gear_type_prey' relation; this attribute is entered if there are at least 4 types of gear used to sample prey; gear information is only entered if potential prey items are collected from the environment
gear_type_prey_id5	FK	INT(10)	This id number corresponds to the 'gear_type_prey' relation;

			this attribute is entered if there are at least 5 types of gear used to sample prey; gear information is only entered if potential prey items are collected from the environment
prey_subsample_id		INT(10)	A unique number that is generated and assigned to prey samples if a source reports the biological parameters of individual prey with no unique identifiers; values are assigned for each source starting from 1 and increasing by a value of 1 each time (e.g., 1,2,3…); not NULL; default value is 0
prey_sex		VARCHAR(11)	Either 'male' or 'female'; not NULL; default value is 'unspecified'
type_diet_data		VARCHAR(100)	The diet metric reported in the source (e.g., percent weight of prey, index of relative importance); not NULL; default value is 'not relevant'
value_diet_data		DECIMAL(13,3)	The value of the diet data metric for a specific prey item and sample
diet_data_units		VARCHAR(45)	The diet data units (e.g., percent); if the diet data are reported as a number then the units are left as blank; units are fully spelled out and plural (e.g., grams instead of gram)
formula		VARCHAR(200)	The formula for the diet metric, if applicable; this is for metrics such as the index of relative importance or the stomach content index because they may be calculated differently in different sources
prey_replicates		INT(10)	The total number of prey replicates per sample
prey_notes		VARCHAR(300)	Any additional comments on the prey

8) 'gear_type_predator' relation

This relation provides information about the gear used to capture the predators. There can be up to five different gear types per predator sample.

Attribute	Key	Data Type	Explanation
gear_type_predator_id	PK	INT(10)	A unique number that is generated and assigned to each predator gear type
gear_type		VARCHAR(45)	The most basic description of the type of gear given in the source (e.g., trawl, gillnet, longline)
gear_length_value		DECIMAL(13,3)	The gear length
gear_length_min		DECIMAL(13,3)	If there are a range of gear lengths, then this attribute represents the minimum length
gear_length_max		DECIMAL(13,3)	If there are a range of gear lengths, then this attribute represents the maximum length
gear_length_units		VARCHAR(45)	The units associated with the gear length; units are fully spelled out and plural (e.g., meters instead of meter)
gear_width_value		DECIMAL(13,3)	The gear width
gear_width_min		DECIMAL(13,3)	If there are a range of gear widths, then this attribute represents the minimum width
gear_width_max		DECIMAL(13,3)	If there are a range of gear widths, then this attribute represents the maximum width
gear_width_units		VARCHAR(45)	The units associated with the gear width; units are fully spelled out and plural (e.g., meters instead of meter)
gear_depth_value		DECIMAL(13,3)	The gear depth; if the gear is reported to be deployed at the surface then the depth value is assigned to 0
gear_depth_min		DECIMAL(13,3)	If there are a range of gear depths, then this attribute represents the minimum depth; if the gear is reported to be deployed at the surface then the depth value is assigned to 0
gear_depth_max		DECIMAL(13,3)	If there are a range of gear depths, then this attribute represents the maximum depth
gear_depth_units		VARCHAR(45)	The units associated with the gear depth; units are fully spelled out and plural (e.g., meters instead of meter)
mesh_size_value		DECIMAL(13,3)	The gear mesh size

mesh_size_min		DECIMAL(13,3)	If there are a range of gear mesh sizes, then this attribute represents the minimum mesh size
mesh_size_max		DECIMAL(13,3)	If there are a range of gear mesh sizes, then this attribute represents the maximum mesh size
mesh_size_units		VARCHAR(45)	The units associated with the mesh size; units are fully spelled out and plural (e.g., millimeters instead of millimeter)
fishing_depth_value		DECIMAL(13,3)	The fishing depth; if fishing is reported to be at the surface then the depth value is assigned to 0
fishing_depth_min		DECIMAL(13,3)	If there are a range of fishing depths, then this attribute represents the minimum fishing depth; if fishing is reported to be at the surface then the depth value is assigned to 0
fishing_depth_max		DECIMAL(13,3)	If there are a range of fishing depths, then this attribute represents the maximum fishing depth
fishing_depth_units		VARCHAR(45)	The units associated with the fishing depth; units are fully spelled out and plural (e.g., meters instead of meter)
tow_speed_value		DECIMAL(13,3)	The gear tow speed
tow_speed_min		DECIMAL(13,3)	If there are a range of gear tow speeds, then this attribute represents the minimum tow speed
tow_speed_max		DECIMAL(13,3)	If there are a range of gear tow speeds, then this attribute represents the maximum tow speed
tow_speed_units		VARCHAR(45)	The units associated with the tow speed; units are fully spelled out and plural (e.g., knots instead of knot)
duration_deployment_value		DECIMAL(13,3)	The gear duration of deployment
duration_deployment_min		DECIMAL(13,3)	If there are a range of gear durations of deployment, then this attribute represents the minimum duration of deployment
duration_deployment_max		DECIMAL(13,3)	If there are a range of gear durations of deployment, then this attribute represents the maximum duration of deployment
duration_deployment_units		VARCHAR(45)	The units associated with the duration of deployment; units are fully spelled out and plural (e.g., minutes instead of minute)

gear_notes		VARCHAR(300)	Any additional comments on the gear

9) 'gear_type_prey' relation

This relation provides information about the gear used to capture prey. This relation is used both if data are collected on "potential" prey items from the environment and if data are collected on prey found in salmon stomachs. For example, if a zooplankton tow was conducted in Pacific salmon habitat and "potential" prey were measured, then the details on the zooplankton tow would be recorded in this relation. If a predator stomach contained prey items that were measured, then the gear type is 'predator. There can be up to five different gear types per prey sample.

Attribute	Key	Data Type	Explanation
gear_type_prey_id	PK	INT(10)	A unique number that is generated and assigned to each prey gear type
gear_type		VARCHAR(45)	The most basic description of the type of gear given in the source (e.g., bongo net); if the prey is a diet item, then the gear_type should be 'predator' to indicate that it was not collected from the environment but instead in a salmon stomach
gear_length_value		DECIMAL(13,3)	The gear length
gear_length_min		DECIMAL(13,3)	If there are a range of gear lengths, then this attribute represents the minimum length
gear_length_max		DECIMAL(13,3)	If there are a range of gear lengths, then this attribute represents the maximum length
gear_length_units		VARCHAR(45)	The units associated with the gear length; units are fully spelled out and plural (e.g., meters instead of meter)
gear_width_value		DECIMAL(13,3)	The gear width
gear_width_min		DECIMAL(13,3)	If there are a range of gear widths, then this attribute represents the minimum width
gear_width_max		DECIMAL(13,3)	If there are a range of gear widths, then this attribute represents the maximum width
gear_width_units		VARCHAR(45)	The units associated with the gear width; units are fully spelled out and plural (e.g., meters instead of meter)
gear_depth_value		DECIMAL(13,3)	The gear depth; if the gear is reported to be deployed at the surface then the depth value is assigned to 0
gear_depth_min		DECIMAL(13,3)	If there are a range of gear depths, then this attribute represents the

			minimum depth; if the gear is reported to be deployed at the surface then the depth value is assigned to 0
gear_depth_max		DECIMAL(13,3)	If there are a range of gear depths, then this attribute represents the maximum depth
gear_depth_units		VARCHAR(45)	The units associated with the gear depth; units are fully spelled out and plural (e.g., meters instead of meter)
mesh_size_value		DECIMAL(13,3)	The gear mesh size
mesh_size_min		DECIMAL(13,3)	If there are a range of gear mesh sizes, then this attribute represents the minimum mesh size
mesh_size_max		DECIMAL(13,3)	If there are a range of gear mesh sizes, then this attribute represents the maximum mesh size
mesh_size_units		VARCHAR(45)	The units associated with the mesh size; units are fully spelled out and plural (e.g., millimeters instead of millimeter)
fishing_depth_value		DECIMAL(13,3)	The fishing depth; if fishing is reported to be at the surface then the depth value is assigned to 0
fishing_depth_min		DECIMAL(13,3)	If there are a range of fishing depths, then this attribute represents the minimum fishing depth; if fishing is reported to be at the surface then the depth value is assigned to 0
fishing_depth_max		DECIMAL(13,3)	If there are a range of fishing depths, then this attribute represents the maximum fishing depth
fishing_depth_units		VARCHAR(45)	The units associated with the fishing depth; units are fully spelled out and plural (e.g., meters instead of meter)
tow_speed_value		DECIMAL(13,3)	The gear tow speed
tow_speed_min		DECIMAL(13,3)	If there are a range of gear tow speeds, then this attribute represents the minimum tow speed
tow_speed_max		DECIMAL(13,3)	If there are a range of gear tow speeds, then this attribute represents the maximum tow speed
tow_speed_units		VARCHAR(45)	The units associated with the tow speed; units are fully spelled out and plural (e.g., knots instead of knot)
duration_deployment_value		DECIMAL(13,3)	The gear duration of deployment

duration_deployment_min		DECIMAL(13,3)	If there are a range of gear durations of deployment, then this attribute represents the minimum duration of deployment
duration_deployment_max		DECIMAL(13,3)	If there are a range of gear durations of deployment, then this attribute represents the maximum duration of deployment
duration_deployment_units		VARCHAR(45)	The units associated with the duration of deployment; units are fully spelled out and plural (e.g., minutes instead of minute)
gear_notes		VARCHAR(300)	Any additional comments on the gear

10) 'environmental_data' relation

This relation contains information on environmental conditions that are relevant for understanding predator and prey interactions in the habitat range of Pacific salmon. Data are reported as either a mean or a value, not both. Data can also be reported as error, min, and max.

Attribute	Key	Data Type	Explanation
environmental_data_id	PK	INT(10)	A unique number that is generated and assigned to each environmental data point
source_id	FK	INT(10)	This id number corresponds to the 'source' relation; not NULL
site_id	FK	INT(10)	This id number corresponds to the 'site' relation; not NULL
environmental_data_type		VARCHAR(100)	The environmental data type reported in the source (e.g., temperature, salinity); not NULL
measurement_depth		DECIMAL(13,3)	The depth associated with the environmental parameter measurement; if the measurement in reported to be at the surface (e.g., sea surface temperature) then the measurement depth value is assigned to 0
depth_units		VARCHAR(45)	The units associated with the measurement depth; units are fully spelled out and plural (e.g., meters instead of meter)
environmental_subsample_id		INT(10)	A unique number that is generated and assigned to environmental samples if there are no other unique identifiers; values are assigned for each source starting from 1 and increasing by a value

			of 1 each time (e.g., 1,2,3…); not NULL; default value is 0
value		DECIMAL(13,3)	The value of the environmental parameter
mean		DECIMAL(13,3)	The mean of the environmental parameters
error		DECIMAL(13,3)	The error associated with the environmental parameter mean
min		DECIMAL(13,3)	If there are a range of values for the environmental parameter this attribute represents the minimum value
max		DECIMAL(13,3)	If there are a range of values for the environmental parameter this attribute represents the maximum value
environmental_units		VARCHAR(45)	The units associated with the environmental parameter; units are fully spelled out and plural (e.g., micrograms per liter instead of microgram per liter)
environmental_notes		VARCHAR(300)	Any additional comments on the environmental parameter measurement

11) 'spatial' relation

This relation contains information on the latitude and longitude coordinates of the sampling locations. Every data point must have geographic coordinates associated with the data, however, if there are no specific coordinates given in the source then ranges are estimated based on maps in the source or from the description of the sampling location. Together with the 'temporal' relation this relation defines the site_id.

Attribute	Key	Data Type	Explanation
spatial_id	PK	INT(10)	A unique number that is generated and assigned to each spatial data point
lat_min		DECIMAL(9,6)	If there is just one value for the latitude then it is found in this attribute; if there are a range of values for the latitude then the minimum value is found in this attribute; values are in decimal degrees format; not NULL
lat_max		DECIMAL(9,6)	If there are a range of values for the latitude then the maximum value is found in this attribute; values are in decimal degrees format
lon_min		DECIMAL(9,6)	If there is just one value for the longitude then it is entered into this attribute; If there are a range of values for the longitude then the minimum value is entered into this attribute; must be in decimal degrees format; not NULL
lon_max		DECIMAL(9,6)	If there are a range of values for the longitude then the maximum value is entered into this attribute; must be in decimal degrees format

12) 'temporal' relation

This relation contains information on the temporal aspect of the site_id. Since sources report the temporal information to varying resolutions, this relation contains many attributes to help define the time in which the sampling takes place. As much information as possible is entered with regards to the year, month, date, and time – meaning if all of these attributes can be assigned values then they have values. However, for season, warm_cool_years, and odd_even_years, these attributes only contain data if the source explicitly uses these attributes to define the time period in which this data was collected and they are required for identifying unique samples in the database. Together with the 'spatial' relation, this relation defines the site_id.

Attribute	Key	Data Type	Explanation
temporal_id	PK	INT(10)	A unique number that is generated and assigned to each temporal data point
year_min		YEAR(4)	YYYY; if there is just one value for the year then it is entered into this attribute; if there are a range of values for the year then the minimum value is entered into this attribute
year_max		YEAR(4)	YYYY; if there are a range of values for the year then the maximum value is entered into this attribute
warm_cool_years		CHAR(4)	Either 'warm', 'cool' or NA; this attribute will only have a value if the samples are explicitly reported as being from a warm versus cool year(s) and can only be uniquely identified this way
odd_even_years		VARCHAR(4)	Either 'odd', 'even or NA'; this attribute will only have a value if the samples are explicitly reported as being from an odd versus even year(s) and can only be uniquely identified this way
season_min		CHAR(6)	Either 'spring', 'summer', 'autumn', or 'winter'; if there is just one value for the season then it is entered into this attribute; if there are a range of values for the season then the minimum value is entered into this attribute; this attribute will only have a value if there is no value for the month/date and the source explicitly defines the temporal sampling period by season
season_max		CHAR(6)	Either 'spring', 'summer', 'autumn', or 'winter'; if there are a range of values for the season then the maximum value is entered into this attribute; this attribute will only have a value if there is no value for the month/date and the source explicitly defines the temporal sampling period by season
month_min		VARCHAR(9)	Month names should be completely spelled out with the first letter capitalized; if there is just one value for the month then it is entered into

			this attribute; if there are a range of values for the month then the minimum value is found in this attribute
month_max		VARCHAR(9)	Month names should be completely spelled out; if there are a range of values for the month then the maximum value is found in this attribute
date_min		DATE	YYYY-MM-DD; if there is just one value for the date then it is entered into this attribute; if there are a range of values for the date then the minimum value is found in this attribute
date_max		DATE	YYYY-MM-DD; if there are a range of values for the date then the maximum value is found in this attribute
time_min		TIME	HH:MM:SS; if there is just one value for the time then it is entered into this attribute; if there are a range of values for the time then the minimum value is found in this attribute
time_max		TIME	HH:MM:SS; if there are a range of values for the time then the maximum value is found in this attribute

13) 'site' relation

This relation combines both the spatial_id and temporal_id to designate a unique site_id. The site_id can be used to link diet information, predator biological parameters, prey biological parameters, and environmental data through space and time.

Attribute	Key	Data Type	Explanation
site_id	PK	INT(10)	A unique number that is generated and assigned to each site, defined by both space (spatial_id) and time (temporal_id)
spatial_id	FK	INT(10)	This id number corresponds to the 'spatial' relation; not NULL
temporal_id	FK	INT(10)	This id number corresponds to the 'temporal' relation; not NULL

14) 'predator_fork_length' relation

This is just one example of a predator biological parameter relation which uses the predator_id from the 'predator' relation and in this way allows the predator biological parameter data to be directly linked to the diet data. However, the predator biological parameter data also does not have to be directly linked to diet data and could be from a different sample of salmon. For example, sometimes sources will report diet data for individual sampling sites but then pool the predator biological parameter data from an area that includes multiple sample sites. In this case the predator biological parameter data will not be directly linked to diet data.

Data are reported as either a mean or a value, not both. Data can also be reported as error, min, and max. There is a different relation for each predator biological parameter and an infinite number of

these relations can be added to the database structure. Each relation will have the same attributes. The updated list of predator biological parameter relations is as follows: predator_age, predator_age_composition, predator_body_length, predator_body_weight, predator_condition_factor, predator_daily_ration, predator_empty_stomachs, predator_female_composition, predator_fork_length, predator_gill_raker_count, predator_gonad_weight, predator_immature_composition, predator_index_of_fullness, predator_gonad_weight, predator_male_composition, predator_male_female_ratio, predator_number_of_females, predator_number_of_males, predator_number_of_scale_circuli, predator_ocean_age_1_composition, predator_ocean_age_2_composition, predator_ocean_age_3_composition, predator_ocean_age_4_composition, predator_scale_radius, predator_stomach_content_index, predator_stomach_content_weight, predator_total_body_length.

Attribute	Key	Data Type	Explanation
predator_id	PK/FK	INT(10)	This id number corresponds to the 'predator' relation
value		DECIMAL(13,3)	The fork length value
mean		DECIMAL(13,3)	The fork length mean
error		DECIMAL(13,3)	The error associated with the fork length mean
min		DECIMAL(13,3)	If there are a range of values for the fork length this attribute represents the minimum value
max		DECIMAL(13,3)	If there are a range of values for the fork length this attribute represents the maximum value
units	PK	VARCHAR(45)	The units associated with the fork length; units should be fully spelled out to avoid problems with losing symbols when uploading data and they should be plural (e.g., centimeters instead of centimeter); not NULL; default value is 'unspecified'
predator_bio_notes		VARCHAR(300)	Any additional comments on the predator biological parameters

15) 'prey_body_length'

This is just one example of a prey biological parameter relation which uses the prey_id from the 'prey' relation and in this way allows the prey biological parameter data to be directly linked to the diet data. However, the prey biological parameter data also does not have to be directly linked to a salmon stomach sample and could be linked to a sample of potential prey items that are directly sampled from the habitat of Pacific salmon.

Data are reported as either a mean or a value, not both. Data can also be reported as error, min, and max. There is a different relation for each prey biological parameter and an infinite number of these relations can be added to the database structure. Each relation will have the same attributes. The updated list of prey biological parameter relations is as follows: prey_body_length, prey_body_weight, prey_body_width, prey_size_index.

Attribute	Key	Data Type	Explanation
prey_id	PK/FK	INT(10)	This id number corresponds to the 'prey' relation

value		DECIMAL(13,3)	The body length value
mean		DECIMAL(13,3)	The body length mean
error		DECIMAL(13,3)	The error associated with the body length mean
min		DECIMAL(13,3)	If there are a range of values for the body length this attribute represents the minimum value
max		DECIMAL(13,3)	If there are a range of values for the body length this attribute represents the maximum value
units	PK	VARCHAR(45)	The units associated with the body length; units should be fully spelled out to avoid problems with losing symbols when uploading data and they should be plural (e.g., millimeters instead of millimeter); not NULL; default value is 'unspecified
prey_bio_notes		VARCHAR(300)	Any additional comments on the prey biological parameters

HOW TO ENTER DATA USING CSV FILES

Data are uploaded into the database using csv files that are imported into R, run through an editing script, and then imported into the MySQL Marine Salmon Diet Database. The sample excel file: "msdd_data_entry_template.xlsx" contains the correct sheets and columns for entering data that will eventually be added to the database. This excel file contains the following seven sheets: 'sources', 'gear_type_predator', 'gear_type_prey', 'diet_data', 'predator_biological_data', 'prey_biological_data', and 'environmental data'. These sheets will be imported into R as separate csv files for data editing purposes. The methods for data entry are outlined below. As a general rule, if there is no information for the column, then it should be left blank. For more detailed explanations about formatting, see the above section: 'Database Relations and Attribute Descriptions'. If the information is available in figure format but not in table format then the data should be extracted using WebPlotDigitizer (https://automeris.io/WebPlotDigitizer/).

1) sources

Column	Explanation
source_id	A unique number that is generated and assigned to each source (e.g., 1, 2, 3…etc.)
publication_year	Publication year in the format: YYYY
title	Title of the source
author1	First author of the source in the following format: M. C. Graham or M. Graham (first and middle names are abbreviated by initials and a period while last names are completely spelled out)
author2	Second author of the source in the following format: M. C. Graham or M. Graham (first and middle names are abbreviated by initials and a period while last names are completely spelled out)
author3	Third author of the source in the following format: M. C. Graham or M. Graham (first and middle names are abbreviated by initials and a period while last names are completely spelled out)
author4	Fourth author of the source in the following format: M. C. Graham or M. Graham (first and middle names are abbreviated by initials and a period while last names are completely spelled out)
author5	Fifth author of the source in the following format: M. C. Graham or M. Graham (first and middle names are abbreviated by initials and a period while last names are completely spelled out)
author6	Sixth author of the source in the following format: M. C. Graham or M. Graham (first and middle names are abbreviated by initials and a period while last names are completely spelled out)

author7	Seventh author of the source in the following format: M. C. Graham or M. Graham (first and middle names are abbreviated by initials and a period while last names are completely spelled out)
author8	Eighth author of the source in the following format: M. C. Graham or M. Graham (first and middle names are abbreviated by initials and a period while last names are completely spelled out)
author9	Ninth author of the source in the following format: M. C. Graham or M. Graham (first and middle names are abbreviated by initials and a period while last names are completely spelled out)
author10	Tenth author of the source in the following format: M. C. Graham or M. Graham (first and middle names are abbreviated by initials and a period while last names are completely spelled out)
url	URL associated with source, if applicable
citation	The full citation for the source; in the format of *Deep Sea Research II* with full journal titles
source_notes	Any additional notes about the source; for example, if the data might be overlapping with another source, this should be indicated in this attribute
data_processing_notes	Notes about how the data were processed – in the lab or field, were quantities measured using scales or visually estimated, etc.
date_entered	The date the source was added to the database in the format: YYYY-MM-DD
entered_by	The full name of the person who entered the data (e.g., Caroline Graham)

2) gear_type_predator

Column	Explanation
gear_type_predator_id	A unique number that is generated and assigned to each unique predator gear type for each source (e.g.,1, 2, 3…etc.)
source_id	This number must correspond with the source_id from the 'sources' excel sheet
gear_type	The most basic description of the type of gear given in the source (e.g., trawl, gillnet, longline)
gear_length_value	The gear length
gear_length_min	If there are a range of gear lengths, then this attribute represents the minimum length
gear_length_max	If there are a range of gear lengths, then this attribute represents the maximum length
gear_length_units	The units associated with the gear length; units should be fully spelled out to avoid problems with losing

	symbols when uploading data and they should be plural (e.g., meters instead of meter)
gear_width_value	The gear width
gear_width_min	If there are a range of gear widths, then this attribute represents the minimum width
gear_width_max	If there are a range of gear widths, then this attribute represents the maximum width
gear_width_units	The units associated with the gear width; units should be fully spelled out to avoid problems with losing symbols when uploading data and they should be plural (e.g., meters instead of meter)
gear_depth_value	The gear depth; if the gear is reported to be deployed at the surface then the depth value is assigned to 0
gear_depth_min	If there are a range of gear depths, then this attribute represents the minimum depth; if the gear is reported to be deployed at the surface then the depth value is assigned to 0
gear_depth_max	If there are a range of gear depths, then this attribute represents the maximum depth
gear_depth_units	The units associated with the gear depth; units should be fully spelled out to avoid problems with losing symbols when uploading data and they should be plural (e.g., meters instead of meter)
mesh_size_value	The gear mesh size
mesh_size_min	If there are a range of gear mesh sizes, then this attribute represents the minimum mesh size
mesh_size_max	If there are a range of gear mesh sizes, then this attribute represents the maximum mesh size
mesh_size_units	The units associated with the mesh size; units should be fully spelled out to avoid problems with losing symbols when uploading data and they should be plural (e.g., millimeters instead of millimeter)
fishing_depth_value	The fishing depth; if fishing is reported to be at the surface then the depth value is assigned to 0
fishing_depth_min	If there are a range of fishing depths, then this attribute represents the minimum fishing depth; if fishing is reported to be at the surface then the depth value is assigned to 0
fishing_depth_max	If there are a range of fishing depths, then this attribute represents the maximum fishing depth
fishing_depth_units	The units associated with the fishing depth; units should be fully spelled out to avoid problems with losing symbols when uploading data and they should be plural (e.g., meters instead of meter)
tow_speed_value	The gear tow speed
tow_speed_min	If there are a range of gear tow speeds, then this attribute represents the minimum tow speed
tow_speed_max	If there are a range of gear tow speeds, then this attribute represents the maximum tow speed

tow_speed_units	The units associated with the tow speed; units should be fully spelled out to avoid problems with losing symbols when uploading data and they should be plural (e.g., knots instead of knot)
duration_deployment_value	The gear duration of deployment
duration_deployment_min	If there are a range of gear durations of deployment, then this attribute represents the minimum duration of deployment
duration_deployment_max	If there are a range of gear durations of deployment, then this attribute represents the maximum duration of deployment
duration_deployment_units	The units associated with the duration of deployment; units should be fully spelled out to avoid problems with losing symbols when uploading data and they should be plural (e.g., minutes instead of minute)
gear_notes	Any additional comments on the gear

3) gear_type_prey

Column	Explanation
gear_type_prey_id	A unique number that is generated and assigned to each unique prey gear type for each source (e.g.,1, 2, 3…etc.)
source_id	This number must correspond with the source_id from the 'sources' excel sheet
gear_type	The most basic description of the type of gear given in the source (e.g., bongo net); if the prey is a diet item, then the gear_type should be 'predator' to indicate that it was not collected from the environment but instead in a salmon stomach
gear_length_value	The gear length
gear_length_min	If there are a range of gear lengths, then this attribute represents the minimum length
gear_length_max	If there are a range of gear lengths, then this attribute represents the maximum length
gear_length_units	The units associated with the gear length; units should be fully spelled out to avoid problems with losing symbols when uploading data and they should be plural (e.g., meters instead of meter)
gear_width_value	The gear width
gear_width_min	If there are a range of gear widths, then this attribute represents the minimum width
gear_width_max	If there are a range of gear widths, then this attribute represents the maximum width
gear_width_units	The units associated with the gear width; units should be fully spelled out to avoid problems with losing

	symbols when uploading data and they should be plural (e.g., meters instead of meter)
gear_depth_value	The gear depth; if gear is reported to be deployed at the surface then the depth value is assigned to 0
gear_depth_min	If there are a range of gear depths, then this attribute represents the minimum depth; if gear is reported to be deployed at the surface then the depth value is assigned to 0
gear_depth_max	If there are a range of gear depths, then this attribute represents the maximum depth
gear_depth_units	The units associated with the gear depth; units should be fully spelled out to avoid problems with losing symbols when uploading data and they should be plural (e.g., meters instead of meter)
mesh_size_value	The gear mesh size
mesh_size_min	If there are a range of gear mesh sizes, then this attribute represents the minimum mesh size
mesh_size_max	If there are a range of gear mesh sizes, then this attribute represents the maximum mesh size
mesh_size_units	The units associated with the mesh size; units should be fully spelled out to avoid problems with losing symbols when uploading data and they should be plural (e.g., millimeters instead of millimeter)
fishing_depth_value	The fishing depth; if fishing is reported to be at the surface then the depth value is assigned to 0
fishing_depth_min	If there are a range of fishing depths, then this attribute represents the minimum fishing depth; if fishing is reported to be at the surface then the depth value is assigned to 0
fishing_depth_max	If there are a range of fishing depths, then this attribute represents the maximum fishing depth
fishing_depth_units	The units associated with the fishing depth; units should be fully spelled out to avoid problems with losing symbols when uploading data and they should be plural (e.g., meters instead of meter)
tow_speed_value	The gear tow speed
tow_speed_min	If there are a range of gear tow speeds, then this attribute represents the minimum tow speed
tow_speed_max	If there are a range of gear tow speeds, then this attribute represents the maximum tow speed
tow_speed_units	The units associated with the tow speed; units should be fully spelled out to avoid problems with losing symbols when uploading data and they should be plural (e.g., knots instead of knot)
duration_deployment_value	The gear duration of deployment
duration_deployment_min	If there are a range of gear durations of deployment, then this attribute represents the minimum duration of deployment

duration_deployment_max	If there are a range of gear durations of deployment, then this attribute represents the maximum duration of deployment
duration_deployment_units	The units associated with the duration of deployment; units should be fully spelled out to avoid problems with losing symbols when uploading data and they should be plural (e.g., minutes instead of minute)
gear_notes	Any additional comments on the gear

4) diet_data

Column	Explanation
source_id	This number must correspond with the source_id from the 'sources' excel sheet
year_min	YYYY; if there is just one value for the year then it is entered into this attribute; If there are a range of values for the year then the minimum value is entered into this attribute
year_max	YYYY; if there are a range of values for the year then the maximum value is entered into this attribute
warm_cool_years	Either 'warm' or 'cool'; this attribute should only have a value if the data are explicitly reported as being from a warm versus cool year(s)
odd_even_years	Either 'odd' or 'even'; this attribute should only have a value if the data are explicitly reported as being from an odd versus even year(s)
season_min	Either 'spring', 'summer', 'autumn', or 'winter'; ff there is just one value for the season then it is entered into this attribute; If there are a range of values for the season then the minimum value is entered into this attribute; this attribute should only have a value if there is no value for the month/date and the source explicitly defines the temporal sampling period by season
season_max	Either 'spring', 'summer', 'autumn', or 'winter'; if there are a range of values for the season then the maximum value is entered into this attribute; this attribute should only have a value if there is no value for the month/date and the source explicitly defines the temporal sampling period by season
month_min	Month names should be completely spelled out; if there is just one value for the month then it is entered into this attribute; If there are a range of values for the month then the minimum value is entered into this attribute

month_max	Month names should be completely spelled out; if there are a range of values for the month then the maximum value is entered into this attribute
date_min	YYYY-MM-DD; if there is just one value for the date then it is entered into this attribute; If there are a range of values for the date then the minimum value is entered into this attribute
date_max	YYYY-MM-DD; if there are a range of values for the date then the maximum value is entered into this attribute
time_min	HH:MM:SS; if there is just one value for the time then it is entered into this attribute; if there are a range of values for the time then the minimum value is entered into this attribute
time_max	HH:MM:SS; if there are a range of values for the time then the maximum value is entered into this attribute
lat_min	If there is just one value for the latitude then it is entered into this attribute; If there are a range of values for the latitude then the minimum value is entered into this attribute; values must be in decimal degrees format
lat_max	If there are a range of values for the latitude then the maximum value is entered into this attribute; must be in decimal degrees format
lon_min	If there is just one value for the longitude then it is entered into this attribute; If there are a range of values for the longitude then the minimum value is entered into this attribute; must be in decimal degrees format
lon_max	If there are a range of values for the longitude then the maximum value is entered into this attribute; must be in decimal degrees format
predator_lowest_taxonomic_level	The lowest taxonomic level reported in the source. If a source reports to the species level then this attribute should include both the genus and species names (e.g., Oncorhynchus nerka); only scientific names are reported, not common names; first letter should be capitalized
predator_life_stage	Either 'juvenile' or 'adult'
predator_life_stage_min	If there are a mixture of juveniles and adults, then 'juvenile' should be entered here
predator_life_stage_max	If there are a mixture of juveniles and adults, then 'adult' should be entered here
predator_freshwater_age	An integer to indicate the number of years spent living in freshwater
predator_freshwater_age_min	If there are a mixture of freshwater ages, then this attribute represents the minimum age; an

	integer to indicate the number of years spent living in freshwater
predator_freshwater_age_max	If there are a mixture of freshwater ages, then this attribute represents the maximum age; an integer to indicate the number of years spent living in freshwater
predator_ocean_age	An integer to indicate the number of years spent living in the ocean
predator_ocean_age_min	If there are a mixture of ocean ages, then this attribute represents the minimum age; an integer to indicate the number of years spent living in the ocean
predator_ocean_age_max	If there are a mixture of ocean ages, then this attribute represents the maximum age; an integer to indicate the number of years spent living in the ocean
predator_maturity	Either 'juvenile', 'immature', 'maturing' 'mature', or 'kelt' (for steelhead)
predator_maturity_min	If there are a mixture of maturity levels, then the minimum maturity level should be entered here
predator_maturity_max	If there are a mixture of maturity levels, then the maximum maturity level should be entered here
predator_length_value_cm	The length of a predator in centimeters (could be either fork length or total length); only reported if there is no other way to determine life stage, or if length or weight categories are the only way to uniquely identify samples of diet data from a source
predator_length_min_cm	The minimum length of a predator in centimeters (could be either fork length or total length) if there are a range of sizes; only reported if there is no other way to determine life stage, or if length or weight categories are the only way to uniquely identify samples of diet data from a source
predator_length_max_cm	The maximum length of a predator in centimeters (could be either fork length or total length) if there are a range of sizes; only reported if there is no other way to determine life stage, or if length or weight categories are the only way to uniquely identify samples of diet data from a source
predator_weight_value_g	The weight of a predator in grams; only reported if there is no other way to determine life stage, or if length or weight categories are the only way to uniquely identify samples of diet data from a source
predator_weight_min_g	The minimum weight of a predator in grams if there are a range of sizes; only reported if there is no other way to determine life stage, or if length

	or weight categories are the only way to uniquely identify samples of diet data from a source
predator_weight_max_g	The maximum weight of a predator in grams if there are a range of sizes; only reported if there is no other way to determine life stage, or if length or weight categories are the only way to uniquely identify samples of diet data from a source
predator_subsample_id	A unique number that is generated and assigned to predator samples if a source reports the diets of individual predators with no unique identifiers; values are assigned for each source starting from 1 and increasing by a value of 1 each time (e.g., 1,2,3…)
predator_sex	Either 'male' or 'female'
hatchery_wild	Either 'hatchery' or 'wild'
predator_replicates	The total number of predator replicates per sample
predator_notes	Any additional comments on the predator
gear_type_predator_id1	This id number corresponds with the gear_type_predator_id from the 'gear_type_predator' excel sheet
gear_type_predator_id2	This id number corresponds with the gear_type_predator_id from the 'gear_type_predator' excel sheet; this attribute is entered if there are at least 2 types of gear used to sample predators
gear_type_predator_id3	This id number corresponds with the gear_type_predator_id from the 'gear_type_predator' excel sheet; this attribute is entered if there are at least 3 types of gear used to sample predators
gear_type_predator_id4	This id number corresponds with the gear_type_predator_id from the 'gear_type_predator' excel sheet; this attribute is entered if there are at least 4 types of gear used to sample predators
gear_type_predator_id5	This id number corresponds with the gear_type_predator_id from the 'gear_type_predator' excel sheet; this attribute is entered if there are at least 5 types of gear used to sample predators
type_diet_data	The diet metric reported in the source (e.g., percent weight of prey, index of relative importance)
diet_data_units	The diet data units (e.g., percent); if the diet data are reported as a number then the units are left as blank; units should be fully spelled out to avoid problems with losing symbols when

	uploading data and they should be plural (e.g., grams instead of gram)
formula	The formula for the diet metric, if applicable; this is for metrics such as the index of relative importance or the stomach content index because they may be calculated differently in different sources
prey_lowest_taxonomic_level1	These column heading should be changed to reflect the different prey taxonomic categories (e.g., copepoda, Limacina helicina, octopus < 10 cm) and these categories should be reported exactly as they are in the text and should include any life stage information reported as well; the cells will contain diet data values for the specific prey item and the diet metric; if the value is 0 then it is left as a blank value
prey_lowest_taxonomic_level2	These column heading should be changed to reflect the different prey taxonomic categories (e.g., copepoda, Limacina helicina, octopus < 10 cm) and these categories should be reported exactly as they are in the text and should include any life stage information reported as well; the cells will contain diet data values for the specific prey item and the diet metric; if the value is 0 then it is left as a blank value
prey_lowest_taxonomic_level3	These column heading should be changed to reflect the different prey taxonomic categories (e.g., copepoda, Limacina helicina, octopus < 10 cm) and these categories should be reported exactly as they are in the text and should include any life stage information reported as well; the cells will contain diet data values for the specific prey item and the diet metric; if the value is 0 then it is left as a blank value

5) predator_biological_data

Column	Explanation
source_id	This number must correspond with the source_id from the 'sources' excel sheet
year_min	YYYY; if there is just one value for the year then it is entered into this attribute; If there are a range of values for the year then the minimum value is entered into this attribute
year_max	YYYY; if there are a range of values for the year then the maximum value is entered into this attribute

warm_cool_years	Either ‘warm’ or ‘cool’; this attribute should only have a value if the data are explicitly reported as being from a warm versus cool year(s)
odd_even_years	Either ‘odd’ or ‘even’; this attribute should only have a value if the data are explicitly reported as being from an odd versus even year(s)
season_min	Either ‘spring’, ‘summer’, ‘autumn’, or ‘winter’; ff there is just one value for the season then it is entered into this attribute; If there are a range of values for the season then the minimum value is entered into this attribute; this attribute should only have a value if there is no value for the month/date and the source explicitly defines the temporal sampling period by season
season_max	Either ‘spring’, ‘summer’, ‘autumn’, or ‘winter’; if there are a range of values for the season then the maximum value is entered into this attribute; this attribute should only have a value if there is no value for the month/date and the source explicitly defines the temporal sampling period by season
month_min	Month names should be completely spelled out; if there is just one value for the month then it is entered into this attribute; If there are a range of values for the month then the minimum value is entered into this attribute
month_max	Month names should be completely spelled out; if there are a range of values for the month then the maximum value is entered into this attribute
date_min	YYYY-MM-DD; if there is just one value for the date then it is entered into this attribute; If there are a range of values for the date then the minimum value is entered into this attribute
date_max	YYYY-MM-DD; if there are a range of values for the date then the maximum value is entered into this attribute
time_min	HH:MM:SS; if there is just one value for the time then it is entered into this attribute; if there are a range of values for the time then the minimum value is entered into this attribute
time_max	HH:MM:SS; if there are a range of values for the time then the maximum value is entered into this attribute
lat_min	If there is just one value for the latitude then it is entered into this attribute; If there are a range of values for the latitude then the minimum value is entered into this attribute; values must be in decimal degrees format
lat_max	If there are a range of values for the latitude then the maximum value is entered into this attribute; must be in decimal degrees format

lon_min	If there is just one value for the longitude then it is entered into this attribute; If there are a range of values for the longitude then the minimum value is entered into this attribute; must be in decimal degrees format
lon_max	If there are a range of values for the longitude then the maximum value is entered into this attribute; must be in decimal degrees format
predator_lowest_taxonomic_level	The lowest taxonomic level reported in the source. If a source reports to the species level then this attribute should include both the genus and species names (e.g., Oncorhynchus nerka); only scientific names are reported, not common names; first letter should be capitalized
predator_life_stage	Either 'juvenile' or 'adult'
predator_life_stage_min	If there are a mixture of juveniles and adults, then 'juvenile' should be entered here
predator_life_stage_max	If there are a mixture of juveniles and adults, then 'adult' should be entered here
predator_freshwater_age	An integer to indicate the number of years spent living in freshwater
predator_freshwater_age_min	If there are a mixture of freshwater ages, then this attribute represents the minimum age; an integer to indicate the number of years spent living in freshwater
predator_freshwater_age_max	If there are a mixture of freshwater ages, then this attribute represents the maximum age; an integer to indicate the number of years spent living in freshwater
predator_ocean_age	An integer to indicate the number of years spent living in the ocean
predator_ocean_age_min	If there are a mixture of ocean ages, then this attribute represents the minimum age; an integer to indicate the number of years spent living in the ocean
predator_ocean_age_max	If there are a mixture of ocean ages, then this attribute represents the maximum age; an integer to indicate the number of years spent living in the ocean
predator_maturity	Either 'juvenile', 'immature', 'maturing' 'mature', or 'kelt' (for steelhead)
predator_maturity_min	If there are a mixture of maturity levels, then the minimum maturity level should be entered here
predator_maturity_max	If there are a mixture of maturity levels, then the maximum maturity level should be entered here
predator_length_value_cm	The length of a predator in centimeters (could be either fork length or total length); only reported if there is no other way to determine life stage, or if length or weight categories are the only way to

	uniquely identify samples of diet data from a source
predator_length_min_cm	The minimum length of a predator in centimeters (could be either fork length or total length) if there are a range of sizes; only reported if there is no other way to determine life stage, or if length or weight categories are the only way to uniquely identify samples of diet data from a source
predator_length_max_cm	The maximum length of a predator in centimeters (could be either fork length or total length) if there are a range of sizes; only reported if there is no other way to determine life stage, or if length or weight categories are the only way to uniquely identify samples of diet data from a source
predator_weight_value_g	The weight of a predator in grams; only reported if there is no other way to determine life stage, or if length or weight categories are the only way to uniquely identify samples of diet data from a source
predator_weight_min_g	The minimum weight of a predator in grams if there are a range of sizes; only reported if there is no other way to determine life stage, or if length or weight categories are the only way to uniquely identify samples of diet data from a source
predator_weight_max_g	The maximum weight of a predator in grams if there are a range of sizes; only reported if there is no other way to determine life stage, or if length or weight categories are the only way to uniquely identify samples of diet data from a source
predator_subsample_id	A unique number that is generated and assigned to predator samples if a source reports the biological parameters of individual predators with no unique identifiers – this number should correspond with the predator_subsample_id from the 'diet_data' excel sheet; values are assigned for each source starting from 1 and increasing by a value of 1 each time (e.g., 1,2,3…)
predator_sex	Either 'male' or 'female'
hatchery_wild	Either 'hatchery' or 'wild'
predator_replicates	The total number of predator replicates per sample
predator_notes	Any additional comments on the predator
gear_type_predator_id1	This id number corresponds with the gear_type_predator_id from the 'gear_type_predator' excel sheet
gear_type_predator_id2	This id number corresponds with the gear_type_predator_id from the 'gear_type_predator' excel sheet; this attribute is

	entered if there are at least 2 types of gear used to sample predators
gear_type_predator_id3	This id number corresponds with the gear_type_predator_id from the 'gear_type_predator' excel sheet; this attribute is entered if there are at least 3 types of gear used to sample predators
gear_type_predator_id4	This id number corresponds with the gear_type_predator_id from the 'gear_type_predator' excel sheet; this attribute is entered if there are at least 4 types of gear used to sample predators
gear_type_predator_id5	This id number corresponds with the gear_type_predator_id from the 'gear_type_predator' excel sheet; this attribute is entered if there are at least 5 types of gear used to sample predators
biological_parameter	The predator biological parameter reported in the source (e.g., total length, fork length, body weight, etc.)
predator_bio_notes	Any additional comments on the predator biological parameters
value	The biological parameter value
mean	The biological parameter mean
error	The error associated with the biological parameter mean
min	If there are a range of values for the biological parameter this attribute represents the minimum value
max	If there are a range of values for the biological parameter this attribute represents the maximum value
units	The units associated with the biological parameter; units should be fully spelled out to avoid problems with losing symbols when uploading data and they should be plural (e.g., centimeters instead of centimeter)

6) prey_biological_data

Column	Explanation
source_id	This number must correspond with the source_id from the 'sources' excel sheet
year_min	YYYY; if there is just one value for the year then it is entered into this attribute; If there are a range of values for the year then the minimum value is entered into this attribute

year_max	YYYY; if there are a range of values for the year then the maximum value is entered into this attribute
warm_cool_years	Either 'warm' or 'cool'; this attribute should only have a value if the data are explicitly reported as being from a warm versus cool year(s)
odd_even_years	Either 'odd' or 'even'; this attribute should only have a value if the data are explicitly reported as being from an odd versus even year(s)
season_min	Either 'spring', 'summer', 'autumn', or 'winter'; ff there is just one value for the season then it is entered into this attribute; If there are a range of values for the season then the minimum value is entered into this attribute; this attribute should only have a value if there is no value for the month/date and the source explicitly defines the temporal sampling period by season
season_max	Either 'spring', 'summer', 'autumn', or 'winter'; if there are a range of values for the season then the maximum value is entered into this attribute; this attribute should only have a value if there is no value for the month/date and the source explicitly defines the temporal sampling period by season
month_min	Month names should be completely spelled out; if there is just one value for the month then it is entered into this attribute; If there are a range of values for the month then the minimum value is entered into this attribute
month_max	Month names should be completely spelled out; if there are a range of values for the month then the maximum value is entered into this attribute
date_min	YYYY-MM-DD; if there is just one value for the date then it is entered into this attribute; If there are a range of values for the date then the minimum value is entered into this attribute
date_max	YYYY-MM-DD; if there are a range of values for the date then the maximum value is entered into this attribute
time_min	HH:MM:SS; if there is just one value for the time then it is entered into this attribute; if there are a range of values for the time then the minimum value is entered into this attribute
time_max	HH:MM:SS; if there are a range of values for the time then the maximum value is entered into this attribute
lat_min	If there is just one value for the latitude then it is entered into this attribute; If there are a range of values for the latitude then the minimum value is entered into this attribute; values must be in decimal degrees format

lat_max	If there are a range of values for the latitude then the maximum value is entered into this attribute; must be in decimal degrees format
lon_min	If there is just one value for the longitude then it is entered into this attribute; If there are a range of values for the longitude then the minimum value is entered into this attribute; must be in decimal degrees format
lon_max	If there are a range of values for the longitude then the maximum value is entered into this attribute; must be in decimal degrees format
predator_lowest_taxonomic_level	The lowest taxonomic level reported in the source. If a source reports to the species level then this attribute should include both the genus and species names (e.g., Oncorhynchus nerka); only scientific names are reported, not common names; first letter should be capitalized
predator_life_stage	Either 'juvenile' or 'adult'
predator_life_stage_min	If there are a mixture of juveniles and adults, then 'juvenile' should be entered here
predator_life_stage_max	If there are a mixture of juveniles and adults, then 'adult' should be entered here
predator_freshwater_age	An integer to indicate the number of years spent living in freshwater
predator_freshwater_age_min	If there are a mixture of freshwater ages, then this attribute represents the minimum age; an integer to indicate the number of years spent living in freshwater
predator_freshwater_age_max	If there are a mixture of freshwater ages, then this attribute represents the maximum age; an integer to indicate the number of years spent living in freshwater
predator_ocean_age	An integer to indicate the number of years spent living in the ocean
predator_ocean_age_min	If there are a mixture of ocean ages, then this attribute represents the minimum age; an integer to indicate the number of years spent living in the ocean
predator_ocean_age_max	If there are a mixture of ocean ages, then this attribute represents the maximum age; an integer to indicate the number of years spent living in the ocean
predator_maturity	Either 'juvenile', 'immature', 'maturing' 'mature', or 'kelt' (for steelhead)
predator_maturity_min	If there are a mixture of maturity levels, then the minimum maturity level should be entered here
predator_maturity_max	If there are a mixture of maturity levels, then the maximum maturity level should be entered here

predator_length_value_cm	The length of a predator in centimeters (could be either fork length or total length); only reported if there is no other way to determine life stage, or if length or weight categories are the only way to uniquely identify samples of diet data from a source
predator_length_min_cm	The minimum length of a predator in centimeters (could be either fork length or total length) if there are a range of sizes; only reported if there is no other way to determine life stage, or if length or weight categories are the only way to uniquely identify samples of diet data from a source
predator_length_max_cm	The maximum length of a predator in centimeters (could be either fork length or total length) if there are a range of sizes; only reported if there is no other way to determine life stage, or if length or weight categories are the only way to uniquely identify samples of diet data from a source
predator_weight_value_g	The weight of a predator in grams; only reported if there is no other way to determine life stage, or if length or weight categories are the only way to uniquely identify samples of diet data from a source
predator_weight_min_g	The minimum weight of a predator in grams if there are a range of sizes; only reported if there is no other way to determine life stage, or if length or weight categories are the only way to uniquely identify samples of diet data from a source
predator_weight_max_g	The maximum weight of a predator in grams if there are a range of sizes; only reported if there is no other way to determine life stage, or if length or weight categories are the only way to uniquely identify samples of diet data from a source
predator_subsample_id	A unique number that is generated and assigned to predator samples if a source reports the diets/biological parameters of individual predators with no unique identifiers – this number should correspond with the predator_subsample_id from the 'diet_data' excel sheet; values are assigned for each source starting from 1 and increasing by a value of 1 each time (e.g., 1,2,3…)
predator_sex	Either 'male' or 'female'
hatchery_wild	Either 'hatchery' or 'wild'
predator_replicates	The total number of predator replicates per sample
predator_notes	Any additional comments on the predator

gear_type_predator_id1	This id number corresponds with the gear_type_predator_id from the 'gear_type_predator' excel sheet
gear_type_predator_id2	This id number corresponds with the gear_type_predator_id from the 'gear_type_predator' excel sheet; this attribute is entered if there are at least 2 types of gear used to sample predators
gear_type_predator_id3	This id number corresponds with the gear_type_predator_id from the 'gear_type_predator' excel sheet; this attribute is entered if there are at least 3 types of gear used to sample predators
gear_type_predator_id4	This id number corresponds with the gear_type_predator_id from the 'gear_type_predator' excel sheet; this attribute is entered if there are at least 4 types of gear used to sample predators
gear_type_predator_id5	This id number corresponds with the gear_type_predator_id from the 'gear_type_predator' excel sheet; this attribute is entered if there are at least 5 types of gear used to sample predators
prey_lowest_taxonomic_level	The lowest taxonomic level reported in the source – exactly as it is reported in the source and should include any life stage information reported as well. If a source reports to the species level then this attribute should include both the genus and species names (e.g., Oncorhynchus nerka); only scientific names are reported, not common names; first letter should be capitalized
prey_subsample_id	A unique number that is generated and assigned to prey samples if a source reports the biological parameters of individual prey with no unique identifiers; values are assigned for each source starting from 1 and increasing by a value of 1 each time (e.g., 1,2,3…)
prey_replicates	The total number of prey replicates per sample
prey_notes	Any additional comments on the prey
gear_type_prey_id1	This id number corresponds with the gear_type_prey_id from the 'gear_type_prey' excel sheet
gear_type_prey_id2	This id number corresponds with the gear_type_prey_id from the 'gear_type_prey excel sheet; this attribute is entered if there are at least 2 types of gear used to sample predators
gear_type_prey_id3	This id number corresponds with the gear_type_prey_id from the 'gear_type_prey

	excel sheet; this attribute is entered if there are at least 3 types of gear used to sample predators
gear_type_prey_id4	This id number corresponds with the gear_type_prey_id from the 'gear_type_prey excel sheet; this attribute is entered if there are at least 4 types of gear used to sample predators
gear_type_prey_id5	This id number corresponds with the gear_type_prey_id from the 'gear_type_prey excel sheet; this attribute is entered if there are at least 5 types of gear used to sample predators
biological_parameter	The prey biological parameter reported in the source (e.g., body length, body weight, etc.)
prey_bio_notes	Any additional comments on the prey biological parameters
value	The biological parameter value
mean	The biological parameter mean
error	The error associated with the biological parameter mean
min	If there are a range of values for the biological parameter this attribute represents the minimum value
max	If there are a range of values for the biological parameter this attribute represents the maximum value
units	The units associated with the biological parameter; units should be fully spelled out to avoid problems with losing symbols when uploading data and they should be plural (e.g., centimeters instead of centimeter)

7) environmental data

Column	Explanation
source_id	This number must correspond with the source_id from the 'sources' excel sheet
year_min	YYYY; if there is just one value for the year then it is entered into this attribute; If there are a range of values for the year then the minimum value is entered into this attribute
year_max	YYYY; if there are a range of values for the year then the maximum value is entered into this attribute
warm_cool_years	Either 'warm' or 'cool'; this attribute should only have a value if the data are explicitly reported as being from a warm versus cool year(s)
odd_even_years	Either 'odd' or 'even'; this attribute should only have a value if the data are explicitly reported as being from an odd versus even year(s)

season_min	Either ‘spring’, ‘summer’, ‘autumn’, or ‘winter’; ff there is just one value for the season then it is entered into this attribute; If there are a range of values for the season then the minimum value is entered into this attribute; this attribute should only have a value if there is no value for the month/date and the source explicitly defines the temporal sampling period by season
season_max	Either ‘spring’, ‘summer’, ‘autumn’, or ‘winter’; if there are a range of values for the season then the maximum value is entered into this attribute; this attribute should only have a value if there is no value for the month/date and the source explicitly defines the temporal sampling period by season
month_min	Month names should be completely spelled out; if there is just one value for the month then it is entered into this attribute; If there are a range of values for the month then the minimum value is entered into this attribute
month_max	Month names should be completely spelled out; if there are a range of values for the month then the maximum value is entered into this attribute
date_min	YYYY-MM-DD; if there is just one value for the date then it is entered into this attribute; If there are a range of values for the date then the minimum value is entered into this attribute
date_max	YYYY-MM-DD; if there are a range of values for the date then the maximum value is entered into this attribute
time_min	HH:MM:SS; if there is just one value for the time then it is entered into this attribute; if there are a range of values for the time then the minimum value is entered into this attribute
time_max	HH:MM:SS; if there are a range of values for the time then the maximum value is entered into this attribute
lat_min	If there is just one value for the latitude then it is entered into this attribute; If there are a range of values for the latitude then the minimum value is entered into this attribute; values must be in decimal degrees format
lat_max	If there are a range of values for the latitude then the maximum value is entered into this attribute; must be in decimal degrees format
lon_min	If there is just one value for the longitude then it is entered into this attribute; If there are a range of values for the longitude then the minimum value is entered into this attribute; must be in decimal degrees format
lon_max	If there are a range of values for the longitude then the maximum value is entered into this attribute; must be in decimal degrees format

environmental_data_type	The environmental data type reported in the source (e.g., temperature, salinity)
measurement_depth	The depth associated with the environmental parameter measurement; if the measurement in reported to be at the surface (e.g., sea surface temperature) then the measurement depth value is assigned to 0
depth_units	The units associated with the measurement depth; units should be fully spelled out to avoid problems with losing symbols when uploading data and they should be plural (e.g., meters instead of meter)
environmental_subsample_id	A unique number that is generated and assigned to environmental samples if there are no other unique identifiers; values are assigned for each source starting from 1 and increasing by a value of 1 each time (e.g., 1,2,3…)
environmental_notes	Any additional comments on the environmental parameter measurement
value	The value of the environmental parameter
mean	The mean of the environmental parameters
error	The error associated with the environmental parameter mean
min	If there are a range of values for the environmental parameter this attribute represents the minimum value
max	If there are a range of values for the environmental parameter this attribute represents the maximum value
environmental_units	The units associated with the environmental parameter; units should be fully spelled out to avoid problems with losing symbols when uploading data and they should be plural (e.g., micrograms per liter instead of microgram per liter)

APPENDIX: LIST OF TERMS

Term	Definition*
database	a collection of data
database management system	software designed to assist in maintaining and utilizing large collections of data
entity relationship diagram (ERD)	a diagram that allows us to describe the data involved in a real-world enterprise in term of objects and their relationships and is widely used to develop an initial database design; in ERDs the squares represent the tables (or entity sets) in a database and the diamonds represents relationships between the tables
entity	an object in the real world that is distinguishable from other objects
entity set	a collection of similar entities; can be thought of as a table in your ERD
attribute	a set of attributes is used to describe an entity; can be thought of as a column in your database
relational model	a collection of high-level data description constructs that hide many low-level storage details
relation	the central data description construct in the relational model, which can be thought of as a table in your database
primary key	a minimal set of attributes whose values uniquely identify an entity in an entity set (or a row in the table)
foreign key	an attribute, or set of attributes that references a primary key
ISA hierarchy	this is symbolized by a triangle with 'ISA' in the middle; this means the attributes of the parent entity are inherited by the child entities (for example, the predator_fork_length entity inherits attributes from the predator entity)
key constraint	a rule in a database management system that limits the type of data that can be inserted
one-to-one	a type of key constraint; as an example, if you had entity sets for departments and managers a one-to-one key constraint would mean that each department can only have one manager and each manager can only be in one department
one-to-many	a type of key constraint; for example, each sample of predators can only have one site but many different samples of predators could be from the same site
many-to-many	a type of key constraint; for example, if you had entity set for students and classes a many-to-many key constraint would mean that each student can take multiple classes and each class can have multiple students enrolled
participation constraint	can either be total or partial; total meaning that entities must participate in the relationship (e.g., all predator samples must have a site), or partial meaning that entities may not always participate in the relationship (e.g., some predators will have reported fork length values and others will not)

*definitions are based on definitions from Ramakrishnan, R., & Gehrke, J. (2003). Database Management Systems (3rd ed.). Singapore: McGraw-Hill.

CITATIONS

Horton, T. *et al.* World Register of Marine Species (WoRMS). (2019). Available at: http://www.marinespecies.org/.

MySQL. (2019). Available at: https://dev.mysql.com/downloads/.

R Core Development Team. R: A language and environment for statistical computing. (2019). Available at: https://www.r-project.org/.

Ramakrishnan, R., & Gehrke, J. (2003). Database Management Systems (3rd ed.). Singapore: McGraw-Hill.

Rohatgi, A. WebPlotDigitizer. (2019). Available at: https://automeris.io/WebPlotDigitizer/.