MAIA Estuaries 1997/1998 Summary Database

The MAIA Estuaries Summary Database contains water quality, sediment, benthic community, and fish data collected by several partners in MAIA Region estuaries in 1997 and 1998. The database consists of sixteen primary data sets and three code tables, which are available as delimited ascii text files from the MAIA Estuaries web site. Each data set has an accompanying metadata file, in PDF format, which provides detailed description of all data elements. The metadata files also provide information on sampling methods, laboratory methods, QA protocols, and references to other related documents.

These data sets are available as SAS export files, in WINZIP format, from the same web site.

The purpose of this document is to provide:

1) An overview of the database
2) Brief descriptions of each data set
3) A discussion of key fields, and guidance on how the different tables can be joined to form an integrated database.
4) A discussion of QA codes and Lab Codes used in the database

Users should consult the metadata files for more detailed information.

1) Database Overview

The data sets that make up the database are:

Date/Location Data:	STATIONS	Sampling Station Location Data
	EVENTS	Station Visit Data
Water Quality Data:	WATRPHYS	Water Quality Physical Measurements Data
	NUTRNTS	Water Quality Nutrients Data
Sediment Quality Data:	SEDGRAIN	Sediment Grain Size Data
	TOXICITY	Sediment Toxicity and Microtox Test Data
	SEDCHEM	Sediment Chemistry Data
Benthic Community Data:	BENGRAIN	Benthic Sample Grain Size Data
	BEN_ABUN	Benthic Abundance Data
	BEN_BIOM	Benthic Biomass Data
	BENSUMRY	Benthic Community Summary Data

Fish Data:	FTRAWL	Standard Trawl Data
	FISHSPEC	Fish Counts by Species per Trawl
	FISHPATH	Fish Length and Pathology Data
	FSH_SPLN	Fish Spleen Exam Results
	TISSCHEM	Fish and Crab Tissue Chemistry Data
Code Tables:	ANALYTES	Chemical Analyte Code Table
	BEN_TAXA	Benthic Taxon Code Table
	FTAXON	Fish Taxon Code Table

2) Data Set Descriptions:

STATIONS - Sampling Station Location Data

This data set contains one record for each planned sampling station, with the planned latitude and longitude. Note that some of these stations were not successfully sampled and, therefore, are not represented in any other data set. The exact sampling location from each visit is recorded in the EVENTS data set, which may differ slightly from the planned locations. However, for analytical purposes, the station locations in the STATIONS data set should be used. The data set contains the following elements:

STATION	Station Name
STA_TYPE	Station Type
ORGSTATN	Sampling Organization Station Identifier
ORG_CODE	Sampling Organization Code
STA_SYS	Station System Code
ESTUARY	Estuary Name
STA_LAT	Latitude (decimal degrees) refer to datum NAD83
STA_LNG	Longitude (decimal degrees) refer to datum NAD83
STA_AREA	Station Surface Area (sq. km)
STASTATE	Station State Code
YEAR	Sampling Year

EVENTS - Station Visit Data

This data set contains one record for each sampling visit to a station. There may be multiple records per station. The actual latitude and longitude for a visit may differ slightly from the planned latitude and longitude included in the STATIONS data set. The two variables STATION and EVNTDATE make up the unique identifier for this data set. The field PARTNER identifies the organization responsible for field sampling. EVENT_ID is an alternative event identifier used by the partners in their own databases. It is included here only for documentation purposes.

STATION	Station Name
EVNTDATE	Date of Sampling Event
EVENT_ID	Sampling Event Identifier
DEPTH	Water Depth (meters)
PARTNER	Sample Collection Crew Identifier
EVNT_LAT	Latitude (decimal degrees) refer to datum NAD83
EVNT_LNG	Longitude (decimal degrees) refer to datum NAD83
YEAR	Sampling Year

WATRPHYS - Water Quality Data-Physical Measurements

This data set contains surface and bottom measurements of temperature, salinity, dissolved oxygen and pH collected in the field during sampling. At shallow stations, the surface and bottom values are identical - these records are identified by a QACODE value of “WTR-A”. See the metadata file for details. Secchi Depth measurements are also present. This data set contains one record per sampling event.

STATION	Station Name
EVNTDATE	Date of Sampling Event
SECCHI_D	Secchi Depth (meters)
SL_TEMP	Surface Layer-Temperature from CTD (deg. C)
SL_SAL	Surface Layer-Salinity from CTD (ppt)
SL_OXY	Surface Layer-Dissolved Oxygen from CTD (mg/l)
SL_PH	Surface Layer-pH (pH units)
BL_TEMP	Bottom Layer-Temperature from CTD (deg. C)
BL_SAL	Bottom Layer-Salinity from CTD (ppt)
BL_OXY	Bottom Layer-Dissolved Oxygen from CTD (mg/l)
BL_PH	Bottom Layer-pH (pH units)
QACODE	QA Qualifier Code
YEAR	Sampling Year

NUTRNTS - Water Quality - Nutrients Data

This data set contains water quality measurements derived from laboratory analyses of surface and bottom layer water samples collected in the field. At shallow stations, the surface and bottom values are from the same sample - these records are identified by a QACODE value of "NUT-A". See the metadata file for details. There is one record per sampling event for stations where nutrients samples were collected.

STATION	Station Name
EVNTDATE	Date of Sampling Event
B_SI	Bottom Dissolved Silica (mg/l as SI)
B_SI_D	Method Detection Limit for Bottom SI
B_NH4	Bottom Dissolved Ammonia (mg/l as N)
B_NH4_D	Method Detection Limit for Bottom NH4
B_NO23	Bottom Dissolved Nitrite & Nitrate (mg/l as N)
B_NO23_D	Method Detection Limit for Bottom NO23
B_NO2	Bottom Dissolved Nitrite (mg/l as N)
B_NO2_D	Method Detection Limit for Bottom NO2
B_PON	Bottom Particulate Organic Nitrogen (mg/l as N)
B_PON_D	Method Detection Limit for Bottom PON
B_TDN	Bottom Total Dissolved Nitrogen (mg/l as N)
B_TDN_D	Method Detection Limit for Bottom TDN
B_TDP	Bottom Dissolved Phosphorus (mg/l as P)
B_TDP_D	Method Detection Limit for Bottom TDP
B_PO4F	Bottom Dissolved Orthophosphate (mg/l as P)
B_PO4F_D	Method Detection Limit for Bottom PO4F
B_PHOS	Bottom Total Partic Phosphorous (mg/l as P)
B_PHOS_D	Method Detection Limit for Bottom PHOS
B_DOC	Bottom Dissolved Organic Carbon (mg/l as C)
B_DOC_D	Method Detection Limit for Bottom DOC
B_POC	Bottom Particulate Org Carbon (mg/l as C)
B_POC_D	Method Detection Limit for Bottom POC
B_TSS	Bottom Total Suspended Solids (mg/l)
B_TSS_D	Method Detection Limit for Bottom TSS
B_CHLA	Bottom Chlorophyll a (ug/l)
B_CHLA_D	Method Detection Limit for Bottom CHLA
B_PHAE	Bottom Phaeophytin (ug/l)
B_PHAE_D	Method Detection Limit for Bottom PHAE
S_SI	Surface Dissolved Silica (mg/l as SI)
S_SI_D	Method Detection Limit for Surface SI
S_NH4	Surface Dissolved Ammonia (mg/l as N)
S_NH4_D	Method Detection Limit for Surface NH4
S_NO23	Surface Dissolved Nitrite & Nitrate (mg/l as N)

S_NO23_D	Method Detection Limit for Surface NO23
S_NO2	Surface Dissolved Nitrite (mg/l as N)
S_NO2	Method Detection Limit for Surface NO2
S_PON	Surface Particulate Organic Nitrogen (mg/l as N)
S_PON_D	Method Detection Limit for Surface PON
S_TDN	Surface Total Dissolved Nitrogen (mg/l as N)
S_TDN_D	Method Detection Limit for Surface TDN
S_TDP	Surface Dissolved Phosphorus (mg/l as P)
S_TDP_D	Method Detection Limit for Surface TDP
S_PO4F	Surface Dissolved Orthophosphate (mg/l as P)
S_PO4F_D	Method Detection Limit for Surface PO4F
S_PHOS	Surface Total Particulate Phosphorous (mg/l as P)
S_PHOS_D	Method Detection Limit for Surface PHOS
S_DOC	Surface Dissolved Organic Carbon (mg/l as C)
S_DOC_D	Method Detection Limit for Surface DOC
S_POC	Surface Particulate Organic Carbon (mg/l as C)
S_POC_D	Method Detection Limit for Surface POC
S_TSS	Surface Total Suspended Solids (mg/l)
S_TSS_D	Method Detection Limit for Surface TSS
S_CHLA	Surface Chlorophyll a (ug/l)
S_CHLA_D	Method Detection Limit for Surface CHLA
S_PHAE	Surface Phaeophytin (ug/l)
S_PHAE_D	Method Detection Limit for Surface PHAE
LABCODE	Contract/Lab Identifier
QACODE	QA Qualifier Code
YEAR	Sampling Year

SEDGRAIN - Sediment Grain Size Data

This data set contains grain size data from the composite samples collected for sediment chemistry analysis. There is one record per event, for events where sediment samples were collected. TOC measures are not present for all events.

STATION	Station Identifier
EVNTDATE	Date of Sampling Event
SAND	Sand Content (%)
SILTCLAY	Silt/Clay Content (%)
MOISTURE	Moisture Content (%)
TOC	Total Organic Carbon (%)
LABCODE	Contract/Lab Identifier
QACODE	QA Qualifier Code
YEAR	Sampling Year

TOXICITY - Sediment Toxicity and Microtox Test Data

This data set contains summary results from two sediment toxicity tests: an ampelisca survival test and a microtox test.

STATION	Station Name
EVNTDATE	Date of Sampling Event
SRVPCCON	Ampelisca Survival as % of Control
SRVPC_SG	Ampelisca Survival - Statistical Significance
ATOX_SIG	Ampelisca Survival - Significance
EC50_MC	Microtox Moisture Corrected Mean EC50 (%)
MTOX_SIG	Microtox Test Significance
OE_EC50	Organic Extract Microtox Mean EC50
OE_SRI	Organic Extract Microtox Sed Ref Index
OE_SIG	Organic Extract Microtox Test Significance
LABCODE	Contract/Lab Identifier
QACODE	QA Qualifier Code
YEAR	Sampling Year

SEDCHEM - Sediment Chemistry Data

This data set contains sediment chemistry measures. There are multiple records for each event - one record for each concentration measured per station visit. Three variables make up the unique identifier for records in this data set: STATION, EVNTDATE, and ANALYTE.

A concentration value is provided for every analyte unless the concentration could not be detected by the lab instruments. In these cases, the detection limit is present (MDL), and the QACODE is set to "CHM-A". If the analyte was detected but at a level below the detection limit, the concentration is reported and the QACODE is set to "CHM-B". The detection limit is provided in this case also.

STATION	Station Name
EVNTDATE	Date of Sampling Event
ANALYTE	Code for Analyte Measured
CONC	Concentration of Analyte in Sample
CHMUNITS	Concentration Unit of Measure
MDL	Method Detection Limit
LABCODE	Contract/Lab Identifier
QACODE	QA Qualifier Code
YEAR	Sampling Year

BENGRAIN - Benthic Sample Grain Size Data

This data set contains grain size data taken from the benthic samples that are included in BEN_ABUN and BEN_BIOM. There is one record for each benthic grab - up to three per station. Three variables make up the unique identifier for records in this data set: STATION, EVNTDATE, and BENGRAB. Note that some MAIA partners did not supply grain size measurements for benthic samples. In these cases, an estimate of the grain size composition can be found in the SEDGRAIN data set.

STATION	Station Name
EVNTDATE	Date of Sampling Event
BENGRAB	Grab Associate with Grain Size Sample (#)
SAND	Sand Content (%)
SILTCLAY	Silt/Clay Content(%)
LABCODE	Contract/Lab Identifier
QACODE	QA Qualifier Code
YEAR	Sampling Year

BEN_ABUN - Benthic Abundance Data

This data set contains benthic abundance measurements from up to three grab samples per event.. It contains one record for each taxon found per grab. Four fields are needed to uniquely identify a record: STATION, EVNTDATE, BENGRAB, and TAXNCODE. The TAXNCODE can be used to look up taxonomic information, including the most current taxonomic name known for each taxon. The variable ID_LEVEL describes the level at which the organism was identified (SPECIES, GENUS, FAMILY, etc.) The field TAX_DSCR provides the partner's original identification of the taxon. This field may contain unusual spellings of taxonomic names, including ancillary information such as life stage (JUV, PUPAE, etc). The most accurate taxonomic name for these taxa is found in the look up table BEN_TAXA.

STATION	Station Name
EVNTDATE	Date of Sampling Event
BENGRAB	Grab Associate with Infauna Sample (#)
TAXNCODE	Taxonomic ID Code
ABUNDANC	Species Abundance in Sample (#)
ID_LEVEL	Level of Taxonomic ID
TAX_DSCR	Taxon Description (from Partner)
LABCODE	Contract/Lab Identifier
QACODE	QA Qualifier Code
YEAR	Sampling Year

BEN_BIOM - Benthic Biomass Data

This data set contains benthic biomass measurements from up to three grab samples per event. It contains one record for each taxon found per grab. (In some cases, species were grouped and weighed together in broader taxonomic groupings) Four fields are needed to uniquely identify a record: STATION, EVNTDATE, BENGRAB, and TAXNCODE. The TAXNCODE can be used to look up taxonomic information, including the most current taxonomic name known for each taxon. The field TAX_DSCR provides the partner's original identification of the taxon. This field may contain unusual spellings of taxonomic names, including ancillary information such as life stage (JUV, PUPAE, etc). The most accurate taxonomic name for these taxa is found in the look up table BEN_TAXA.

STATION	Station Name
EVNTDATE	Date of Sampling Event
BENGRAB	Grab Associated with Infauna Sample (#)
TAXNCODE	Taxonomic ID Code
BIOMASS	Species Ash Free Dry Weight in Sample (mg)
BM_ABUND	Abundance Contributing to Biomass
TAX_DSCR	Taxon Description (from Partner)
LABCODE	Contract/Lab Identifier
QACODE	QA Qualifier Code
YEAR	Sampling Year

BENSUMRY - Benthic Community Summary Data

This data set contains one record for each visit to a station where benthic samples were collected. These records summarize the benthic abundance and biomass findings from all samples (up to three) collected at the station. The field B_IND94B is an index of the overall benthic community based on EMAP analyses. An index value greater than zero indicates a healthy benthic community. A value less than zero indicates a degraded benthic community.

STATION	Station Name
EVNTDATE	Date of Sampling Event
A_SAMPS	Number of Grabs with Abundance Data
INF_ABU	Mean Abundance per Grab, All Infauna
EPI_ABU	Mean Abundance per Grab, All Epifauna
SPIONID	Spionid Polychaetes, Mean Abundance per Grab
TUBIFIC	Tubificid Oligochates, Mean Abundance per Grab
B_SAMPS	Number of Grabs with Biomass Data
MN_BIOM	Mean Biomass per Grab, All Species
TSINFCNT	Total Number of Infauna Species
TSEPICNT	Total Number of Epifauna Species
MSINFCNT	Mean Number of Infauna Species per Grab
MSEPICNT	Mean Number of Epifauna Species per Grab
SHANNON3	Shannon-Wiener Index- All Taxa
GLEASON3	Gleason's D- All Taxa
BOT_SAL	Bottom Salinity used in Benthic Index
PEXP_GL3	Percent Expected Gleason's D
PEXP_TUB	Percent Expected Tubificid Abundance
B_IND94B	EMAP VA Province Benthic Index
LABCODE	Contract/Lab Identifier
QACODE	QA Qualifier Code
YEAR	Sampling Year

FTRAWL - Standard Trawl Data

This data set contains one record for each standard fish trawl. A standard trawl is approximately 10 minutes in duration at 1 to 3 knots. Only one standard trawl was performed per station visit. Additional, non-standard trawls were often performed to catch fish for chemical and pathology exams but data from those trawls are not recorded in this data set. To compare fish communities, only standard trawls can be used. Two fields are needed to uniquely identify a record: STATION and EVNTDATE.

STATION	Station Name
EVNTDATE	Date of Sampling Event
FTRAWLID	Trawl Identifier
FTRL_DUR	Trawl Duration (minutes:seconds)
FTRL_SPD	Average Speed over Bottom (knots)
FSPECCNT	Number of Unique Species in Trawl
FISHCNT	Number of Individual Fish of all Species in Trawl
YEAR	Year of Sampling

FISHSPEC - Fish Counts by Species per Trawl

This data set contains one record for each species of fish caught in a standard trawl, and provides a count of individual fish, and their mean fork length. Four fields are needed to uniquely identify a record: STATION, EVNTDATE, FTRAWLID, and TAX_CODE.

STATION	Station Name
EVNTDATE	Date of Sampling Event
FTRAWLID	Trawl Identifier
TAX_CODE	Taxonomic Code
TAX_CNT	Number of Fish of This Species
AVG_LEN	Average Fork Length (mm)
YEAR	Year of Sampling

FISHPATH - Fish Length and Pathology Data

This data set contains one record for each of the first 30 individual fish of each species caught in each standard trawl, along with records for certain fish caught in non-standard trawls. Fish from non-standard trawls were saved for specific analytical purposes and not for the purpose of comparing fish community structures. The variable, TRLTYPE, indicates whether the fish came from a standard or non-standard trawl. The records provide fish fork lengths and gross pathology findings. Note that not every fish caught in the standard trawls was measured and recorded here. To find the total number of fish caught in standard trawls, the FISHSPEC data set must be used. Five fields are needed to uniquely identify a record: STATION, EVNTDATE, FTRAWLID, TAX_CODE, AND FSEQNUM.

STATION	Station Name
EVNTDATE	Date of Sampling Event
FTRAWLID	Trawl Identifier
TAX_CODE	Taxonomic Code
FSEQNUM	Fish Sequence Number
F_LENGTH	Fish Fork Length (mm)
LUMPS	Fish Pathology: Lumps (Y/N)
LUMPLOC	Locations of Lumps
GROWTHS	Fish Pathology: Growths (Y/N)
GRTHLOC	Locations of Growths
ULCERS	Fish Pathology: Ulcers (Y/N)
ULC_LOC	Locations of Ulcers
FINROT	Fish Pathology: Fin Erosion (Y/N)
FROTLOC	Locations of Fin Rot
TRLTYPE	Standard or Non Standard Fish Trawl (S/N)
YEAR	Year of Sampling

FSH_SPLN - Fish Spleen Exam Results.

This data set contains one record for each fish spleen examined in the histopathology lab. High Percent Macrophage Aggregates is an indicator of internal pathology. Five fields are needed to uniquely identify a record: STATION, EVNTDATE, FTRAWLID, TAX_CODE, and FSEQNUM.

STATION	Station Name
EVNTDATE	Date of Sampling Event
FTRAWLID	Trawl Identifier
TAX_CODE	Taxonomic Code
FSEQNUM	Fish Sequence Number
FSPL_PMA	Fish Spleen Percent Macrophage Aggregates.
YEAR	Year of Sampling

TISSCHEM - Fish and crab tissue chemistry data.

The data set contains one record for each analyte measured per tissue type (edible, inedible, and whole body; crab or flounder) at a station. Fish composite samples were made up of approximately two to ten summer flounder, caught at the same station. Crab composite samples were made up of two to ten blue crabs caught at the same station. One composite of only edible tissue, and one composite of all remaining inedible material were made from each group of fish or crab. The two composite samples (fish or crab) were analyzed separately. Whole body concentrations of the analytes found in the fish or crab were then computed as mass-weighted means of the edible and inedible results. Four fields are needed to uniquely identify a record: STATION, EVNTDATE, TISSUTYP, and ANALYTE.

A concentration value is provided for every analyte unless the concentration could not be detected by the lab instruments. In these cases, the concentration (CONC) is set to zero, the detection is provided (MDL), and the QACODE is set to "CHM-A". If the analyte was detected, but at a level below the detection limit, the concentration is reported and the QACODE is set to "CHM-B". If the analyte was detected, but there was some interference in the testing, the concentration is reported and the QACODE is set to "CHM-C". In all these cases, the detection limit is provided (MDL).

The QACODE for the whole body fish or crab is based on the QACODES of the edible and inedible composite samples. If they agree, the QACODE for the whole body record is the same. If the edible and inedible records are different, the whole body record is reported with a QACODE of "CHM-D".

Tissue type codes identify the type of tissue analyzed. These codes are FE (edible fish tissue), FI (inedible fish tissue), FW (whole body fish), CE (edible crab tissue), CI (inedible crab tissue, and CW (whole body crab). TAX_CODE identifies the species of the fish or crab analyzed.

The TISSCHEM data set contains the following elements:

STATION	Station Identifier
EVNTDATE	Date of Sampling Event
SAMPLEID	Sample Identifier
TAX_CODE	Taxonomic Code of Organism Analyzed
TISSUTYP	Tissue Type Code (FE, FI, FW, CE, CI, CW)
NUM_HOM	Number of Individuals in Homogenate
MN_WGHT	Mean Weight of Individuals in Homogenate
MN_SIZE	Mean Length (Fish) or Width (Crabs) in Homogenate (mm)
PCTMOIST	Percent Moisture Content of Tissue
PCTLIPID	Percent Lipid Content of Tissue
ANALYTE	Code for Analyte Measured
CONC	Concentration of Analyte in Sample
CHMUNITS	Concentration Units of Measure
MDL	Method Detection Limit
QACODE	QA Code
LABCODE	Contract/Lab Identifier
YEAR	Year of Sampling

FTAXON- Fish Taxon Code Table

This look-up table provides the common name, scientific name, and ITIS taxonomic serial number of the fish species identified in the data sets by TAX_CODE.

TAX_CODE	Fish Species-Taxon Code
FCOMNAME	Fish Species-Common Name
FSCINAME	Fish Species-Scientific Name
TSN	ITIS Taxonomic Serial Number

ANALYTES - Chemical Analyte Code Table

This look-up table provides the full chemical name and the CAS number of the analyte identified in the data sets by ANALYTE.

ANALYTE	Chemical Analyte Code
CHEMNAME	Full Chemical Name
CASNUM	CAS Number

BEN_TAXA - Benthic Taxon Code Table

This look-up table provides the full taxonomic name, the family tree, and the ITIS taxonomic serial number of the benthic organism identified in the data sets by TAXNCODE.

TAXNCODE	Taxonomic ID Code
TAXNAME	Taxon Name
TSN	ITIS Taxonomic Serial Number
PHYLUM	Phylum Name
CLASS	Class Name
ORDER	Order Name
FAMILY	Family Name
GENUS	Genus Name
SPECIES	Species Name

3) Key Fields

All data sets (except the Code tables) contain the field STATION, a station identifier which can be linked to the STATIONS table to find descriptive and location data for the sampling station. All data sets containing measurement data also contain the field EVNTDATE, which identifies the date the samples were collected.

The data sets EVENTS, WATRPHYS, NUTRNTS, SEDGRAIN, TOXICITY, FTRAWL and BENSUMRY, have, at most, one record per event. These data sets are in a horizontal structure - they have many fields on one record. The two fields STATION and EVNTDATE can be used to merge these data sets. (Not all events have data in each data set, so there may not be a one-to-one match.)

The data sets SEDCHEM, BENGRAIN, BEN_ABUN, BEN_BIOM, FISHPATH, FISHSPEC, FSH_SPLN, and TISSCHEM all have multiple records per sampling event. These data sets require additional key fields to uniquely identify each record.

The SEDCHEM and TISSCHEM data sets are in a vertical structure instead of having many chemical measurements on a single record (a horizontal structure). This structure is the most efficient way of storing these data.

SEDCHEM has a separate record for each analyte measured. Every analyte measured is identified by the field ANALYTE. This field can be used to look up the full name and CAS number of the chemical in the code table ANALYTES. The three fields, STATION, EVNTDATE and ANALYTE uniquely identify records in the SEDCHEM data set.

TISSCHEM has a separate record for each analyte measured for each type of fish/crab (edible, inedible, whole body). Each type of fish/crab is identified by the field TISSUTYP, with the analyte measured identified by the field ANALYTE. The field, TAX_CODE can be used to look up the full name and ITIS number of the species in the code table FTAXON. The four fields, STATION, EVNTDATE, TISSUTYP and ANALYTE uniquely identify records in the TISSCHEM data set.

The BENGRAIN data set can have up to three records per sampling event, since up to three different benthic samples were collected at many stations. Records in this data set are uniquely identified by the fields STATION, EVNTDATE, and BENGRAB.

The two benthic data sets BEN_ABUN and BEN_BIOM also contain the field BENGRAB. These data sets contain results of analyses of up to three benthic samples collected at a station. These data sets are also in a vertical structure - they each contain multiple records per grab sample. Each benthic taxon identified in a sample is recorded on a separate record. These two data sets have four key fields to uniquely identify a record: STATION, EVNTDATE, BENGRAB, and TAXNCODE.

Users should take care when merging records from the different data sets. The horizontally structure data sets with one record per STATION and EVNTDATE may be joined using these two fields. Fish data sets, FSH_SPLN, FISHSPEC, and FISHPATH should always be joined with at least four fields: STATION, EVNTDATE, FTRAWLID, and TAX_CODE. FISHPATH and FSH_SPLN should be joined using these four plus a fifth field, FSEQNUM. Benthic data sets (including BENGRAIN) should always be joined with at least three fields: STATION, EVNTDATE, and BENGRAB. BEN_ABUN and BEN_BIOM should be joined using these three plus a fourth field, TAXNCODE. (Note that there is not always a one-to-one match).

Records in the SEDCHEM data set may be joined to the ANALYTES table on the field, ANALYTE to get a descriptive name and the CAS number for analytes measured.

Records in BEN_ABUN and BEN_BIOM can be joined to the BEN_TAXA table on the field, TAXNCODE to get taxonomic information on benthic organisms.

Records in FISHSPEC, FISHPATH, FSH_SPLN, AND TISSCHEM can be joined to the FTAXON table on the field, TAX_CODE to get taxonomic information on fish and crab.

4) Lab Codes and QA Codes

All data sets with measurement data contain the field QACODE. This field is blank if there are no quality assurance issues associated with the data. The field may contain one or more codes to document QA issues. QA qualifier codes are described in detail in the metadata files. Data with significant QA problems were excluded from this database. All data in these data sets have been reviewed and deemed acceptable for MAIA analyses.

All data sets with laboratory measurements contain the field LABCODE. This code identifies the laboratory and/or the contracting partner responsible for the analyses. Any differences in analytic methods used by different labs are documented in the metadata files. The LABCODES are useful in explaining missing values - the different MAIA partners did not all have the same analyses performed. Only data elements that are comparable across different records are included in the database.