



ANIMIDA III
 Sampling Results Database
 Entity Relationship Diagram
 Revised: 2016-11-21
 By Tim Whiteaker and Xing Zheng
 The University of Texas at Austin

DATA DICTIONARY

This data dictionary describes the tables representing sampling results from the ANIMIDA III Study. The project uses a modified version of the CUAHSI Observations Data Model (ODM) version 1.1 (<http://his.cuahsi.org/odmdatabases.html>) as its database schema. The schema is modified to include a Taxonomy table for associating sample values with a classification of an organism from which a given sample was taken or to which a given sample applies. CUAHSI provides detailed documentation on the tables in ODM as well as rationale and example uses of this database design; see the design specifications from the page linked above, or use the direct link: <http://his.cuahsi.org/documents/ODM1.1DesignSpecifications.pdf>. The reader is encouraged to view the ODM design specifications if further information beyond the descriptions provided in the sections that follow is desired for tables and fields used in this project database.

To improve ease-of-use of the database, which is implemented in Microsoft SQL Server version 2008 R2, data are exported to more widely used formats including ArcGIS file geodatabase and comma-separated values (CSV) text file. The exported version of the database joins key fields from several ODM tables together to reduce the number of tables that the user must work with, and to also eliminate the need to deliver empty tables that are not utilized by this project in ODM. For example, the DataValues table described below includes the quality control level definition from the ODM QualityControlLevels table, eliminating the need to include that table in the output. Furthermore, several fields from related tables are duplicated in the DataValues table purely for convenience of the end user, e.g., the SiteName field from the Sites table. The idea is to include some basic metadata for each data value in a single table, and let the user explore relationships to other tables if more information is desired. This makes it easier for users who are not familiar with database concepts like relationships to utilize this dataset. Thus, the design of this product should be familiar enough to users experienced with ODM, while still being accessible to those unfamiliar with ODM.

The description of tables for the preceding entity relationship diagram is provided in alphabetical order in the sections below, followed by a section that describes the relationships between these tables.

For questions or comments about the project entity relationship diagram or data dictionary, or about the project database, please contact:

Dr. Tim Whiteaker
The University of Texas at Austin
whiteaker@utexas.edu
512-471-0570

DATAVALUES

This table contains numerical values representing sampling results.

Field	Description	Type
SiteID	References the site at which the observation was measured. This links data values to their locations in the Sites table.	Long Integer
VariableID	References the variable that was measured. This links data values to their variable in the Variables table.	Long Integer
TaxaID	References the taxonomic classification of the organism (if any) associated with the measurement. This links data values to their taxonomic record in the Taxonomy table.	Long Integer
DataValue	The numeric value of the observation.	Double
Qualifier	A data qualifying comment that may accompany the data value.	Text
CensorStatus	Indicates whether the data value is censored, e.g., below the detection limit of the sensor.	Text
OffsetValue	Distance from a datum or control point to the point at which a data value was observed. If not given, the OffsetValue is inferred to be 0, or not relevant/necessary.	Double
OffsetUnits	Units of measure used for the offset.	Text
OffsetDescription	Describes the offset type.	Text
SiteName	Full text name of sampling site from the Sites table, duplicated here for convenience.	Text
Latitude	Latitude in decimal degrees from the Sites table, duplicated here for convenience.	Double
Longitude	Longitude in decimal degrees from the Sites table, duplicated here for convenience.	Double
VariableCode	Unique variable code for the variable from the Variables table, duplicated here for convenience.	Text
VariableName	Name of the variable from the Variables table, duplicated here for convenience.	Text
Units	Units of measure used for the data value.	Text
SampleMedium	The medium of the sample from the Variables table, duplicated here for convenience.	Text
LocalDateTime	Local date and time at which the data value was observed.	Date
ValueAccuracy	Numeric value that describes the measurement accuracy of the data value. If not given, it is interpreted as unknown.	Double
MethodID	References the method used to generate the data value. This links data values to their methods in the Methods table.	Long Integer
LabSampleCode	Text code assigned to the sample (if any) associated with the data value.	Text
SampleType	Controlled vocabulary specifying the sample type.	Text
LabMethodID	References the lab method used to analyze the sample. This links data values to their lab method in the LabMethods table.	Long Integer
QualityControlLevel	Level of quality control that the value has been subjected to.	Text
SourceID	References the record in the Sources table giving the source of the data value.	Long Integer
ContactName	The contact name from the Sources table, duplicated here for convenience.	Text

LABMETHODS

This table contains descriptions of the laboratory methods used to analyze physical samples for specific constituents.

Field	Description	Type
LabMethodID	Unique integer identifier for each laboratory method.	Long Integer
LabName	Name of the laboratory responsible for processing the sample.	Text
LabOrganization	Organization responsible for sample analysis.	Text
LabMethodName	Name of the method and protocols used for sample analysis.	Text
LabMethodDescription	Description of the method and protocols used for sample analysis.	Text
LabMethodLink	Link to additional reference material on the analysis method.	Text

METHODS

This table contains the methods used to collect the data and any additional information about the method.

Field	Description	Type
MethodID	Unique integer ID for each method.	Long Integer
MethodDescription	Text description of each method.	Text
MethodLink	Link to additional reference material on the method.	Text

SITES

This table provides information giving the spatial location at which data values have been collected. The table is intended to be visualized as a point feature class in a geographic information system, i.e., a set of point locations in a map.

Field	Description	Type
SiteID	Unique identifier for each sampling location.	Long Integer
SiteName	Full name of the sampling site.	Text
Latitude	Latitude in decimal degrees.	Double
Longitude	Longitude in decimal degrees. East positive, West negative.	Double
PositionalAccuracy_meters	Value giving the accuracy with which the positional information is specified in meters.	Double
Comments	Comments related to the site.	Text
Datum	Spatial Reference System of the latitude and longitude coordinates.	Text

SOURCES

This table describes the source organization and contact which originated the data values.

Field	Description	Type
SourceID	Unique integer identifier that identifies each data source.	Long Integer
Organization	Name of the organization that collected the data.	Text
SourceDescription	Full text description of the source of the data.	Text
SourceLink	Link for more information about the data source.	Text
ContactName	Name of the contact person for the data source.	Text
Phone	Phone number for the contact person.	Text
Email	Email address for the contact person.	Text
Address	Street address for the contact person.	Text
City	City in which the contact person is located.	Text
State	State in which the contact person is located.	Text
ZipCode	US Zip Code or country postal code.	Text
Citation	Text string that gives the citation to be used when the data from each source are referenced.	Text
SourceFile	Name of the original data file submitted by the scientist from which data values were loaded	Text

TAXONOMY

This table contains the taxonomy information to describe organisms in the study.

Field	Description	Type
TaxaID	Unique ID for each taxonomic classification.	Long Integer
TSN	Taxonomic Serial Number, from itis.gov. Negative TSN values are used when an appropriate TSN could not be obtained from itis.gov.	Long Integer
Kingdom	Scientific Kingdom name.	Text
Subkingdom	Scientific Subkingdom name.	Text
Infrakingdom	Scientific Infrakingdom name.	Text
Superphylum	Scientific Superphylum name.	Text
Phylum	Scientific Phylum name.	Text
Subphylum	Scientific Subphylum name.	Text
Class	Scientific Class name.	Text
Subclass	Scientific Subclass name.	Text
Infraclass	Scientific Infraclass name.	Text
Superorder	Scientific Superorder name.	Text
Order_	Scientific Order name.	Text
Suborder	Scientific Suborder name.	Text
Infraorder	Scientific Infraorder name.	Text
Superfamily	Scientific Superfamily name.	Text
Family	Scientific Family name.	Text
Genus	Scientific Genus name.	Text
Subgenus	Scientific Subgenus name.	Text
Species	Scientific Species name.	Text
Subspecies	Scientific Subspecies name.	Text
CommonName	Common name.	Text
Synonyms	Common synonyms.	Text
TaxaLink	Hyperlink to the taxa report on itis.gov or other authoritative website.	Text
TaxaComments	Comments on the taxonomic classification.	Text

VARIABLES

This table contains the full descriptive information about what variables have been measured.

Field	Description	Type
VariableID	Unique integer identifier for each variable.	Long Integer
VariableCode	Text code used by the organization that collects the data to identify the variable.	Text
VariableName	Full text name of the variable that was measured, observed, modeled, etc.	Text
Speciation	Text code used to identify how the data value is expressed (e.g., total phosphorus concentration expressed as P).	Text
UnitsName	Full text name of the units of the data values associated with the variable.	Text
SampleMedium	The medium in which the sample or observation was taken or made.	Text
ValueType	Text value indicating what type of data value is being recorded.	Text
DataType	Text value that identifies the data values as one of several data types.	Text
GeneralCategory	General category of the data values.	Text
NoDataValue	Numeric value used to encode no data values (i.e., null or missing values) for this variable.	Double

RELATIONSHIPS

These relationships define ancillary information about data values.

Name	Table	Field	Type	Field	Table
<i>Locate</i>	DataValues	SiteID	* <-> 1	SiteID	Sites
<i>Characterize</i>	DataValues	VariableID	* <-> 1	VariableID	Variables
<i>Create</i>	DataValues	MethodID	* <-> 1	MethodID	Methods
<i>Analyze</i>	DataValues	LabMethodID	* <-> 1	LabMethodID	LabMethods
<i>Identify</i>	DataValues	TaxaID	* <-> 1	TaxaID	Taxonomy
<i>Generate</i>	DataValues	SourceID	* <-> 1	SourceID	Sources