California Freshwater Species Database, Version 2.0.7

Kirk Klausmeyer[1], Kurt Fesenmyer[2], Jeanette Howard[1], Scott Morrison[1]

The Nature Conservancy, California

May 21st, 2015

Abstract

This document describes the California Freshwater Species Database, Version 2.0.7, and the process steps, components and sources used for its creation. We hope these data will be a useful resource for studying the distribution of freshwater biodiversity and improving conservation efforts in California.

Usage and Restrictions

This database is provided for informational purposes "as-is" and without warranty of any kind. The Nature Conservancy (TNC) is not responsible for the accuracy, completeness, quality or legal sufficiency of this information, nor is TNC responsible for updating the information set forth in this map.

Whenever listing the element names for a given geography, please include the original data provider source citation as referenced in the “AU_v_Elm” and “Sources” tables (see below). This database may be redistributed as long as the original source data is included.

The database is a snapshot of observations submitted to various sources described in the “Sources” table. These data have been compiled and reviewed as of April 13, 2015 and serve as an update to previous versions of the database, including the *California Freshwater Species Database, Version 1.0.* Published results using these data should be cited as follows:

Kirk Klausmeyer, Kurt Fesenmyer, Jeanette Howard, Scott Morrison. *California Freshwater Species Database, Version 2.0.7.* The Nature Conservancy, California and Trout Unlimited. 2015. San Francisco, California.

Introduction

[1] The Nature Conservancy, 201 Mission St., 4th Floor, San Francisco, California, 94105.
[2] Trout Unlimited, 910 W. Main Street, Suite 342, Boise, ID 83702 USA.

Californians face profound decisions regarding the management of their state's increasingly limited water supply. Information about the plants and animals that also rely on California's freshwater resources to survive is critical for decision-making. To provide that information, The Nature Conservancy created the California Freshwater Species Database —a comprehensive spatial database of California's freshwater species compiled from numerous disparate sources. The database is now available to help improve conservation efforts across California. For more information, please visit http://www.scienceforconservation.org/projects/freshwater.

Database Organization

The California Freshwater Species Database consists of several related tables provided in comma separated values (.csv) format, and a geospatial analysis unit file in ESRI shapefile (.shp) format. Since these data are relational, the individual files need to be imported to a database management system of your choice. We used ESRI's file geodatabase format for processing and compilation, however we deliver as csv and shp formats to provide maximum flexibility of use in other software packages. Figure 1 provides a graphical representation of the tables and the relationships between the tables. Appendix 1 provides a description of each table and field in the tables.

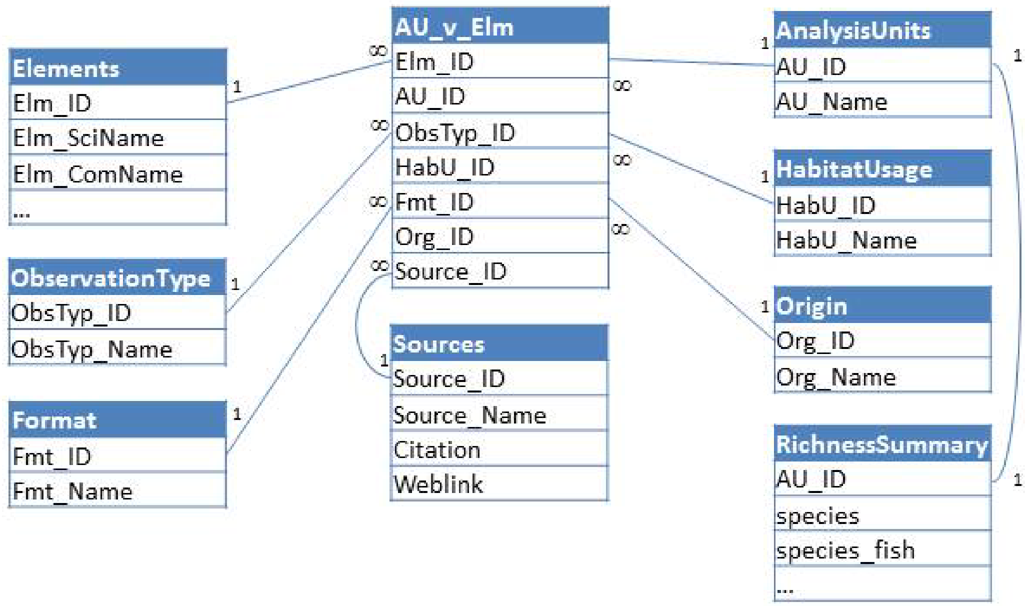

Figure 1: Entity Relationship Diagram for California Freshwater Species Database.

The primary table "AU_v_Elm" links the observation of freshwater life elements ("Elm", which include taxa at the family, genera, species, sub-species, or variety level) to an analysis unit ("AU"). This table provides information about the observation data, including the type of observation (e.g., post 1980 observation, modeled habitat, range data), the source of the original data, and the geometry of the original data (i.e., point, lines, or polygons). This table also provides information (where available) about how the element uses the freshwater habitats (e.g., spawning, migration) and the origin of the element (e.g., if the observation is within its native range or has been introduced outside its native range). Note that this table can include multiple records for one element in one analysis unit if the observation was recorded from several data sources or if the element uses the habitats in the analysis unit during different life stages.

A richness summary table, "RichnessSummary" is provided that includes summaries of the count of elements by the analysis unit watersheds. Different categories of richness are provided, including the count of all species, endemic species, fish species, etc. A full list of the summaries provided is included in Appendix A.

The remaining tables are lookup tables to provide more information about the codes in the "AU_v_Elm" table. The "Elements" table is the largest, as it contains information for all of the freshwater life elements that are native to the study area according to NatureServe as of February, 2011 and further refined and expanded after extensive review by taxonomic experts. This table contains records at the sub-species or variety level, species level, genus level, or family level. The remaining tables and fields are described in Appendix A.

The database also includes an Esri format shapefile, "AnalysisUnits", which covers the entire study area and includes all of the analysis units. The coordinate system of this shapefile is "NAD_1983_California_Teale_Albers"; WKID: 3310; Authority: EPSG; linear unit: Meters. Shapefiles can be viewed in any number of free software packages (e.g. Esri ArcGIS Explorer, QGIS, uDIG, R, etc). The analysis units were extracted from the 12-digit hydrologic unit code watersheds from the Watershed Boundary Dataset (USDA - NRCS et al. 2009). The shapefile contains metadata in the "AnalysisUnits.shp.xml" file. The "AU_ID" or analysis unit ID is a 12-digit code that is made up from 6 nested 2-digit codes. By removing 2-characters from the right of the code, you can generate larger watersheds and hydrologic units to aggregate the data at larger scales.

Use Examples

(Note: These examples were written for Esri ArcGIS Desktop users, however they should be replicable in any number of GIS software packages.)

1. Count the Freshwater Taxa or Species in California. The "Elements" table lists all the freshwater elements known in California, but it includes several taxonomic levels of organization. The total count of rows in the table is 4,774 (excluding the header). This includes 868 records at the genus level. Exclude these records by filtering out all rows with "Species_ID" = -999. This results in 3,906 species, subspecies, Evolutionary Significant Units, and Distinct Population Segments (hereafter reffered to as "Taxa"). Multiple taxa can be included in one species. To get a list of the unique species in California, filter "Species_ID" and remove duplicates. This can be done in MS Excel by creating a pivot table from the Elements table, and dragging "Species_ID" to the Row Labels box. This will result in 3,727 unique species in the database.

2. List Elements in a Watershed. In order to get a list of the freshwater elements observed in a watershed, open the “AnalysisUnits” shapefile in ArcMap to determine the AU_ID for the watershed of interest. Add the “AU_v_Elm”, “Elements”, and “ObservationType” tables to ArcMap. Use a definition query to limit the records in “AU_v_Elm” to the AU_ID of interest. Then join the “Elements” table to “AU_v_Elm” based on Elm_ID and the “ObservationType” table based on ObsTyp_ID. Use additional queries to filter out elements or observation types you are not interested in (e.g., historical observations, range data, plants). Then use the summarize tool to get a list of the elements of interest by scientific or common name.

3. Make a map of species richness.

OPTION 1: The “RichnessSummary” table includes a variety of species counts at the watershed scale. Add “AnalysisUnits” and “RichnessSummary” to ArcMap and then join the table to the shapefile using the “AU_ID” field. Then symbolize the map based on the “species” field.

OPTION 2: In order to calculate your own richness summaries, import the “AU_v_elm” and “Elements” csv tables into a file geodatabase and “add them to ArcMap (the options for working with .csv files in ArcMap are limited). Join the “Elements” table to the “AU_v_elm” table using “ELM_ID” field. Select out observations of species inside their native range (AU_v_elm.Org_ID = 0) and remove genus level records (Elements.Species_ID is not -999). Use the “Frequency” geoprocessing tool to create a temporary table that provides one record for each element and AU_ID combination by dissolving on AU_ID and Species_ID (note that this step is necessary since the same species could be noted in a given watershed by multiple sources). Use the “Frequency” tool again on the temporary table, and dissolve on AU_ID only. Add the “AnalysisUnits” shapefile to ArcMap, and join the new table to the AnalysisUnits and use the resulting count to generate a map of the richness of elements by watershed.

Database Generation Methods

Study Area: We used the state boundaries of California to define our study area. We selected the 12-digit HUC watershed in the Watershed Boundary Dataset (n=4,450 in the study area) as the unit of analysis for this assessment because it was the smallest scale watershed available in this nested national dataset.

Taxa List: We collaborated with NatureServe (http://natureserve.org/) to generate an annotated list of all elements that are known to occur in California. Because NatureServe collects and manages information for only a subset of species, we supplemented that list with regional and taxonomic reviews and checklists. Experts reviewed that list to identify the taxa which require freshwater habitats for some or a portion of their life cycle and generated a final list of freshwater dependent taxa. We further attributed each element with information related to higher level taxonomy, endemism in California, and conservation status (Appendix A). Additional details on the generation of the taxa list and sources are provided in a peer-reviewed scientific journal article currently in review.

Spatial Data Collection: We acquired spatial data (GIS format or tabular data easily converted to GIS format) related to the occurrence or distribution of target elements primarily through web searches for original data. In most cases, we were able to download data directly over the web. Other data sources required specific requests to individuals, particularly for data referenced in scientific literature.

Data Management: Once we identified a data source, we used a series of general GIS data management steps to insert the appropriate data into the geodatabase. We downloaded primary data and archived it by source in its native format. We then joined these original data to the Elements table by element name to identify the records within the source data that corresponded to species on the assessment target list. Next, we added a new field to the source data and attributed this field to equal the unique element ID within the Element table. Since some scientific and common names have changed over time for an individual species, we repeated this join and attribute calculation for all available names in the source data and Element table. Similarly, we added new fields to the original data source related to habitat use type, observation type, and format type and attributed these corresponding to information contained in the original data. These newly calculated fields directly corresponded to the geodatabase schema and allowed us to append the new data to the appropriate geodatabase feature classes. We also appended additional information from the original data, including data quality, accuracy, and any original identification or coding information, and added a field to identify the source data; the "Sources" table lists data sources in order of the unique source ID assigned. After we added all the data to the geodatabase, we intersected each of the three feature classes (point, line, and polygon) with the AnalysisUnits feature class and summarized the occurrence of elements in each subwatershed to generate the AU_v_elm table.

Appendix A: Table and Field Name Descriptions

Table/*Field Name* *Description*

AU_v_elm	**Table that lists all observations by analysis unit**
Elm_ID	Unique ID of life element
AU_ID	Analysis unit ID
ObsTyp_ID	Observation Type ID
HabU_ID	Habitat Usage ID
Fmt_ID	Format ID
Org_ID	Origin ID
Source_ID	Source ID

Elements	**List of the life elements described in the study area (list is not exhaustive)**
ELM_ID	Unique ID to life element (see ElmID sheet for breakdown)
ELM_SCINAM	Scientific name (G_name in NatureServe)
ELM_COMNAM	Common name
GROUP_	Taxonomic grouping
TAX_LIST	Taxonomic list(s) providing element
G_Rank	Global conservation status rank (From NatureServe)
S_Rank	California conservation status rank (From NatureServe)
ELM_SCIN_1	First alternate scientific name
ELM_SCIN_2	Second alternate scientific name
ELM_SCIN_3	Third alternate scientific name
ELM_SCIN_4	Fourth alternate scientific name
Kingdom	Kingdom
Phylum	Phylum
TaxClass	Class
TaxOrder	Order
Family	Family
Genus	Genus
Species	Species
Subsp_Var	Sub-species, Variety, Evolutionary Significant Unit, or Distinct Population Segment
Kingdom_ID	ID for each unique Kingdom
Phylum_ID	ID for each unique Phylum
TaxClass_I	ID for each unique Class
TaxOrder_I	ID for each unique Order
Family_ID	ID for each unique Family
Genus_ID	ID for each unique Genus (-999 for elements at the family level)
Species_ID	ID for each unique Species (-999 for elements at the genus or family level)

Elements (con't)	List of the life elements described in the study area (list is not exhaustive)
Fed_list	Status on Federal Endangered Species List as of April 13, 2015
State_list	Status on California Endangered Species or Sensitive Species lists as of April 13, 2015
Other_list	Status on other sensitive species lists as of April 13, 2015
MgtAg_list	Status on land management agency (USFS, BLM) sensitive species lists as of April 13, 2015
Listed	Present on Federal or California Endangered Species lists as endangered or threatened (1 = yes)
Vulnerable	Present on Federal, State, other, or management agency status list, but not federally or state listed as endangered or threatened (1 = yes)
Endemic	Endemic to California (1= yes)
Common	Common or not vulnerable (1 = yes)
Not_evaluated	Not evaluated for vulnerability (1 = yes)
Extinct	Extinct (1 = yes)
Status	Text description of status

HabitatUsage	If provided, the types of uses the element makes of the habitat
HabU_ID	Habitat Usage ID
HabU_Name	Habitat Usage Name

ObservationType	The type of the field observation of the element
ObsTyp_ID	Observation Type ID
ObsTyp_Name	Observation Type Name

Format	The format of the geospatial data provided by the source
Fmt_ID	Format ID
Fmt_Name	Format Name

Sources	List of all the sources used to assemble the database
Source_ID	Source ID
Source_Name	Short name for source
Citation	Full citation (in general, author, publication year, full name of source, publisher, agency, date accessed, etc.)
Weblink	Hyperlink for where data were accessed

Origin	The origin of the observation of the element
Org_ID	Origin ID
Org_Name	Descriptor of the origin of observation

AnalysisUnits	Shapefile of the analysis units in the study area
AU_ID	Analysis unit ID, equivalent to the Watershed Boundary Dataset HUC12
AU_Name	Analysis unit Name

RichnessSummary	Summary of element richness by analysis unit
HUC_12	Analysis unit ID (Same as AU_ID)
species	All Species
species_fish	Fishes
species_crust	Crustaceans
species_herps	Herps
species_inverts	Insects & other inverts
species_mollusks	Mollusks
species_plants	Plants
species_birds	Birds
species_mammals	Mammals
species_mollusks_crust	Mollusks and Crustaceans
species_endemic	Endemic
species_endemic_fish	Endemic Fishes
species_endemic_crust	Endemic Crustaceans
species_endemic_herps	Endemic Herps
species_endemic_inverts	Endemic Insects & other inverts
species_endemic_mollusks	Endemic Mollusks
species_endemic_plants	Endemic Plants
species_endemic_birds	Endemic Birds
species_endemic_mammals	Endemic Mammals
species_vulnerable	Vulnerable
species_listed	Listed
species_endemic_vulnerable	Endemic Vulnerable
species_endemic_listed	Endemic Listed
genus	All Genera
family	All Families
tax_order	All Orders
tax_class	All Classes
phylum	All Phyla
species_current	Current Observations
species_historical	Historical Observations
species_other	Other Observations

References

USDA - NRCS, USGS, and US EPA. 2009. Watershed Boundary Dataset. http://www.nrcs.usda.gov/wps/portal/nrcs/main/national/water/watersheds/dataset/.