**README: Figshare**

*Created by Kaija Gahm on 17 August 2020*

This folder is intended to be a mirror of the contents of the Figshare repository, beginning with the August 2020 Figshare update.

**Basic Structure**

The MFE database has a hierarchical structure (see the schema). The highest-level table is LAKES, followed by SITES, \*\_SAMPLES (SAMPLES, FISH\_SAMPLES, etc.), and then the tables containing data/measurements. Each data table is linked to the PROJECTS table, which describes the project for which the samples were collected. Both the SAMPLES table and the data tables are linked to the METADATA table. The ISOTOPE\_SAMPLES tables (red outlines on the schema) are on the same level as the other data tables and are linked to the SAMPLES table. The ISOTOPE\_RESULTS table is linked only to each of the ISOTOPE\_SAMPLES tables. The ISOTOPE\_BATCHES table links to the ISOTOPE\_RESULTS table.

Fish data (blue outlines on the schema) and creel survey data (yellow outlines on the schema) are somewhat separated. There is a FISH\_SAMPLES table, which contains info about the fish sampling event (lake, date, time, gear, etc.)––this is the equivalent of the SAMPLES table, but specifically for fish. The other fish tables are linked to FISH\_SAMPLES (with the exception of FISH\_YOY, which is kind of off doing its own thing––this table will be more consistently linked to the others in a future database update.) Similarly, the CREEL\_SAMPLES table gives details for each unique creel sampling/surveying event, with the other CREEL tables linked to it. Creel surveys were performed as part of the Fishscapes project, to understand “the social-ecological dynamics of recreational fishery landscapes.”

**Other Tables**

You should also be aware of some of the non-primary tables, since they’re important to navigating through the database.

* OTU: gives operational taxonomic units (otu's) for organisms whose data show up in the database. An operational taxonomic unit, in this case, is the most specific taxonomic level to which the organism could be identified.
* METADATA contains metadata information about sampling schemes, experimental procedures, etc., as well as pointing to file names when there is a separate metadata file. Metadata files are available in the “Metadata Files” folder. Each metadata file corresponds with a unique metadataID in METADATA. *NOTE: while the metadataID is sometimes the same as the name of the corresponding metadata document, this is not always the case. If you’re trying to find the metadata document that corresponds to a given metadataID, make sure you look up the correct fileName in METADATA.*
* UNITS lists column/variable names for all database tables—it is the “data dictionary” for the database. The file “columnNames\_allDatabaseTables.csv” is the same as UNITS, just in CSV format for slightly easier viewing.
* UPDATE\_METADATA describes updates made to various database tables, including the date, updater initials, and details about what was changed and why. Each database table contains an “updateID” column, corresponding to the data in UPDATE\_METADATA. *NOTE: for brevity, the updateID column is not listed in UNITS.*
* PROJECTS gives a general overview of projects for which the database data were collected, and the names of the people in charge.

**Database Schema**

The Google Slides presentation used to make the database schema is editable here: <https://docs.google.com/presentation/d/1SzoGPQkLM8qGpXo0WuKjdDhSu45BK0FARBA3sHauno0/edit?usp=sharing>. This is heavily based on Alex Ross’s previous database schema, a version of which is stored in his archive as a PDF and a Powerpoint, but the Powerpoint file there was corrupted, so I have recreated the schema from scratch.

There are two versions of the schema. “dbSchemaKeys.20200820.pdf” shows the primary and foreign keys of each table, with arrows and colors indicating their relationships. “dbSchemaDetails.20200820.pdf” provides brief descriptions of what each table contains, still showing the relationships between the tables with arrows.

Italicized variable names in “dbSchemaKeys.20200820.pdf” are primary keys. If there are two italicized variables in one table, then they form a composite primary key. Non-key variables are not shown; they can be viewed in UNITS.