

Methods for Monitoring Fish Communities of Buffalo National River and Ozark National Scenic Riverways in the Ozark Plateaus of Arkansas and Missouri, Version 2.0

Standard Operating Procedure (SOP) 9: Data Management, Version 2.0

Revision History Log:

Previous version number	Revision date	Author	Changes made	Reason for change	New version number
1.0	6/17/2016	Rowell, Dodd, Hinsey, Williams	Revisions in database and Quality Assurance Plan	Inventory and Monitoring Division standards (2015-16)	2.0

This SOP explains procedures for data management of BUFF and OZAR fish community monitoring data. The fish community database (called HTLNFish) described in this SOP was created from modifications of the river fish database (RMFish) described in SOP 9 of Petersen *et al.* (2008) and combining this river fish database with the stream fish database (STFish) described in SOP 8 of Dodd *et al.* (2008). This SOP includes a general description of the design and procedures for data recording, data entry, data verification and validation, and data integrity for the primary fish sampling data. More detailed step-by-step procedures on data entry will be kept on file in the HTLN Aquatics Laboratory located at Missouri State University, Springfield, Missouri.

Data management can be divided into (1) the initial design phase that involves defining the data model, its entities and their relations and (2) the procedures necessary to implement the database. Microsoft (MS) Access is the primary software used for maintaining fish data. Water quality data will be stored in the National Park Service (NPS) water quality database (see Water Quality SOP 4). Environmental Systems Research Institute (ESRI) ArcGIS is used for managing spatial data associated with field sampling locations. Data products derived from this project will be available at the NPS Integrated Resource Management Applications (IRMA) Data Store and EPA water quality National Data Warehouse. QA/QC guidelines in this document are based on the NPS Inventory and Monitoring document *Certification Guidelines for Inventory and Monitoring Data Products* (NPS 2016).

Design

The data model and relational database system for this project was created several years before the development of requirements and standards for monitoring data systems at the NPS Inventory and

Monitoring Division (Frakes *et al.* 2015), however the design follows the earlier Natural Resource Database Template design (NPS 2006).

Generally, the database model is developed from the appropriate field forms based on the relevant fish sampling protocols. Prior to table development, the data manager coordinates with project staff to determine data types (byte, text, numeric, for example), precision, and range of values. Acquiring field data forms assists with developing data logic (that is, how data components relate) and identifying indexes and primary keys.

A generalized model of the fish community database includes two primary tables for sampling events and locations (Figure 1 and Table 1). These two core tables contain general information pertaining to the field sampling occasion (the when and where of the sample). This includes information such as date and time, reach ID and Universal Transverse Mercator (UTM) coordinates, and park/project codes. The fish community tables serve as the organizing hub for fish data. Other tables primarily address habitat, weather or water quality conditions. See Figure 1 and Table 1 for a cross reference of the fish field forms to the corresponding tables.

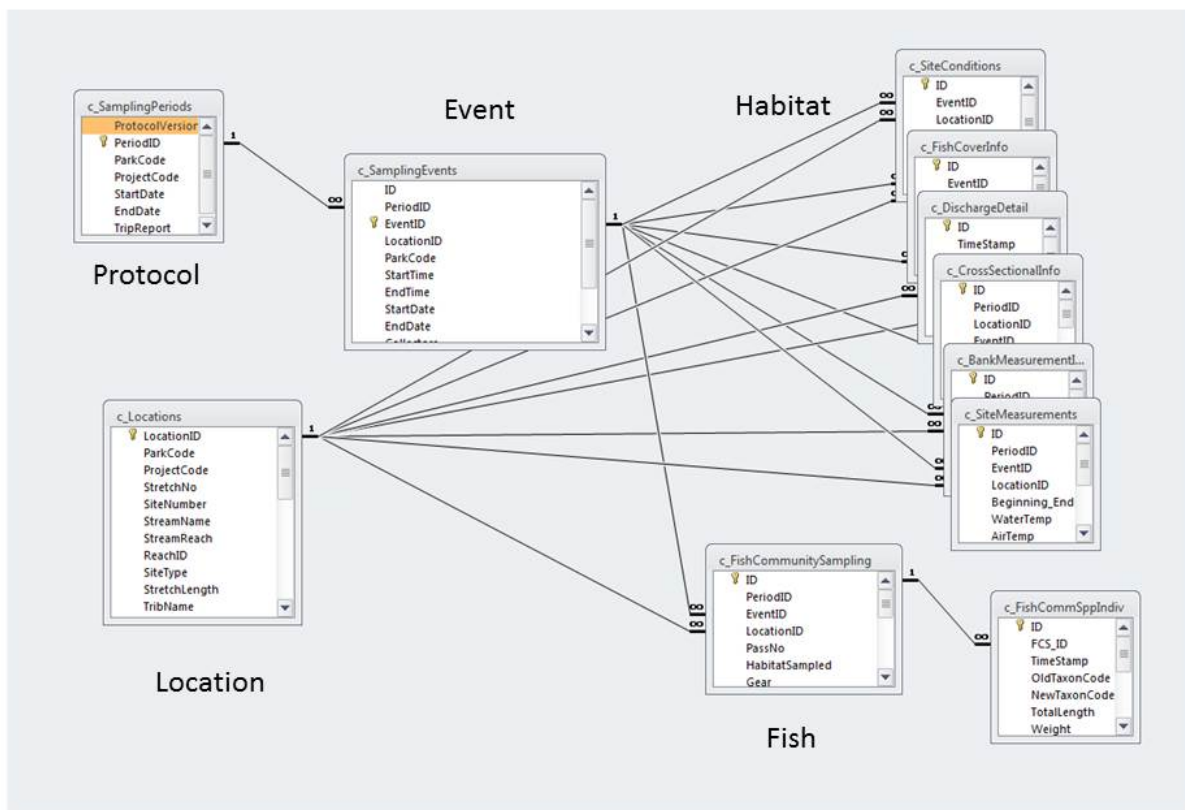


Figure 1. Entity relationship diagram of the HTLNFish Database. Project data tables consist of fish data and associated habitat data.

Data Type	Fish Field Form	Form Reference	Data Description	Fish Database Table Name
Protocol	Determined based on the dates of all events during the sample season	n/a	Beginning and ending dates of the sample season	c_SamplingPeriods
Event	Determined based on the first date of the sample event	n/a	A specific date a sample event occurred	c_SamplingEvents
Location	Determined based on original GRTS analysis	Narrative, Table 1 and App 3	A specific location (site) of a sample event	c_Locations
Fish	Fish Community Data	SOP 5, Fig 3	Channel type sampled, gear and effort, crew and individual fish length/weight data	c_FishCommunitySampling, c_FieldObservers, c_FishCommSpplndiv
Habitat	Weather and Reach Conditions	SOP 6, Fig 1	Weather conditions and discrete water quality measurements	c_SiteConditions, c_SiteMeasurements
Habitat	Instream Habitat Assessment Form	SOP 6, Fig 4	Instream habitat data per transect	c_CrossSectionalInfo
Habitat	Fish Cover Form	SOP 6, Fig 6	Bank and instream fish cover per transect	c_FishCoverInfo
Habitat	Bank Measurement Form	SOP 6, Fig 9	Bank stability and cover per transect	c_BankMeasurmentInfo
Habitat	Discharge	SOP 7, Fig 1	Interval, depth, and velocity measurements	c_DischargeDetail

Table 1. Cross reference table for fish database tables and corresponding field forms.

Data Recording

Quality Assurance/Quality Control (QA/QC) procedures related to data recording, verification, validation and processing are important components of any project sampling data, including sample methods, effort, weather, water quality conditions, species abundance data, and habitat data that are recorded on the field form (see Fish Field Form and Form Reference columns in Table 1). QA are procedures that ensure data collection, products, and services have minimal errors. QC are procedures to ensure that data or products meet established data quality standards (see *Integrating QA/QC into the IMD Pipeline*, NPS 2016). To ensure data is recorded properly on field sheets and to reduce data recording errors, consistency among the persons recording data is necessary and any new data recorder should be trained by the project leader on how to properly fill out the field sheets. During the training process, the project leader (Fisheries Biologist) or HTLN staff that have been trained (Aquatic Ecologist) by the project leader in data recording should periodically check the quality of data recording by the new crewmember while the data is being collected at a site. Field forms are checked for completeness by the project leader or trained HTLN staff member after all data has been recorded and before leaving a site or within 24 hours of data recording. This will aid in verification and validation of the data after entry into the database. Verification is the process to ensure data maintain accuracy and completeness throughout the data management cycle. Examples include: spot checks and audit of field forms and database tables, double-entry of field forms, or double observations of data. Validation is the process to evaluate legitimacy of the data throughout the data management cycle. Examples of validation procedures include: range checks based on

environmental limits, comparison of data to previous data, and comparison of data to expected data or modeled expectations.

To prevent the complete loss of field form data because of unforeseen circumstances (for example, fire or flood in the workplace), all field sheets are photocopied and a hard copy located in a separate location from the original. Field sheets are scanned into a computer server and electronic copies of the data sheets are located in at least three physical locations. This will ensure that at least one copy of the field sheets is available for data entry and verification.

Data Entry

To reduce data entry errors and assist in the verification process, data entry will be completed by a staff member who collected the data or is familiar with the field collection process. Upon opening the database (HTLNFish), the user is presented with a switchboard that can be used to access forms for data entry (see Figure 2). A preliminary set of forms define the sampling occasion and requires the input of location, period, and event IDs prior to entry of additional data. The other forms address the details of fish occurrence, habitat, and observers, etc. Once all fields for the preliminary set of forms have been completed, data can be entered for the remaining forms.

Several features are “built-in” to form properties that enable the user to maximize data entry efforts while minimizing error. These include: data input masks for ease of viewing multi-part data (such as park/project codes and date in PeriodID, LocationID, or EventID), “fill-in-as-you-type” to automatically complete a field, default values to autopopulate common values, limiting input values to known ranges (or restricting null values) or providing “drop-down boxes”, and tab indexes to control the order of data entry. Forms also contain fields that require data input or are constrained to properties and integrity of related tables.

More detailed step-by-step procedures on data entry will be kept on file in the HTLN Aquatics Laboratory located at Missouri State University, Springfield, Missouri.

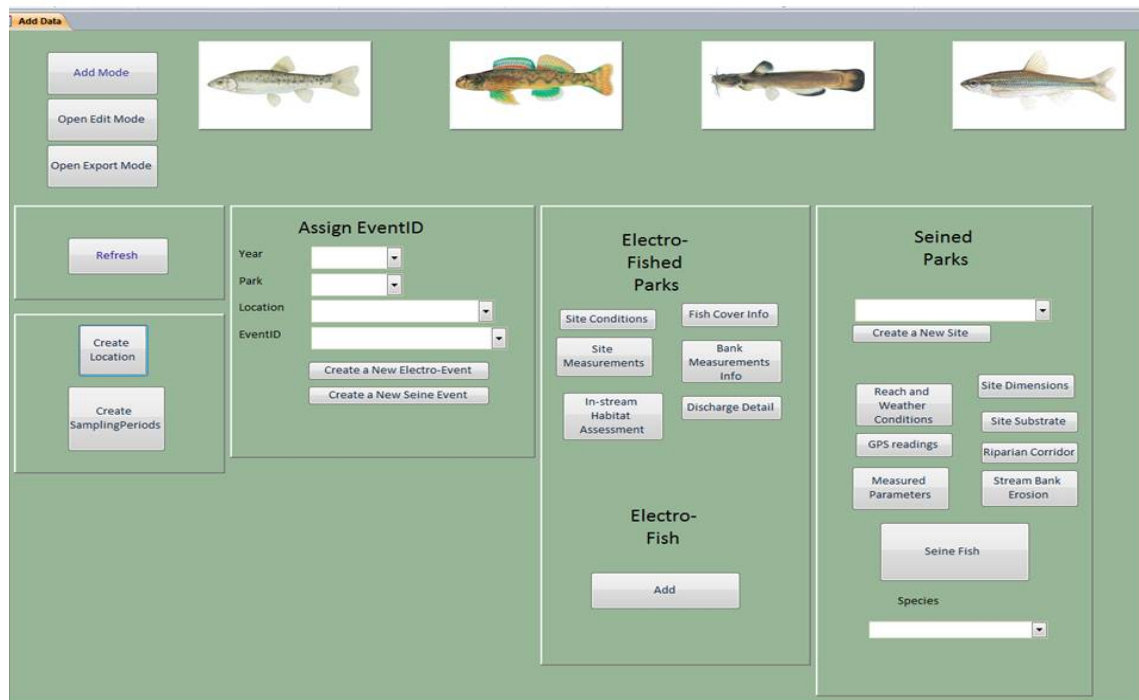


Figure 2. HTLNFish Database Switchboard in “Add Data” mode.

Entering Water Quality Data

Discrete Core 4 water quality data collected by hand-held meters are entered from the field sheets (see SOP 4, Figure 1) by selecting the appropriate LocationID and EventID and entering data into the c_SiteMeasurement table (Table 1).

Unattended Core 5 water quality data collected by unattended data loggers or sondes (see SOP 4) are uploaded from the logger using the manufacturer’s accompanying software program and stored on the HTLN server. Data then are edited to correct for any missing data due to logger maintenance (downtime), corrected for drift since last calibration in accordance with Wagner *et al.* (2006), and validated to determine if the data meet the expected range requirements or critical limits. Unattended CORE 5 water quality summary data then are entered into the NPS water quality database. Coordinate data for logger locations are collected in accordance with the current HTLN spatial data collection techniques and stored in the NPS water quality database. Metadata are then entered for each characteristic/parameter. The water quality data is sent annually to the NPS WRD staff for initial QA/QC and subsequent upload into the WRD master copy of the EPA water quality National Data Warehouse database. See the Water Quality SOP 4 for details on data download, correction, and import.

Data Verification

Data verification immediately follows data entry and involves checking the accuracy of computerized records against the original source, usually paper field records. Data tables are queried

to produce hard copy reports of specific sets of data (for example, fish community data and bank measurement data). These reports then can be compared to the original field data sheets to identify missing, mismatched, or redundant records. A 100% verification between field sheets and the database tables will be conducted by a staff member familiar with data collection but who did not enter the data. Any corrections will be completed by the project leader. Following these procedures, a 10% verification between field sheets and the database tables will be conducted by the project leader. If errors are present in the 10% verification processes, the data in question will undergo another round of 100% verification.

Data Validation

The purpose of data validation is to ensure the correct values of data are based on logical, expected or known prior values. Although data may be correctly transcribed from the original field forms, they may not be accurate or logical. For example, field crews collect data per occasion (location and event) and the resident data should reflect this. At any given occasion, fish sampling is conducted in each reach during the sampling event and can be validated by querying the ID number for these values. A query of these data should reflect these conditions and confirm the coincident (relational) nature of the database.

As data are amassed in the database, validation is conducted by query and comparison among years to identify gross differences. For example, species “A” may be recorded at a location in a specific year, but not in previous years, thus representing a possible new locality or potentially a mis-identification. In addition, range checks should be completed to ensure values on the field forms and in the database are within typical environmental limits. For example, a pH value of 13 may be recorded on the field sheet and in the database, but is likely an error in data recording or is an instrument error and should be flagged as such. Validation should be conducted by the project leader, and any questionable values should be confirmed with the data recorder and/or crewmember that collected the data.

Once verification and validation are complete, the dataset is turned over to the HTLN Data Manager for archiving and storage. These data then can be used for all subsequent data activities.

Spatial validation of sample coordinates can be accomplished using the ArcMap component of ArcGIS (ESRI 2015). Coordinate data of site locations are maintained in the database and can be added to an ArcMap project and compared with existing features (park boundaries, USGS Digital Ortho Quarter Quadrangles, National Hydrography Dataset hydrography, previous sample sites) to confirm that coordinate data are valid.

File Organization

The various databases, reports, and GIS coverages used and generated by HTLN create a large number of files and folders to manage. Poor file organization can lead to confusion and data corruption. As a standard data management technique, files pertaining to the project are managed in their own folder (RMFish project folder for data collected in this protocol) with subfolders: Analysis (for data analysis), Data (for copies of archived data as well as data sheets), Documents (for supporting materials related to the project), and Spatial Info (for various spatial data and GIS

coverages). The database (HLTNFish) is managed in a Databases folder and contains prior versions and backups in subfolders.

Version Control

Prior to any major changes of a dataset, a copy is stored with the appropriate dataset version number. This allows for the tracking of changes over time. With proper controls and communication, versioning ensures that only the most current version is used in any analysis. Versioning of archived datasets is handled by adding a floating-point number to the file name, with the first version being numbered 1.00. Each major version is assigned a sequentially higher whole number. Each minor version is assigned a sequentially higher number incremented by 0.1. Major version changes include migrations across software versions and complete rebuilds of front-ends and analysis tools. Minor version changes include bug fixes in front-end and analysis tools. Frequent users of the data are notified of the updates and provided with a copy of the most recent archived version.

Secure data archiving is essential for protecting data files from corruption. Once a dataset has passed the QA/QC procedures specified in the protocol and standards as identified in the Quality Assurance Plan (NPS 2016), the dataset will be considered “certified” and become broadly available to the general public. Once certified, the data is designated as “read only” for all users except the Project Leader and Data Manager.

References Cited

- Dodd, H. R., D. G. Peitz, G. A. Rowell, D. E. Bowles, and L. M. Morrison. 2008. Protocol for monitoring fish communities in small streams in the Heartland Inventory and Monitoring Network. Natural Resource Report NPS/HTLN/NRR—2008/052. National Park Service, Fort Collins, Colorado.
- ESRI. 2015. GIS and mapping software. accessed March 3, 2016 at URL <http://www.esri.com/software/arcgis/index.html>.
- Frakes, B., S. Kingston, and M. Beer. 2015. Inventory and Monitoring Division database standards: September 11, 2015. Natural Resource Report NPS/NRSS/NRR—2015/1035. National Park Service, Fort Collins, Colorado.
- National Park Service. (NPS) 2006. Inventory and Monitoring Natural Resource Database Template Version 3.1 documentation: Fort Collins, Colorado, Natural Resource Program Center, Office of Inventory, Monitoring, and Evaluation, 28 p.
- National Park Service. (NPS) 2016. Certification guidelines for Inventory and Monitoring data products. National Park Service, Inventory and Monitoring Division, Fort Collins, Colorado. 5p.
- Petersen, J. C., B. G. Justus, H. R. Dodd, D. E. Bowles, L. W. Morrison, M. H. Williams, G. A. Rowell. 2008. Protocol for monitoring fish communities of Buffalo National River and Ozark National Scenic Riverways in the Ozark Plateaus of Arkansas and Missouri: Version 1.0. U.S. Geological Survey, Little Rock, AR. Open-File Report 2007-1302.

Wagner, R. J., R. W. Boulger Jr., C. J. Oblinger, and B. A. Smith. 2006. Guidelines and standard procedures for continuous water-quality monitors: station operation, record computation, and data reporting. U.S. Geological Survey, Reston, VA, accessed December 2015 from <http://pubs.usgs.gov/tm/2006/tm1D3/pdf/TM1D3.pdf>.