**State of California**

**The Resources Agency**

**DEPARTMENT OF FISH AND WILDLIFE**

# Coastal Salmonid Monitoring Database Management:

# Status Report and Concepts for Success[1]

**2015 Administrative Report**

**By**

Douglas Burch, Sean P. Gallagher, Seth Ricker, and Russ Bellmer
January 29, 2015

[1] This report should be cited as: Burch, D. S.P. Gallagher, S. Ricker, and R. Bellmer. 2015. Coastal Salmonid Monitoring Database Management: Status Report and Concept for Success. California State Department of Fish and Wildlife, Redding California. 2015 Administrative Report. 49 Pages

# Contents

## ABSTRACT

This document describes the progress and accomplishments of the California Coastal Salmonid Monitoring Plan's (the plan) Database Management Team, a subgroup of the Technical Team, between January 2011 and January 2015. We provide an overview of the plan and related field data collection efforts. This document also details activities relating to the plan's data management and the scheme (the current system) being employed also known as the Coastal Monitoring Plan Aquatic Survey Program. It finishes with a discussion of our vision for data management activities and products over the next two to five years. We employ a storyboard method to illustrate our data management elements, concepts, processes, and products for users of tools we developed or propose. User stories have been included to illustrate the activities or experiences of the different individuals (contributors) collectively involved with the data and give examples of the terms they encounter within the current system. The current system is hierarchical, and contributors may be associated with one or more of its elements. For instance field staff members are primarily concerned with the data collection process, and we describe how they use the current system during a typical day, whereas an analyst or primary consumer of processed data, such as a manager, has different needs and thus we use other scenarios to describe their use of these tools. Examples of tools or utilities for which the current system provides are as follows: (1) surveyor interactive data validation is provided in order to prevent errors during both field and desktop data entry; (2) a custom data rule validation utility has been included which provides on/off data rule reporting capability, such as enforcing entry value patterns or identifying when entry values are missing required associations with other data elements; (3) utilities for importing and exporting of data are included in the current system; (4) functionality has been added for easily finding data within the desktop component's database; (5) capabilities exist which facilitate seamless merging of data from discrete monitoring projects or efforts; and, (6) the current system provides an overall meaningful and logical organizational structure to the data, one which facilitates better data management and ensures that we can derive the values needed for analysis and reporting requirements and works well with our spatial framework and sampling methods or strategy.

## INTRODUCTION

The mission of California Department of Fish and Wildlife (CDFW) is to manage California's diverse fish, wildlife, plant resources, the habitats upon which they depend on for their ecological values, and for their use and enjoyment by the public." As part of this mission, the CDFW is implementing the California Coastal Salmonid Population Monitoring Plan (CMP – Adams et al. 2011) to evaluate and monitor the status and trends of coastal Chinook salmon, coho salmon, and steelhead populations. The CMP strives to provide information needed for recovery of these threatened and endangered species by determining status and trends for abundance, productivity, diversity, and spatial structure. The plan is a scientific survey design that converts field data into valuable information that can be used for making management level decisions, and the management of this data is a critical aspect of this complex statewide effort.

> ***"data"***
>
> Succinctly defined variables and their values collected following standardized protocols. These are things measured, estimated, or observed and recorded in the field, for instance the length of a fish, a count of a particular species, the occurrence of a salmon nests, or the median axis dimension of a streambed cobblestone.

***"metrics"***

Variables that get their values from measurable observations that can be used to make quantitative evaluations for a variety of conditions, e.g., fish abundance, mortality, and recruitment. These values in turn can be used to gain scientifically meaningful information, i.e., useable for statistical tests or mathematical calculations.

The location and date of each field observation are critical elements that link field data to the monitoring design, allowing for the conversion of field data into useful information, e.g., status and trends. To produce reliable information from data collected under the plan's monitoring design a well-crafted, and scientifically based, data management scheme was necessary.

***"information"***

For our purposes, information is knowledge received or gained through scientific study, research, or communication, e.g., fish monitoring efforts. As it relates to this document, this is information that can be used at various organizational levels for making scientific and management decisions.

***"redd"***

A redd is the nest of trout or salmon.

The CMP is a scientific survey design whereby data is collected in the field and processed to create summarized values for various metrics as well as inferential statistics, and from this scientists can derive actionable information, e.g., biological, physical, spatial, and temporal measurements, are converted into numbers of fish per kilometer of stream, number of pools per watershed area, etc., which are summarized to provide qualitative and quantitative information, such as population status and trends over time. As a result of the large amounts of information collected in the field which makes this all possible, a well-designed data management system is an invaluable component of the overall CMP strategy.

***Digital Handheld Data Collection Device (handheld device)***

For the purpose of this document, this is a catch-all term for any handheld digital electronic data collection device that is used primarily for collecting tabular data in the field and may include spatial data collection as well.

The underlying goal of the CMP is to provide scientifically reliable information on the status and trends of California's coastal salmonid populations. As part of the effort to achieve this goal field data is collected with handheld device's then transferred to the current system's database, i.e., the Coastal Monitoring Plan Aquatic Survey Program database application. These data are collected following published protocols and prescribed methods using a spatially balanced random sample design which has a temporal component.

**The Teams**

There are three primary teams under the CMP: 1) the Management Team, 2) the Policy and Leadership Team (the Policy Team), and 3) the Technical Team.

The objective of the Management Team is to support the Policy Team and to provide the necessary policies and direction to implement the CMP as a program (the program). In this role, the Management Team provides recommendations to the Policy Team on membership in committees, teams, and groups formed to implement the Program. The Management Team also reviews and approves reports, plans, budgets, and other documents as necessary to implement all aspects of the Program.

The objective of the Policy Team is to support the Management Team and provide leadership and policy for the program. In addition, the Policy Team acts on recommendations from the Management Team and provides the necessary policies and leadership to implement the program. It reviews and approves reports, plans, budgets, and other documents from the Management Team as necessary to implement the Program. Specifically, it carries out the following:

- Demonstrating the significance of CMP within the National Marine Fisheries Service (NMFS) and CDFW

- Demonstrating the NMFS and CDFW's commitment to the program and to the executive and legislative branches of both State of California and Federal government agencies

- Achieving the necessary policy and financial support for establishing and maintaining the plan

- Pursuing and achieving collaboration with the other agencies and organizations critical to a successful implantation of the plan.

The objective of the Technical Team is to serve the Management Team and to provide science support for the Program. The Technical Team accomplishes this through a combination of the following: reviewing technical documentation; drafting or acquiring scientific reports, plans, and other documents as necessary to provide scientific bases for all aspects of the program; and, providing recommendations on membership in any of the necessary subgroups formed to implement the program.

**Purpose of this Document**

1. To record the Data Management Team's accomplishments to date; this team is a subgroup of the Technical Team.

2. In regard to the current system's database, describe the field data collection and data management schema and processes needed to distill field data into usable and actionable information for the CMP, both in precise technical detail and in concise, simplified storyboard format.

3. In terms of data management, detail our vision for success; this includes a description of our strategy, over the next two to five years, for implementation of a functioning data collection, management, and information dissemination system.

**Need**

This document provides and concentrates information from multiple levels of the data management effort that would otherwise need to be gleaned from numerous undocumented sources; it is a progress report which can be used as an indicator of how successful the Data Management Team's actions have been thus far; it can be used to explain to others, including upper management, what the team is attempting to accomplish; and, it provides a vision, based on tried and tested data management strategies, that can be implemented into the future. From the information contained within this document, resource managers can understand how the current data management system works while enabling developers of data management tools to scale the current design up to a statewide enterprise system.

**History**

The effort to develop our current system evolved slowly over several years, but received a substantial boost to the process at the Spawner Survey Protocol and Data Management Meeting which took place on December 12, 2007 at the CDFW Northern Region Fortuna office. The impetus behind this effort is explained in Fish Bulletin 180 (Adams et. al. 2011 with much influence from Boydstun and McDonald 2005), i.e., California Coastal Salmonid Population Monitoring: Strategy, Design, and Methods. This is a specific edition of a bulletin series that is essentially a separate scientific paper, i.e., part of the CDFW Journal which has been around since the Fish and Game Commission's early days (1800's) – An item too large (82 pages) to include in the journal so was published separately. It has been peer reviewed by some renowned fisheries experts, from private, academic, and other institutions (…Adams called it "the most reviewed document in the world," taking over five years to draft and review). Fish Bulletin 180 describes the overall strategy for meeting the monitoring needs of California coastal salmonids, and it uses the Viable Salmonid Population (VSP) concept as a framework for development of a plan (McElhany et. al., 2000, describes in detail the criteria for a viable salmonid population). So, why is this effort underway? It is because coastal steelhead and salmon are listed species which are managed under the guidelines of various recovery plans, and, as such, the CDFW has a vested interest in, and in-part funds, CMP efforts, i.e., these listings require that both the CDFW and the National Marine Fisheries Service (NMFS) develop recovery strategies that will conserve, protect, restore, and enhance listed species (NOAA 2010, 2012). In addition, Fish Bulletin 180 summarizes and updates earlier efforts of Boydstun and McDonald, 2005 (The California Plan – Action Plan for Monitoring California's Coastal Anadromous Salmonids), and was prepared for use in developing the CMP (Adams et. al. 2011).

As mentioned already the Data Management Team is a subgroup of the Technical Team; it was established in May of 2011 to support the Technical Team in an effort to adopt the current system with the expectation that it would eventually be managed centrally and utilized by all stakeholders; the Technical Team in turn was established to support the Management Team and provide science support for the program. The Data Management Team currently consists of Douglas Burch (CDFW, RA, lead - Redding), Seth Ricker (CDFW, ES, Arcata), Sean Gallagher (CDFW, ES, Fort Bragg), David Dixon (CDFW, RA, Sacramento), Rod Gonzalez (CDFW, RA, Sacramento), Gretchen Umlaaf (NOAA, Sacramento), Joe Ferreira (CDFW Statistician, Sacramento), Joel Casagrande (NOAA, Santa Rosa), Dana McCanne (CDFW, ES, south coast), Russ Bellmer (CDFW, provides updates to the Policy team - Sacramento), Jennifer Nelson (CDFW, ES, central coast).

**Fish Monitoring**

To provide a high level description, and for the purpose of this report, this section describes what is meant by "Fish Monitoring" as it relates to the CMP effort.

***Data Gatherer***

Any person who goes out into the field and collects the raw-data, typically with handheld devices capable of collecting both tabular (database) and spatial (GPS) data, e.g., portable digital assistants, GPS, flow meters, data loggers, PITT tag readers, etc. These tend to be crews of two seasonal technicians. Crews may collect data using more than one methodology during a survey season and using more than one method while at a location or site; they would likely use different tools for the various methodologies, and revisit any given survey site multiple times during a sampling season.

***Field***

This is any outdoor location where raw-data is collected by the data gatherer and a coarse level of data quality control takes place, e.g. visual review of the raw-data, required database fields, forced value ranges, entries limited to value lists, automated custom scripts, etc.

***Field Office***

This is any base of operations where individuals return to from the field. It is typically where raw-data begins the process of compilation, secondary quality control, and centralization.

***Fish Monitoring***

*Fish monitoring* is a scientific survey design whereby data are collected in the field, e.g., fish species observed, spawning redd measurements, fish habitat indicators, spatial and temporal information, etc. Data is processed to create usable information as it relates to various metrics, e.g., numbers of fish per kilometer of stream or number of pools per watershed area, which are then synthesized into information that can be acted upon, e.g., population status and trends over time.

**Products**

There are essentially three fundamental types of products that we obtain from the monitoring process, i.e., (1) raw data, (2) information for various metrics evaluated from the raw data, and (3) summarized information or statistics used for management purposes. In other words, we collect raw data in the field in the form of observations, such as redd characteristics, fish measurements, tag history, etc., all tied to date and location (temporal and spatial, respectively). Through management, storage, and analysis we turn this raw data into information that can answer questions, such as fish abundance, mortality, recruitment, etc. Finally, with the knowledge gained by answering these questions we get actionable information needed for management of a species.

For the purpose of identifying who the product users are, the gray text areas depict how each user participates in the creation or use of the above CMP products, or defines related terminology.

***Project Lead***

This person leads a data gathering and/or data analysis effort for a given project or area of study; typically, this is the project biologist or Environmental Scientist assigned to a study effort that is based in part on the geographic survey extent of a scope of work within a program area, e.g., anadromous salmonid restoration, monitoring, and recovery for Big-Navarro-Garcia watershed.

***Tabular Data Manager***

"*Tabular Data Manager*" is a catch-all term for any person that develops, supports, or assists with a CMP project database. If it is a distributed database then this role might be shared among the other various CMP participants, e.g., a data analyzer or project lead. In a centralized system this role would go to one or two lead information technology specialists, similar to the GIS Data Manager.

***GIS Data Manager***

Lead person for providing program-wide GIS services, products, and tools that support identifying and defining geographical considerations that relate to establishing the CMP spatial sample-framework, e.g., standardized stream based routed hydrography, hydrological and topological tools for helping determine stream site suitability, and long-term management of any value added attributes related to site locations, such as the spawning suitability of a survey sample site. The incumbent of this role works closely with the Tabular Data Manager.

***Global Positioning System (GPS)***

For the purpose of this document, this is a catch-all term for any handheld device that is used primarily for collecting spatial data in the field.

***Data Inquirer***

This is the person who takes the raw-data and runs it through one or more queries in order to perform a secondary level of quality control in order to produce "cleansed" data (to some degree this can be automated). In many cases this is done first by a seasonal technician (data gatherer) who collects data in the field and next by an Environmental Scientist (project lead).

***Statistician***

Sometimes referred to as the "data analyst," this role involves taking cleansed data and performing high level analyses, e.g., running the data through various analytical models or programs in order to generate inferential information that can later be used for management level decision making; this should be performed by a trained statistician, and the statistician should be able to explain the statistical tests fully.

***Data Analyzer***

The *data analyzer* may fill multiple and overlapping roles and must work closely with the statistician. More often than not this function will be performed by the project lead having a professional science background. In the early phases of the data flow process, somewhere between when the data inquirer performs a secondary level of data validation and a trained statistician has worked on the data, the *data analyzer* may perform both a higher level of data cleansing and a lower level of statistical analysis, e.g., looking for outlying data points and reviewing the data to identify values for the various metrics being monitored (occupancy, abundance, survival, etcetera). In later phases of the data flow process the *data analyzer* may use the results from inferential statistics, derived by the statistician, to provide information and recommendations, e.g., regarding the status and trends of fish populations.

***Report Viewer***

The audience for CMP information, e.g., federal and state endangered species managers, academics, non-governmental organizations, and the public who will need to see some CMP data product, other than the raw data. This would typically be analyzed/summarized data, suitable for inclusion into an agency report, i.e., they would see only analysis results and not the raw-data. The report viewer generally gets information for various metrics derived from values that were generated by the data analyzer and collected by the data gatherer.

***Report***

This is any of a variety of reports that result form the analyzed/summarized CMP data, e.g., user-administrative, management, ESAs, public, NGOs (TNC Report Card), and CDFW web publications (utilized by legislators and aides).

## METHODOLOGY

### Data Management

For the purpose of this document data management methodology refers to how the CMP data is currently collected, organized, and managed. As with most other data collection and data management efforts, the CMP system has a structure to (1) organize the raw-data for program needs and (2) best represent the information collected. There has been a lengthy process to the development of the CMP database structure. Under earlier monikers it began as an effort to create a model or template that could relatively easily be implemented by CDFW Environmental Scientists at the onset of data gathering projects. The principles are fairly simple when compared to many other database models. The idea is to create a data structure that lends itself to (1) ease of use in the field data collection scenario, and (2) scaling-up relatively easily in order to represent the data in a highly normalized, enterprise system.

**Observation-based Approach**

In developing our current system we drew heavily from an earlier CDFW concept referred to as the "observation-based approach" to collecting biological data (Burch 2007) that is based on an even earlier CDFW data standardization effort which defines an observation as follows:

*A circumstance, finding, situation, or activity that is executed, referenced, or noted in the field, about which the Department has a mandate or desire to record or report*

Uniform Field Observation Model Level I: Specifications for minimum uniformity requirements…
(Burch D., Gaul P. A., Haney E., Davis K., and Kauffman E. 2002)

In short, a database that adheres to these principals has three fundamental data components which make up the minimum requirements, i.e., (1) unit of observation, (2) standardized locations, and (3) project level metadata: the unit of observation refers to information, collected and maintained in a standardized format, that answers the "who," "what," "when," and "where" of the observation (Figure 1) as well as any categorical information or data that is required by specific survey types.

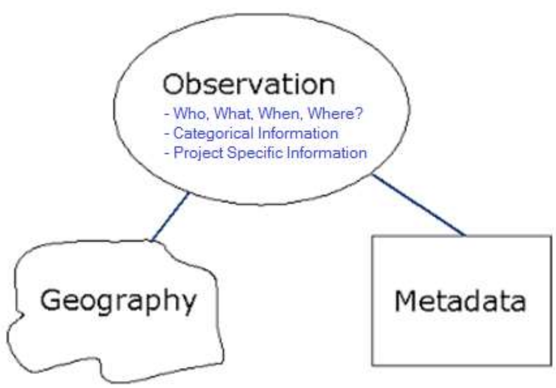


Figure 1. The diagram shows minimum requirements for observation-based data and relationships between its entities from *Observation-based Approach to Developing Databases* (Burch 2007).

In addition to the above, observations that have a relationship to other observations, i.e., are of the same observation type and are part of the same effort, can be grouped together by way of a survey component (Figure 2).

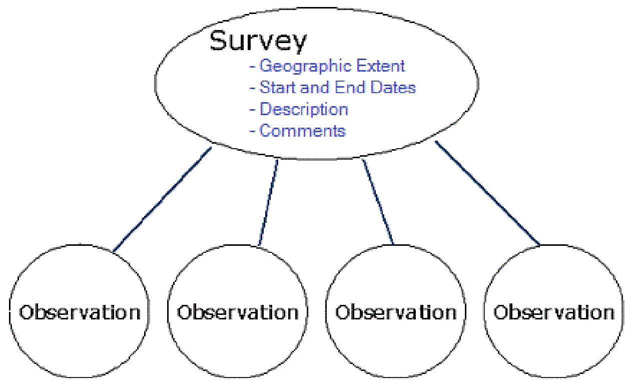


Figure 2. The diagram from *Observation-based Approach to Developing Databases* (Burch 2007) shows how common observation-based entities can be grouped together and associated with each other by way of a survey entity.

**Field Data Collection Database Template**

Our current system evolved from the observation-based approach, but is based on a later departmental concept, i.e., the Field Data Collection Database Template (template), described in more detail in *Field Data Collection Database Template Introduction* (Burch 2007). Excerpts from that document are below:

*An inherent goal behind use of the template is to move data from its point of origin – typically a field survey data collection scenario – to a central repository, such as a secured web-enabled data server. The additional goal of not encumbering field data collectors with complicated database architecture – while forcing data integrity necessary for eventual inclusion of field data into a central, secured data repository – is a challenge. It may not be the solution in every case, but the template does provide a workable approach for bridging the gap between needing simplicity in the field data collection scenario and scalability, i.e., the ability to normalize raw field data so that it can be stored more efficiently in an all-inclusive datastore.*

*…in a fisheries study a unit of observation might be an event like the "pulling of a fish trap" out of the water, and information about the trap itself would be a data object that is classified as the "observation event" while what is observed in the trap are the details of that observation event, or its sub-data objects or components.*

*Since these "core" aspects of the template structure have been established and standardized, any number of sub-data classes (tables) and properties (columns) can be added to this foundation, e.g., in the case of our example above a "tag" table could be added as a "child" table to the Individual table, its "parent," and "tag-identifier," "tag-type," and "tag-color" columns could be added to capture properties or attributes of each tag observed on or in an individual.*

*While this "add-on" capacity of the template provides for a flexible quality to the data model, the foundation or "core" aspects adheres to a strict standard and is not subject to modifications or customization – in other words, the core structure (Figure 3) does not change, allowing for trend analysis quality data that can be useful across a broad landscape.*

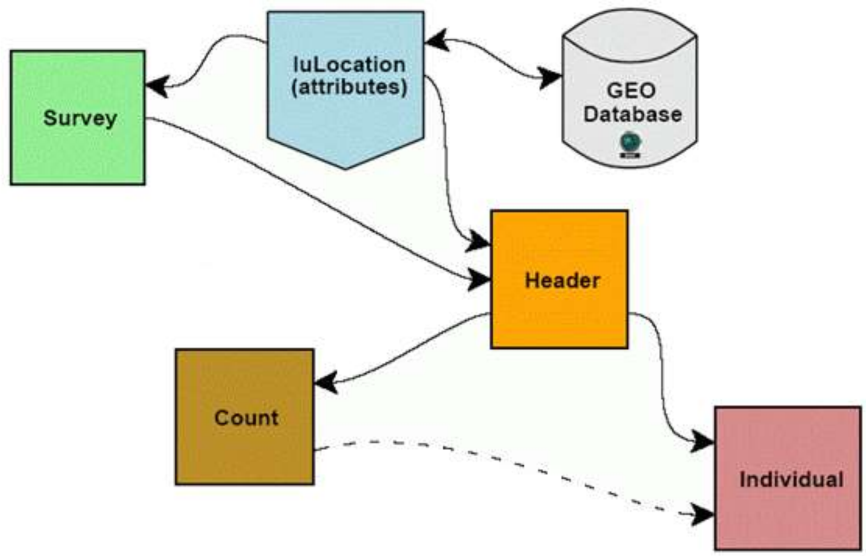


*Figure 3. The diagram above is a simplified illustration of some of the template's core tables along with their relationships to each other. Solid lines indicate relationships enforced by the database design while the dashed line implies that the linkage is conditional and enforced by programmatic data rules within the database's frontend application (its user interface tier).*

*Taking a step back and looking at the template design at a higher level, the organizational hierarchy has three more nodes or relationship levels above the Survey table. These high level nodes in the structure allow for maintaining data from multiple projects and programs in a single database.*

*Illustrated in Figure 4, the "Observation" node of the template model is equivalent to the Header table, and the "Details" node represents its sub-data tables. The Project level consists of a simple lookup table which is a list of possible projects that represents purpose or scope of the field data collection effort; all data must be a part of an established project within an overarching statewide program, for example, coastal versus inland monitoring programs. Since this is for ultimate inclusion into a statewide database then for the purpose of the template, the State is the highest level of the organizational hierarchy. A couple other aspects of the diagram worth pointing out are that (1) nodes above the Survey node are based on business management entities, such as a management unit assigned to a specific project lead and defined spatially by something like a district or management boundary. And, (2) there will be only one record in the Metadata table per project, however if methodology has been adopted according to programmatic protocols then there should be little difference between the metadata of one project and another within the same Program node.*

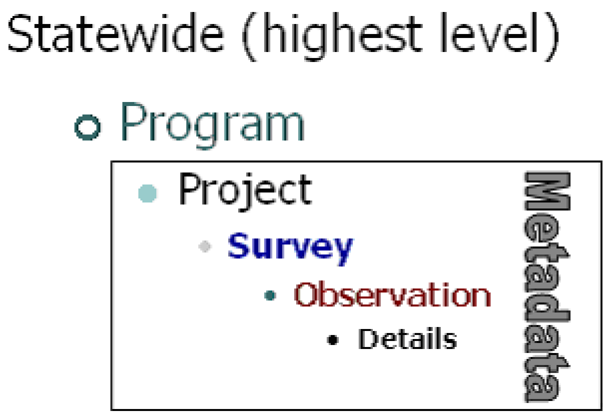


*Figure 4. The above graphic illustrates positions of the various hierarchical nodes within the template's organizational structure.*

**The Current System**

Two of the current system's fundamental data classes have already been mentioned in the *Observation-based Approach* section, i.e., the Survey and Observation, and several additional classes have been added to that in the Field Database Data Collection Template section, e.g., Program, Project, Header, Inividual, Count, etc. As with the template, the current system is based on hierarchal classes translating into data objects or tables in the database structure. Each of these tables is titled and described in language common to the field data collection scenario, but also adheres to specific definitions under the current system. For the purpose of database design and throughout this document we use capitalization to indicate that words, such as "Header," "Individual," or "Count" are being used as proper pronouns, e.g., when the word "Individual" is capitalized, it represents a data object class or table.

For the purpose of identifying what these data classes or tables are, in the below gray text areas that follow we define the data classes and/or table names that this document will refer back to when describing how the current system works.

***Program data class***

This term is loosely synonymous with its dictionary counterpart, however for the purpose of the current system, *Program* is the highest level data class and its entities define the overall plan of action that is to be accomplished, e.g., the CMP effort itself is a *Program* level entity.

***Project data class or luProject lookup table***

This term is loosely synonymous with its dictionary counterpart, however for the current system's purpose, *Project* is a high level data class, represented by the "luProject" lookup table; its parent is the Program data class, and its entities or records represent or define overall plans of action that are to be accomplished for given projects; to be a *Project* entity or record usually requires a plan of action with a sustained level of funding, equipment, field staff, and involves a concerted effort, e.g., "Humboldt Bay Spawning Surveys" is a distinct Project entity.

***Metadata data table***

Maintains information about the data itself, e.g., a catalog card is metadata related to a book warehoused in a library, because it describes a publication. In other words, it is information about a body of information. As it relates to the current system, a single *Metadata* entity or record is information about one unique project, i.e., the details relevant to its scope and purpose.

***The luLocation lookup table***

This lookup table maintains unique location entities and their associated attribute information for survey geographical extent and observation event locations. This is but one of the many lookup tables used by the current system; its attributes are similar to those of the sample-frame's attribute table, and records within this lookup table are maintained for quick and easy reference during field data collection and data analysis; since the sample-frame is established prior to the data collection effort, and sample locations are static after that, these fixed locations can be easily referenced by way of a lookup table, and there is no need to generate them during the field data collection process. That being said, location entities that are not known until field-time, e.g., redd locations, fish sightings, etc., are captured by GPS in the field and not stored in the luLocation lookup table.

***Survey table***

This term is loosely synonymous with its dictionary counterpart, however for the current system's purpose, *Survey* is a high level data class or table; the purpose of its entities or records is that of defining the spatial area and temporal bounds of a study. In addition, a record in this table typically has a short handle, or title, and a more lengthy description, and/or verbose comments about the survey effort; its parent is the Project data class or luProject lookup table.

***Header table***

A pivotal table in the database structure, the term "*Header*" is loosely synonymous with its computer science counterpart, i.e., information which precedes a sub-data component and is the same for all of its sub-data entities or records, however for the current system's purpose the *Header* data class or table is roughly equivalent to the observation-based approaches' Observation data class, i.e., it maintains high level entities, in the form of *Header* table records, that store information common to the details of an observation event, such as the "who," "what," "when," and "where" of the event; its parent is the Survey data class or table. In practice, it may take multiple records in the Header table for representation of an observation event as by definition a header record is unique by the combination of date and location while an actual observation event can occur over multiple days. Future versions of the current system would likely include an actual "Observation" table in the database schema, however for simplicity and ease of use in a field data collection scenario the "Observation" table has been left out of the current database structure in favor of deriving Observation level information by way of the data analysis process.

***Individual table***

This term is loosely synonymous with its dictionary counterpart, however for the purpose of the current system, *Individual* is a detail level table that maintains entities or records of biological information collected while an organism is figuratively "in-hand" during the observation event, i.e., field staff can collect as much of the required information about the specimen as needed, such as measurements, tissue samples, tag information, etc.; its parent is typically the Header data class or table, but conditionally can be the Count table.

***Count table***

This term is loosely synonymous with its dictionary counterpart, however for the purpose of the current system, *Count* is a detail level table that maintains records that represent batch counts of Individual table records whether or not Individual records exist for any or all of the fish represented by the batch count, i.e., *Count* table records are comprised of data elements that do not require "in-hand" observation as with the Individual table. In some cases both Individual and *Count* table data may be collected for the same fish depending on the protocols of a particular project; its parent is the Header table.

***Redd table***

This table is named after its dictionary counterpart, however for the purpose of the current system, *Redd* is a detail level table that maintains information about a nest of trout or salmon. Records within this table uniquely identify a fish redd as well as track measurements of the fish redd and other elements of data specific only to fish redds, such as tailspill length, pot size, and distance from the stream bank; its parent is the Header table.

***Snorkel table***

This table is named after fisheries sampling practices whereby snorkel gear is used to make fish observations; for the current system, *Snorkel* is a detail level table that maintains information collected during systematic-diving which is conducted for the purpose of estimating fish abundance or presence. The *Snorkel* table can be used in conjunction with the Count table to describe aspects of a batch count of fish that are unique to observations made when using systematic-diving methods; its parent is the Header table.

***Efish table***

This table is named after fisheries sampling practices whereby electro-fishing gear is used to make fish observations; for the current system, *Efish* is a detail level table that maintains information collected while electro-fishing, conducted for the purpose of estimating fish abundance or presence. The *Efish* table can be used in conjunction with the Count table to describe aspects of a batch count of fish that are unique to observations made when using electro-fishing methods; its parent is the Header table.

***StreamHabitat table***

This table is named after fisheries sampling practices whereby stream habitat survey methods are used; for the current system, *StreamHabitat* is a detail level table that maintains information collected while conducting stream habitat surveys for the purpose of making qualitative evaluations of fish habitat; its parent is the Header table.

***ClipMark table***

This table is named after fisheries sampling practices whereby fin clipping and/or body marking techniques are used; for the current system, *ClipMark* is a detail level table that maintains information collected while conducting mark and recapture studies for the purpose of evaluating a wide range of metrics, such as fish abundance, gear capture efficiency, and more. The *ClipMark* table can record information about both marks and clips that are applied to, or observed on, fish during an observation event; its parent tables are the Individual and Count.

***Tag table***

This table is named after fisheries sampling practices whereby tagging techniques are used; for the current system, *Tag* is a detail level table that maintains information collected while conducting tagging studies for the purpose of determining a wide range of metrics, such as fish abundance, migration patterns, and more. The *Tag* table can record information about tags that are applied to fish, or observed on/in fish, during an observation event; its parent is the Individual table.

***FishOnRedd table***

This table is named after fish behavior and observation practices whereby an observer makes note of the fish that are associated with a redd or fish nest in a stream; for the current system, *FishOnRedd* is a detail level table that maintains cross reference information collected as an adjunct to conducting fish redd studies, i.e., each *FishOnRedd* table record identifies the redd, already recorded in the Redd table, that one or more fish are associated with, and recorded separately in the Individual table; its parent is the Individual table.

***Sample table***

This table is named after collection practices whereby samples are kept and cataloged in relation to individual fish that were observed; for the current system, *Sample* is a detail level table that maintains records which document the samples collected during an observation event, such as tissue and other body parts that one or more fish, already documented in the Individual table, are associated with; its parent is the Individual table.

***Data Package***

For the purpose of the CMP effort this term is used to describe a discrete collection of data that is specific to a given project; it must adhere to all the requirements of the current system, e.g., geography, metadata, and uniquely identifiable observation entities. An example of a "*data package*" would be all of the data, along with subordinate detail information, for a given Survey event; this would include the metadata and spatial information; additionally, multiple Survey events for a project could be made available in one large data package.

***Distributed Dataset***

For the purpose of the CMP effort this term refers to collections of data, or data packages, typically complete at the Project level, that are specific to given Projects and maintained at discrete locations, e.g., a CMP project field office.

***Datastore***

For the purpose of the CMP effort, this term is used to describe a central location where CMP data packages can be pooled and maintained as a single collection of data, typically accessible by way of the web; the current data infrastructure does allow for direct submittal to a central datastore from the field.

***Raw-data***

Raw-data is data that has not been processed for use, sometimes called source data or atomic data (TechTarget.com).

***Data Products***

For the current system we are referring to *data products* as information derived through data processing or documentation of the raw-data. Although raw-data has the potential to become "information" in the form of *data products*, it requires selective extraction, organization, and sometimes analysis and formatting for presentation (TechTarget.com). For organizational purposes we are grouping information derived from the raw-data, or about the raw-data, under a catchall "*data products*" category, e.g., reports, statistical analyses, photos, and metadata.

Also as with the template, in the current system the Project class of data supersedes or is a parent of the Survey class of data, and the Survey class of data supersedes or is a parent of the Observation class of data. In other words, for each project there can be zero to many Survey events, or records in the Survey table, and for each Survey event, or record in the Survey table, there can be zero to many Observation events, or records in the Header table. An example of this would be a 10 week long fisheries survey that takes place within the tributaries (the geographic extent) that flow into Humboldt Bay. Any Observation events taking place within that 10 week period and geographic extent would be a hierarchic subordinate or child entity of that Survey event or record, and in turn the Survey event would be a child of the given project effort or data entity that we have defined as a record in the "luProject" lookup table. Using our Humboldt Bay tributaries example, the *Humboldt Bay Tributaries – 2011 Season* Survey event, or record in the Survey table, is a subordinate or child of the *Coastal Salmonid Monitoring Plan (CMP): Redwood Creek-Humboldt Bay Regional Spawning Survey* project, having a many-to-one relationship with that

Project entity. There can be a multiple of these subordinate Survey events or records, and they can refer to multiple time periods and/or multiple geographic survey extents as long as each of the subordinate Survey records relate directly back to the same project and do not coincide temporally nor spatially; that being said, these Survey entities can overlap temporally and spatially. It should also be reiterated here that it is at the Project entity level that raw field data must be appropriately documented in the Metadata table (Figure 4), i.,e., one Metadata table record per project.

For the most part we have used observation-based approach concepts to describe what has been termed the Observation event, i.e., the "who," "what," "when," and "where" of a discrete field data collection sampling effort (Refer to the Header table explanation in one of the gray text box above.). With the current system the "Observation" terminology has been replaced more or less with the term "Header," i.e., a table within the current system that is roughly equivalent to the Observation class or data object. As mentioned earlier, Header is a top level table used for capturing information common to the unit of observation in the form of its table records, and there are potentially a large number of sub-data table records that are subordinate to each Header table record. We refer to these subordinate records collectively as observation "details." These details are represented by sub-data objects, such as the "Individual," "Count," "Redd," "ClipMark," "Tag," "FishOnRedd," and "Sample" tables. Together the various data tables, inclusive of their records and organizational structure, comprise what the current system defines as a "data package" which can exist as distributed, discrete datasets (usually located at one of the various project field offices), or these discrete datasets can be pooled then merged into a multi-project datastore and centrally maintained for better efficiency and security. Because the current system's data tables maintain both locally, unique long integer identifiers and 38 character alpha-numeric globally unique identifiers (GUID data-types), the current system allows for merging discrete project level datasets without the possibility of data table key field conflicts.

In addition to ease of merging discrete datasets, the current system is also flexible in its ability to scale-up when normalizing the data, i.e., the process of efficiently organizing data in a database. The current system's structure can also easily be scaled-down to a simple, "de-normalized" design that works well on a handheld device. And, because its high level attributes meet all core design requirements of the CMP, and are at the basis of its design, the integrity of existing data would not be affected if additional detail tables, and/or project specific attributes or columns, were to be added in the future.

**Reference Lists (lookup tables)**

The above discussion briefly covered some of the basics of the current system, e.g., high level tables (luProject, Survey, and Header), low level tables (observation details), the hierarchical relationships between these database structures, and some of the minimum database requirements, such as standardized geographic features and project level metadata; this is the "data" portion of the current system, i.e., the raw-data relating to information gathered in the field and reference-data necessary to support the current system, like standardized locations and species codes. The majority of the above discussion focused on raw-data, however important also to the current system is the later, or reference-data, maintained in the current system by way of a group of tables that are lists of unique data elements that do not change, e.g., the "luSpecies" lookup table is a list of species codes that are static and can be referenced during the data collection process. Although usually not specifically referred to as the CMP data, this reference information is also a type of data; this section will only briefly touch on a few of the most important of these reference-data tables for the purpose of explaining by example, and because of their unique roles within the current system. For a more complete explanation of these reference lists or lookup tables, refer to Appendix 1. Note that lookup tables also differentiate from other data tables by way of the naming convention whereby the prefix "lu" is added which stands for "lookup."

Of the five standardized reference lists that will be mentioned here, the luSpecies lookup table will be covered first since it is fairly straight forward and best represents the many other lookup tables that will not be described in this section.

**luSpecies**

The luSpecies lookup table simply maintains a standardized collection of species related attributes, such as taxonomic identifiers as well as common and scientific names, i.e., one unique species code or identifier per entity or record. The current system uses standardized species key identifiers based on the Species Explorer (CDFW 2014) system, i.e., each species entity has a unique "TaxonID" for species representation. In addition to the unique CDFW identifier provided for each species of interest, this lookup table also provides the option of an alternate international system species identifier for cross-reference of the species entity, i.e., the Integrated Taxon Information System (ITIS) Taxonomic Serial Number (TSN) identifier. In the field we recommend that common name be used, although each program has specific needs and histories such that the Taxon and ITIS identifiers as well as a project specific species code attribute, can provide flexibility and multiple avenues for consistent coding. With the exception of the following lookup tables, the other lookup tables in the current system are about this straight forward, i.e., they are no more abstract and have the same or fewer attributes.

***"Sample-frame"***

The sample-frame is the sample universe; it is the collection of all possible one to three kilometer stream segments in all streams in coastal California that contain fish of interest.

**luLocation**

The luLocation lookup table warrants a discussion all its own as it not only represents a list of referenced location entities or table records in the current system, but it can also serve as spatial data in a GIS, and in most cases a GIS spatial layer can be developed from records in the luLocation lookup table. The luLocation lookup table maintains static entities or records that are referenced by both the Survey and Header tables. The Survey table's location entities are the geographic extents of the different survey bounds or scopes of study, and the Header table's location entities are the specific sub-locations or sites where the observations or field data collection work takes place. Records for both the Survey and Header location entities reside in the luLocation lookup table; however and even though these differing types of luLocation lookup table records can exist side by side, there is an inherent hierarchy between them and this hierarchy is maintained by protocols and programmatic data rules that are built into the current system. One example of these rules stipulates that a Header record's location entity can only exist within and out to the bounding geographic extent of a Survey record's location entity. In addition to recording descriptive information about a location, the luLocation lookup table record maintains attributes that place it spatially and include both qualitative and quantitative information, such as latitude-longitude coordinate pairs, references to other luLocation lookup table records, GIS metadata, and verbose comments about the location entity itself, e.g., how was it digitized, GPS specifications, and names of those involved in the location's spatial feature development. In the case of stream "reach" type location entities, the current system's luLocation lookup table maintains a "snapshot" of the sample-frame's GIS attribute table. The main differences are that the sample-frame's GIS attribute table has additional attributes that are specific to the CMP while the luLocation lookup table is more generic and can work for a variety of monitoring efforts at the program level, such as inland versus coastal monitoring programs. The sample-frame's GIS attribute table also maintains a more dynamic set of attributes that are used simply for development of the sample-frame. In effect, once the sample-frame has been developed for a particular study area then the current system's luLocation lookup table adopts a snapshot of the sample-frame's GIS layer's attribute table, and this becomes the static set of lookup table records to be deployed for use by the current system, i.e., for stream reach type location entities only. In other words, the luLocation lookup table identifies all referenced locations within the CMP effort, such as stream reaches within the spatial sample-frame, and in doing so can be used to link field data to the CMP's survey design, tying it all together.

To point out the degree to which the luLocation lookup table is a flexible way for storing location information in a database's tabular format, monitoring efforts unrelated to the CMP could use the current system's luLocation lookup table for maintaining location information related to those specific programs, even though unrelated monitoring efforts would likely have their own spatial model, e.g., relevant to lakes, estuaries, or terrestrial zones. etc.

Prior to the onset of monitoring efforts there are a number of processes that must be implemented in order to establish an overall framework so that data gathering can begin. These steps should be planned out and take place well before routine survey work begins. Among these are developing metadata for each project and defining reaches for which potential survey work might take place. For the later, there are three lookup tables that will be discussed next, "luLocationLink," "luReachSchedule," and "luReachLifeStage."

**luLocationLink**

The "luLocationLink" lookup table's purpose is to define geographic linkages, or associations, between specific site locations, i.e., referenced in a Header table record, and the more generally defined geographic extent of the study, i.e., referenced in a Survey table record. This relationship information can be used in a variety of ways, e.g., creating cascading lookup lists in the various data entry applications, validating data for correct entries, mapping locations for the evaluation of various metrics, and more.

***"GRTS"***

The acronym and short name for the statistical methodology referred to as "Generalized Random Tessellation Stratified" sampling (McDonald 2003) and implemented as a strategic component of the plan.

***"panel"***

A collection of stream segments sampled through the GRTS process which have a similar time visitation schedule

***LCMS***

"Life Cycle Monitoring Stations" streams with fixed adult counting structures and smolt traps which produce estimates of freshwater and marine survival, estimate spawner to redd ratios, and are places where methodologies and equipment can be tested and calibrated in a controlled setting.

**luReachSchedule**

The luReachSchedule lookup table maintains rotating panel information in a format that can be referenced by the current system. Defined earlier a panel is a collection of stream segments sampled through the GRTS process which have a similar time visitation schedule so a rotating panel design is one such that in any given year the reaches, visited for data collection, represent observations made based on a schedule which includes consecutive one, three, 12, and 30 year reoccurring visitation for trend detection (Adams et. al. 2011). A more detailed explanation of this methodology can be found in Appendix 2.

The purpose of this lookup table is primarily for use in the survey planning stage and typically would not be referenced in the field. It is a normalized, less user friendly, version of what is in Figure 5.

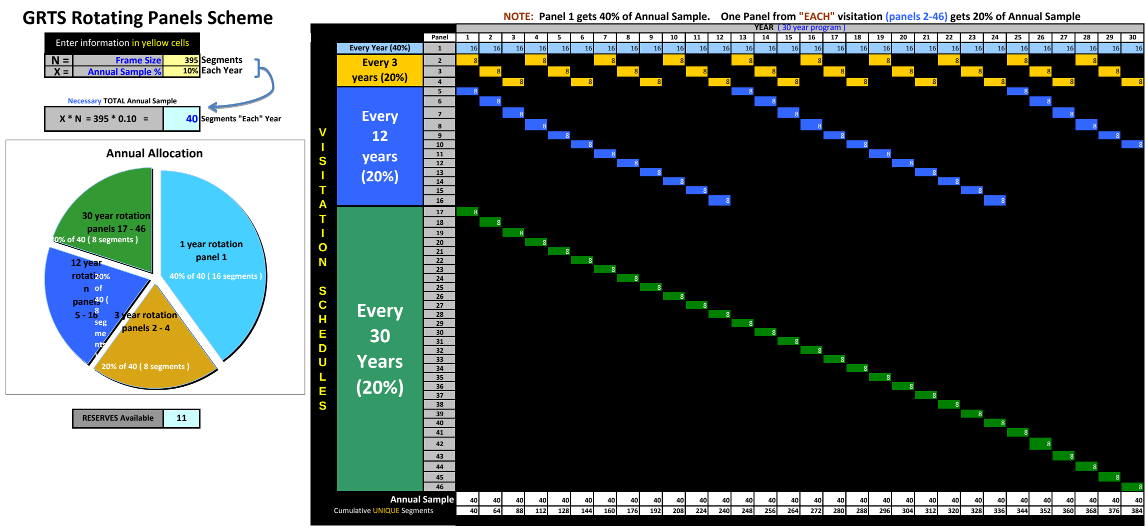


Figure 5. Rotating panel assigns reaches to the temporal design.

The advantage of having this lookup table is that, once the information has been entered, various helpful reports can be generated using this table in conjunction with other lookup tables, and the information can then be presented in whatever format is most useful for a particular planning effort. A brief overview on how this table works is as follows:

Each record represents a unique Location-Year-Revisit-Panel combination. The location represents a unique reach in the sample-frame; year represents the time period that the rotating panel sample occurred; revisit is a numeric integer value that is a code for the resample plan, e.g., 0 = 'Not revisited', 1 = 'Revisit under revisit design if year is...'; and, panel is a numeric value that represents rotating panel assignments for individual reaches or GRTS stream sections, i.e., the numeric value is a code which represents the set of reaches that will be sampled on the given year under the indicated revisit plan for that location.

The luReachSchedule lookup table should be populated prior to implementation of the proposed project, and it is the project lead who takes the primary role in providing this information.

### luReachLifeStage

The luReachLifeStage lookup table maintains information which describes life stage occurrence for each species by reach. Each record represents a unique Location-TaxonID-Stage combination. The location represents a unique reach in the sample-frame; TaxonID represents the species of interest; Stage is a code that represents a life stage found within the given reach for that species, e.g., adult, sub-adult, etc.

As with the luReachSchedule lookup table, the luReachLifeStage lookup table should be populated prior to implementation of the proposed project, and it is the project lead who takes the primary role in providing this information, however this would typically include assistance from the GIS Data Manager in the form of specialized GIS products which would help determine likely reaches for occurrence based on geographical features associated with those reaches, e.g., barriers, stream gradient, flow, etc.

## Spatial and Temporal Data

The current system combined with CMP protocols provides scientifically credible information as data are collected within a viable, however complex, spatial and temporal study design with the location entity, i.e., its unique record in the luLocation lookup table, providing the necessary spatial linkages. For stream reach type locations based on the CMP sample-frame, these entities are the spatially balanced random annual sample locations within the total sample-frame universe, in other words these selected subset locations, within the overall study area where data are to be collected each year, constitutes the spatial design, and how these sub-locations are assigned and sampled each year is the temporal design. Below briefly describes methodology based on the CMP sample-frame strategy:

Field data that is collected in association with "fixed-point" type locations, such as a weirs, fish ladders, and rotary screw traps, are spatially referenced to strategically defined coordinates within a stream while data associated with sample-frame based "non-point" locations, such as that collected from adult spawner surveys, is referenced to "sampling reach" type locations. Sampling reaches are prescribed lengths of stream which include all locations inclusive of that stream length, from the downstream to upstream demarcation points and from bank to bank; each sampling reach is a primary sample unit (PSU) within our spatially balanced random sample design. The methodology for this design is commonly referred to as GRTS sampling. The PSU is drawn from a sampling universe called a "sample-frame." The selected PSUs are fit into temporally based rotating "panels" since each PSU is a sampling reach within the sample-frame (Figures 6 and 7).

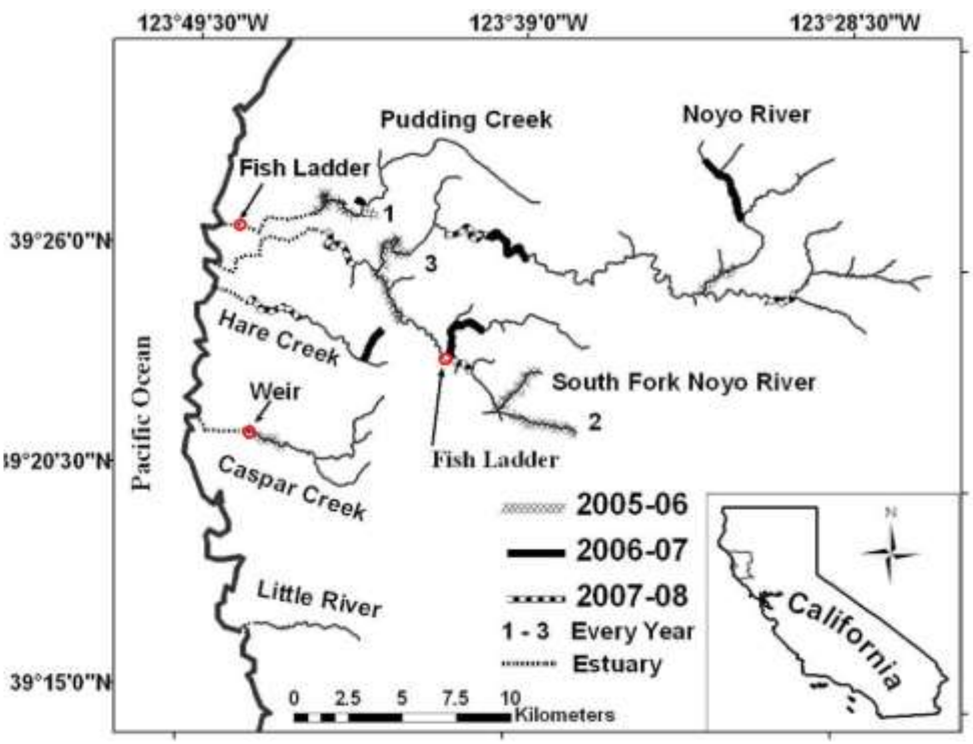


Figure 6. Locations of the study streams in Mendocino County and the 10% regional GRTS ordered steam reaches sampled each year of the study from Gallagher and Knechtle 2003. (Fixed-point locations, non-sample-frame based, are circled in red.)

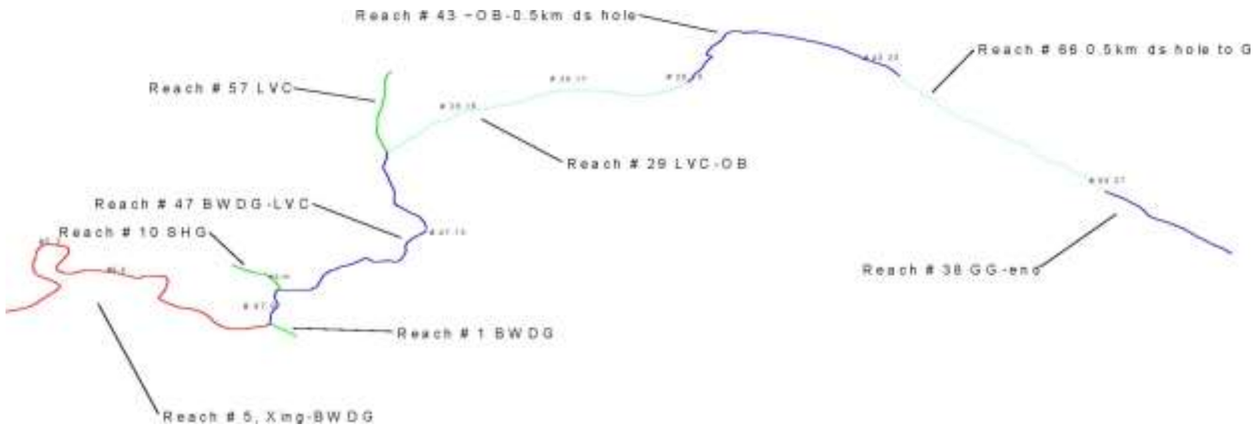


Figure 7. Pudding Creek spawning ground survey reach stratification from Gallagher and Knechtle 2003. Reach numbers, e.g., #5, are based on GRTS order; sub-reach numbers, e.g., #5.3, are the 33% sampling rate selections.

Values for each metric, according to protocol, are linked to the spatial design. By way of this standardized spatially and temporally referenced information, maintained through the CMP sample-frame's GIS layer and the current system's database tables, data points for the various metrics can quickly be drawn from to provide information or inferences related to the participating CMP project study areas. The values for these metrics are processed in respect to the statistical non-biased spatial and temporal component of the sampling design in order to provide information on the status and trends of salmonid populations at various spatial scales and during temporal periods going as far back in time as there is data (see Adams et al. 2011 pages 39-46). This sampling design has been vetted and is currently being implemented as part of the plan's monitoring strategy; its spatial links being maintained in the database through relational joins to the luLocation lookup table and temporally through the "Survey" and "Header" data tables. The luLocation lookup table derives its spatial component to the CMP survey design through use of the sample-frame's GIS layer's attribute table. As touched on already the sample-frame is the sample universe, i.e., it is the collection of all possible one to three kilometer stream segments in all streams in coastal California that contain fish of interest.

Sample locations are grouped into what the current system refers to as "geographical survey extents" which spatially bounds data for a given overall study location. This association is accomplished by way of the luLocation, luLocationLink, and luReachSchedule lookup tables. A location becomes part of a survey when assigned boundaries in time, e.g., the November 1, 2014 to March 30, 2015 late fall-run survey. As part of the monitoring design locations are also grouped into panels that are sampled during the same survey period. The pattern of visits to a panel of PSUs is called the revisit design and refers to the schedule and scheme by which PSUs are visited and sampled through time, i.e., the membership of these PSU locations to a panel and their specified revist schedule, such as annual visits or visits for a prescribed number of years consecutively after skipping some number of years before repeating, etc., and the rotating panel schedule and scheme is prescribed by referencing the luReachSchedule lookup table (See McDonald 2002 for review of monitoring design terminology). Because not all salmonid species utilize the same locations at their respective life stages, the membership of a location for a particular species and at a specific life stage is recorded and referenced by way of the luReachLifeStage lookup table.

For only stream reach or PSU type locations, data as it relates to these metrics are synthesized through the statistical spatial and temporal sampling design to provide information on the status and trends of salmonid populations at various spatial scales (see Adams et al. 2011 pages 39-46). Appendix 2 provides examples of how this spatial and temporal sampling design works.

It should be restated here that the current system does not always use design based sampling, such as GRTS or some other algorithm, e.g., some LCM metrics are generated by conducting a spatial census of locations or other biological criteria while some projects are only CMP "candidates" and do not yet adhere to the plan's sample-frame strategy, and as mentioned previously fixed-point locations, such as weir and trap sites, are not tied to the sample-frame, i.e., they are their own "sample-universe."

**Practical Use of Current Methods**

While the current system is not a long-term solution so far it has been effective in the near-term, operating by way of distributed databases. The reason for using this type of "distributed database approach" is because a distributed system can be implemented relatively quickly compared to a statewide centralized, or "enterprise," system. The current system relies on two database application components, i.e., a desktop MS Access database application and a Pendragon Forms handheld device application. We chose MS Access for its ease of use; development times can be fairly fast and flexible; its capabilities are robust; and, for many years now this software has been the CDFW standard for file server type databases. Pendragon Forms has some of the same advantages as MS Access: the learning curve is relatively short; rapid development times; and, its capabilities are robust, e.g., it has built-in data validation and some scripting capability. In addition to these software choices, at the time of writing this report the preferred hardware has been ruggedized handheld data collection devices capable of running the Pendragon Forms software – ruggedized features meaning waterproof and shock resistant. We have also favored the less expensive hardware that usually lacks robust geographic capabilities as the current system has minimal need for spatial functionality in the field since locations are known and can simply be referenced by way of the luLocation lookup table. For instances when coordinate data is required, e.g., a redd's latitude-longitude, then a low cost GPS unit can be used in conjunction with the handheld device.

The basic operation of the current system is for data to be collected in the field on handheld devices then synchronized with MS Access on project workstations in regional field offices. Data in these distributed databases can later be pooled as needed to produce a complete universe of data that eventually resides in a central datastore, and this can be made available to others by way of the web. Figure 8 illustrates the various ways that field data can "flow" from its collection location to where it will ultimately reside.

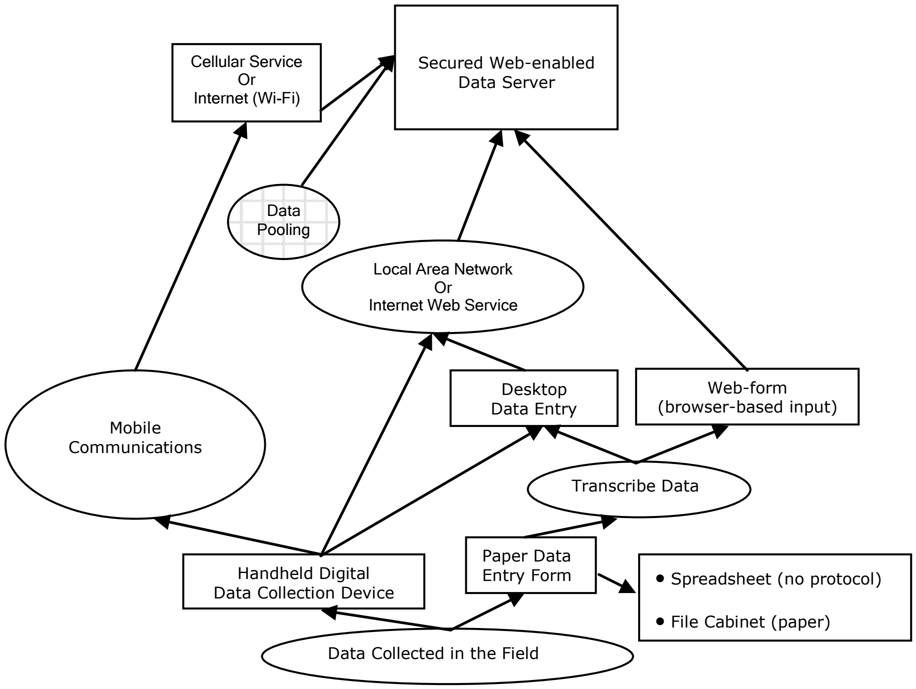


Figure 8. A flow chart illustrating some of the more common options implemented when considering the one directional flow of raw field data to a secure environment where it can be stored and managed most efficiently (Burch 2007).

## Current System Summary

The preceding sections provide a general picture of how the current system operates. Obviously, there is more than what has been covered here, and for details appendix 2 provides working examples of how the "Coastal Mendocino Regional Salmonid Monitoring Surveys" project has been implementing the current system since 2009. Also, appendix 3 provides a step by step walkthrough of the entire process using a series of simple "user stories," short dialogs describing the various experiences of its "users." It has been our goal that this document illustrates at a high level what the current system has accomplished, and to some degree the reason for development of the current system. We are also trying to make the point that there is already substantial history behind this design, and considerable logic to how it was constructed, in terms of gathering and managing monitoring data. It is for another document to describe the various CMP products that have been developed from the data captured over the last several seasons since its first use, not to mention the host of potential products that could be developed. We will limit ourselves here to say that the number of products has been impressive. We also need to mention that what we have thus far has been extensively tested in the field and has been a workable solution not to be lightly modified. Primarily, this success has happened because of a couple of design choices, 1) simplicity to the design so that it does not take a database administrator to understand and manage the data, and 2) flexibility so that as monitoring research evolves the model can take on new study "modules," such as various methodologies for stream habitat surveys.

## DISCUSSION

### Handheld Devices

We see several clear advantages to data management systems that leverage handheld devices, and in particular ones that use mobile technologies, such as tablet devices, and this argument may extent to using Smartphones as field data collection devices as well. Much of this is based on the success that the current system has had thus far with use of the Aceeca Measura handheld devices. While tablet PC's and SmartPhones would work a bit differently than the Aceeca Meazura handheld devices, i.e., the older portable digital assistant technology being implemented by the current system, they do have the common feature of providing a method of computerized handheld data entry in the field. The idea behind using handheld devices for field data collection has multiple years of tried and tested research as a basis for our confidence in them. The advantages of devices that collect data in the field are fairly straight forward: (1) they save time, because data only has to be entered once; (2) they improve integrity of the data in two ways, (a) errors resulting from data entry are not multiplied when transcribing from paper to database (Johnson, C., Temple, G., Pearsons, T., and Webster. 2009); (b) in-field data validation in combination with some limited scripting capability can be used to automate data input, enforce some basic validation rules, and customize error responses; and, (3) they are lightweight and highly portable, e.g., they can weigh about the same as an average field clipboard while supporting multiple applications, including GIS tools.

It should be mentioned here that the obvious disadvantages to handheld devices are (1) challenges which can arise when equipment malfunctions and (2) the additional learning curve associated with relying on a technological solution. With the current system the field crew brings along paper data forms as an insurance policy against the possibility of malfunctions; and, we feel that despite the complications that technology brings to the process, the advantages of greater efficiency and increased data quality more than makes up for any additional training that might be required.

With the tablet and Smartphone scenarios data can be collected in a "disconnected" fashion, i.e., even though these devices may not have a cable, wireless router (Wi-Fi), cellular, or any other live connection to the web, they can still be used for data collection in remote field locations. When these disconnected devices become "connected," e.g., when a tablet is brought within range of cellular communication towers or Wi-Fi hotspots, synchronization with a central data server on the web is automatic, including "pushing" out to the handheld devices any application or database design updates that occurred while the devices were disconnected. Data priority levels are not such that CMP data gatherers would need to synchronize in a real-time fashion so a disconnected scenario like this would work well for these sorts of monitoring efforts and substantial cost savings would be realized, because a more expensive real time connection, such as cellular, is not essential – a low cost, or even free, Wi-Fi connection will do the job, e.g., most Wi-Fi systems can provide synchronization of multiple devices at the same time therefore field crew would not have to vie for perhaps limited cabling hardware and workstations. In addition, this Wi-Fi only setup is typically a much better value than requiring all field devices be capable of cellular communications; also, consider that cellular coverage is not available in many of these field locations so in most cases there is little advantage to cellular versus a landline or Wi-Fi setup.

The handheld device development software that is currently being used by most CMP projects is Pendragon Forms "Industrial" version. This software uses a hardware cable to connect and synchronize with departmental workstations where data can be further validated then cleansed and centralized. While this solution has many advantages over paper forms and transcription of the data into a computer application, it comes up somewhat lacking in the computer systems environment of today that includes such technological breakthroughs as ubiquitously accessible data by way of the web, e.g., cloud computing, wireless connections, and rapid application development (RAD) tools that make development of web-based data entry forms quick, easy, and doable with minimal if any writing of computer code; Pendragon Forms "Universal" version is an example of this type of RAD tool – It can quickly generate the field data entry forms once the back-end database structure has been setup, and can push the field data entry forms out to tablet or Smartphone devices by way of the web.

This document refers to third party hardware and software to better illustrate methods used by the current system or how we think the system might work better in the future with newer technologies; however, we are only suggesting that further research and understanding of these third party products would be prudent. What we are promoting is a concept or methodology that frees us from being completely dependent on a particular type of hardware or software product, or development team for that matter. What we are looking for are solutions that preserve the valuable aspects of the current system, like its data model, while leveraging the benefits of emerging technologies. Our focus has been on the constants: (1) we have been primarily concerned with the data model itself, and its scientific viability; (2) we also have a strong focus on whether the data model is able to allow us the ability to measure all of the metrics necessary for determining the answers to key questions which result from the mandates placed on CMP efforts as well as perhaps questions that have not been thought of yet; (3) we need a model that is going to be a workable solution, considering the constraints of our various field data collection scenarios and conditions; (4) we want the hardware, software, methods, and procedures, that support and facilitate our data management system to limit data collection activities as little as possible so that budgets and timelines can be met without unreasonable constraints upon field operations. As far as the technology goes, we envision something like a system of tablets and Smartphone devices that communicate, when not in disconnected mode, directly with a web-enabled data server and its underlying secured, central database structure. In other words, the data server handles requests from the remote handheld devices by way of modern web protocols and interacts with the client by displaying a dynamic geographical user interface (GUI). This means that there would be no intermediary desktop application to act as "middle-ware" as is the mode with the current system. Once synchronization occurs with the data server, the client-side GUI receives any updates which occurred while the client was disconnected. If the data must be accessed, e.g., for editing, viewing, or downloading, then it can be revisited at this central, web accessible data server, and editing would be transacted on the server-side through a web enabled GUI displayed on a field office workstation as opposed to on the handheld devices where the data was originally entered. What we are ultimately striving to achieve is a system whereby the focus can be placed on the data itself, and to the greatest extent possible our future data management system would have a high degree of freedom from proprietary software and hardware as well as not being dependent on development teams that have a proprietary or inside knowledge as to how the system was developed or works.

**Data Analysis and Data Products**

As part of our vision and touched on already, some auto-generated, common data products would be available from a central, web accessible data server through a web enabled GUI displayed on field office workstations, or on any remote workstation for that matter, where departmental security policies could be enforced; however, many of the frequently requested data products would not likely be auto-generated directly from the central data server. We foresee that the majority of information derived from the raw-data for various metrics will be a result of custom products created by project leads and their support staff, i.e., at this time we do not see automated reports for every conceivable product that could come from the raw-data. This is currently the case, and it is estimated that there will be requests and requirements for hundreds of unique data products perhaps useful to one specific group or another. We think that the best way to facilitate this is to maintain something similar to the current system, i.e., whereas the data itself should be centralized, however the products that come from that data are best left to the scientists that collect it. In other words, centralize the raw-data, but provide tools which allow those who were involved in its collection to easily get the raw-data back, usually in discrete project specific, filtered datasets. The typical scenario would be as follows: the field staff collect the data; the data is synchronized with the central data server and made available by way of a state web server; if needed the data can be further edited through a web enabled GUI displayed on a field office workstation and transacted on the server-side; based on one's permissions level, the raw-data can be filtered and downloaded in a readily usable format, such as MS Excel or comma separated value text (CSV); project leads, data analyzers, and statisticians use tools, such as MS Access and R, for their workup and analysis of the data; the final summarized reports are made available by way of an online informational-portal that is secured.

Our vision does not include a system whereby every single CMP data product ever imagined can be generated by simply clicking on a button or web link. Although, a system that could generate a wide variety of reporting products would have its advantages, we realize that the resources and cost of developing a system capable of outputting every conceivable CMP data product would be prohibitive. It may very well be impossible to do as well, because we likely will not know every question that we want to ask upfront. As an alternative to this "does everything" approach, we have leaned towards a tandem system that consists of (1) raw-data and (2) informational-data. In other words, the two components are related, however they are developed and maintained separately, and we see a distinction between what we will call the CMP "Raw-data Portal" and the CMP "Informational-data Portal," i.e., we have approached maintenance of the raw-data and informational-data as two separate, but parallel data management efforts. This concept is further explained in the sections below.

**Raw-data Portal**

The raw-data portal as we envision it is the central repository for all of the "raw" data being generated by the various CMP projects, i.e., by personnel who are the leads for these projects, and their field staff who do much of the raw-data collection. This data currently resides in distributed MS Access databases up and down the California coast, but eventually must make its way into a centralized CDFW datastore.

Regardless of the technical details, this raw-data is what is used to generate a variety of derived data products required for making decisions related to the resources being monitored. In some cases one project lead might share raw-data with another, but typically if this raw data were accessible beyond that personnel level it could be easily misinterpreted and/or misused. The point being that the raw-data is not directly usable to the general CMP stakeholder and should not be lumped into the CMP Informational-data Portal as an in kind product. On the other hand, what is made available on the CMP Informational-data Portal should be everything relevant to the CMP effort that is not the raw-data, and this can include some broad categories, e.g., summarized datasets, reports, white papers, query tools, narrative about the CMP effort or specific projects, metadata (including methodology), photos, contacts, links to other sites, etc., and the list goes on.

**Informational-data Portal (everything else)**

We envision this aspect of the effort being something that should be maintained as a separate web presence than that of the Raw-data Portal, and perhaps could be managed most efficiently through a third party organization that has been successful at managing data oriented websites for multi-agency efforts. One option is going through a quasi-governmental, nonprofit organization, like Pacific States Marine Fisheries Commission (PSMFC), for either the near-term or long-term management of CMP informational-data published to the web. Our technical team has suggested that a new CMP workgroup be created just to deal with Informational-data Portal issues. This new workgroup would be tasked with developing a strategy for how best to manage, and publish to the web, the myriad of potential CMP informational-data items that will require diligent organization and review prior to their being made available online; this would also include developing and implementing well organized and professional website layout and design preferences.

## CONCLUSION

### Establish Core Attributes

As part of our overall strategy, the Technical Team must identify what core attributes are needed and therefore should be retained from the current database structure. To a degree much of this work has already been accomplished. The next step in this process is primarily to establish common protocols, or data rules, for using these attributes, e.g., all CMP projects would be required to use the same measurement units and the same method of measuring fish. Other issues have to do with attributes that one particular project lead might find useful while other project leads may choose not to utilize. At this point we are fairly certain that all of the attributes in the current structure are useful; however, not all project leads feel the need to collect data for every attribute within the current data structure. The challenge will be to convince each of them that data for all attributes in the current data structure should be collected, and these would then collectively become the core attributes. Or, members of the team might be convinced to drop certain attributes from the distinction of a "core" attribute; projects could still collect these attributes, but they would need to be maintained outside the central or enterprise data structure that CDFW's Data and Technology Division would be tasked with developing. In any case, each project while fulfilling the same programmatic needs has slightly different directives because of existing infra-structures and partnerships, e.g., LCM surveys and fish life history monitoring, such that meeting the requirement of establishing core attributes across projects may not be possible so the future enterprise data management system will require some degree of flexibility.

### Vision for Future Methods

This topic encompasses both field techniques as well as database management systems in general, including considerations such as hardware, software, costs, and support. In future scenarios field techniques and methodology would remain much as described above, however these concepts might be extended to whole other areas of monitoring, such as stream habitat surveys involving Columbia Habtat Monitoring Program (CHaMP) methods. As an example, we would treat CHaMP as a separate, stand-alone, data module complementary to the current aquatic surveys monitoring module; that being said, data managed through these separate modules would be compatible and comparable, spatially and temporally, with existing and/or future data modules. By taking this modular approach new data modules or components should be, and we think can be, developed without "breaking" existing ones.

How we envision a future data management system to some extent revolves around current and emerging technologies which to a degree are a moving target, but some clear trends are becoming apparent, and we hope to leverage on these new technological developments. It is also important that any new data management system adopted be adaptable to new technological developments, specifically regarding software and hardware so that our department is not locked-in to using outdated systems.

### Strategies for Success

Our strategies have taken into consideration multiple approaches so that we have options if one of them is not working out to our satisfaction. The focus has primarily revolved around three efforts that relate back to the Raw-data Portal and Informational-data Portal concepts. For the Raw-data Portal, i.e., the raw tabular data, (1) we are currently working with CDFW's Data and Technology Division (DTD) to develop a new data model that will be the basis for a centralized data structure capable of maintaining statewide raw-data for the core CMP attributes; most – but not all as of yet – of these core attributes have been determined. For CMP's spatial data that is of the reach or PSU location type (2) we are working with DTD's Biogeographic Data Branch (BDB) in developing the spatial sample-frame which will be used to establish the luLocation lookup table records that will be referenced by the CMP raw-data. In addition to assisting with development of the spatial sample-frame, BDB will also be responsible for maintaining and editing this spatial data layer, and BDB will work with DTD in order to provide information necessary for managing CMP core attribute data. In tandem with that, (3) we have approached PSMFC regarding maintaining a web portal for CMP's informational-data needs.

**Keep the Focus Moving Forward**

Some of the tasks ahead, that will make this vision happen, seem straight forward enough; however, there will likely be multiple junctures along the way where the process can stall if the Technical Team is not making persistent efforts towards this goal, e.g., establishing a common set of core attributes will require some difficult decisions and a consensus within the group. In addition, the Technical Team will need to communicate on multiple levels, and to multiple points of view. On one level we must work with DTD's database development team, and fulfill their requirements, however there are other groups within DTD that focus on different issues, such as hardware-software standards, and in-field IT systems specialists; we must establish those working relationships as well, in part so that we meet their requirements, but also so that we can communicate our unique IT needs and ensure that those needs are being satisfied; and, there will need to be a degree of give and take with this process. We will need to communicate with management at another level to ensure that our needs are appropriately represented and so that management can best help with providing programmatic resources. On a whole other level, we will be working within the scope of each of the projects, on a day to day basis, using the database system and providing feedback.

**Monitoring Efforts in General**

There are a lot of commonalities that take place across multiple categories of monitoring efforts, and to a large part that is what the Data Management Team has tried to address. We are well aware that the challenges are not isolated, and other groups have gone through exactly the same process as the CMP in providing coordinated standards and structures to data collection, maintenance, retrieval, and documentation, i.e., proper data management (Kolb, T.L., et. al.). The Data Management Team has built on the work of others, and tried to come up with a model that not only fits with project requirements, but can relatively easily scale-up to something which resembles an enterprise solution. What we have done so far establishes a tried and tested method for efficiently capturing monitoring data electronically in the field. The current system reduces redundancies in data handling efforts, and provides multiple points at which data errors can be identified and corrected before data is ready for centralizing. The distributed database application approach has allowed us to get the process going where it otherwise might have stretched out indefinitely. One of the caveats of our current process and methodology is that project leads, and in some cases their field staff, are the data managers for their own projects, at least for the active field season. While this is not the most ideal scenario, it has allowed for moving ahead within an environment of tight budgets and limited staff resources. Moreover, there is something to be said for project leads and their field staff taking a greater role in managing the data that they are responsible for collecting, e.g., it promotes better awareness of data integrity issues that affect our resource management efforts; that being said, this effort has matured to where the distributed database application approach presents some serious limitations for using the data in a seamless way and managing it with the level of security which it deserves.

As part of our goal to keep moving forward, and because of the need for an improved data management model, we are at the point where the existing system needs to be reengineered, scaled-up, and, to some extent, revamped, and to accomplish this challenge a more highly technical team must be brought in to continue the database management side of our work. We see this as a multi-part team approach with expert IT staff providing data management support, and the field scientists now taking on more of a client-feedback role, as well as providing the technical advice needed to continue identifying field data system needs and interpret/evaluate the data.

**REFERENCES**

Adams, P. B., L. B. Boydstun, S. P. Gallagher, M. K. Lacy, T. McDonald, and K. E. Shaffer. 2011. California coastal salmonid population monitoring: strategy, design, and methods. California Department of Fish and Game Fish Bulletin 180.

CDFW. 2014. Species Explorer. California Department of Fish and Wildlife. [Cited 2014 December 30]. Available from: https://nrm.dfg.ca.gov/taxaquery/Default.aspx

Boydstun, L. B., McDonald, T. 2005. Action plan for monitoring California's coastal salmonids. WASC-3-1295. Final report to NOAA Fisheries, Santa Cruz, CA. 78 p.

Bjorkstedt, E. P. 2003. DARR (Darroch analysis with rank-reduction) A method for analysis of stratified mark-recapture data for small populations, with application to estimating abundance of smolts from out-migrant trap data. Administrative Report SC-00-02.National Marine Fisheries Service, Southwest Fisheries Science Center, Santa Cruz, California, USA Administrative Report SC-00-02.

Bouwes, N., J. Moberg, N. Weber, B. Bouwes, C. Beasly, S. Bennett, A. Hill, C. Jordan, R. Miller, P. Nelle, M. Polino, S. Rentmeester, B. Semmens, C. Volk, M. B. Ward, G. Wathen, and J. White. 2011. Scientific protocol for salmonid habitat surveys within the Columbia Habitat Monitoring Program. Prepared by the Integrated Status and Effectiveness Monitoring Program and published by Terraqua, Inc., Wauconda, Washinton, USA.

Burch, D. Data Flow Diagram. California Department of Fish and Game; 11/12/2007. [Cited 2014 December 18]. Available from: https://nrm.dfg.ca.gov/FileHandler.ashx?DocumentID=92336

Burch, D., Gaul, P. A., Haney, E., Davis, K., and Kauffman, E. Uniform Field Observation Model Level I: Specifications for minimum uniformity requirements - version 1.0c. California Department of Fish and Game; 11/1/2002. [Cited 2014 December 8]. Available from: https://nrm.dfg.ca.gov/FileHandler.ashx?DocumentID=91757

Burch, D. Field Data Collection Database Template Introduction. California Department of Fish and Wildlife; 11/12/2007. [Cited 2014 December 18]. Available from: https://nrm.dfg.ca.gov/FileHandler.ashx?DocumentID=92335

Burch, D. Observation-based Approach to Developing Databases. California Department of Fish and Game; 11/12/2007. [Cited 2014 December 18]. Available from: https://nrm.dfg.ca.gov/FileHandler.ashx?DocumentID=92314

Flosi, G., S. Downie, J. Hopelain, R. Coey, and B. Collins. 1998. California Salmonid Stream Habitat Restoration Manual. Third Edition. Inland Fisheries Division. California Department of Fish and Game. Sacramento, CA. 495 pp.

Gallagher, S. P. and M. Knechtle. 2003. Coastal northern California salmonid spawning survey protocol. California Department of Fish and Game, Fort Bragg, California, USA.

Gallagher, S. P., and C. M. Gallagher. 2005. Discrimination of Chinook and coho salmon and steelhead redds and evaluation of the use of redd data for estimating escapement in several unregulated streams in Northern California. North American Journal of Fisheries Management 25:284-300.

Gallagher, S. P., D. Wright, B. Collins, and P. Adams. 2010. A Regional Approach for Monitoring Salmonid Status and Trends: Results from a Pilot Study in Coastal Mendocino County, California. North American Journal of Fisheries Management 30:1075-1085.

Gallagher, S. P, P. K. Hahn, and D. H. Johnson. 2007. Redd Counts. Pages 197–234 in D.H. Johnson, B. M. Shrier, J. S. O'Neal, J. A. Knutzen, X.Augerot, T. A. O'Neil, and T. N. Pearsons. Salmonid field protocols handbook: techniques for assessing status and trends in salmon and trout populations. American Fisheries Society, Bethesda, Maryland.

Gallagher, S. P., P. B. Adams, D. W. Wright, and B. W. Collins. 2010a. Performance of spawner survey techniques at low abundance levels. North American Journal of Fisheries Management 30:1086-1097.

Gallagher, S. P. and D. W. Wright. 2012. Coastal Mendocino County salmonid life cycle and regional monitoring: monitoring status and trends for 2011. California State Department of Fish and Game, Coastal Watershed Planning and Assessment Program, 1487 Sandy Prairie Court, Suite A, Fortuna, CA 95540. 34 pp.

Garwood, J. and S. Ricker. 2011. Spawner survey sample frame development for monitoring adult salmonid populations in California. California Department of Fish and Game Anadromous Fisheries Resource and Monitoring Program. Arcata, CA. 21 pp.

Johnson, C.L., G. Temple, T.Pearsons, T. Webster. 2009. An Evaluation of Data Entry Error and Proofing Methods for Fisheries Data. Transactions of the American Fisheries Society 138:593-601

Johnson, D. H., B. M. Shrier, J. S. O'Neal, J. A. Knutzen, X. Augerot, T. A. O'Neil, and T. N. Pearsons, editors. 2007. Salmonid field protocols handbook: techniques for assessing status and trends in salmon and trout populations. American Fisheries Society, Bethesda, Maryland.

Kolb, T. L., E. A. Blukacz-Richards, A. M. Muir, R. M. Caramunt, M. A. Koops, et. al. 2013. How to Manage Data to Enhance Their Potential for Synthesis, Preservation, Sharing, and Reuse—A Great Lakes Case Study. American Fisheries Society Journal, Bethesda, Maryland. Volume 38, Number2.

McDonald, T. 2003. GRTS for the average Joe: A GRTS sampler for windows. West Inc., Cheyenne, Wyoming. 10pp. Available: www.west-inc.com/reports/grts.pdf.

National Marine Fisheries Service (NOAA). 2012. Public draft recovery plan for southern Oregon Northern California coast coho salmon (*Oncorhynchus kisutch*) Evolutionarily Significant Unit. National Marine Fisheries Service, Southwest Region, Arcata, California, USA.

National Marine Fisheries Service (NOAA). 2010. Public draft recovery plan for central California coast coho salmon (*Oncorhynchus kisutch*) evolutionarily significant unit. National Marine Fisheries Service Southwest Region, Santa Rosa, California, USA.

**APPENDIXES**

Appendix 1 – Matrix of major reference lists, or lookup tables, used by the current system.

Appendix 2 – Coastal Mendocino Regional Salmonid Monitoring Surveys Examples

Appendix 3 – User Stories

***Appendix 1 – Matrix of major reference lists, or lookup tables, used by the current system***

| Lookup Table | Description |
|---|---|
| luAgeClass | Manages fish age class indicators or distinctions. |
| luAgeClassStructure | Lookup table for length range information, i.e., cross reference for minimum-maximum length ranges used during a given survey. |
| luClipMark | Maintains a list of standardized lookup codes for possible clip-mark-position combinations. |
| luContributor | Lookup table for maintaining information about contributors to the database, i.e., persons who collect, enter, or use the data. |
| luCoordSys | Lookup table of possible spatial coordinate systems used when geo-referencing locations. |
| luLengthRange | Lookup table for length range information, i.e., cross reference for minimum-maximum length ranges used during a given survey. |
| luLocation | Lookup table for maintaining unique location entities and their associated attribute information (for Survey and Header table records). |
| luLocationLink | Lookup table for relating Survey and Header locations to each other. |
| luMethod | Lookup table for maintaining codes which represent methods used during field observation events. |
| luObsEvent | Maintains a list of codes representing procedures used when making the observation, e.g., independent observers, multiple skill levels, etc. |
| luProject | Lookup table for maintaining unique project level entities; each project should be documented by way of its own Metadata table record. |
| luPunch | Maintains a list of possible shapes or patterns used when an opercle punch has been applied. |
| luReachLifeStage | Lookup table that references information which describes life stage occurrence for each species by reach. |
| luReachSchedule | Lookup table that references information which defines the revisit periodicity for each reach by survey year. |
| luReddAge | Maintains a list of codes used to represent different redd age categories. |
| luSpecies | Lookup table for maintaining unique species identifiers. |
| luStage | Lookup table for maintaining life stage codes used in reference to observed or counted individuals. |
| luStatus | Catch-all lookup table for maintaining a variety of status codes, i.e., elements that indicate qualitative distinctions about the individual, count, or sample, such as overall health or disposition. |
| luStrHabType | Maintains codes and descriptions for a variety of stream habitat type categories. |
| luSurveyType | Manages codes for referencing different types of surveys conducted during a field data collection effort. |
| luTagColor | Maintains a list of possible tag colors that can be used to describe a tag. |
| luTagType | List of possible tag types that can be used. |
| luUnit | Maintains a list of possible unit abbreviations used when measurement data are collected. |
| luWeather | Lookup table for maintaining codes used to describe weather conditions during an observation event. |

### *Appendix 2 – Coastal Mendocino Regional Salmonid Monitoring Surveys Examples*

By way of examples this section will illustrate important concepts behind the current CMP spatial and temporal sampling design; but first, some details on spatial distribution of various salmon populations and their life histories will be covered, e.g., what spatial data is needed and why. Second, there will be a brief discussion regarding information needs of the CMP with some specific examples to illustrate the sample-frame's overall importance. Third, as it relates to spatial and temporal aspects of the study design, some examples and discussion will be provided. Fourth, with actual data collected along the coastal areas of Mendocino County, California, using current CMP methodology and design, we will work through a final example. It is our hope that through this illustration we will provide the reader with a general or high level understanding of how the underlying statistical design is implemented as well as the importance placed on the spatial and temporal components of the design. These examples should clarify that much of what makes the CMP design useful, across regional boundaries and as it relates to a statewide data management system, is the key role played by the spatial unit, i.e., the GRTS stream section (the sampling "reach") or otherwise known as the PSU, and the specific time reference for which a piece of information was collected.

Coho salmon range from the Anadyr River in Russia to Aptos Creek in Santa Cruz County, California (Figure 9). However, most salmon tend to spawn in streams in which they were born and thus there are many "populations" throughout their range. A salmon population is a collection of fish in the same geographic area that are more likely to breed among themselves than they are with fish of other areas. Salmon are managed in the US under the Federal Endangered Species Act by way of spawning aggregations called evolutionarily significant units; like a population, the idea is that fish in any given evolutionary significant unit (ESU) are more likely to breed among themselves than they are with fish from different ESUs (Figure 10). Groups of salmon populations, that span the diversity and distribution that currently exists or historically existed within the ESU, we refer as diversity strata (Figure 11) to reflect our primary focus on the issue of diversity, broadly defined as the basis for delineating these groups. ESUs are composed of collections of diversity strata; and, like with a population and ESU, the idea is that fish in one diversity strata are more likely to breed among themselves than they are to breed with fish in other diversity strata (Figure 12). Therefore, to understand and manage California's coastal salmon populations we need a monitoring plan that can provide information on their status and trends at these different spatial scales. By way of implementing a complex spatial and temporal sampling scheme, this is what the CMP design does. The sample-frame with assigned GRTS order shows the annual and rotating panel samples for the Mendocino coast, consisting of two diversity strata in the central California coastal (CCC) ESU. The spatial scale is based on each population's annual sample estimate for the two diversity strata (Figure 13). Repeated annual sites are sampled to best evaluate trends over time; adding sites through a rotating panel where revisit repetition intervals vary based on species life histories adds to trend detection and incorporates the variation in species life histories (Figure 14).

Figure 9. Coho salmon range map (from NOAA Fisheries, http://www.nmfs.noaa.gov/pr/species/fish/cohosalmon.htm)

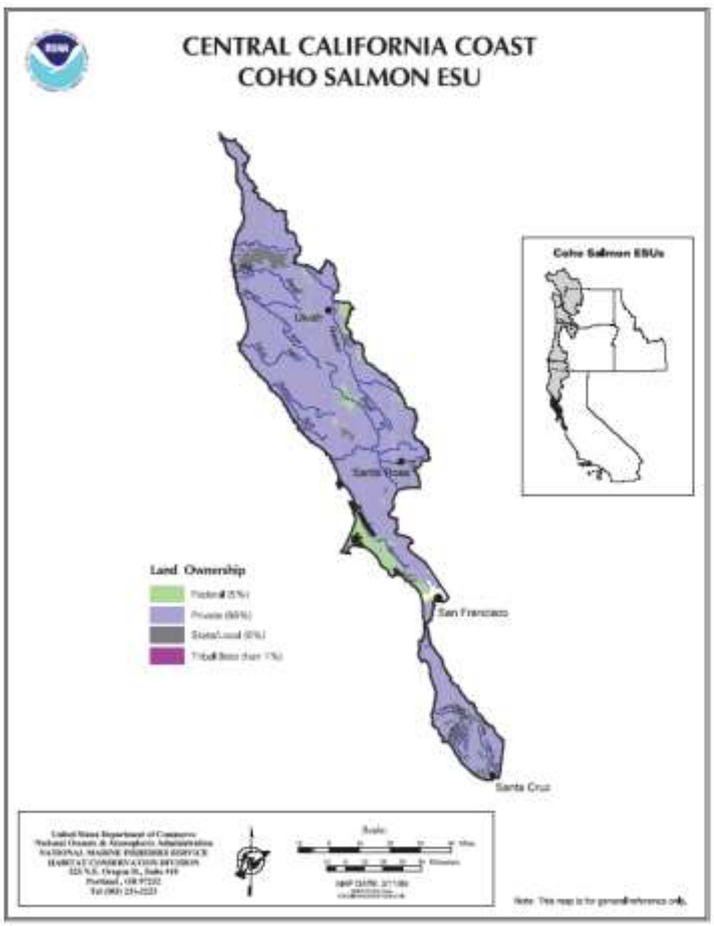


Figure 10. Coho salmon CCC ESU. Inset shows other coho salmon ESU’s to the north.

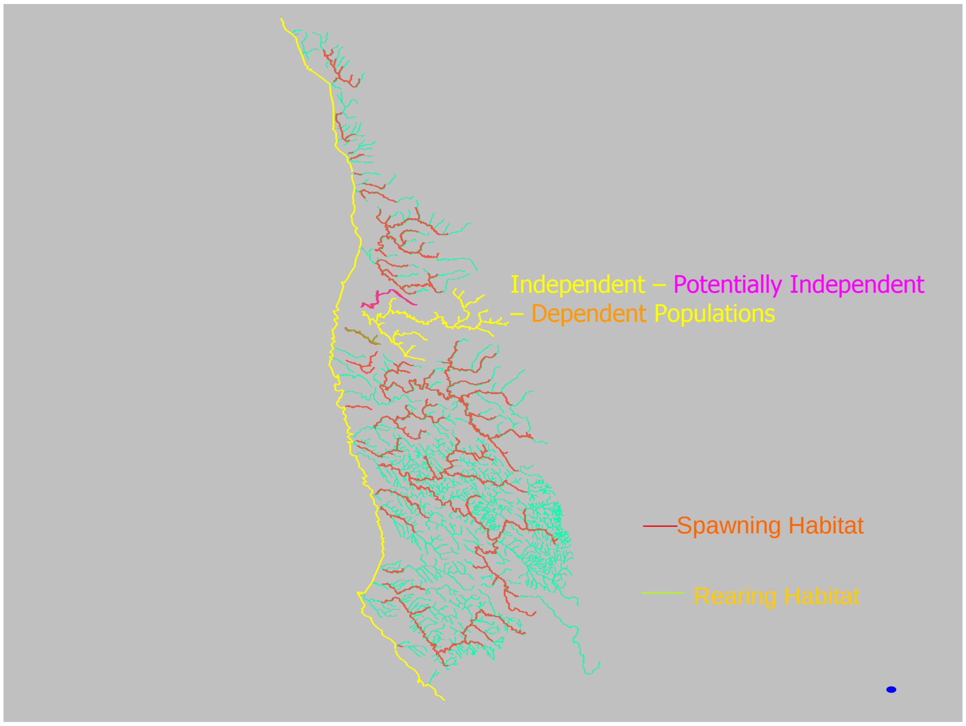

Figure 11. Coho salmon populations in the Lost Coast diversity strata within the CCC ESU.

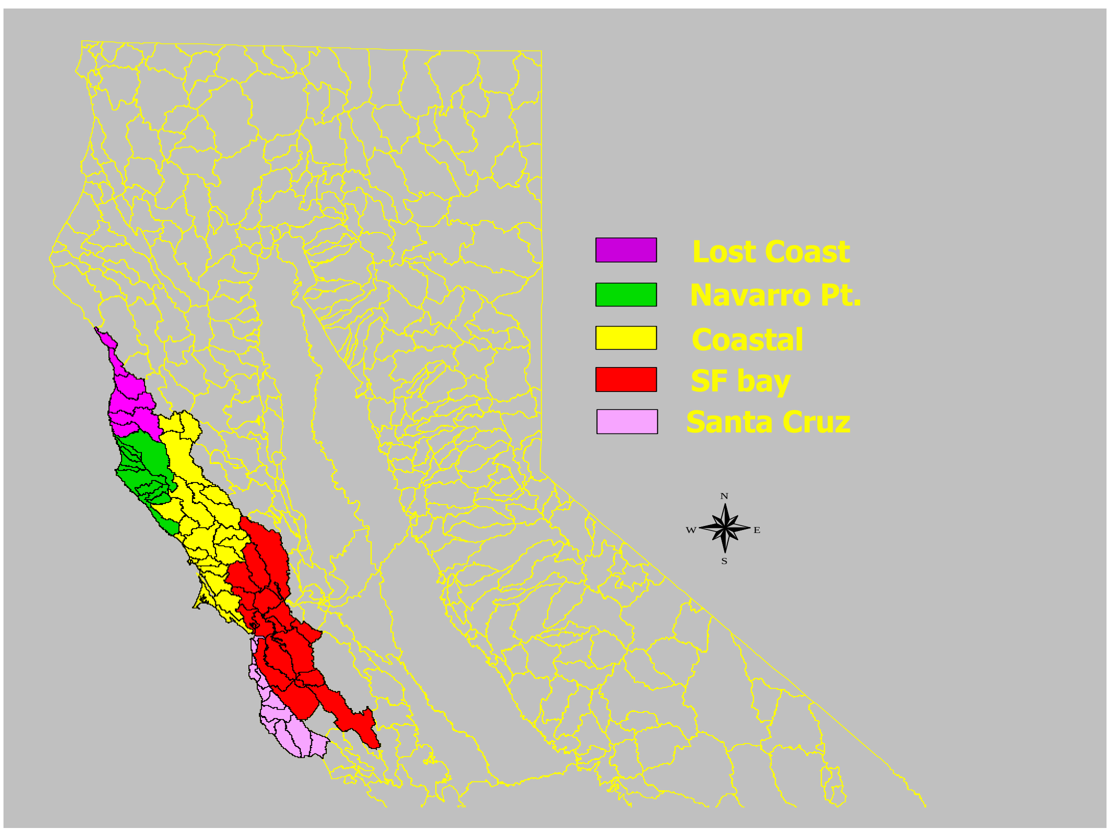

Figure 12. Diversity strata within the CCC coho salmon ESU.

The above figures (from 9 to 12) illustrate coho salmon distribution; steelhead and Chinook salmon have similar population spatial structures, although all three species have considerable variation with respect to their life histories. Coho salmon generally live three years, one in freshwater and two in the ocean, before returning to their natal streams to spawn and die, however some individuals out-migrate to the ocean after less than one year while others may spend two years in freshwater before outmigration, remaining for less than one to more than two years in the ocean. Chinook salmon generally spawn at four or five years of age and die after spawning. Steelhead mostly spawn at four years of age, first spending two in freshwater and then two in the ocean; they return to the ocean after spawning and can live to spawn many times. It is because of these varied life histories that the monitoring plan requires a temporal component which is prescribed by way of the rotating panel design and implemented through the CMP sampling strategy. This design is tied to the luLocation and luLocationReach lookup tables of the CMP database structure such that a collection of sampling reaches in any one year represents sampling for one, three, 12, and 30 year trend detection (Adams et al. 2011)

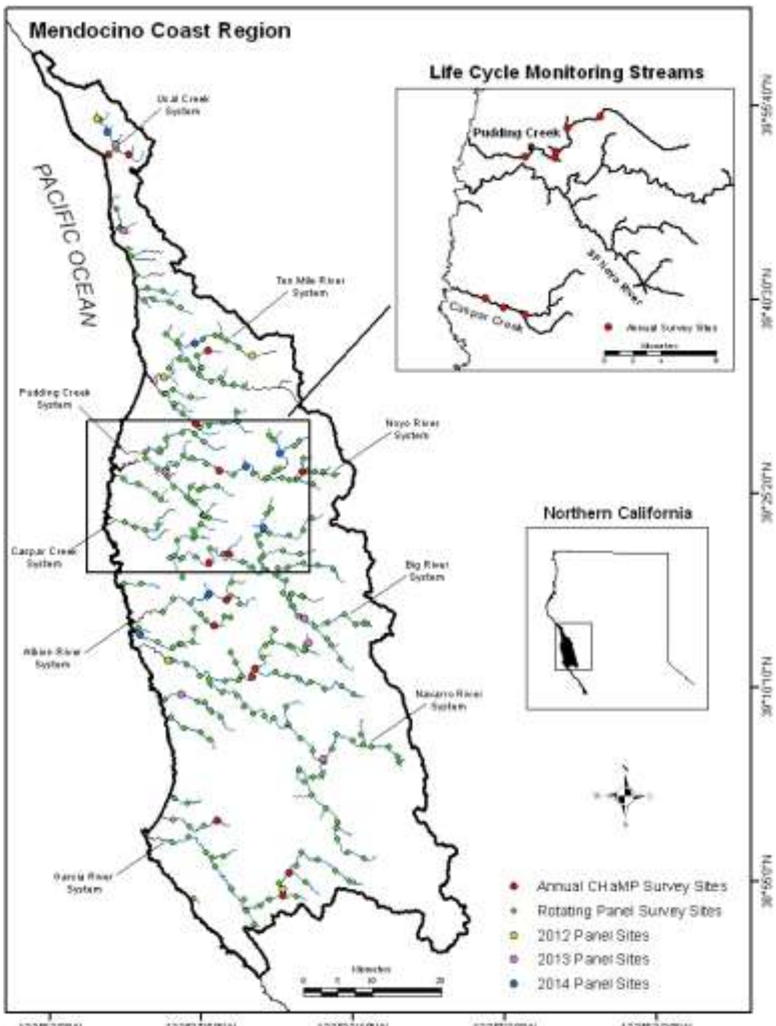


Figure 13. Example rotating panel design for coastal Mendocino County.

**Site Stratification**

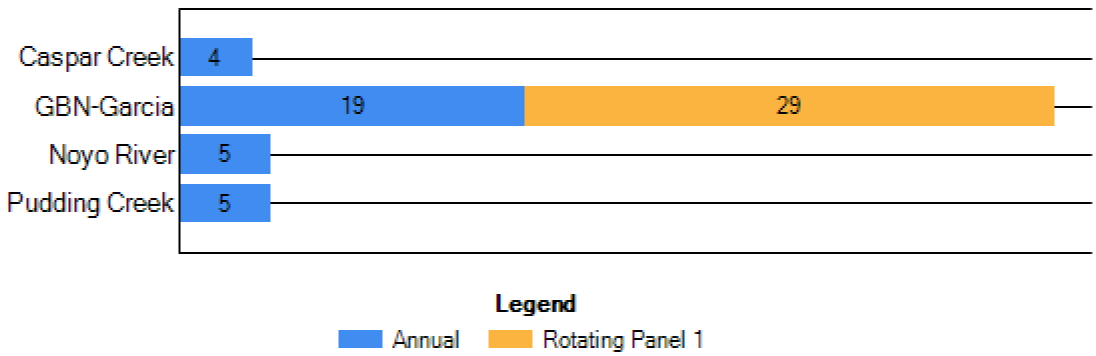


Figure 14. Spatial design consists of all sites visited each year, 40% annual and 60% in rotating panels.

The following example describes the CMP design framework and how it is leveraged in order to estimate annual salmon abundance and trends. The first step is to create the sample-frame; this is a GIS process which removes tributaries and estuary areas that are not freshwater habitat for salmon. The red lines in Figure 15 represent known spawning habitat and the light blue stream areas are what is removed. This is a complicated process involving sophisticated, program modeling, species presence and stream barrier data analysis, combined with expert opinion (Garwood and Ricker 2011) and carried out using a GIS. Once the sample universe has been established then it is broken into one to three kilometer reaches, numbered recursively, and a computer process is run, using a statistical software package called SDraw (McDonald 2003), to assign each reach a spatially balanced random sample order or draw number. Any sequentially numbered collection of reaches from this GRTS order constitutes a spatially balanced random sample. The entire collection of GRTS ordered sample reaches are assigned to a rotating panel (refer to Figure 5 in the luReachSchedule sub-section).

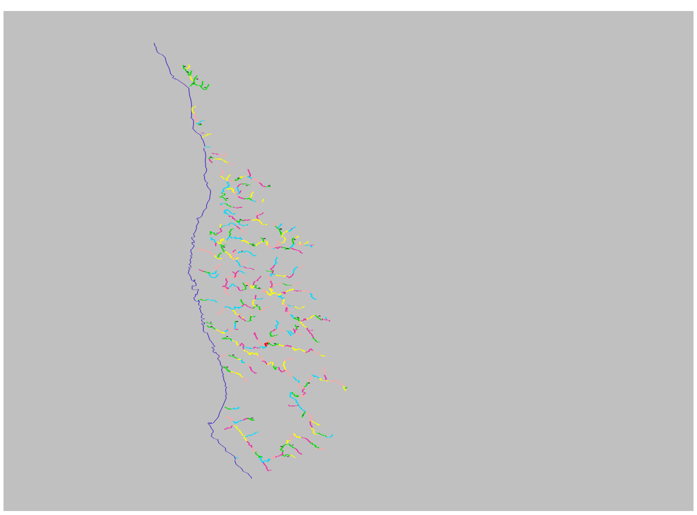

Figure 15. GRTS ordered sample frame for coastal Mendocino County consisting of 339 potential sampling reaches.

To estimate escapement redds are counted during spawning ground surveys in selected reaches following strict field protocols. For our example, 41 of 339 reaches were sampled in year one (Figure 16). Estimating salmon abundance from this is relatively straight forward; once the sample-frame has been defined the reaches are assigned a GRTS order and are "fit" to panels. The number of redds counted in each selected reach is given a weighted average, and this average is multiplied by the total number of reaches in the sample-frame so if we have an average of 10 redds in our sample of 41 reaches the total estimate would be 3,390 (10 redds per reach multiplied by 339 reaches). The redd counts are converted to fish numbers using spawner to redd ratios (see Gallagher et al. 2010). In this way the CMP Aquatic Surveys design provides estimates of the number of salmon returning to spawn per diversity strata (Table 1). Each annual estimate is an index of population status, and after a few years of sampling we can begin to evaluate trends in these populations (Figure 17). Moreover, because of the importance placed on providing this critical information to program managers and the public, data must be collected according to this prescribed, albeit complex, design strategy, and by way of these examples we hope that this has become clearer, i.e., critical data elements, mapped to a randomly balanced spatial and temporally based design are key to the sampling efforts that are employed in order to develop usable information needed for management of endangered California coastal salmon populations. Therefore, it is critical that our data management solution incorporate the spatial and temporal aspects of the CMP Aquatic Surveys design.

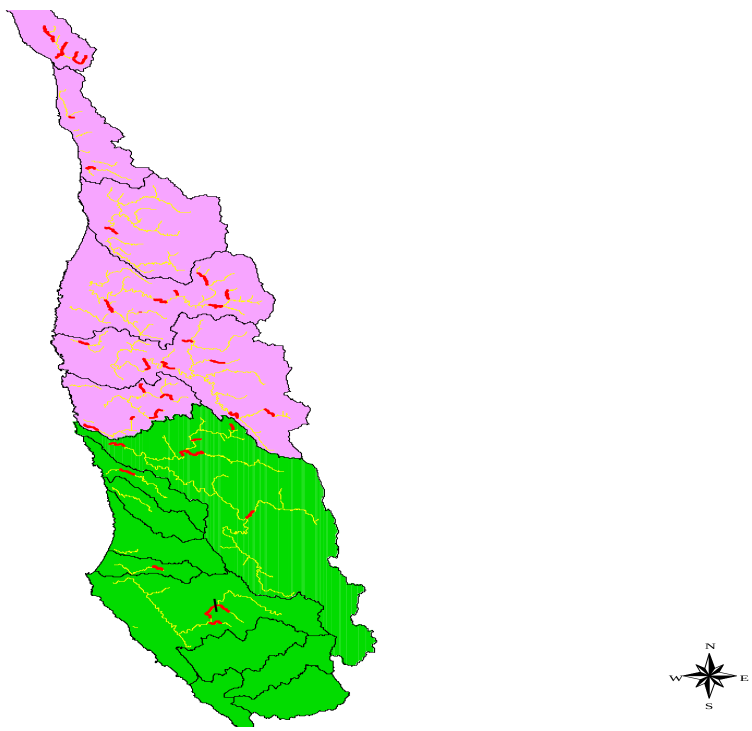


Figure 16. Selected sample reaches for year one regional escapement estimates in coastal Mendocino County: red lines indicate selected reaches; yellow lines are the entire sample frame; pink and green areas represent the two diversity strata.

Table 1. Coho salmon escapement estimates (95% confidence limits) for coastal Mendocino County California, 2009 to 2011: ns = not surveyed, na = not available, and DS = diversity strata. Precision is the 95% confidence limit half widths relative to the mean as a composite of the three years shown.

| **Stream** | **N** | **Number of Adults** | | | **Precision** |
|---|---|---|---|---|---|
| | | 2009 | 2010 | 2011 | |
| Mendocino Coast | 41 | 887 (415 to 1545) | 898 (555 to 1308) | 1575 (534 to 2947) | 61% |
| Lost Coast DS | 32 | 672 (295 to 1083) | 1059 (515 to 1711) | 1318 (328 to 2700) | 69% |
| Navarro Point DS | 9 | 158 (41 to 342) | 513 (108 to 989 | 176 (18 to 369) | 94% |
| Albion River | 3 | 8 (0 to 22) | 0 | 99 (0 to 297) | 148% |
| Big River | 6 | 80 (0 to 210) | 134 (20 to 214) | 147 (0 to 435) | 122% |
| Big Salmon Cr. | 2 | 0 | ns | ns | na |
| Brush Cr. | 1 | 0 | 0 | 0 | na |
| Caspar Cr. | 6 | 6 | 5 (3-9) | 30 | na |
| Cottaneva Cr. | 1 | 0 | 0 | ns | na |
| Garcia River | 3 | 69 (0 to 206) | 9 (0 to 18) | 65 (13 to 130) | 166% |
| Greenwood Cr. | 1 | 9 | ns | ns | na |
| Little River | 2 | 4 | 2 | 2 | na |
| Navarro River | 6 | 124 (18 to 124) | 452 (159 to 790) | 137 (0 to 420) | 103% |
| Noyo River | 10 | 294 (82 to 573) | 286 (58 to 650) | 494 (24 to 583) | 79% |

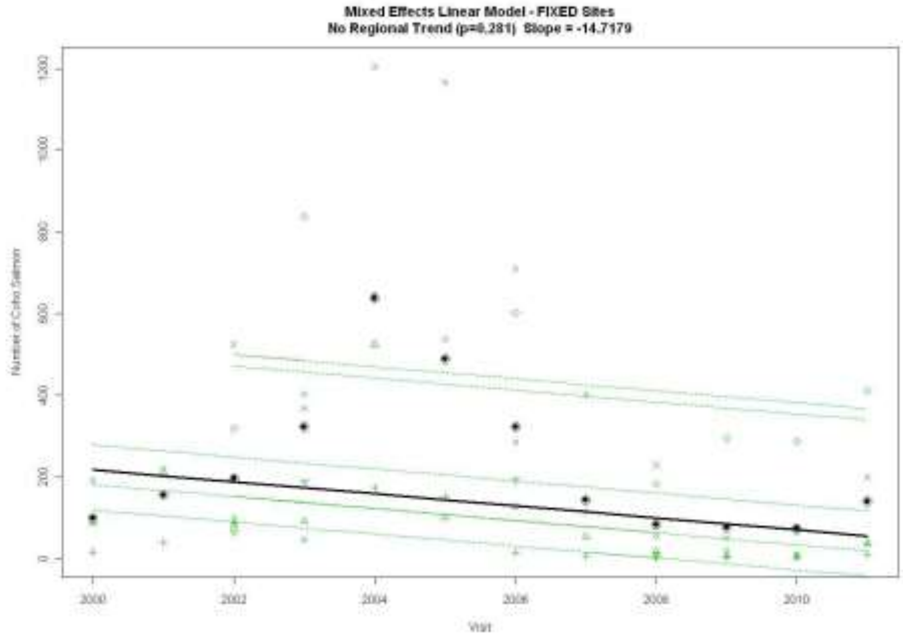


Figure 17. Coho salmon abundance trends in coastal Mendocino County lifecycle monitoring streams, 2000 to 2011 (from Gallagher and Wright 2012).

***Appendix 3 – User Stories***

The following are high level stories related to the "user" experience when using the current system, including protocols and techniques. Stories 1 to 4 are described in detail in Crain et al (in review). Field methods for spawning surveys are described in Johnson et. al. (2007).

**User Story 1 – Create the Spatial Information**

The GIS Data Manager develops stream based routed hydrography for the area of study. Various GIS tools and methodologies are used to edit existing hydrography, digitize and route new stream line work where spatial data is missing, and coarsely identify site suitability based on best available base data and practices. Site locations, whether or not ultimately selected by project leads, are assigned universally unique identifiers.

**User Story 2 – Review the Potential Site Locations**

Project leads review locations provided by the GIS Data Manager. Site visits are made to ground truth and establish reach specific characteristics, such as whether there is spawning habitat and suitability for a particular age class of salmon.

**User Story 3 – Setup the Survey Sample Scheme**

After standardized stream based hydrography is available and site locations selected, a rotating panel of site locations (stream reaches) is drawn and a rotating panel visit matrix is created, i.e., the "survey scheme." This is a joint effort of the GIS Data Manager, project lead, and statistician. When complete, survey schedule matrices are merged and maintained by the GIS Data Manager.

**User Story 4 – Establish the Survey Entity**

Prior to the onset of a survey season, i.e., an approximate 10 week period when crews visit and sample sites within a particular spatially balanced rotating panel and collect data in the field, project leads define the Survey entity, i.e., the time-range or period, location or geographical survey extent, and provide brief notes/comments about any particulars of the survey. The survey extent is a fixed, known location entity, previously defined in the GIS, that is referenced by way of a relationship between the survey entity and location entity.

**User Story 5 – Prepare for Field Data Collection**

Data gathers or surveyors prepare for survey sample site visits by putting together a "package" of equipment and survey materials, e.g., tools would include a GPS unit, handheld device, and various measuring tools, and materials, or "reach packet," would include specifics about the survey site or reach, driving directions, landowner contact information, details on gates and combinations, special considerations, e.g., call before entering, two vehicles needed for pickup or drop off, etc.; reach maps, and lists of special tools, such as ATV's, kayaks, dry suits, etc.

**User Story 6 – Initiate the Field Event**

Based on the project's predetermined rotating panel survey schedule the project lead advises surveyors of which site or reach location they will visit for the day; at that time they will be provided with any additional instruction and specific details on how to get there.

**User Story 7 – Go into the Field**

Once surveyors know where to go then they are assigned a vehicle; they put together any additional personal gear and load the needed field tools from their package of equipment and survey materials (Table 2). On their way to the site they fill the field vehicle up with fuel if needed.

Table 2. Spawning survey equipment checklist.

| | | |
|---|---|---|
| ☐ Spawning survey protocols | ☐ Chest waders and rain gear | ☐ Cell or satellite phone |
| ☐ PDA | ☐ or Dry suit | ☐ Contact and emergency numbers |
| ☐ Reach maps | ☐ Wadding boots | ☐ GPS Unit |
| ☐ Pencils, pens, and sharpies | ☐ Hat | ☐ Swift water safety gear |
| ☐ Field notebook | ☐ Polarized sunglasses | ☐ Machete |
| ☐ Tissue and scale collection materials | ☐ Field vest or back pack | ☐ Brush Axe |
| ☐ Knife with sheath | ☐ Calibrated Wading staff with gaff | ☐ Chain saw |
| ☐ Forceps for scale collection | ☐ Measuring tape (mm) | ☐ Food and water |
| ☐ Scale envelopes | ☐ Flagging | ☐ Extra clothing |
| ☐ Compass | ☐ Watch | ☐ Other |

**User Story 8 – Locate the Survey Site**

Surveyors use a combination of maps, textual instructions or directions, and GPS to find the downstream end of the prescribed sample survey site or reach which is on average about two miles in length. Once at the data collection location surveyors collect both course and detailed survey information that relates to that specific stream reach location and observation time instance.

During the data gathering effort much of the process can be automated through the handheld device, or by post processing after returning to the field office. This is possible, because some of the information collected is known in advance or can be assumed, e.g., who entered the data, date of data entry, measurement units used, such as meters or inches, type of measurement made, such as total length or fork length. In addition, the handheld device is capable, to a degree, of automating data processes, e.g., population of current date-time, numeric value minimum and maximums, etc. Other information options can be simplified by way of pick-lists, checkboxes, etc., that limit entry options based on what can be expected, or is appropriate, at a given site during a particular time period, e.g., species options.

**User Story 8 – The Header Table**

Before the surveyors begin heading upstream, they record the location, date-time, names of the data gatherers or observers, method of spawner survey used (this is usually "pedestrian"), standardized code for the weather conditions, water and air temperatures, and comments particular to the event; these could be brief or verbose. All of these attributes would be recorded as records in the Header table, and there would be only one Header record for a given reach or site location and observation date-time instance. These sample sites are fixed, known location entities, previously defined in the GIS or sample-frame, and they are referenced by way of a relationship between the Header entity and Location entity.

**User Story 9 – Collecting the Header Information**

Now begins collection of information specific to the Header event or record. Surveyors work their way upstream looking for live and dead salmon as well as redds until they get to the upstream end of the survey sample reach. For each instance they use a GPS device to determine the latitude and longitude to about five or six decimal places (plus or minus two to three meters). Getting this level of accuracy may require remaining at that location for several minutes, however new redd instances are not as common as revisits to prior observed redd locations, and GPS coordinates for fish observation locations are not always collected nor used by all projects. When a coordinate pair is determined then a decimal value error factor is also obtained.

When live or dead fish are spotted then information is collected about the individual fish, and this would be recorded as part of detail attributes. This detailed information resides at various sub-levels to the Header entity. The first, and most prominent, sub level discussed here is the Individual data object or table.

**User Story 10 – The Individual Table**

When surveyors observe fish, the following information is determined about each fish and recorded as records in the Individual table, having an immediate sub-data table relationship to a record in the Header table, and an optional sub-data table relationship to a record in the Count table: an optional unique fish identifier, or fish record number, is created in the field by the person recording the information – this unique record number is the concatenation of several bits of information and consists of the day, month, and number (n + 1 from 001 to 999); as described above, GPS coordinates; identifier for the distance upstream in meters from the beginning of reach to where the individual was observed, i.e., meter marker number (a flag or marker with this number is placed on the adjacent stream bank); condition of the fish, i.e., code which indicates the condition or general health of the fish; species identification; observer's certainty of the species, e.g., sure, maybe, or unknown; length and weight of the fish; gender of the fish, i.e., male, female, or unknown; observer's certainty of the individual's gender recorded as sure, maybe, or uncertain; whether fish is alive; whether fish previously was, or has currently been tagged, clipped, or marked; whether fish originated from a hatchery; whether fish was migrating upstream or downstream; injuries that the fish has sustained; whether the fish has the blackspot disease or not; and last, brief notes about any issues specific to the individual fish.

**User Story 11 – The Redd Table**

When surveyors observe a redd, the following information is determined about each redd and recorded as a record in the Redd table, having an immediate sub-data table relationship to a record in the Header table: unique redd identifier, or redd "record number" is created in the field by the person recording the information. This identifier is the concatenation of several bits of information, i.e., the day, month, and number (n + 1 from 001 to 999); as described above, GPS coordinates; bearing and distance is recorded, i.e., from where the redd is to where the surveyors tied flagging tape which marks the redd location; species of fish that built the redd is recorded as well as measurements for the redd dimensions. The redd structure is called a "pot," and its length, width, and depth are determined and recorded. The downstream end of the pot is called the "tail spill," and its length and width is also recorded as well as information about the substrate material composition. Surveyors determine the redd’s age in terms of standardized age codes and whether it was re-measured or not; and last, surveyors record brief notes regarding any particulars of the redd, i.e., one to three sentences.

**User Story 12 – The StreamHabitat Table**

When surveyors stream habitat survey they use classification techniques in order to describe specific habitat within a stream (Flosi et. al. 1998). In the CMP model this information has an immediate sub-level relationship to a record in the Header table; while classifying habitat type and measuring habitat units, surveyors might also collect other parameters, such as gravel quality, cover, shade canopy, pool depth, and bank condition; in addition, fish counts from biological surveys, such as snorkel and/or electrofishing, might be collected in conjunction with stream habitat data for analysis of habitat diversity and suitability for salmon and steelhead; and last, surveyors would record comments regarding any particulars of the of the stream habitat unit survey event.

**User Story 13 – The Snorkel Table**

When surveyors snorkel dive a reach, they visually count adult salmon and steelhead prior to spawning. This class of data, collected using underwater survey methods, comprises estimates that are used as an index of fish abundance; having an immediate sub-data table relationship to a record in the Header table; this information is not used for absolute counts. In addition to estimating counts of fish, surveyors record the stream habitat unit number (unique for a given survey event or effort), number of persons in water actively observing, percentage of dive unit surveyed during this effort, and comments regarding any particulars of the of the dive event.

**User Story 14 – The Efish Table**

When surveyors electrofish (e-fish) a reach, they use electricity to stun fish before capturing them with dip nets. This class of data, collected using in-stream survey methods which include backpack and boat mounted electrofisher equipment, comprises estimates that are used to determine abundance, density, and species composition – information having an immediate sub-data table relationship to a record in the Header table which can correlate with counts from other catch techniques, such as snorkel surveys. In addition to collecting counts of captured fish: surveyors record the stream habitat unit number (unique for a given survey event or effort); a number of other parameters, such as pulse rate, intensity of the electric field, and water conductivity in order to determine effectiveness of the capture effort; as well as comments regarding any particulars of the of the electrofishing event.

**User Story 15 – The Count Table**

When surveyors do not have time to collect information about each individual fish, or in addition to collecting information about each fish, batch count information may be collected and recorded in the Count table, having an immediate sub-data table relationship to a Header record. The following information is recorded each time a batch count is made: as described above, GPS coordinates; species identification; life stage, e.g., adult, sub-adult, etc.; actual or estimated number of individual fish counted; and last, brief notes about any issues specific to the batch count.

While collecting information about an individual fish, other tables are used to collect information as well, i.e., information that has a sub-data table relationship to an Individual record. These tables are for clips and/or marks, the “ClipMark” table; tags or the “Tag” table; fish that are observed on redds, or the “FishOnRedd” table; and, fish biological collection samples, e.g., scales, otoliths, tissue, etc., recorded in the “Sample” table.

**User Story 16 – The ClipMark Table**

When surveyors clip or mark a fish, or observe clips or marks on a fish, the following information is determined about each clip or mark and recorded as records in the ClipMark table, having an immediate sub-data table relationship to a record in the Individual table: code which indicates the clip or mark type as well as position, e.g., adipose clip, left maxillary punch, etc.; code which indicates shape of opercle punch if one was used; status of clip or mark for this instance, i.e. applied, recaptured, or unknown. Higher resolution status information recorded at this level supersedes any status information recorded in the Individual table; and, optional brief notes about the clip or mark.

**User Story 17 – The Tag Table**

When surveyors tag a fish, or observe tags on or in a fish, the following information is determined about each tag and recorded as records in the Tag table, having an immediate sub-data table relationship to an Individual record code that indicates the type of tag, i.e., spaghetti, Disc, PITT, CWT, etc.; identification number associated with the tag; tag color; descriptor for location on the body where tag was observed, i.e., right, dorsal, etc.; status of tag for this instance, i.e. applied, recaptured, or unknown. Higher resolution status information recorded at this level supersedes any status information recorded in the Individual table; and, optional brief notes about the clip or mark.

**User Story 18 – The FishOnRedd Table**

If a fish is seen associated with a redd then additional information – that has an immediate sub-data table relationship to an Individual record – is recorded. This single element of data is the unique redd identifier, i.e., the redd record number from the above mentioned Redd table. This identifier must have already been created as a previously recorded Redd record, and is merely being referenced here. For each Individual table record there can be multiple FishOnRedd records, i.e., a redd can have more than one fish associated fish.

**User Story 19 – The Sample Table**

While observing and recording information about individual fish if biological samples are collected from the fish then additional information, that has an immediate sub-data table relationship to a record in the Individual table, is recorded. This information should indicate what type of specimen sample was collected, e.g., scales, tissue, otoliths, head, stomach, whole body, and other. In addition, if the collected biological samples are to be cataloged in a specimen collection library then codes must be provided to both identify the sample, e.g., PITT tag number, user-defined information, etc., and indicate how the collection item is organized at its storage location.

In addition to the above tables, for each Header record a list of one or more surveyors are recorded in a Surveyors table. This list of surveyors allows for drawing a relationship between the collected information and who contributed to the effort.

**User Story 20 – Post-processing of Field Information**

After information has been gathered in the field, crews return to the field office and synchronize the handheld device with their project's distributed or local database. The information already has undergone some coarse level validation in the field, but once in the local database data analyzers quality control the information by way of a series of query tools and makes any necessary edits prior to the data being queried for further analysis.

**User Story 21 – Generating Useable Information**

After the raw field information that was gathered has been quality controlled, the project lead executes a series of queries to extract useable information for various metrics; this will be used for further analysis.

**User Story 22 – Statistical Analysis**

With usable information provided by a project lead, the statistician uses a series of routines to estimate the parameters of interest, such as annual abundance, and through analysis identifies trends in the data, determining significance if any.

**User Story 23 – The Final Products**

The project lead uses both the usable information generated earlier and the statistical analysis provided by the statistician to produce various summary datasets and reports. These final products are made available to both public and private entities on a product by product basis. Furthermore, at this point the data gathering season has ended and the distributed project level datasets can be merged for centralization and long-term secured storage.

These user stories describe user activities based on classes of data or tables within the current system, i.e., how each table is populated with data; however, in addition to these "class" database objects which represent themselves in the form of discrete tables, it should be noted that the database allows for the extraction of data based on a "Survey Type" which may include classes of information derived from a combination of these tables, e.g., "Out-migrant trapping" is a type of survey, and would include information from the following tables: Project, Survey, Header, Individual, Count, ClipMark, Tag, and Sample.