

**State of California**

**The Resources Agency**

**DEPARTMENT OF FISH AND WILDLIFE**

**Coastal Salmonid Monitoring Database Management:**

**Metadata and Protocol  
for the Aquatic Surveys  
Sampling-frame Attribute Table**

**Version 2.2 using CalStreams<sup>1</sup> Hydrography  
and the Aquatic Surveys<sup>2</sup> Database  
March 31, 2015**

**By**

David Dixon, Tom Christy, Douglas Burch, and Roderick Gonzalez  
Minor updates, July 9, 2015

---

<sup>1</sup> "CalStreams" is a California stream-based routed hydrography dataset based on the NHD 1:24K linework.

<sup>2</sup> The Aquatic Surveys Database, also referred to as the Aquatic Survey Program (ASP) is the current database system used by the Coastal Monitoring Plan (CMP) effort (Burch et. al. 2015).

## Contents

- **Introduction**
- **Data Dictionary**
  - **Comparison of the luLocation Lookup table and Sampling-frame Attributes**
- **Protocols**
  - **Incorporating Candidate Project Reaches into the Sampling-Frame**
- **Walkthrough**
- **Getting Help**
- **References**

## Introduction

The Sampling-frame is the stream sample universe for fisheries surveys; it is the collection of all possible, approximately, one to three kilometer stream segments (also referred to as reaches) in all coastal California streams that contain fish of interest and, in time, the sampling-frame may be expanded to cover all salmonid bearing California Streams both coastal and inland. In addition, the sampling-frame “universere” consists of a series of individual “Frames<sup>3</sup>” that are defined based on a species and/or a restricted geographic area, e.g., Eel River, California Central Coast Environmentally Significant Unit (CCC ESU), South Central and Southern California watersheds, etc.

The purpose of this document is to establish and define a set of core attributes that the CMP Technical Team’s Sampling-frame working group has determined should be inherent to the CMP Sampling-frame’s GIS layer. By “inherent” we mean a part of the feature class itself and not related by way of links to other tables, i.e., these are properties, represented by fields or columns, of the Sampling-frame GIS layer’s feature class. In addition to these inherent core attributes, there can be a variety of additional attributes, in related tables, that provide important information; however, we felt that the related information could be better managed by way of linked or relational attribute tables, e.g., location suitability for various species life stages (luReachLifeStage) and the location revisit schedule (luReachSchedule); these related attributes can be used to construct distribution maps, and for other analysis. For the purposed of this document, we not only use the term “related” as is it is defined for relational databases, but we also imply that these related attributes are “join-able” to the Sampling-frame’s attribute table. Our goal here is to use this document to establish an agreed upon set of core attributes for the Sampling-frame’s feature class. Once agreed upon, we give it a version number. A given version may be developed over several drafts, but once a versioned format has been adopted by the Technical Team we can begin working with these attributes as they are defined, and can assume that they will not change until the Technical Team’s Sampling-frame subgroup is in agreement with adopting a new or updated version.

Although, this document is informational in nature, and intended for a wide-range of ASP Database audiences and stakeholders, it is primarily a tool for facilitating a division of workload and agreement between the ASP Database GIS Data Manager and Tabular Data Manager (See Burch et. al. 2015 for detailed explanation of terminology related to the ASP Database.)

## Data Dictionary

**ObjectID** – [Geometry] Assigned by ArcGIS and created in the GIS. (BLOB data type)

---

<sup>3</sup> “Frame” implies a subset of the Sampling-frame universe specific to a single project, and this shorthand terminology is most often used when referring to portions of the Sampling-frame that are still under development.

**Shape** – [Geometry] Assigned by ArcGIS and created in the GIS. (BLOB data type)

**LocationID** – [Primary Key] Identifier for maintaining global uniqueness of the location entity, e.g., a stream reach. This attribute is populated by the GIS Data Manager when the feature is created in the GIS at the onset of a project or when a project adopts use of the Sampling-frame; once created, this value can never be changed or reused. If a reach is split then two or more new location entities would need to be created with their own unique identifiers. If a reach is expanded then new location entities would need to be created, or at least one for the extended portion. These new location entities, created in a GIS would then be appended to the ASP database's luLocation lookup table by either the Tabular Data Manager or the project lead. (GUID data type)

**ProjID** – [Required] Unique integer identifier for the CMP project, e.g., Mendocino coast streams, etc. This value must be created in the CMP database before it can be referenced by the Sampling-frame GIS layer. This value gets created in the CMP database by the Tabular Data Manager when a project is added to the CMP, i.e., prior to the onset of inclusion of the project into the CMP database effort. (Short Integer data type)

**Description** – [Required] Brief description of the location. This can be created in the CMP database or in the GIS. However, once defined in the CMP database it should only be updated there. These location descriptions can be created and/or updated by way of a secured web interface, by the Tabular Data Manager or project leads. (Variable Text, 255 character data type)

**FrameOrder** – Sample reach recursive numbering order, north to south and mouth to headwater. Like the LocationID, this value is created as part of the GIS process by the GIS Data Manager at the onset of a project or when the project adopts the GRTS<sup>4</sup> Sampling-frame methodology. These values become the LocationCode values when the [project] Frame has been finalized. (Floating Decimal data type)

**DrawOrder** – Sample reach order output from the GRTS S-Draw program. These values are created by the GIS Data Manager when the S-Draw program has been run on the FrameOrder values. The "draw" is the actual execution of this methodology to develop a set of random draw order numbers and is performed by the GIS Data Manager. (Long Integer data type)

**TempOrder** – Temporary sort index field used during development or modification of the Sampling-frame's FrameOrder or DrawOrder numbers. Older index numbers or in-development numbers can temporarily use this field, but, ultimately, no longer used or needed sort index values would be archived to a crosswalk table -- see "Crosswalk and Supplemental Attribute Tables" document<sup>5</sup> for more information on archiving deprecated index numbers. (Floating Decimal or Double data type)

**GeoArea** – Name of the "watershed" that contains the sample reaches. This value may be derived from one or more US Geological Survey Watershed Boundary Dataset HUC area delineations. The value is then used to populate the ASP Database's GeoArea attribute. (Variable Text, 255 Characters data type)

**StreamName** – Name of the stream that contains the sample reach. This value is derived from the CA\_Streams GIS layer. If the stream has no name in the CA\_Streams GIS layer, the DFGWaterID will be used instead. The value is then used to populate the ASP Database's GeoUnit attribute. (Variable Text, 255 Characters data type)

---

<sup>4</sup> "GRTS" is 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 coastal monitoring effort.

<sup>5</sup> An addendum to this document

**DFGWaterID** – Unique CDFW identifier for the stream that contains the sample reach. This value is derived from the CA\_Streams GIS layer's DFGWaterID. The value is then used to populate the ASP Database's GeoUnitID attribute. (Long Integer data type)

**BegGeoUnitMeas** – Begin measure for the downstream end of the linear section or reach, i.e., the demarcation point. This value is calculated in the GIS then used to populate the BegGeoUnitMeas attribute in the ASP Database's luLocation lookup table. Eventually this value is replaced by use of the DnLocationID attribute; once the DnLocationID attribute has been populated in the ASP Database's luLocation lookup table then it is assumed that the DnLocationID is a more accurate representation of the downstream demarcation point's location, and the BegGeoUnitMeas value can be disregarded or even deleted. (Floating Decimal data type)

**EndGeoUnitMeas** – End measure for the upstream end of the linear section or reach, i.e., the demarcation point. This value is calculated in the GIS then used to populate the EndGeoUnitMeas attribute in the ASP Database's luLocation lookup table. Eventually this value is replaced by use of the UpLocationID attribute; once the UpLocationID attribute has been populated in the ASP Database's luLocation lookup table then it is assumed that the UpLocationID is a more accurate representation of the upstream demarcation point's location, and the EndGeoUnitMeas value can be disregarded or even deleted. (Floating Decimal data type)

**BegXCoord** – Beginning X coordinate or longitude, i.e., decimal degrees value for the downstream demarcation point of the linear stream section or reach. Generated by the GIS Data Manager at the onset of a project or when the project adopts the Sampling-frame methodology. Eventually field crew should ground truth this using a GPS, or carefully digitize the demarcation point coordinate from a program such as Google Earth. With the ground truthed data, "marker" type location entities should be created for the ASP database's luLocation lookup table. These actual ground-truthed demarcation coordinates will thereafter be referenced by way of the ASP Database's reach location records, i.e., using the DnLocationID attribute; in the luLocation lookup table the XCoord value can be disregarded or even deleted after the DnLocationID attribute has been populated. (Double data type to 6 decimal places)

**BegYCoord** – Beginning Y coordinate or longitude, i.e., decimal degrees value for the downstream demarcation point of the linear stream section or reach. Generated by the GIS Data Manager at the onset of a project or when the project adopts the Sampling-frame methodology. Eventually field crew should ground truth this using a GPS, or carefully digitize the demarcation point coordinate from a program such as Google Earth. With the ground truthed data, "marker" type location entities should be created for the ASP database's luLocation lookup table. These actual ground-truthed demarcation coordinates will thereafter be referenced by way of the ASP Database's reach location records, i.e., using the DnLocationID attributes; in the luLocation lookup table the YCoord value can be disregarded or even deleted after the DnLocationID attribute has been populated. (Double data type to 6 decimal places)

**EndXCoord** – Beginning Y coordinate or longitude, i.e., decimal degrees value for the upstream demarcation point of the linear stream section or reach. Generated by the GIS Data Manager at the onset of a project or when the project adopts the Sampling-frame methodology. Eventually field crew should ground truth this using a GPS, or carefully digitize the demarcation point coordinate from a program such as Google Earth. With the ground truthed data, "marker" type location entities should be created for the ASP Database's luLocation lookup table. These actual ground-truthed demarcation coordinates will thereafter be referenced by way of the ASP Database's reach location records, i.e., using the UpLocationID attribute. (Double data type to 6 decimal places)

**EndYCoord** – Ending Y coordinate or longitude, i.e., decimal degrees value for the upstream demarcation point of the linear stream section or reach. Generated by the GIS Data Manager at the onset of a project or when the project adopts the Sampling-frame strategy. Eventually field crew should ground truth this using a GPS, or carefully digitize the demarcation point coordinate from a program such as Google

Earth. With the ground truthed data, "marker" type location entities should be created for the ASP Database's luLocation lookup table. These actual ground-truthed demarcation coordinates will thereafter be referenced by way of the ASP Database's reach location records, i.e., using the UpLocationID attribute. (Double data type to 6 decimal places)

**Notes** – Brief notes or succinct metadata for the location entity, i.e., the reach. This text value can be used to initially populate the Notes attribute in the ASP Database's luLocation lookup table. (Variable Text, 255 Character data type)

**UpdDate** – Date that location attributes were created or last updated. (Date-time data type)

**LengthKM** – Sample reach length in kilometers. This value is derived from the CA\_Streams GIS layer. (Floating Decimal data type)

**Chinook** – Chinook salmon expected to be found in the given sampling reach, based on historic data and local knowledge. This field is used to help define the sample universe, and information contained within it is used by the project lead to populate the ASP Database's luReachLifeStage lookup table. (Short Integer)

**Coho** – Coho salmon expected to be found in sample reach, based on historic data and local knowledge. This field is used to help define the sample universe, and information contained within it is used by the project lead to populate the ASP Database's luReachLifeStage lookup table. (Short Integer)

**Steelhead** – Steelhead expected to be found in sample reach, based on historic data and local knowledge. This field is used to help define the sample universe, and information contained within it is used by the project lead to populate the ASP Database's luReachLifeStage lookup table. (Short Integer)

**LifeStage** – Life stage of fish expected to be found in the sample reach, based on historic data and local knowledge, e.g., adult, sub-adult, etc. Information contained within this field is used to help define the sample universe, and it is used by the project lead to populate the ASP Database's luReachLifeStage lookup table. (Short Integer)

**Juvenile** – Juveniles expected to be found in the sample reach, based on historic data and local knowledge. This field is used to help define the sample universe. This information is used by the project lead to populate the ASP Database's luReachLifeStage lookup table. (Short Integer)

**Adult** – Adult salmon/steelhead, i.e., potential spawners, that are expected to be found in the sample reach, based on historic data and local knowledge. This field is used to help define the sample universe. This information is used by the project lead to populate the ASP Database's luReachLifeStage lookup table. (Short Integer)

## **Comparison of the luLocation Lookup table and Sampling-frame Attributes**

There are attributes that we only use in the Sampling-frame's GIS layer while in the luLocation lookup table there are attributes which would not be useful outside of the tabular realm. Attributes, or fields, in the luLocation lookup table are used for referencing known locations; these "referenced" locations are opposed to locations that are created, on the spot, in the field, such as coordinates for fish redds or carcasses. For example, fish carcass locations would not be known until observation time and could change from one visit to the next. Where there is overlap between the tabular and spatial, to the degree possible, we have used the same attribute names, datatypes, and field definitions so that there is parity between the database and GIS; we feel that this minor duplication is necessary and can be managed by way of a protocol that establishes which attribute is the "master" and which is the "copy." For Sampling-frame reach locations only, the feature gets created in the Sampling-frame layer by the GIS Data

Manager. When this process has been finalized the information is provided to the Tabular Data Manager who populates the ASP Database’s luLocation lookup table, and from that point forward updating of reach information is managed by the Tabular Data Manager and/or project staff by way of the luLocation lookup table so the GIS Data Manager is only in the development picture at the onset of Sampling-frame development and for reach type locations specifically

The tables below illustrate similarities between the ASP Database’s luLocation lookup table and the Sampling-frame’s attribute table; analogous fields have been color-coded to get a better visual representation of similarities between attributes or fields:

<b>luLocation (ASP database)</b>
<b>LocationID</b>
<b>ProjID</b>
LocationType
<b>Description</b>
<b>LocationCode</b>
<b>DrawOrder</b>
IsSurveyExtent
<b>GeoArea</b>
<b>GeoUnit</b>
<b>GeoUnitID</b>
<b>BegGeoUnitMeas</b>
<b>EndGeoUnitMeas</b>
GeoMeasUnits
<b>XCoord</b>
<b>YCoord</b>
CoordSys
DnLocationID
UpLocationID
GISLayer
<b>Notes</b>
<b>UpdDate</b>

<b>Sampling-frame (attributes table from GIS)</b>
<b>ObjectID</b>
<b>Shape</b>
<b>LocationID</b>
<b>ProjID</b>
<b>Description</b>
<b>FrameOrder</b>
<b>DrawOrder</b>
<b>TempOrder</b>
<b>GeoArea</b>
<b>StreamName</b>
<b>DFGWaterID</b>
<b>BegGeoUnitMeas</b>
<b>EndGeoUnitMeas</b>
<b>BegXCoord</b>
<b>BegYCoord</b>
<b>EndXCoord</b>
<b>EndYCoord</b>
<b>Notes</b>
<b>UpdDate</b>
<b>LengthKM</b>
<b>Chinook</b>
<b>Coho</b>
<b>Steelhead</b>
<b>LifeStage</b>
<b>Juvenile</b>
<b>Adult</b>

## Protocol

Once reach locations are created and finalized in the Sampling-frame, and universally unique LocationID values assigned, these location entities then become permanent members of the system regardless of whether the locations are ever surveyed or not. Exceptions to this protocol are as follows: location entities can be “changed” if data have yet to reference them, i.e., in-field re-definitions of the location are what we would consider a changed location, like splitting or lengthening of reaches. That being said, this does not include location accuracy improvements, such as demarcation point coordinates. During the

development period project leads with the help of others will make decisions regarding whether a reach is suitable for inclusion into the Sampling-frame: decisions regarding whether a reach should be removed from the Sampling-frame; whether the precision of coordinate information should be updated; whether the reach needs to be split or created anew; and, whether a reach should be added or not. The GIS Data Manager is there to facilitate this process, but is otherwise not the one that ultimately decides where a reach begins and ends. The LocationID values themselves can never change, and once created they are the only identifier that can be considered universally unique for a given location entity, and referenced as such. Another exception is that a LocationID and its associated reach attribute information can be removed from the Sampling-frame and ASP Database if no data is associated with it, e.g., a reach location can be removed from the Sampling-frame if it is later determined to be above a barrier and inaccessible to fish; however, its LocationID and FrameOrder values would be "archived" from the system rather than deleted, i.e., flagged and not used again. On the other hand, the LocationCode and DrawOrder can be re-indexed as needed during the development and testing period. The goal is to have unique FrameOrder values for an entire, statewide Sampling-frame.

Once the Tabular Data Manager is provided with a copy of the Frame's GIS attribute table, fields in the luLocation lookup table can be populated. From this point on refinements to the Frame's sampling reach attributes happens in the ASP Database's luLocation lookup table first, and if necessary this information can be downloaded by the GIS Data Manager and used to update attributes in the Sampling-frame's GIS layer.

It is here that we begin the process of obtaining accurate coordinate pairs for the downstream and upstream reach demarcation points. This will require new location records in the luLocation lookup table: new records will need to be created for both the downstream and upstream demarcation points; LocationType values for stream demarcation points should always be set to a value of "marker" unless the coordinate is also used as an observation location, such as a weir, and "reach" type location records in the luLocation lookup table will reference the marker location records by way of the DnLocationID and UpLocationID fields.

- It is primarily the project lead's responsibility to create these new records in the luLocation lookup table. However, it is expected that there will be significant help from the Tabular Data Manager, a Regional CMP Coordinator<sup>6</sup>, and in some cases the GIS Data Manager. This information is important for developing accurate fisheries maps and making spatial and temporal comparisons of fish observations, e.g., distribution and trend analysis.
- At this point in developing the Sampling-frame locations, the relational attribute, or supplemental, tables can be populated, and this will primarily be the project lead's responsibility, e.g., creating new records in the luReachSchedule and luReachSpecies lookup tables. However, it is expected that there will be significant help from the GIS Data Manager, a regional CMP coordinator, and in some cases the Tabular Data Manager.

### **Incorporating Candidate Project<sup>7</sup> Reaches into the Sampling-frame**

For new projects being incorporated into the Sampling-frame, the GIS Data Manager, with assistance from the project lead, can use existing project sampling sites, i.e., legacy reaches, adapting the new

---

<sup>6</sup> The "Regional CMP Coordinator" role is typically filled by a CMP project lead that has a good working understanding of the system and can assist other project leads within close proximity.

<sup>7</sup> Candidate Projects are projects that would benefit the statewide programmatic, effort and are currently using the Aquatic Surveys Database but have not yet fully adopted use of its Sampling-frame methodology as well as other protocols of the CMP.

Frames to existing reach boundaries where appropriate and necessary. The GIS Data Manager does this by “fitting” what they already have to the new Frame, i.e., making adjustments to reach demarcation coordinates as needed.

## Walkthrough of the Sampling-frame Development Process

1. When initially setting up a new frame, the CalStreams 1:24K streams based hydrography will be used (derived from the NHD). General guidelines are as follows and adhere to the Sampling-frame development protocol outlined in *Spawner Survey Sampling-frame Development for Monitoring Adult Salmonid Populations in California* (Garwood and Ricker 2011):
  - a. Reaches will be the distance from one confluence to the next.
  - b. Reaches will typically be from one to two kilometers in length.
  - c. The GIS Data Manager and project lead have the discretion to modify these boundaries, for example:
    - i. ...in order to exclude regions of the stream not suitable to fish
    - ii. ...in order to create optimal sampling reach lengths
    - iii. ...in order to agree with access, put-in, and take-out points along the stream
2. Developing a frame is an iterative process of one to four attempts using “test” reaches and “test” draw order numbers.
  - a. The project lead will begin collecting data using these test reaches.
    - i. In some cases the test reaches will be removed from the list of sampling reaches due to lack of suitability.
3. In other cases, some longer reaches may be split into two or more reaches.
4. In order to create reach breaks where there are good landmarks, such as bridges, to facilitate navigation “to” and “from” the sampling area as well as while sampling is occurring, reach length may be adjusted.
5. When ready to begin the process, a project lead must want to use the GRTS methodology for a given survey area, i.e., a “spatial survey extent,” such as the “lower Eel river.”
6. To begin, the project lead provides the GIS Data Manager with GIS file(s) for the stream(s).
7. The project GIS files are compared with CalStreams and other GIS layers.
8. At that point, the GIS Data Manager provides a project lead with a “first attempt” at the new Frame with the DrawOrder column populated from the GIS Layer’s TempOrder column.
9. If the Frame’s linework, reach delineations, and frame order are accepted by the project lead then the GIS Data Manager runs the final draw and...
  - a. ...provides the project lead with the final frame. (The FrameOrder values cannot change at this point.)
  - b. ...provides Tabular Data Manager with the information for population of the luLocation lookup table.
  - c. ...the FrameOrder values can be used for LocationCode values if the reach has been GRTS drawn.
    - i. The Tabular Data Manager can execute this update if requested.
10. From this step onward, the Tabular Data Manager maintains the location data.
11. The project lead uses the luLocation lookup table to provide “start” and “end” demarcation points for all GRTS reaches, and these demarcation point locations are referenced by the reach location records using the DnLocationID and UpLocationID fields.



12. The project lead uses luReachSchedule lookup table to track what reaches were sampled in which year.
13. In addition, the project lead uses luReachLifeStage lookup table to track what reaches were for various species-life-stage combinations.
14. When the Frame is initially developed by the GIS Data Manager, stream route information might be provided in the GIS attribute table, e.g., measurements up the stream to where the reach begins and ends and the name of the hydrography (GIS layer) used to develop the measured stream line which represents the sampling reach. In some cases this information will be highly accurate, and this should be so stated in the metadata; however, in most cases coordinate information for downstream and upstream demarcation points (developed later and not inherent to the hydrography) is more accurate and supersedes the routed line measure.
  - a. The GIS Data Manager may want to later use the demarcation point coordinates to update the GIS layer.
  - b. If they exist the coordinate values for DnLocationID and UpLocationID should be used to define the precise reach "begin" and "end" demarcation points.

## Getting Help

Sampling-frame development and implantation is carried out through the CMP Technical Committee which is one of several groups that comprise the overall CMP effort, and more specifically executed by way of the Technical Committee's Sampling-frame Team. Because members change, roles are redefined, and responsibilities shift, in the matrix below we provide a list of current members (as of the date of this document) that may be contacted for help related Sampling-frame questions as well as overall support when the decision has been made to implement the Sampling-frame for surveying reaches using the GRTS methodology.

Member	E-mail	Agency
Bellmer, Russ (mgmt.)	Russ.Bellmer@wildlife.ca.gov	CDFW - Sacramento
Burch, Douglas (GIS)	doug.burch@wildlife.ca.gov	CDFW - N. Coast
Casagrande, Joel (fisheries)	joel.casagrande@noaa.gov	NOAA - S. Cruz
Christy, Tom (GIS)	Tom.Christy@wildlife.ca.gov	CDFW N. Coast
Dixon, David (GIS)	David.Dixon@wildlife.ca.gov	CDFW - Sacramento
Ferreira, Joe (statistician)	Joe.R.Ferreira@wildlife.ca.gov	CDFW - Sacramento
Gallagher, Sean (lead)	Sean.Gallagher@wildlife.ca.gov	CDFW - N. Coast
Gonzalez, Rodrick (GIS)	Roderick.Gonzalez@wildlife.ca.gov	CDFW - Sacramento
McCain, Dana (fisheries)	Dana.McCanne@wildlife.ca.gov	CDFW - S. Coast
Ricker, Seth (fisheries)	Seth.Ricker@wildlife.ca.gov	CDFW - N. Coast
Umlaaf, Gretchn (fisheries)	Gretchen.Umlauf@noaa.gov	NOAA - Sacramento

## References

Dixon, D., Christy, T., Burch, D. Coastal Salmonid Monitoring Database Management: Crosswalk and Supplemental Attribute Tables. California Department of Fish and Wildlife; 3/31/2015. [Cited 2015 March 31]. Available from: <https://nrm.dfg.ca.gov/FileHandler.ashx?DocumentID=100934>

Burch, D., Gallagher, S. Ricker, S., Bellmer, R. Coastal Salmonid Monitoring Database Management: Report and Concept for Success. California Department of Fish and Wildlife; 1/29/2015. [Cited 2015 March 31]. Available from: <https://nrm.dfg.ca.gov/FileHandler.ashx?DocumentID=92766>

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.

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](http://www.west-inc.com/reports/grts.pdf).