California Recreational Fisheries Survey (CRFS) Data Access

FY 2015 Proposal

Connie Ryan
Created: 05/13/2015

1. Overview

1.1. Sponsor

Ed Hibsch, MRIP IMT sponsor

1.2. Focus Group

Information Management

1.3. Background

The marine waters off California's 1,100 miles of coastline are home to a diverse array of fish species. Many of these species are targeted by recreational anglers who take more than four million fishing trips annually in California's marine waters. The CRFS collects data on these trips and the resulting catch using eight field surveys and a telephone survey. Each year CRFS interviews over 60,000 angling parties in the field, identifies over 200,000 fish to species and measures about 120,000 fish. In addition, about 26,000 licensed anglers are interviewed via a telephone survey. The CRFS software system produces estimates of total catch in numbers of fish and by total weight, and effort by month and by six geographic districts. The CRFS data and estimates are used by a number of federal agencies, State agencies, and regulatory bodies to manage marine fisheries. It is critical that Pacific RecFIN and data users have access to documentation of all data elements (e.g., meaning, origin, usage, relationships, level and formats) so that data are not misinterpreted or used inappropriately in the data tables created by Pacific RecFIN or in stock assessments, monitoring of annual catch limits, regulatory development and other analyses. The CDFW submits the CRFS data and estimates to Pacific RecFIN monthly. Pacific RecFIN makes data and estimates from Washington, Oregon and California available to the public on its website, www.recfin.org. In 2013, about 60 percent of the records on the Pacific RecFIN public website were from CRFS with additional confidential CRFS records (e.g., telephone survey data and fishing location data) on a private (confidential) Pacific RecFIN website. The CRFS data is held in a fully relational Microsoft SQL Server Enterprise database maintained by CDFW. In addition to supporting the collection and estimation of the survey data, the database also aids in making monthly draws, importing data from a telephone survey, and providing data to assist the CRFS staff in tracking assignments and performing data quality checks. Currently, there are approximately one hundred data tables in the CRFS database. In addition, there are over forty tables that provide an audit history by automatically generating a backup record whenever data is added, edited, or deleted in the CRFS primary data tables. Although we have started the process of creating the prescribed data dictionary elements, there is a significant amount of work remaining to document all of the tables and columns in the database. The development of the full data dictionary will assist all current and future users of the CRFS data. The dictionary will provide a definitive source for the meaning, rules, and documentation associated with each data element. Currently, Pacific RecFIN stores data in SAS files, but they are in the process of converting their system to a Microsoft SQL Server database. The switch from SAS to Microsoft SQL Server by Pacific RecFIN will require CDFW to develop new procedures (and documentation of those procedures) for the transmission of data to the Pacific RecFIN. These procedures will involve developing data extracts and transformations that allow the data to be efficiently and reliably transmitted and used by Pacific RecFIN. In addition, CDFW staff will need to revise, review and test all transmissions of data. Coordination between CDFW and Pacific RecFIN is critical as Pacific RecFIN creates its new database so that data elements don't have the same name but different meanings, or have data elements that appear to hold the same data but instead hold aggregated sums of data. This project will facilitate coordination and will provide Pacific RecFIN will clear documentation for all the data elements that CDFW provides to Pacific RecFIN.

1.4. Project Description

The project consists of two major tasks: (1) development of a comprehensive data dictionary for the CRFS data to assist all users in understanding the meaning, scope, and rules associated with each of the data elements, and (2) development and documentation of new procedures for transmission of CRFS data and estimates to the Pacific Recreational Fisheries Information Network (Pacific RecFIN). The data dictionary will store the metadata for the data collected and the estimates generated for CRFS. The survey data are currently stored in a fully relational Microsoft SQL Server Enterprise database. We propose to create and store the metadata in a set of Microsoft SQL Server tables. These tables and their contents will adhere to California Department of Fish and Wildlife (CDFW) data dictionary standards. The data dictionary can be easily shared with Pacific RecFIN and data users, and will describe the tables in the CDFW CRFS Microsoft SQL Server database and the relationship between tables. Information about each data element will include: (1) The data element name; (2) A brief description of the meaning of the data element; (3) A longer description of the data stored in this data element (e.g., origin or how it was collected, the survey instrument, usage); (4) Range of acceptable values; (5) Validations performed to ensure data integrity; (6) Relationship to other data elements; (7) Comments about the data element. The comments generally will not be shown to users of the database (e.g., users accessing the data through a cube or data warehouse), but will be made available through an online tool, or in a written format. These comments would include links to the CRFS Methods Documents, descriptions of how the variable may have been derived or calculated. The software scripts created as part of this project will extract and transform the CRFS data into formats that are conducive for reporting and transmission. This will allow CDFW to efficiently and reliably submit the data and estimates to Pacific RecFIN on a monthly basis.

1.5. Public Description

1.6. Objectives
(1) Provide data users with documentation on all CRFS data and estimates by completing a comprehensive data dictionary for the CDFW fully relational Microsoft SQL Server database that houses CRFS data and estimates. (2) Facilitate coordination with Pacific RecFIN as Pacific RecFIN develops a Microsoft SQL Server database. (3) Efficiently and reliably submit CRFS data and estimates to Pacific RecFIN when the new Pacific RecFIN database is completed by creating CRFS data extracts and transformations.

1.7. References

2. Methodology

2.1. Methodology
The data dictionary is stored in a Microsoft SQL Server database. The CDFW have been developing an online tool for entering data into a data dictionary. This project will use that tool, or a simple SQL script for entering data. The project will involve developing short, user-friendly names and descriptions for each of the data elements, and developing longer comments for each element. The longer comments will provide links to other forms of documentation such as our CRFS methodology papers, and descriptions of the data collection and estimation procedures. The transformations will be developed as Microsoft SQL Server scripts, and the documentation will be in both the scripts and in the data dictionary. The data dictionary and associated transformations will cover all survey data and all data supporting the collection and reporting of the survey data and estimates. An Application Software Specialist will be hired for nine months to work directly with the CRFS Programmer/Analyst in the CDFW Data and Technology Division in Sacramento, California. The current CRFS Programmer/Analyst will guide and monitor all work by the consultant, on a daily basis. The final decision on all data dictionary elements will be made by the CRFS staff, and will be transmitted to them bi-weekly for their approval.

2.2. Region
Pacific

2.3. Geographic Coverage
California

2.4. Temporal Coverage
Not a data collection project

2.5. Frequency
Not a data collection project

2.6. Unit of Analysis
Not a data collection project

2.7. Collection Mode
Not a data collection project

3. Communication

3.1. Internal Communication
COMMUNICATION AND PROJECT TRACKING: The Team Leader and Application Software Specialist (who will be hired specifically for this project) will work in the same office and will communicate daily, and the Team Leader will track progress weekly. Progress and blockages will be addressed in the CDFW CRFS data team weekly check-in calls. The entire team for the CRFS Data Access Project will have a project call once per month. The Team Leader and Application Software Specialist will contact other team members for input as needed via e-mail and telephone calls. The Application Software Specialist may make one trip to Portland to coordinate with the Pacific RecFIN programmer/analyst (who is a project team member). SHARING AND DISTRIBUTING INFORMATION AND PRODUCTS: The primary means of distribution will be through e-mail. Files that are too large to share via e-mail will be placed on an ftp site. The Team Leader will provide data dictionary and script updates to the Pacific RecFIN programmer/analyst on a regular basis.

3.2. External Communication
A monthly report will be submitted to the MRIP Operations Team using the MRIP reporting system. A final report will be posted on the MRIP reporting system and distributed to the Pacific RecFIN Technical Committee.

4. Assumptions/Constraints

4.1. New Data Collection
N

4.2. Is funding needed for this project?
Y

4.3. Funding Vehicle
Pacific RecFIN Grant awarded to Pacific States Marine Fisheries Commission.

4.4. Data Resources
CRFS database/ Microsoft SQL Server will be fully available to the Application Software Specialist who is hired with the funds for this project

4.5. Other Resources
CDFW will provide staff to lead and assist the Application Software Specialist, and will provide the required office space.

4.6. Regulations
No new regulations are required.

4.7. Other
(1) Assume that we will be able to find a Application Software Specialist (programmer) with the requisite skills willing to work on a short-term project (about nine months) in Sacramento, California. (2) Assume funding will be available by September 2015.

5. Final Deliverables

5.1. Additional Reports
A final report containing the primary data elements of the CRFS system

5.2. New Data Set(s)
MS SQL Server tables containing CRFS data dictionary & MS SQL Server scripts for data transmission

5.3. New System(s)
none

6. Project Leadership

6.1. Project Leader and Members

First Name	Last Name	Title	Role	Organization	Email	Phone 1	Phone 2
Ed	Hibsch	RecFIN Programmer /Analyst	Team Member	Pacific States Marine Fisheries Commission	EHibsch@psmfc.org	503-595-3100	
Kevin	Hitchcock	Research Analyst (GIS)	Team Member	California Department of Fish and Wildlife	kevin.hitchcock@wildlife.ca.gov	707-576-2865	
Jeanne	Rimpo	CRFS Programmer /Analyst	Team Leader	California Department of Fish and Wildlife	jeanne.rimpo@wildlife.ca.gov	916-327-8767	

First Name	Last Name	Title	Role	Organization	Email	Phone 1	Phone 2
Connie	Ryan	Senior Environmental Scientist	Team Member	California Department Fish and Wildlife	connie.ryan@wildlife.ca.gov	650-631-2536	
Ashok	Sadrozinski	Environmental Scientist	Team Member	California Department of Fish and Wildlife	ashok.sadrozinski@wildlife.ca.gov	650-631-2535	
TBD	TBD	Application Software Specialist-hired with MRIP funds for this project	Team Member	Pacific States Marine Fisheries Commission			

7. Project Estimates

7.1. Project Schedule

Task #	Schedule Description	Prerequisite	Schedule Start Date	Schedule Finish Date	Milestone
1	Develop position description, identify interview team, and develop interview questions		08/03/2015	08/31/2015	
2	Advertise position for Application Software Specialist, interview candidates, make selection	1	09/01/2015	09/30/2015	
3	Hire Application Software Specialist and begin project	1,2	10/01/2015	10/30/2015	Y
4	update data dictionary and develop scripts for transmitting data to Pacific RecFIN	1,2,3	11/02/2015	06/30/2016	Y
5	write final report and submit final report and data sets	1,2,3,4	07/01/2016	07/29/2016	

7.2. Cost Estimates

Cost Name	Cost Description	Cost Amount	Date Needed
Overhead	overhead at 11.16%	$8365.00	09/01/2015

Cost Name	Cost Description	Cost Amount	Date Needed
Salary	9 months salary for Application Software Specialist	$60514.00	09/01/2015
Benefits	9 months of benefits for Application Software Specialist	$11442.00	09/01/2015
Travel	Travel with in CA to meet with CRFS staff and to Portland, OR to meet with Pacific RecFIN staff	$1000.00	09/01/2015
Equipment	Computer etc for Application Software Specialist	$2000.00	09/01/2015
TOTAL COST		$83321.00	

8. Risk

8.1. Project Risk

Risk Description	Risk Impact	Risk Probability	Risk Mitigation Approach
Funds for the project not available by September 1, 2015.	Would not start the project until funds were available, and thus, the project timeline would need to be adjusted.	Medium	Delay the project start date and adjust the timeline accordingly.
A project team member is unavailable for a significant period of time (e.g., a month).	The impact would depending on the expertise of the project team member. In the worst case scenario, the project might be delayed.	Low	If the project member is the only person with the needed expertise, then the timeline would be adjusted. If other team members or other staff outside the team have the expertise, then add a team member with the needed expertise.
Delay in the development of the Microsoft SQL Server database at Pacific RecFIN. The development of data export transformations for transmitting data to Pacific RecFIN depend on the completion of the Microsoft SQL Server database at Pacific RecFIN.	Impacts the transmission scripts, but not the Data Dictionary. The risk can be mitigated.	Medium	If the development of the Pacific RecFIN SQL Server database is delayed, we will attempt to develop “generic” transformations that can be modified later when their database structure is complete.
Not able to hire an Application Software Specialist with sufficient expertise.	This may delay the project completion, or may result in focusing on documenting the primary tables and to a lesser extent the lookup tables in the CDFW CRFS database.	Low	If an Application Software Specialist with sufficient expertise cannot be hired, we would train a less qualified individual.

9. Supporting Documents