MARINE RECREATIONAL INFORMATION PROGRAM

IMT: FY 2012 Project Proposal

Development of an Oracle System for the Southeast Region Headboat Survey

June 2012

Table of Content

To create Table of Content, right click this line then select 'Update Field'!
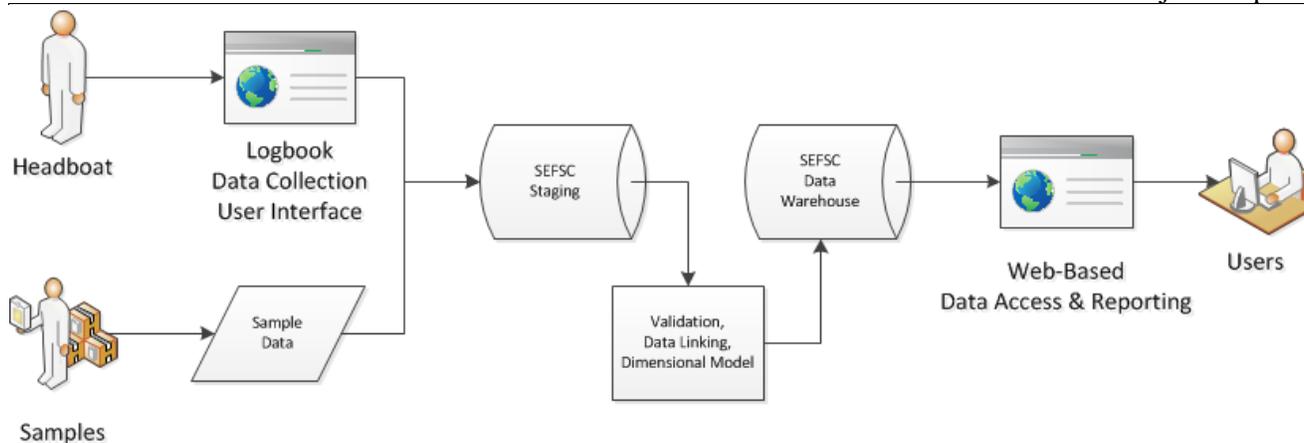
1. Overview

1.1. Background

As part of the ongoing effort to address the increasing demands on data collection programs for improved and timelier recreational data, the MRIP Operations Team approved FY12 funding for *Pilot Project, Phase II: Survey-Wide Implementation of Electronic Logbook Reporting on Headboats Operating in the U.S. South Atlantic and Gulf of Mexico.* This project will expand electronic logbook reporting to the entire universe of headboats in the Southeast Region Headboat Survey (SRHS) currently reporting on paper logbooks. The funding for the project will support the development of an internet-based software system with mobile applications for the South Atlantic and Gulf of Mexico headboat fishery.   Although this system will significantly increase the efficiency of collecting catch and effort data, it does not address the current deficiencies for storage and accessibility of incoming or archived headboat logbook data, biological data and other source data maintained at the NOAA Beaufort Laboratory.   Since 1981 the SRHS has maintained data files in two basic formats, ASCII (dta) and Dbase (dbf).   In recent years these formats have presented numerous problems, including compatibility with operating system updates, data extraction and limited storage capacity.   As data managers strive for consistency between management structures and data availability, these issues will impede the ability of the SRHS to achieve these objectives.   Furthermore, gains that are realized from electronic reporting may be offset by these deficiencies due to the lag time required to convert the data to the present formats.

1.2. Project Description

It is the goal of this project to transition the SRHS database to a complete Oracle system which will reside at the Southeast Fisheries Science Center (SEFSC), Miami, FL. This system would be a fully relational system incorporating the multiple components of the SRHS (logbook reports of catch and effort, reports of effort activity, biological data records and in time the new dockside and at sea catch effort interviews for verification of log reports), auditing and other quality control components, including data extracting (downloading) capabilities for offline analysis and use. Another goal of the proposed system would be to provide catch and effort information compatible (to the extent possible) for the MRIP estimates of catch and effort; thus the information would be electronically available for incorporation in national reports on recreational fisheries.

The project will address deficiencies in the storage and accessibility of incoming or archived headboat logbook data. The project will focus on developing an Oracle database repository of headboat data which will be hosted at the SEFSC. The data will be part of the SEFSC Data Warehouse. The new data will be integrated into the SEFSC Data Warehouse, which employs a modern underlying architecture and technology supporting a database that can produce gains in process efficiency and data security, as well as improved capabilities for reporting, analysis, and integration with other data sources.

1.3. Objectives

The main objectives of the project are:

1. Loading of data from Electronic Logbook Reporting system
2. Loading of data from port-side sampler data storage (include measuring board data)
3. Build data staging area at SEFSC for hosting raw data from data providers
4. Build data processing to validate, link, and merge data from sources
5. Load data into SEFSC data warehouse
6. Provide web-based data access for reporting and extraction (download local copy)

1.4. References

During project execution, the development team will seek the participation of data managers, business users, and IT staff to assist with the resolution of issues, design decisions, and coordination of activities. In addition, the team will refer to the following sources of information as needed:

- SE-HMS Data Integration Project - Approach & Design
- SE-HMS Data Integration Project - Data Storage Specifications
- Available documentation, coding, reference lists, and data files from the current SRHS system

2. Methodology

2.1. Methodology

This section describes the methodology to be employed by the development team during the execution of the project, in order to complete the requirements of: analysis, design, and data processing.

Activities will begin with a review of the project requirements and scope, with emphasis on user requirements and deliverables around a set of project goals that are common to all project stakeholders. Requirements gathering sessions and conference calls will be conducted as needed with users and stakeholders to confirm the functional requirements of the system.

The project team will document the data flow processing to serve as the basis for identifying required data fields and types. A workshop will be conducted with project team members to demonstrate and critique the design approach.

A checkpoint review of project scope will be conducted upon completing the first two project activities:

- Review/confirm project requirements
- Data flow processing design

With the project scope confirmed, the team will begin the integration effort by designing and developing the technical components that will support the data processing requirements. Users and stakeholders are kept informed on project progress and are updated as areas of the system are developed.

Historic data from the prior system (dBase database) will be migrated to the new system platform. Project progress and deliverables will be communicated through a project collaboration website accessible by all team members. Progress will be reported in ad-hoc meetings, broadcast emails, and the collaboration website.
Major phases of the project include:

Requirements Review

Project sponsors will confirm project requirements.

Data Processing Design

During this phase the team will design high-level specifications of the data loading, validation, and processing system.

Data Model

During this phase the team will document and entities of the database for staging area as well as SEFSC data warehouse.

Build Data Processing Procedures

During this phase the team will build and test the data loading, validation, and processing system components. As areas of the system are completed they will be made available for demonstrations.

Build Data Access User Interface

During this phase the team will build a web-based user interface for data access, online reporting, and data extraction (downloading).

2.2. Geographic Coverage

Southeast U. S through the Gulf of Mexico (North Carolina through Texas).

2.3. Temporal Coverage

The headboat survey has operated along the east coast of the U.S. since 1972 and began operations in the Gulf of Mexico in 1986 and continues to the present. Years when area coverage began are: North Carolina and South Carolina-1972; NEFL 1976; SEFL and the Florida Keys-1978; SWFL to South Padre Island TX-1986.

2.4. Frequency

It is anticipated that headboat data collected in electronic logbooks will be available for retrieval on a daily basis. Data originated in paper logbook forms, as well as biological sampling data collected by port agents requires submission to a forms processing vendor and is expected to be available on a monthly basis at the longest.

2.5. Unit of Analysis

Data for SRHS originates from biological samples taken by port agents (dockside intercepts), and from electronic logbook forms filled out by boat captains. The electronic logbook contains data field's specific to the catch and effort of individual headboat fishing trips.

2.6. Collection Mode

SRHS electronic logbooks are filled out by boat captains. The SRHS system will provide a data feed to the SEFSC data processing staging area.

Sample data will be provided to the SEFSC data processing staging area by the Beaufort Lab.

The data feed/loading processing from both sources will be automated.

3. Assumptions and Constraints

3.1. Data Resources

The Southeast Region Headboat Survey is comprised of the following sources of data:

Biological Sampling Data: The dockside intercept component of the SRHS can best be described as a systematic opportunistic design. Each port agent responsible for a particular area is asked to sample their vessels in a systematic rotation.   Upon attaining an angler's fish for sampling the port agent will measure and weigh individual fish using an electronic fish measuring board connected to an electronic balance. All measurements are recorded into computer memory for later download and editing. Another job the port agent has is to obtain otoliths, spines or gonads from selected species.

Biological sampling data is composed of individual species:

- Length
- Weight
- Date & location of harvest
- Sex

Logbooks (entered electronically): Headboat logbook forms collect information about numbers and total weight of individual species caught, total number of passengers, total number of anglers, location fished, trip duration (half, ¾, full or multiday trip), and species and numbers of released fish with their disposition (dead or alive).

Logbooks contain the following information:

- Catch composition
- Numbers of fish/species harvested and released alive\dead
- Vessel
- License #
- Number of passengers/anglers
- Trip type
- Departure/arrival time
- Location fished

3.2. Other Resources

Besides data from logbooks and biological sampling, the SRHS database contains a list of vessels that is maintained with information provided by port agents and the permits office at SERO. Individual information for each vessel in the list includes the vessel name, vessel ID, location, and owners.

3.3. Regulations

50 CFR Part 622.5 - FISHERIES OF THE CARIBBEAN, GULF, AND SOUTH ATLANTIC

4. Final Deliverables

4.1. Additional Reports

Reporting will be provided through the online (web-based) application of the SEFSC Data Warehouse.

4.2. New Data Sets

A completely redesigned data model including documentation on the entity-relationship diagram (ERD) will be delivered as part of the project.

4.3. New Systems

The new system is comprised of the following high-level modules:

- Data Loading, Validation, and Linking (SEFSC Data Warehouse Staging Area)
- Headboat Oracle Database (part of the SEFSC Data Warehouse)
- Online (web-based) data access and reporting (Application for queries, reporting, and downloading)

5. Project Leadership
5.1. Project Leader and Members
Table 1: Project Members

Project Role	Name	Organization	Title	Email	Phone 1	Phone 2
Team Leader	Patrick Cope	SEFSC,Miami	IT Specialist	Patrick.cope@noaa.gov	305-361-4251	
Team Member	Kenneth Brennan	NMFS Beaufort Laboratory	Coordinator Southeast Region Headboat Survey	Kenneth.Brennan@noaa.gov	252-728-8618	
Team Member	Erik Williams	NMFS Beaufort Laboratory	Supervisory Research Fishery Biologist	Erik.Williams@noaa.gov	252-728-8603	
Team Member	David Gloeckner	SEFSC, Miami	Chief, Fisheries Monitoring Branch, SEFSC	David.Gloeckner@noaa.gov	305-361-4257	252-646-7334
Team Member	Steve Turner	SEFSC, Miami NMFS Laboratory	Division Chief	Steve.turner@noaa.gov	305-361-4482 x 482	
Team Member	Kelly Fitzpatrick	NMFS Beaufort Laboratory	Operations Research Analyst	Kelly.Fitzpatrick@noaa.gov	252-728-8760	

6. Project Estimates
6.1. Project Schedule

Table 2: Project Schedule Major Tasks and Milestones

#	Schedule Description	Planned Start	Planned Finish	Prerequisites
1	Confirm project scope and requirements	Mon 7/2/12	Fri 7/13/12	
2	Design High Level Data Flow Specifications	Mon 7/16/12	Fri 8/10/12	1
3	Data Modeling	Mon 8/13/12	Fri 8/31/12	2
4	Build and Test Data Processing Procedures	Mon 9/3/12	Fri 10/26/12	3
5	Build and Test Data Access User Interface	Mon 10/29/12	Fri 12/21/12	4

6.2. Cost Estimates
Table 3: Cost Estimates

Project Need	Cost Description	Date Needed	Estimated Cost
Confirm project scope and requirements	Analysis/Requirements	Fri 7/13/12	$7,570
Design High Level Data Flow Specifications	Design	Fri 8/10/12	$15,141
Data Modeling	Design	Fri 8/31/12	$17,033
Build and Test Data Processing Procedures	Development	Fri 10/26/12	$30,282
Build and Test Data Access User Interface	Development	Fri 12/21/12	$30,282
TOTAL			$100,000

Total Rounded from $100,308