Washington Marine Recreational Sampling SQL Server Migration

FY 2015 Proposal

Ed Hibsch, Phillip Weyland
Created: 05/13/2015

1. Overview

1.1. Sponsor
Edward Hibsch

1.2. Focus Group
Information Management

1.3. Background
Collection of reliable fisheries data is essential to the effective conservation, management, and scientific understanding of fishery resources and is mandated by the Magnuson-Stevens Fishery Conservation and Management Act (MSFCMA, 16 USC 1801 et. seq.). In Washington State, marine recreational data collection is composed of two major units, the Ocean Sampling Program (OSP) and the Puget Sound Sampling Unit (PSSU). Together, these programs collect, manage, analyze, and report all marine recreational catch and effort estimates for Washington. On the coast, OSP conducts access point intercept surveys and boat counts to provide total effort and catch estimates along with supplemental surveys to capture additional data (e.g. tag, mark-status, and biological). In Puget Sound, in addition to the access point intercept “baseline” survey, Washington Department of Fish and Wildlife (WDFW) utilizes a telephone survey (see Lee et al.’s draft 2010 document) to provide catch-per-unit-effort (CPUE) estimates in conjunction with multiple “intensive” surveys including aerial and boat surveys, Voluntary Trip Reports (VTRs), and test fishing used for in-season management. Historically, data from these surveys has been collected with pen and paper then entered and maintained in a number of MS Excel workbooks and MS Access databases partitioned by program and time. These workbooks and databases are then stored in multiple locations. Usually, a separate Excel file or Access database has been added incrementally as new surveys are incorporated, which has now led to an oversized collection of disparate datasets. At regular intervals (or as required for reporting), these datasets are moved and compiled into intermediary datasets, before producing final estimates. This workflow and data storage procedure lends itself to several deficiencies: (1). Data Inconsistency:Data inconsistency is likely to occur whenever there is data redundancy (in this case redundant over multiple devices) and creates conflicting versions of the same data. Time and resources have to be spent to resolve these conflicts.(2). Reduced Data Quality:Current data files and local databases have little enforced constraints on data input. Instead of setting limitations during initial data input, this is a separate task and the responsibility data stewards to review the data in regular intervals. Many of these constraints should be automated.(3). Analysis More Complicated and Error Prone:Most analysis of the survey data requires components from multiple datasets. Much time is spent compiling and merging data instead of actually analyzing the data. For example, it is common for a research scientist to spend upwards of a day or two compiling data for the entire time series before data analysis can begin. Not only is this not timely, but the manual process is also inherently error prone. The compilation needs to be done automatically so analysis can begin within a few moments. (4). Insecure Data Storage:Technicians’ and biologists’ computers may be compromised by viruses or environment corrupting and/or destroying data. Each new device used to store a subset of survey data creates a point of possible failure. Enforcing a data backup and recovery plan on multiple devices is not feasible.As surveys are developed and additional complexities are added to data collection requirements, WDFW recognizes the need to store this data in a single relational database management system. This change would significantly enhance WDFW’s ability to efficiently access, query, maintain, and report recreational fisheries data for management purposes and provide a robust backend to allow the development of electronic data collection in the future not possible with the current system. Our proposed project aims to implement MS SQL Server as the database to store and retrieve our marine recreational sampling data. WDFW hopes to transition to electronic data collection in the near future and views the core database work proposed here as setting the key groundwork for such a transition.

1.4. Project Description
We seek funding to construct a SQL Server database to accommodate the existing OSP and PSSU sample datasets and estimation programs. The construction of this database will improve many facets of our sampling and management program as detailed in below objectives.

1.5. Public Description

1.6. Objectives
(1). Enhance the integrity of the data and incorporate rules for data acceptance.(2). Maintain all marine recreational survey data in a single repository using current technology and architecture standards providing consistent and secure data.(3). Expedite the process of adding new data and making it accessible to fishery managers, state biologists, and the Pacific Recreational Fisheries Information Network (RecFIN).(4). Enable users to quickly gather the information they need for decision-making.(5). Create a foundation to allow incorporation of electronic data capture.

1.7. References
Lee Y., Chang W., Kraig E., Tsou T., Palsson P. 2010. Procedures for Estimating Catch and Effort for the Puget Sound Recreational Marine Fisheries with Telephone Survey Information since 2003.

2. Methodology

2.1. Methodology

A project Information Technology Specialist 4 will be hired on a temporary basis to work in collaboration with the state's data management and fisheries management personnel to design, develop, and build a MS SQL database to support the needs of the OSP and PSSU.This project will compile the following datasets into a single SQL Server database:OSP(1). Boat Count- effort data (2). Access Point Intercept- catch, angler effort, descending device use, depth, and biological data (3). On-water Observation- catch and biological data(4). Volunteer Trip Report- salmon mark status dataPSSU(1). Access Point Intercept- catch and biological data(2). Volunteer Trip Report- salmon mark status data(3). Test Fishing Survey- biological and mark status data(4). Aerial Survey- effort data(5). Boat Survey- effort dataAn analysis of current databases and information requirements will set the foundation to which we will model the new Marine Sampling Program (MSP) comprehensive database for all marine recreational surveys. The desired design of this system will completely evaluate the entire workflow, from data entry to QA/QC, and finalize data to use in reports to state managers and RecFIN. A new data entry and data management interface will be developed alongside the new MSP database. To fit in the project's time frame and budget, we plan to develop this in MS Access. A separate OSP project (see supporting document "OSP MRIP Proposal Electronic Data Collection") is funded to test methods addressing the electronic data capture recommendation that came out of MRIP's 2010 review of the WDFW OSP. That project will be utilizing the cloud based, mobile application development platform, iFormBuilder. In order to support this effort and to fully integrate with the new MSP database, this project will develop integration services to automate data extracts and loads between MSP database and iFormBuilder.

2.2. Region

Pacific

2.3. Geographic Coverage

All of Washington's marine ports and access points (Puget Sound and Coastal).

2.4. Temporal Coverage

The database improvements would be ongoing and cover data needs for all sampling activities.

2.5. Frequency

NA

2.6. Unit of Analysis

NA

2.7. Collection Mode

NA

3. Communication

3.1. Internal Communication

A one-time kickoff meeting will review project goals, project management approach and expectations. This will be a face to face meeting led by the project manager. Agenda and minutes will be placed on SharePoint. Project team meetings will be held weekly through face to face or web conferences aimed at reviewing the status of the project with the team and led by the project manager. Minutes will be posted to SharePoint.Technical meetings will be face to face meetings held as needed led by the technical lead to discuss and design technical solutions for the project. Minutes and decision tables will be placed on SharePoint.Project status reports will communicate the status of the project on a monthly basis including progress, timeline, and activities to stakeholders and project team. These will be communicated through email and archived on SharePoint.

3.2. External Communication

Monthly reports will be submitted to the MRIP Operations Team through MDMS.A detailed final report will be submitted to the MRIP Operations Team through MDMS at the conclusion of the project.

4. Assumptions/Constraints

4.1. New Data Collection

N

4.2. Is funding needed for this project?

Y

4.3. Funding Vehicle
RecFIN grant

4.4. Data Resources
Existing OSP and PSSU sample datasets and estimation programs.

4.5. Other Resources

4.6. Regulations

4.7. Other

5. Final Deliverables

5.1. Additional Reports
Monthly progress reports will be provided, as well as a final report and metadata record.

5.2. New Data Set(s)

5.3. New System(s)
The new system that would be produced by the project would be the MSP MS SQL Database

6. Project Leadership

6.1. Project Leader and Members

First Name	Last Name	Title	Role	Organization	Email	Phone 1	Phone 2
Mark	Baltzell	Puget Sound Sampling Unit Manager	Team Member	WDFW			
Wendy	Beeghley	Ocean Sampling Program Manager	Team Member	WDFW			
Brodie	Cox	Bio Data Systems Manager	Team Member	WDFW			
Are	Strom	IT Specialist	Team Member	WDFW			
Theresa	Tsou	Research Scientist	Team Member	WDFW			
Phil	Weyland	IT Specialist	Team Leader	WDFW			

7. Project Estimates

7.1. Project Schedule

Task #	Schedule Description	Prerequisite	Schedule Start Date	Schedule Finish Date	Milestone
1	Hire Project IT		04/01/2015	05/01/2015	

Task #	Schedule Description	Prerequisite	Schedule Start Date	Schedule Finish Date	Milestone
	Specialist				
5	Analysis of current system	2,3,4	05/01/2015	07/01/2015	Y
2	Review current workflow	1	05/01/2015	06/01/2015	
6	Create ER Data Model	5	07/01/2015	08/15/2015	
7	Verify Data Model	6	08/15/2015	09/01/2015	
10	Verify Migration	9	10/15/2015	11/01/2015	
13	Develop Data Entry/Collection Interface	11	11/01/2015	02/01/2016	
3	Analysis of current datasets and inadequacies	1	05/01/2015	07/01/2015	
8	Create MSP Database	6,7	07/01/2015	09/01/2015	Y
11	Migrate legacy data	9,10	09/01/2015	11/01/2015	Y
12	QA/QC legacy data		05/01/2015	11/01/2015	
9	Transform and load scripts for legacy data	8	09/01/2015	11/01/2015	
15	Programming of data entry and UI.	11	11/01/2015	02/01/2016	Y
16	Testing and acceptance	15	02/01/2015	03/31/2016	Y
14	Develop Data Management and Reporting Interface	11	11/01/2015	02/01/2016	
4	Map current business process	1	06/15/2015	07/01/2015	

7.2. Cost Estimates

Cost Name	Cost Description	Cost Amount	Date Needed
Software Licensing	Licensing costs associated with cloud based mobile development platform.	$10000.00	05/01/2015
WDFW IT Specialists	Assistant Programmers ($7,876/month salary x 3 staff months)	$23628.00	05/01/2015
Temporary IT Specialist	Lead Programmer ($7,876/month salary x 11 months)	$86636.00	05/01/2015

Cost Name	Cost Description	Cost Amount	Date Needed
WDFW Indirect Costs	WDFW Agency Indirect Rate 25.76%	$30980.00	05/01/2015
PSMFC Indirect Costs	PSMFC Agency Indirect Rate 1.6%	$2419.00	05/01/2015
TOTAL COST		$153663.00	

8. Risk

8.1. Project Risk

Risk Description	Risk Impact	Risk Probability	Risk Mitigation Approach
Decision delays impact project.	Project and/or milestones not completed on time.	Low	Establish guidelines for decision turnaround time.
Scope Creep	Project and/or milestones not completed on time.	Low	Any change to requirements and features will need approval through weekly project team meetings.
Database fails to pass WDFW governance processes.	Additional time required to gain approval.	Low	Refer to published governance standards during initial design process. If needed, redesign database to conform to architecture standards.
Users reject the new interface.	Potential, significant rework of interface.	Low	Provide frequent prototypes to users.
Inaccurate estimates.	Project and/or milestones not completed on time.	Low	An uncertainty factor was added to expand total project and deliverables timeline. Re-prioritize assistant programmers' workloads to this project.
Resource performance issues.	Inexperience or lack of knowledge held by technical lead could have a significant impact on this project's success.	Medium	Hire for Information Technology Specialist(ITS) 4 position instead of ITS3. Look internally for resources with sufficient skills and experience. Re-prioritize assistant programmers' workloads to this project.

9. Supporting Documents