Development Of A Shared AYK Salmon Database

By:

Helen Hamner,

Shawna Karpovich,

And

Shane StClair

Regional Information Report[1] 3A03-23

Alaska Department of Fish and Game
Division of Commercial Fisheries
Arctic-Yukon-Kuskokwim Region
333 Raspberry Road
Anchorage, Alaska 99518-1599

December 2003

[1] The Regional Information Report Series was established in 1987 to provide an information access system for all unpublished divisional reports. These reports frequently serve diverse ad hoc informational purposes or archive basic uninterpreted data. To accommodate timely reporting of recently collected information, reports in this series undergo only limited internal review and may contain preliminary data; this information may be subsequently finalized and published in the formal literature. Consequently, these reports should not be cited without approval of the author or the Division of Commercial Fisheries.

AUTHORS

Helen Hamner is an Analyst/Programmer for the Alaska Department of Fish and Game, Division of Commercial Fisheries, 333 Raspberry Road, Anchorage, AK 99518.

Shawna Karpovich is a Fishery Biologist for the Alaska Department of Fish and Game, Division of Commercial Fisheries, 333 Raspberry Road, Anchorage, AK 99518.

Shane StClair is a Research Analyst for the Alaska Department of Fish and Game, Division of Commercial Fisheries, 333 Raspberry Road, Anchorage, AK 99518.

ACKNOWLEDGMENTS

The U.S. Fish and Wildlife Service, Office of Subsistence Management, provided $31,900 in funding support for this project (FIS02-069) through the Fisheries Resource Monitoring Program, under agreement number 701812J442. We thank Carlos Monsivais for checking escapement data, scanning data sheets, and keying data, and Amanda Orzechowski for error-checking and keying biological data. We also thank Linda Brannian, Holly Moore, Susan McNeil, and Toshihide Hamazaki of the Alaska Department of Fish and Game, and Karen Hyer and Polly Wheeler of the U.S. Fish and Wildlife Service Office of Subsistence Management for reviewing the report.

OEO/ADA STATEMENT

Project Report Summary Page for the US Fish and Wildlife Service, Office of Subsistence Management

Title: Develop Shared Fishery Database

Study Number: FIS02-069

Investigator: Helen H. Hamner, Analyst/Programmer, Alaska Department of Fish and Game, Division of Commercial Fisheries, 333 Raspberry Road, Anchorage, Alaska 99518
Phone: 907-267-2120 FAX: 907-267-2442
E-mail: Helen_Hamner@fishgame.state.ak.us

Management Regions: Arctic/Kotzebue/Norton Sound, Yukon River and Kuskokwim River

Study Cost: $31,900

Study Duration: July 2002 through January 2003

Geographic Area: Arctic/Kotzebue/Norton Sound, Yukon River, and Kuskokwim River

Federal Conservation Unit (FCU): Data collected from salmon stocks migrating or spawning in waters flowing through or adjacent to FCUs of the Arctic, Kotzebue, Norton Sound, Yukon River, and Kuskokwim River

Information Type: Stock status and trends

Issue Addressed: This is a continuation project and includes the next phase in the development of a comprehensive data management system for use by governmental and public entities involved in fisheries in the Arctic/Kotzebue/Norton Sound, Yukon River, and Kuskokwim River federal subsistence fisheries management regions.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

The Alaska Department of Fish and Game, Division of Commercial Fisheries, Arctic Yukon Kuskokwim (AYK) Region is creating a salmon database management system. The goal of the database is to provide managers, researchers, and the public involved in salmon fisheries in the AYK Region with a system to enter and process new data, as well as to retrieve historical data. Funding for this effort has been provided by numerous sources. The U.S. Fish and Wildlife Service helped initiate this program by funding creation of a data inventory, in the process of which many problems with historical data were identified. Through this effort staff began aggregating electronic data files for salmon in the AYK Region (Office of Subsistence Management project FIS 00-016). Additional funding, as project FIS 02-069, allowed the department to continue data rescue, aggregation, validation, and standardization from March 2002 through January 2003. As part of FIS 02-069, approximately 200 additional electronic age, sex, and length (ASL) files were found while transferring files from obsolete to current storage media and operating systems. A parsing program was written to recognize and load raw ASL files into an Access database. A preliminary database structure for biological data was developed and implemented. A total of 769 electronic ASL files containing 205,244 individual fish samples were loaded into databases. A total of 336 handwritten ASL files containing 106,348 individual salmon samples were keyed into a database. One hundred OPSCAN ASL files containing 27,990 individual salmon samples were rescanned and stored as digital files. Fifteen years of Yukon River subsistence and permit data were converted from obsolete to modern software and 695 escapement data sets were checked for errors. Work on this project is continuing under project FIS 04-701.

Key Words: Arctic-Yukon-Kuskokwim Region, AYK Region, salmon, fisheries management, Kuskokwim River, Kuskokwim Area, Yukon Area, Yukon River, Norton Sound, Port Clarence, Kotzebue, chinook, chum, subsistence, fisheries database, salmon age, sex, size

Citation: Hamner, H., S. Karpovich, and S. StClair. 2003. Development of a shared AYK salmon database. Alaska Department of Fish and Game, Division of Commercial Fisheries, Regional Information Report 3A03-23, Anchorage.

INTRODUCTION

The effectiveness of fishery management is enhanced when full use is made of existing information. Timely access to critical information from a wide variety of sources is needed to make informed decisions. Furthermore, with increasing participation by public entities and federal agencies in both data collection and fishery management decision-making, it is imperative that all parties have access to the same information. Using funds provided by both the U.S. Fish and Wildlife Service (USFWS) Office of Subsistence Management (OSM) and the Alaska Department of Fish and Game (ADF&G) from state general funds and other federal grants, the Arctic Yukon Kuskokwim (AYK) Region of ADF&G has begun to create a database management system in which all data will be centralized, have a standard format, and be shared among various management agencies and the general public.

AYK region salmon fisheries are managed by both state (all uses) and Federal staff (federally qualified subsistence users only). This region encompasses over 70% of the landmass in Alaska (Figure 1) and includes Norton Sound, Kotzebue Sound, Port Clarence, and two large drainages, the Yukon and Kuskokwim Rivers. It is characterized by a diversity of commercial, subsistence, sport, and personal-use fisheries harvesting freshwater, anadromous, and marine fish species. ADF&G, Division of Commercial Fisheries has divided the region into four areas for salmon management purposes: (1) the Kuskokwim, (2) the Yukon, (3) the Norton Sound/Port Clarence and (4) the Kotzebue Sound areas (Figure 1).

Data are collected annually to support fishery management decision-making and data management is specific to particular areas within the region. Paper records and computer spreadsheets are most commonly used, making it difficult to share information among areas or respond to specialized requests. Existing data summaries often lack the ability to display and summarize data in alternate ways. Also, various types of related data are often stored in various formats and locations (ADF&G area offices in Nome, Bethel or Fairbanks and seasonally in Emmonak and Kotzebue) making questions requiring integration of multiple data sources difficult to address. In addition, the dangers of losing data during staff turnover and degradation in electronic media are a concern. A goal of this project is to address standardization of data collection and management and thereby address these concerns.

This report summarizes progress and the current status of development of a salmon database management system for the AYK Region. It represents the final report for project FIS 02-069 funded by the USFWS OSM, which was a continuation of FIS 00-016 (Hamner et al. 2002). Included in this report is an overview of all funding sources being used to develop the AYK salmon database management system, defined tasks, and the timeline for completion. Also included are recommendations for improving data collection, data storage, and data processing in AYK.

Project Background

Recognizing the need to develop a mechanism for standardizing the collection, storage, and analysis of fisheries information in January 2000 (Figure 2), ADF&G AYK staff proposed a 5-year, $300,000 project to create a salmon fisheries management database for AYK to USFWS OSM. This proposal was rejected due to its cost and duration and instead a one year $60,000 pilot project (FIS 00-016) was funded. The specific objectives of project FIS 00-016 were to (1) create an inventory of existing salmon data sources, and, (2) to conduct a preliminary needs-assessment survey of fisheries managers and researchers in the AYK Region. This salmon data inventory and results of the survey are published in *Shared Information for Fishery Management in AYK* (Hamner et al. 2002). As part of this effort, historical and current AYK salmon projects were described and primary contact information (agency, project leader, etc.) was included. In addition, an inventory of project associated data was created. This includes electronic filename, file type, project type, location, species, and data descriptions. Major data categories identified in the inventory included subsistence and commercial harvests, escapements, and biological information (age, sex, and length). As part of the project a summary of the results of an informal survey of state and federal fisheries management and research staff, was included. This summary identified data sets and data summaries required to make more informed fishery management decisions. The inventory and survey were intended to be used as the first steps in creating a database of historical and current salmon data. In the early stages of the project it became apparent that a thorough inventory could not be completed within the allotted timeframe and funding, mostly due to the discovery of large quantities of disorganized data.

The following year, AYK Region staff again submitted a multi-year proposal to OSM requesting over $300,000 to create a database management system for historical fisheries data. This request exceeded available funding and a reduced proposal (for $150,000) to rescue, aggregate and standardize data was submitted by AYK staff to OSM. Further modifications to the investigation plan were requested to greatly lower the cost and change some activities. Ultimately, only $30,000 was approved for the project. It covered the period March 15, 2002 through January 31, 2003 and the goal was to continue organizing historical data (project FIS 02-069).

Other Funding Sources and Related Projects

Supplemental funding was sought to assist in continuing work on the salmon database management system for AYK. Sources included federal disaster relief funds appropriated for the purpose of restoration and rehabilitation of Norton Sound salmon populations, National Oceanic and Atmospheric Administration (NOAA) funds for age, sex, and length (ASL) data rescue, and the AYK Sustainable Salmon Initiative (AYK SSI) representing a portion of the federally appropriated Pacific Salmon Recovery fund. A proposal was also submitted to OSM and subsequently approved for funding in 2004 ($135,000 for July 1, 2004 through June 30, 2007, project FIS 04-701).

The Steering Committee for the Norton Sound Salmon Research and Restoration fund approved a five year project ($250,000) for the development of a Norton Sound Salmon Information

database which will also include ASL data (Hamner et al. 2003). In the first year of the project, a data inventory was updated (see overlap with FIS 00-016 in Figures 2 and 3) and completed for the Norton Sound area. Additional ASL electronic files for Norton Sound salmon were aggregated and transferred to modern storage media. In this second year, ASL data not found as electronic files were reentered from handwritten forms into the ASL database. Although these funds are allocated specifically for Norton Sound, applications and database structures developed for the Norton Sound Area are useful in other areas of the region since data collection follows a standard protocol across the region.

Additional funding was requested after discovering that a significant number of ASL electronic files had been lost due to changes in staff and degradation in aging electronic media. In response, the National Oceanic and Atmospheric Association funded ($20,000) a concurrent project to transfer data in obsolete media to current systems and aggregate, standardize, correct and load chum salmon ASL data for the Kuskokwim, Norton Sound, and Kotzebue areas into a database. A similar proposal focusing on species and areas not covered by the other funding sources was also funded by the AYK Sustainable Salmon Initiative ($15,000).

The work involved in data inventory, rescue, editing, and loading data into databases exceeded the amount of funding available from any one source. Combined funding was sufficient to continue the development of an AYK salmon database management system. Projects spanned through time (Figure 2), sharing many overall objectives, though differing in duration or specific focus (i.e. Norton Sound or a subset of salmon species). At times several funding sources were needed to complete tasks (Figure 3).

Database Development Process

The following steps have been defined for development of a salmon database management system for AYK Region.

1. Create an inventory of available salmon data.
2. Conduct a preliminary needs assessment survey.
3. Transfer important data that are in imminent danger of being lost due to degradation in media to a more permanent storage media.
4. Aggregate, standardize, and edit (missing identifier codes) data.
5. Create intermediate databases in Microsoft Access for reporting and editing historical data and entering new data.
6. Write a software development plan for an integrated SQL Server (Microsoft 2000 version) database with web-access.
7. Develop prototype; show to users; get feedback.
8. Develop software according to plan and user feedback; load data from Access databases (from step 5).
9. Continue modifying database using feedback from users.
10. Add additional datasets as requested by users.

These database development steps have been and continue to be funded by several sources (Figure 3). Project FIS 00-016 addressed the first two steps in the database development process while FIS 02-069 focused on steps 3, 4, and 5. NOAA and AYK SSI projects will fund completion of steps 4 and 5. Norton Sound and FIS 04-701 will fund completion of steps 6 through 10.

Historical Data and Scope of Work

An enormous quantity of historical salmon data exists for the AYK region. More than 150 projects that monitor salmon harvests, escapements, and measure stock abundance and biological attributes have been conducted in the region in the last 40 years (Hamner et al. 2002). Subsistence salmon harvest information has been collected throughout the AYK Region (Burkey et al. 2001, Brase and Hamner 2002, Banducci et al. 2003, Georgette et al. 2003). Data from all commercial landings made in the region's salmon fisheries since 1969, consisting of over nine million commercial sales receipts (fish tickets), are available in electronic format. Salmon observations have been made on more than 315 streams in the AYK region and include over 70,000 records of daily and annual estimates of abundance before 1999. Although a substantial increase in the number of new projects has occurred since 1995, many such as the Fishing Branch Weir, which has been in operation for thirty years, or the Kogrukluk Weir with thirty-five years of fish passage data, are long standing projects. Biological information has been collected for over forty years from harvests and escapements to estimate age, sex and length (ASL) composition of these salmon populations. This information is stored in individual files for each project, gear, species, and year combination. Over 3,400 raw data ASL files were inventoried during project FIS 00-016.

The disorganized state of the large volume of historical data pointed to the need for a region-wide system of archiving electronic data. During the era of mainframe computers, data processing for most major data sets was centralized and controlled and the use of standard codes and formats was enforced. More recently, personal computers have allowed biologists greater independence in controlling and analyzing project data. However, as a result of this independence, data coding, error-checking, and storage have diversified or have been nonexistent and some data residing with project leaders have been lost during staff turnover. Currently, historical fisheries data are stored in hardcopy or various electronic formats such as text (ASCII), spreadsheet and word processing files. These data are stored in various locations by project leaders using diverse and often obsolete storage media (digital tapes or 5 ½" diskettes with various formatting). For a variety of reasons such as staff transitions, degradation in storage media, missing data identifiers and/or a disorganized system of file management and archiving, valuable historic data could be lost.

Efforts to standardize existing data collection and storage are a critical foundation for building a database. The authority to require standardization is the first step. AYK database projects were approved by the AYK Regional Supervisor for the Division of Commercial Fisheries (ADF&G) prior to submission for funding. When funded, these projects represented the region's

commitment to the necessary steps in database development and the authority for project investigators to begin standardizing data collection. In this instance AYK staff needed to develop policy for standard data collection and storage as required by project objectives. Presentations were made to ADF&G staff at their annual staff meetings about this and other projects and the need to standardize data collection and storage for ASL and escapement data. These two data categories represent data sets that had been centralized during the era of mainframe computer data processing and represented categories where standard codes and formats had been enforced until the early 1990s. The Regional Supervisor then empowered the region's programmers to develop and enforce standards after working with affected staff for data entry screens, computer editing routines, and reporting that will create standard data collection and storage procedures. As development of the AYK database management system continues annual updates and meetings with affected staff have occurred. When current year data are entered into the AYK salmon database management system the need for individual files and dissimilar formats will be gone.

Data Categories

During the initial inventory, four major categories of salmon data were identified: (1) biological (ASL) information, (2) escapement data, (3) subsistence harvest data from surveys and (4) commercial salmon catch and effort data from "fish tickets". Priority was given to data that were in immediate danger of being lost, that being ASL and subsistence harvest data. The majority of the project time was spent on ASL data represented by thousands of individual ASCII files in various locations. In addition, virtually all of these files contained data coding errors, including inaccurate or missing header information needed to identify the data. Conversely, historical commercial catch and effort data are much better organized, preserved and archived annually, and did not require rescue efforts during this project.

Age, Sex, and Length

Age, sex, and length data are collected annually from salmon sampled from commercial harvest, escapement, run timing and abundance monitoring projects in the AYK Region. Scales are collected primarily to determine the age of fish, but may also be examined for growth patterns. Since the distance between scale annuli represents the growth of a fish in one year, scales are a permanent record of annual growth over the lifetime of an individual salmon. With the recent decline of some western Alaska salmon stocks, interest in examining fish growth and survival as depicted by scale characteristics has increased. Salmon length is generally represented by a measurement in millimeters from mid-eye to fork of tail. Sex of the salmon is determined from either external characteristics or internal inspection of reproductive products.

Age, sex, and length data have been collected in the Yukon Area since 1960, in the Kuskokwim Area since 1961, and in the Norton Sound-Kotzebue Area since 1962. All salmon species have been sampled but the emphasis has been on chum and chinook salmon. Scales collected from salmon are stored on gum cards along with an acetate impression which is used to determine age.

Both are organized into files by year, species, and project type. They are stored in cabinets located in the Anchorage and Nome ADF&G offices. Paper copies of ASL data are filed in the same locations. In most years, ASL data were converted to electronic data files located on either mainframe or personal computers. However, no formal archiving system was established and much of these electronic data have been lost.

Between 1960 and 1983, ASL data were recorded in the field on handwritten forms. Beginning in 1984, a machine (OPSCAN 90/20) that scans and records data from ASL forms (referred to as OPSCAN or mark-sense forms) was used to improve data processing. All ASL data collected between 1984 and 1998 were recorded on forms that can only be read by the OPSCAN reader. Although the OPSCAN machine was more accurate than hand entry, many new types of errors were introduced when data were omitted or miscoded on mark-sense forms. Static information for each project (e.g. project location, project and gear type, and target species) was often omitted.

Electronic files are in fixed-field text (ASCII) format, usually with one file for each year, species, project, and gear combination. Data output varied as data processing programs evolved over the years, creating seven identified formats. Currently, all inventoried files are located and organized on the Anchorage AYK file server and backed up on tape nightly.

As part of FIS 00-016 project, an electronic inventory was created based on the salmon scales and data forms located in the scale archives (file cabinets) in the Anchorage and Nome ADF&G offices. Next, an effort was made to find the corresponding electronic data (age, sex, and length measurements), for each species, project and year combination. Due to the volume of scales and data collected over the past 40 years, this task took a large amount of time. As files were found and added to the inventory, little additional effort was required to aggregate and transfer files to modern media such as compact discs (CD) and the AYK Region's file server located in the Anchorage office.

Electronic data could not be located for every inventoried scale project. In some cases, paper data were never transferred to electronic media, in which case the data needed to be hand-entered. Other electronic data have been lost due to the lack of a regional data archival system. As part of project FIS 00-016, approximately 1,900 5¼" floppy disks were collected from the Bethel, Nome, Fairbanks, and Anchorage ADF&G offices. A local data recovery company was contracted to transfer files from these disks, many of which were hard-sectored and formatted for an obsolete, pre-DOS operating system called CP/M, to CDs. Approximately 25,000 miscellaneous files were recovered and set aside to be searched for missing electronic ASL data.

Escapement

The abundance of salmon returning to spawn in rivers throughout AYK is monitored using aerial, ground, and boat surveys, weirs, counting towers, fish wheels, sonar, and test fisheries. Raw counts of salmon "escaping harvest" (i.e. escapement data), such as hourly passage counts through weirs or past towers, or the number of salmon counted from aircraft or boats, are generally entered and stored in distinct spreadsheets for each project and year. These files are

created and maintained by ADF&G biologists responsible for the monitoring project. The raw data for large sonar projects such as Pilot Station on the Yukon River are stored in database software such as Rbase or Access. Area managers may also combine current and historical data from numerous escapement projects into integrated spreadsheets for inseason management use. The most complex of these integrated spreadsheets is the Yukon Area 'spread web' which includes historical averages and daily counts for key Yukon Area escapement projects.

In 1996 NOAA's Earth System Data and Information Management System Program, (RFP 52ABNF600096) funded ADF&G to develop a database on the status of Alaska salmon populations (SASPOP). The goal was to develop a geo-referenced database, which included links to the Anadromous Waters Catalog (AWC)[1], of salmon escapement data collected statewide between 1960 and 1998. The SASPOP database was initially developed for southeast Alaska and the fields and structure were most suited to the types of data collected in that region. Later, it was modified to incorporate escapement data for the entire state. Consequently, many of the fields are not applicable to the AYK Region. Some portions of the project could not be completed because agreement was not reached on tasks such as creating a standard method of computing indices of escapement or a system of usage codes defining the appropriate use of these data. Lastly, this was a multiple year project and funding for it was not renewed after completion of the database for southeast Alaska. Work to extend SASPOP to a statewide application ended prior to completion in 1999. The Oracle software version of SASPOP is no longer maintained or updated by ADF&G

Many spreadsheets of daily passage counts from AYK were standardized for inclusion in the SASPOP database prior to project termination. Other data were keyed from paper copies of project reports into a spreadsheet template created for the SASPOP project. These data were imported into database client-server software (Oracle) and were also available in Access. The SASPOP project was terminated before all AYK data were edited, reformatted and incorporated into the database. An example of such uncorrected data is aerial, foot and boat survey data from Norton Sound, Kotzebue and Port Clarence area. These data are maintained in separate spreadsheets for each surveyed stream with a total of over one hundred separate spreadsheets. These data were not included due to the extensive reformatting required to standardize them. Early aerial survey and other data retrieved from a now obsolete Honeywell mainframe computer were also not incorporated into the database. An Access version of SASPOP with AYK-only data resides with AYK programming staff (hereafter referred to as the SASPOP escapement database). This database will form the model and will be built upon to become the AYK database for escapement data to include data through the current year.

Only a preliminary editing of most AYK data in SASPOP was completed prior to the end of the project in 1999. A more thorough review and edit was needed. The Access version of the SASPOP database does not have data entry, editing or reporting capabilities. Data collected since the project ended in 1999 have not yet been included. Also, because escapement projects in the database are geo-referenced, projects that have changed locations (e.g. moved to the opposite river bank) are treated as distinct projects although managers treat them as the same project.

[1] The AWC is a catalog of waters important for the spawning, rearing or migration of anadromous fishes of Alaska, maintained by ADF&G Division of Sport Fish and accessed at, http://www.sf.adfg.state.ak.us/SARR/FishDistrib/anadcat.cfm

Subsistence Harvest

As of 2002, approximately 9,050 households annually harvest salmon for subsistence in AYK Region. In general households are surveyed postseason to collect subsistence data. The first survey of subsistence harvests took place on the Yukon River in 1919; however, harvests were not reported by village until 1931. Systematic surveys have been conducted annually since 1960 in the Kuskokwim Area, 1961 in the Yukon Area, 1963 in the Norton Sound Area, and 1967 in the Kotzebue Area. Surveys during the early years only documented harvests of chinook and small salmon (all other salmon species combined). Most surveys included the number of fishing families and type of gear, and number of dogs per household. An overview and comparison of the subsistence methods used in each area and the history of data collection were reviewed and summarized by the statewide subsistence harvest assessment working group (Caylor 2000) funded by project FIS 00-017.

Detailed information on harvests by species, number of households, dogs in each household, and other information is collected and maintained by Subsistence Division of ADF&G for the Kuskokwim, Norton Sound, Kotzebue, and Port Clarence areas; by ADF&G Division of Commercial Fisheries for the Yukon area, and by Canada Department of Fisheries and Oceans for the Canadian portion of the Yukon drainage. Annual harvests by community are included in a statewide database maintained by Subsistence Division. The results from subsistence harvest survey projects are reported in Division of Commercial Fisheries annual management reports (Ward et al. 2003, Banducci et al. 2003, and Vania et al. 2002) and in a separate Regional Informational Report (RIR) for the Yukon River (Brase and Hamner 2002). From 1988 through 2002, subsistence household survey and permit data for the Yukon River were stored in separate files for each year in what is now obsolete software (Rbase). Yukon River subsistence salmon harvest data noted for work by this project are collected by a single project. Data format and coding standards have been enforced since 1988.

Though subsistence salmon harvest was identified as a data category, data for areas not surveyed by the Division of Commercial Fisheries will continue to reside in the centralized database maintained by Subsistence Division in Anchorage. It has yet to be decided whether higher level harvest summaries will reside in the AYK salmon data management system or if a link between databases will be developed. A link between databases is the preferred method unless database architecture or accessibility dictates inclusion of higher level summaries. This data category will be added or database links developed only after sufficient progress has been made for ASL and escapement data. Yukon River subsistence harvest data will also reside in the AYK database management system until such time that a link with the Subsistence Division is agreed upon and developed.

Commercial Harvest

Generally, a sales receipt ("fish ticket") is issued each time salmon are sold by fishers participating in Alaska's commercial fisheries. Electronic records of these fish tickets from 1969

to the present are archived by the Division of Commercial Fisheries, Computer Services Section, in the ADF&G office in Juneau. Fish ticket data have been entered inseason since 1981 in Emmonak and 1984 in Fairbanks, Nome, Bethel, and Kotzebue. Both hardcopy fish tickets and electronic data are archived postseason in Juneau. Beginning in 2000, fish ticket information was entered and archived in a centralized Oracle database located on a server in the Computer Services Section office in Juneau. An ongoing project imports and corrects historical data for the years 1969 through 1999 into the new data system and data back to 1985 are currently converted. Commercial harvest data are not currently accessible to the general public through the internet and are available through the State's wide area network (WAN) only to ADF&G staff with fish ticket application software loaded onto their personal computer and only back to 1985. All other data requests must be directed to the Computer Services Section office in writing.

Though commercial harvest was identified as a data category during the initial inventory (Hamner et al. 2002), fish ticket data will continue to reside in the centralized database maintained by Computer Services Section office in Juneau. Data back to 1985 can be extracted from that database by the AYK salmon DBMS for reporting. Architectural and consistency problems may necessitate that higher level catch summaries for years prior to 1985 reside in the AYK salmon database management system. This data category will be one of the last to be added. Commercial harvest data are currently collected, entered, stored and reported in a standardized manner following policy developed by the Computer Services Section in consultation with fisheries managers throughout the division.

OBJECTIVES

An overall goal for the AYK Region is to provide managers, researchers and public entities involved in salmon fisheries in AYK a system to enter and process new data as well as retrieve historic data for salmon ASL, escapement, and harvest. This goal will be addressed through a series of projects from various funding sources. Projects will have complementary objectives and focus tasks pertinent to the funding agency (Figure 2 and 3). The primary goal of this project (FIS 02-069) was to continue to develop the database management system beginning with data aggregation, error-checking, and standardization through to development of intermediate data entry, editing and reporting programs. Objectives specified in FWS Agreement Number 701812J442 for OSM project FIS 02-069 include:

1. aggregate diverse sources of fishery data,

2. error-check and correct historical data as necessary,

3. begin standardizing data formats, where necessary, for inclusion into a centralized database,

4. develop intermediate data entry, editing and reporting programs for area staff so that more thorough error-checking, editing and a standard format of data can begin as soon as possible.

METHODS

During this project, attention was generally focused on data needing rescue. Because of the large volume and disorganized status of age, sex and length data, nearly all of this project's funds were spent on this type of salmon data. Of secondary priority was Yukon area subsistence harvest data, which were the only subsistence data requiring rescue. Remaining time was spent editing escapement data. Commercial harvest data were being well maintained by the Computer Services Section of Commercial Fisheries Division and consequently were omitted from this project. Therefore objectives 1-4 were directed primarily at ASL data for the AYK Region.

Age, Sex, and Length Data

CD archives containing over 25,000 files from 1,900 floppy disks were visually scanned to find missing ASL data. The archives were created under project FIS 00-016 but the search process was not completed. During project FIS 00-016 and again in this project, inconsistent electronic file naming conventions have caused problems with locating and recognizing files. All files had to be opened to accurately identify their contents. Numerous identical raw data and summary files were named differently, depending on the purpose of the file (e.g. age or length tables). Some file names were consistent but their extensions were changed (e.g. Bt94k40s.AWL, Bt94k40s.AS, Bt94k40s.L where AWL = raw data, AS = age composition, L = length by age). On occasion, determining which files had been updated most recently was difficult because the file's date of last modification was dependent on an individual computer's internal clock.

A computer program was written to recognize and parse each of the seven different historical ASL electronic file formats. An Access database with a prototypical structure was developed to receive the parsed data. A research analyst loaded previously recovered ASCII files into databases using the parsing and data loading program; data from different areas were loaded into separate databases to facilitate editing. Numerous error-checking protocols were developed to validate the data. Header records, which link sample data to location, project, gear type, mesh size, date, and species data, were carefully checked for valid or missing codes. The loader program flagged errors such as unacceptable alpha-numeric characters, duplicate files, or duplicate scale card numbers within a file. Error-checking queries were developed in the prototype database that identify lengths and ages outside of realistic ranges; files with more than one location, project, or species code; and duplicate header or fish data.

The file cabinets in the region's scale archives were searched for paper copies of missing data. These paper copies include handwritten forms (1960-1983) and OPSCAN scanner forms (1984-

present). Technicians searched both the Anchorage and Nome office archives.

OPSCAN forms for ASL data not found in electronic format needed to be scanned to create electronic files. Older OPSCAN forms (1984-1994) had to be scanned using an OPSCAN 90/20 reader which is no longer in production and for which replacement parts and maintenance cannot be obtained. Attempts to scan these obsolete forms with newer scanners have proved unsuccessful. Using working parts of each, two different machines were combined into a working model, but scanning the forms still required skill and patience. The scan process would abort frequently, sometimes without any apparent reason, and the computer receiving data was prone to freezing up or ceasing to record incoming data during a scan. In addition, the data recorded on the mark-sense forms were often erroneous or incomplete. Each file had to be manually error-checked and corrected; many had to be scanned a second or third time.

A data entry form (electronic screen) was created so that technicians could enter data recorded on older handwritten forms into a database. Data were keyed and maintained in separate databases to allow for specialized error-checking procedures. A research analyst and an analyst programmer wrote database queries to detect and correct data entry-associated errors.

A query to generate a dynamic inventory of loaded data was created. The query builds a table describing the number of individual salmon samples by species, year, district, and subdistrict. The database structure for ASL data was also revised to optimize data storage efficiency and compatibility with hypothetical future data structures.

Escapement Data

ADF&G staff questioned the accuracy and completeness of the AYK escapement database in SASPOP. They could not be assured that only final data were included or that all project data had been located. Therefore, escapement data in the database were checked against season escapement totals for each project site and species as recorded in annual regional information reports. Discrepancies were noted for further investigation. No actual changes were made to the data. A presentation to staff of the discrepancies needs to occur and staff need to agree upon a solution (change or acceptance).

Subsistence Harvest Data

Only subsistence harvest data from the Yukon Area as maintained by the AYK Region of Commercial Fisheries Division was in need of rescue. Subsistence data from other areas within AYK are maintained by ADF&G Subsistence Division. Annual harvest data for the Kuskokwim, Norton Sound/Port Clarence and Kotzebue areas have been archived in current software (Access) and are stored on the Subsistence Division server in the Anchorage office.

Yukon Area subsistence harvest Rbase databases, which include both survey and permit data, needed to be transferred from floppy diskettes and aggregated on the AYK Region file server. The databases were first updated from an obsolete DOS-based version of Rbase to a more recent Windows version. The data were then transferred to Access via a database linking program.

RESULTS

Progress was made in aggregation, correction, and standardization of the three data types in most need of rescue; ASL data, escapement data, and subsistence harvest data (Table 1). Commercial harvest data were not addressed by this project. Project objectives were nearly accomplished for ASL data, as data were aggregated, error checked, and standardized for inclusion in intermediate databases in Access. Lacking completion was the data entry and reporting programs for area staff to enter future data in a standard format (objective 4). Objectives one through four were also accomplished for Yukon subsistence harvest data. Some progress was made in escapement data (Table 1). Funding from several projects was needed to accomplish these results (Figures 2 and 3).

Age, Sex, and Length Data

A total of 769 files representing 205,244 salmon were loaded into area-specific databases using the loading program. These consist of 188 Norton Sound Area files representing ASL data for 42,632 salmon of all species, and 581 Yukon Area files representing 162,612 chinook salmon. This included approximately 200 ASL files that were recovered from the archival CDs.

A total of 100 OPSCAN mark-sense forms representing 27,990 salmon samples were successfully scanned (Table 2). All of these files were from the Kuskokwim or Norton Sound/Kotzebue Areas. Most represented samples of chum and chinook salmon.

Technicians keyed a total of 336 handwritten files containing 106,348 individual salmon samples. Another 425 handwritten files remain to be entered (Table 3), mostly from the Yukon area. The majority of the remaining missing ASL data are in these handwritten files; the recovery of which will bring the ASL data aggregation effort to approximately 86% completion (Table 3).

To date, a significant number of electronic files for 1986 and 1987 Yukon chinook data have not been located. Virtually all Kuskokwim area files created before 1982 remain missing. Gaps in electronic files for the Yukon Area occur between 1972 and 1982 for chinook and between 1971 and 1986 for chum and coho salmon. Many historical files of all Norton Sound and Kotzebue Districts salmon species collected before 1988 are missing.

Aggregating, validating and standardizing ASL data has taken most of the time for this stage of the project because of both the large volume of disorganized data and the number of erroneous or omitted identifying codes in the electronic files. A variety of problems still persist in ASL data (Table 4). The sources of these problems include incomplete data recording, improper mark-sense form coding, scanner errors, and possible degradation of storage media.

Escapement Data

Each escapement data set consists of a unique year, species and project combination. Of 695 data sets, 96 (14%) were discovered to contain discrepancies with the season totals reported in the ADF&G Regional Information Reports (Table 5). Season summary tables could not be found for 163 data sets (23%) contained in the database, so accuracy of these data could not be checked.

Subsistence Harvest Data

Yukon Area subsistence survey and permit database files for all years, 1988–2002, were successfully transferred to Access databases, with one database file for each year. All files are currently stored on the AYK Region server. In addition, these Access databases were provided to the Division of Subsistence to be included in their Alaska Subsistence Fisheries Database (Caylor and Walker 2003).

DISCUSSION

Significant progress was made in fulfilling the objectives of this project which included aggregating, error-checking, standardizing, and developing intermediate data processing applications. Emphasis was placed on ASL data, escapement data, and subsistence harvest data from the Yukon River. Work towards completion of objectives for these data sets since the end of project FIS 02-069 continues using other project funding and state of Alaska general funds. Data rescue for Yukon subsistence data is complete, ASL data rescue is nearing completion, and the rescue of Norton Sound escapement data (that data not in SASPOP) is the furthest along of AYK escapement data. The point that additional ASL data sets are difficult and expensive to find has been reached and priorities have switched to loading and editing newly found data and completing the editing and reporting applications. Data standards are complete for ASL data and will be used when incorporating 2004 data.

Funds from NOAA (Hamner et al. 2004) and AYK SSI (St Clair and Hamner *In Press*) have been used to recover more ASL data and aggregate these data in a regional ASL database. One objective not completed was to develop intermediate data entry, editing, and reporting programs

for area staff. While a computer program to load ASL files into the database was developed, this program needs to be broadened to include automatic error-checking.

Recently found OPSCAN files require scanning and this task will be completed as time and funding permit. Technicians are continuing to manually enter the remaining pre-OPSCAN handwritten forms. Two employees are also reviewing electronic data for missing information, invalid codes and outliers. Suspect data are being checked against the original paper forms and scale cards to validate its accuracy. During the entry of the handwritten files, a thorough re-check of the scale card inventory, tally of borrowed scales, and otolith count should be conducted to ensure that any remaining "missing" files are indeed missing rather than nonexistent. Then, the team can move into the final recovery stages of contacting project managers and area offices in an attempt to locate any remaining missing data. Some physical data such as vertebrae and otoliths remain in the Fairbanks office and their corresponding electronic files are not yet grouped with the others. ADF&G supervisors need to decide where these files should be housed.

A challenge in the storage of escapement data will be to incorporate updates of historical observations into the AYK database. During the comparison of salmon counts and review of the SASPOP escapement database, historical numbers for the main stem Yukon River sonar project were found to have recently been changed (T. Hamazaki, ADF&G, personal communication). This was due to a revision of the species apportionment methodology resulting in a change to all historical numbers of fish. In addition, historical counts for Kwiniuk Tower in the Moses Point Subdistrict of Norton Sound had been revised to correct errors introduced by faulty data transcription and count expansion discovered by Hamazaki (2003). Because of these revisions in the sonar and Kwiniuk tower data sets, these data will have to be reloaded into the database. Once the database is established, procedures will need to be to be developed to incorporate these types of changes.

The slight differences found between reported season totals (in RIRs) and seasonal totals of daily data from the escapement database are probably due to the accumulation of small rounding errors. For example, daily count data from tower projects are calculated, non-integer values, but were entered as integers in the database. Since errors were not quantified (only the presence or absence of errors was indicated) rounding errors cannot be distinguished from more serious errors. Further analyses should examine whether discrepancies are due to accrued rounding or significant data problems. Changes will not be made to the database until area biologists can agree upon a policy of either re-entering data as non-integers or accepting small rounding errors and finalize data as entered. Area staff will need to respond to more serious errors on a case by case basis.

Now that the historic Yukon subsistence harvest data are in modern and commonly-used database software (Microsoft Access), these data are accessible and preserved. Data collected during the 2003 survey were entered directly into a newly developed SQL server database. Data from 2002 were also transferred from Access to the SQL database. The SQL database has allowed both the Anchorage and Fairbanks AYK staff involved in the harvest survey project to have access to the same centralized database through the web, greatly enhancing data entry, editing, and summarization for 2003. Ultimately all years will be entered into the SQL database

and will represent a storage collection standard. The programming sections of both commercial and subsistence divisions have also discussed minimizing duplication of data storage and possibly moving towards a singe repository for subsistence harvest data.

Outlook for Project Completion

A database management system does not have a "final" product or a clear end point, rather it results from an iterative and ongoing process. New data are entered annually, historic data are corrected as errors are noted, and funding initiatives may allow for the addition of previously unavailable data sets. New database queries and reports are created or modified in response to user needs. Additionally, the science of database management is not static and as a result the database system may change in response to software enhancements and changes in technology.

AYK Region had defined 10 steps in the development of a database management system for its salmon fisheries. This report discusses progress to date involving steps 1-5, (see page 4) partially funded by project FIS 02-069. Progress will continue on the remaining steps with funding from the Norton Sound Salmon Research and Restoration fund and a new project funded by OSM (FIS 04-701) to begin in 2004. A recognizable end point will be reached when fishery biologists and the public can access ASL, escapement, and subsistence harvest data through the internet (SQL server database). An intermediate product will consist of Access databases with data through 2003, including intermediate data entry, editing, and reporting programs. This will be recognized when area staffs begin to use these databases to process current year data. Ultimately Access databases will become merely a front end interface to the SQL server database. Current year data as entered into Access or the front end interface of the SQL database will represent the standardization of data collection and storage for AYK Region. Data entry screens, editing for allowable codes and formats will ensure data collection and storage standards.

Funds from the Norton Sound Salmon Research and Restoration fund are targeted for completion of the salmon database management system for Norton Sound. Since projects, data collection, and reporting are similar for all areas of the AYK Region, databases and computer programs developed for Norton Sound can be applied to other areas. The next phase in the recovery of data for the Norton Sound area is the creation of an aerial survey database in Access with data entry and reporting capabilities and the reformatting and loading of data currently stored in nearly 100 Excel and Lotus spreadsheets. Because all areas use the same data collection form for aerial surveys, the rest of the region will also use this database.

Programs to standardize and load other types of data such as escapement data from projects other than aerial surveys need to be developed. The region added two fulltime programmers (one as the unit leader) to its information technology (IT) section to help develop these programs and progress in this area. IT staff need to investigate escapement data discrepancies between the Access database (SASPOP) and project RIR reports. Many of these differences were due to rounding errors and IT staff need to discuss with project biologists how to reconcile these discrepancies.

The lead programmer for AYK Region has begun a software development plan for an SQL server database with web access, which will be the final repository of AYK salmon fisheries data. Technical expertise now exists within the AYK IT section to complete the database management system. In response to review comments of past proposals (AYK SSI and Norton Sound), metadata according to standards of the USGS National Biological Information Infrastructure (NBII) will also be included in the database. IT staff have received training on the NBII standard for metadata and software options.

CONCLUSIONS

Over the last two decades salmon returns in the AYK region have declined sharply. Population dynamics of several salmon species appear to be undergoing substantial change, perhaps in response to large-scale environmental influences. Residents of western Alaska are heavily dependent on both subsistence and commercial salmon fisheries. In order to effectively manage these fisheries and maintain healthy salmon stocks, managers of both subsistence and commercial fisheries need timely access to historical stock and abundance information. They also need to collect new data in a standardized manner. The AYK Region has made substantial progress in fulfilling its goal to develop an electronic information storage and retrieval system that is comprehensive in scope, flexible in use, and available to all parties involved in fisheries management and research. In addition fishery biologists and technicians require an efficient means of processing and storing the large quantity of new data collected annually. Data from new stock assessment projects created by federal subsistence fisheries management can also be addressed by this project. An enormous amount of historical data exists for the region and it needs to be related and placed in context with the new data collected each year. Projects FIS 00-016 and FIS 02-069 have provided funds for the beginning stages in developing the AYK salmon database management system. Funds and staff expertise now exist to reach this goal.

RECOMMENDATIONS

With funding secured and full staffing to complete the database management system for AYK salmon the following is recommended.

1. Complete intermediate databases in MS Access for ASL and escapement data (step 5, page 4, in the development of a database) which include data through the current year and intermediate data entry, editing, and reporting applications.

2. Present data entry, editing, and reporting applications to AYK staff and formalize the data collection standards needed to successfully maintain the region's database management system.

3. Incorporate as many of the SASPOP structure and tables, particularly geographic links, as possible in the escapement database development to maintain compatibility statewide.

4. Begin to incorporate metadata requirements according to standards of the NBII to include choice of software tools, structure, file definitions, etc. for inventoried data.

5. Complete a software plan for an SQL server database with web access, which will be the final repository of AYK salmon fisheries data. Upon plan approval complete the database and import data from Access databases.

6. Work with Subsistence Division to minimize duplicative storage of subsistence salmon harvest data.

With the realization that lost data are becoming increasingly difficult and expensive to find the following is recommended for the data rescue portion of this project.

7. Conduct a thorough re-check of the scale card inventory, tally of borrowed scales, and otolith count to ensure that any remaining "missing" files are indeed missing rather than nonexistent with an emphasis on the Bethel office. This will be the last step in the final recovery stage followed by contacting project managers and area offices in an attempt to locate any remaining missing data.

8. Correct data in the SASPOP database only for errors not associated with rounding.

The following changes are recommended for data processing in AYK region.

9. Develop a system of archiving project and ASL data and scales in a central location. Enforce use of standard codes and format. Enforcement can be accomplished through programming to minimize impact on biologists.

10. Include data processing methods in project operational plans.

11. Add data processing and technical support costs in submitted proposals.

12. Involve other regions of the state in database development to minimize redundancy, maximize efficiency and promote standardization.

13. Develop a procedure whereby changes to historical data resulting from new analysis or otherwise are also incorporated into the AYK salmon database management system.

LITERATURE CITED

Banducci, A., E.L. Brennan, W. Jones, and J. Menard. 2003. Annual management report 2002 Norton Sound – Port Clarence – Kotzebue. Alaska Department of Fish and Game, Division of Commercial Fisheries, Regional Information Report 3A03-30, Anchorage.

Brase, A.L.J., and H.H. Hamner. 2002. Subsistence and personal use salmon harvests in the Alaska portion of the Yukon River drainage, 2001. Alaska Department of Fish and Game, Division of Commercial Fisheries, Regional Information Report 3A02-32, Anchorage.

Burkey, Jr. C., M. Coffing, J. Menard, D.B. Molyneaux, P. Salomone, and C. Utermohle. 2001. Annual management report for the subsistence and commercial fisheries of the Kuskowkim Area 2000. Alaska Department of Fish and Game, Division of Commercial Fisheries, Regional Information Report 3A01-34, Anchorage.

Caylor, D.A. 2000. Subsistence salmon harvest assessment program overview: Yukon Area. In Alaska subsistence fisheries database, Version 3.10, 10/11/02. Alaska Department of Fish and Game, Division of Subsistence, Anchorage.

Caylor, D. A., and R. J. Walker. 2003. Alaska subsistence fisheries database. Version 3.2 for Microsoft Access 2000. Alaska Department of Fish and Game, Division of Subsistence, Juneau.

Georgette, S., D. Caylor and S. Tahbone. 2003. Subsistence salmon harvest summary, Northwest Alaska, 2002: Norton Sound District, Port Clarence District, and Kotzebue. Alaska Department of Fish and Game, Division of Subsistence, Juneau.

Hamazaki, T. 2003. Kwiniuk River salmon counting tower project review and variance estimation. Alaska Department of Fish and Game, Division of Commercial Fisheries, Regional Information Report 3A03-20, Anchorage.

Hamner, H.H., R.R., F. West, and K. Clark. 2002. Shared information for fishery management in AYK. Alaska Department of Fish and Game, Division of Commercial Fisheries, Regional Information Report No. 3A02-43, Anchorage.

Hamner, H.H., S. Karpovich, and S. StClair. 2003. Norton Sound salmon file inventory and problem review. Alaska Department of Fish and Game, Division of Commercial Fisheries, Regional Information Report No. 3A03-01, Anchorage.

Hamner, H.H., S. StClair, and H. Moore. 2004. An inventory of age, sex, and length data for Norton Sound, Kotzebue, and Kuskowkim chum salmon. Alaska Department of Fish and Game, Division of Commercial Fisheries, Regional Information Report No. 3A04-06, Anchorage.

StClair, S., and H.H. Hamner. *In Press*. Kuskokwim, Yukon and Kotzebue data rescue, correction and standardization. Alaska Department of Fish and Game, Division of Commercial Fisheries, Southeast Sustainable Salmon Fund Final Report for Approved Projects 45225, Juneau.

Vania, T., and six coauthors. 2002. Annual management report Yukon and northern areas 2000. Alaska Department of Fish and Game, Division of Commercial Fisheries, Regional Information Report No. 3A02-29, Anchorage.

Ward, T.C., M. Coffing, J.L. Estensen, R.L. Fisher, and D.B. Molyneaux. 2003. Annual management report for the subsistence and commercial fisheries of the Kuskokwim area 2002. Alaska Department of Fish and Game, Division of Commercial Fisheries, Regional Information Report No. 3A03-27, Anchorage.

Table 1. Tasks accomplished for OSM project FIS 02-069 pertaining to salmon data in the AYK Region.

1. Approximately 200 electronic ASL data files were recovered from archives (transferred from 3.5” floppy diskettes to ADF&G server with tape backup).
2. Parsing program written to recognize and load electronic ASL files into an Access database.
3. Preliminary ASL database structure developed and implemented in Microsoft Access.
4. 769 electronic ASL files containing 205,244 individual fish samples loaded into databases.
5. 336 handwritten ASL files containing 106,348 individual fish samples keyed into a database.
6. 100 OPSCAN ASL files containing 27,990 individual fish samples rescanned and stored as digital files.
7. Error-checking protocols developed and 1,272 ASL files preliminarily error-checked.
8. 695 datasets representing 40 projects and 33 years of escapement data error-checked.
9. Fifteen years of subsistence harvest files updated from obsolete (Rbase for DOS) to modern software (MS Access 2000).

Table 2. Rescanned age, sex, and length OPSCAN files.

Area[a]	*Species*	*Years*	*Files*	*# of fish*
Kuskokwim	Chinook	1986-1996	9	1,072
	Chum	1986-1992	5	1,340
	Coho	1985-1995	4	2,187
	Sockeye	1986-1995	10	1,819
	Pink	No data	0	0
		Area total	***28***	***6,418***
Norton Sound	Chum	1985-2000	21	6,614
Kotzebue	Chum	1984-1987	25	12,036
Norton Sound/ Kotzebue	Chinook	1984-2000	17	1,478
	Coho	1985-2000	6	1,438
	Sockeye	1984-1987	3	6
	Pink	No data	0	0
		Area total	***72***	***21,572***
	Total scanned files		**100**	**27,990**

[a]Yukon OPSCAN data forms for data not found in electronic format, could not be located for rescanning.

Table 3. Estimated number of age, sex, and length files by area and species.

Area	*Salmon Species*	*Years*	*Number of Files by Type*					
			Total [a]	*Located Electronic* [b]	*Missing Electronic* [c]	*% Located as Electronic*	*Handwritten, to be Keyed* [d]	*% As Electronic Files After Data Entry Complete*
Kuskokwim	Chinook	1964-1999	257	236	21	92%	0	92%
	Chum	1964-1999	229	203	26	89%	0	89%
	Coho	1961-1999	143	132	11	92%	0	92%
	Sockeye	1964-1999	179	162	17	91%	1	91%
	Pink	1968-1994	11	2	9	18%	0	18%
		Area Total	***819***	***735***	***84***	***90%***	***1***	**90%**
Yukon	Chinook	1960-2001	888	584	304	66%	159	84%
	Chum	1961-2001	908	584	324	64%	206	87%
	Coho	1964-2001	158	106	52	67%	43	94%
	Sockeye	1979-1998	7	2	5	29%	3	71%
	Pink	1970	1	1	0	100%	0	100%
		Area Total	***1,962***	***1,277***	***685***	***65%***	***411***	**86%**
Norton Sound	Chum	1962-2000	159	127	32	80%	6	84%
Kotzebue	Chum	1962-2000	173	135	38	78%	6	82%
Norton Sound/ Kotzebue	Chinook	1966-2000	99	77	22	78%	0	78%
	Coho	1963-2000	56	49	7	88%	1	89%
	Sockeye	1963-1999	12	10	2	83%	0	83%
	Pink	1965-1979	11	1	10	9%	0	9%
		Area Total	***510***	***399***	***111***	***78%***	***13***	**81%**
	Total Files		**3,291**	**2,411**	**880**	**73%**	**425**	**86%**

[a]Total number of ASL files of scale card collections grouped by year, species, gear, and project.
[b]Number of file for which electronic data has been aggregated and inventoried.
[c]Number of ASL files for which electronic data cannot be found.
[d]Number of handwritten ASL files that must be keyed into electronic format. Number is included in the "Missing electronic" column.

Table 4. Specific problems with age, sex, and length data files.

Category	Problem	Problem Details	Solution
Missing/Incorrect Header Codes[a]	Location Codes	Location codes often missing, incorrect, inconsistent, or improperly entered (location in stream code, no sub-district, etc.)	Match files to scale inventory using other data, fix location codes
	Species	Species code missing or incorrect	Match files to scale inventory using other data, fix species code
	Project	Project code missing or incorrect	Match files to scale inventory using other data, fix project code
	Dates	File dates often do not match scale card, sometimes day data not entered	Match files using other data, interpolate missing dates
	Other Codes	Gear codes, mesh sizes, length types sometimes omitted	Match files to inventory using other data, fix data
Card number	Type C files	4 Cards per page, so headers jump from 1 to 5 to 9, etc. Could make matching data to scales difficult.	Manually insert new headers with correct card numbers
Data redundancy	File combinations	Data combined from several locations/projects for summary purposes sometimes mistaken as unique data	Look for inconsistent location codes/dates, find original files and delete combination or extract and separate combined data
File type	Data in Excel spreadsheets	Loader programs can only extract data from text files for loading into database	Export data from spreadsheets into text files, name appropriately
Mark Sense Forms	Miscoding	Coded data differs from values written in margins or data entered twice in same column	Establish better post-season data sheet review practices, interpolate missing data or delete miscoded samples
	Scanner Errors	OPSCAN "stutters" on last card, adds garbage characters and repeats some samples on last card	Manually review files and remove garbage characters and redundant data
		OPSCAN or DOS software breaks transmission of data; scan continues but no further data is recorded	Manually review files and ensure all cards are recorded, rescan as needed

[a]Indicates primary data used to sort and inventory files and will indentify data in the database. Errors in these data categories could have major consequences if not corrected.

Table 5. Escapement data error-checking summary.

Region	Project	Years Operated	Species Counted					Number of Datasets Located		
			King	Chum	Sockeye	Coho	Pink	Spreadsheets	Escapement Reports[b]	Discrepancies[c]
Kuskokwim	Kogrukluk Tower	1969-70, 72-78	x	x	x	x		28	28	0
	Middle Fork Goodnews Tower	1981-1990	x	x	x	x	x	47	47	1
	Kanektok Tower	1996-97		species were pooled				6	3	0
	Kwehtluk Tower	1996-97	x	x	x	x	x	9	0	n/a
	Takona Tower	1995-98	x	x				7	0	n/a
	Kogrukluk Weir	1976-1997	x	x	x	x		81	55	9
	South Fork Salmon River Weir	1981-82	x					2	0	n/a
	Middle Fork Goodnews Weir	1991-98	x	x	x	x	x	40	10	0
	George Weir	1996-97	x	x	x	x	x	10	0	n/a
	Kwehtluk Weir	1992	x	x	x	x	x	5	0	n/a
	Tulusak Weir	1991-94	x	x	x	x	x	20	15	1
	Tatlawisksuk Weir	1998	x	x				2	0	n/a
	Aniak Sonar	1980-94, 96-98		x				18	14	0
	Kanektok Sonar	1982-87	x	x	x	x	x	19	19	2
	Kuskokwim Sonar	1993-95	x	x	x	x	x	15	10	0
	Area Total							**309**	**201**	**13**
Yukon	Pilot Station Sonar	1986-97	x	x				24	4	4
	East Fork Andreasfsky Sonar	1981-85		species were pooled				5	5	4
	Anvik Sonar	1979-97		x				19	19	4
	Melozitna Sonar	1981-83		x				3	0	n/a
	East Fork Andreasfsky Weir	1994-98	x	x	x	x	x	22	14	14
	Giasa Weir	1994-97	x	x	x		x	13	8	1
	South Fork Koyukuk Weir	1996-97	x	x				4	4	2
	Beaver Weir	1996-97	x	x				4	2	1
	East Fork Andreafsky Tower	1986-88	x	x			x	9	6	2
	Anvik Tower	1972-76	x	x			x	15	10	5
	Kaltag Tower	1994-97	x	x				8	5	0
	Nulato Tower	1994-97	x	x				8	8	4
	Clear Tower	1995-97		x				3	3	0
	Chena Tower	1993-97	x	x				10	10	7
	Salcha Tower	1993-97	x	x				10	10	7
	Area Total							**157**	**108**	**55**
Norton Sound/Kotzebue	Kwiniuk Tower	1965-98	x	x		x	x	98	98	17
	North River Tower	1972-74,84-98	x	x		x	x	31	31	2
	Niukluk Tower	1979, 95-98	x	x		x	x	18	18	0
	Eldorado Tower	1995-98	x	x		x	x	16	12	0
	Nome Tower	1993-95	x	x		x	x	12	12	5
	Snake Tower	1995-98	x	x		x	x	14	12	0
	Shaktoolik Tower	1996-98	x	x		x	x	12	12	2
	Pilgrim Tower	1997	x	x		x	x	4	4	1
	Noatak Sonar	1981-83,91-94		x			x	12	12	1
	Nome Weir	1996-98	x	x		x	x	12	12	0
	Area Total							**229**	**223**	**28**
	Region Total							**695**	**532**	**96**

[a]A dataset was defined as salmon counts for one species/year/project
[b]Escapement reports are stored in hard copy format in the Anchorage Fish and Game office
[c]Salmon escapement counts from the spreadsheet were compared to the counts in the escapement reports

Figure 1. AYK Region management areas.

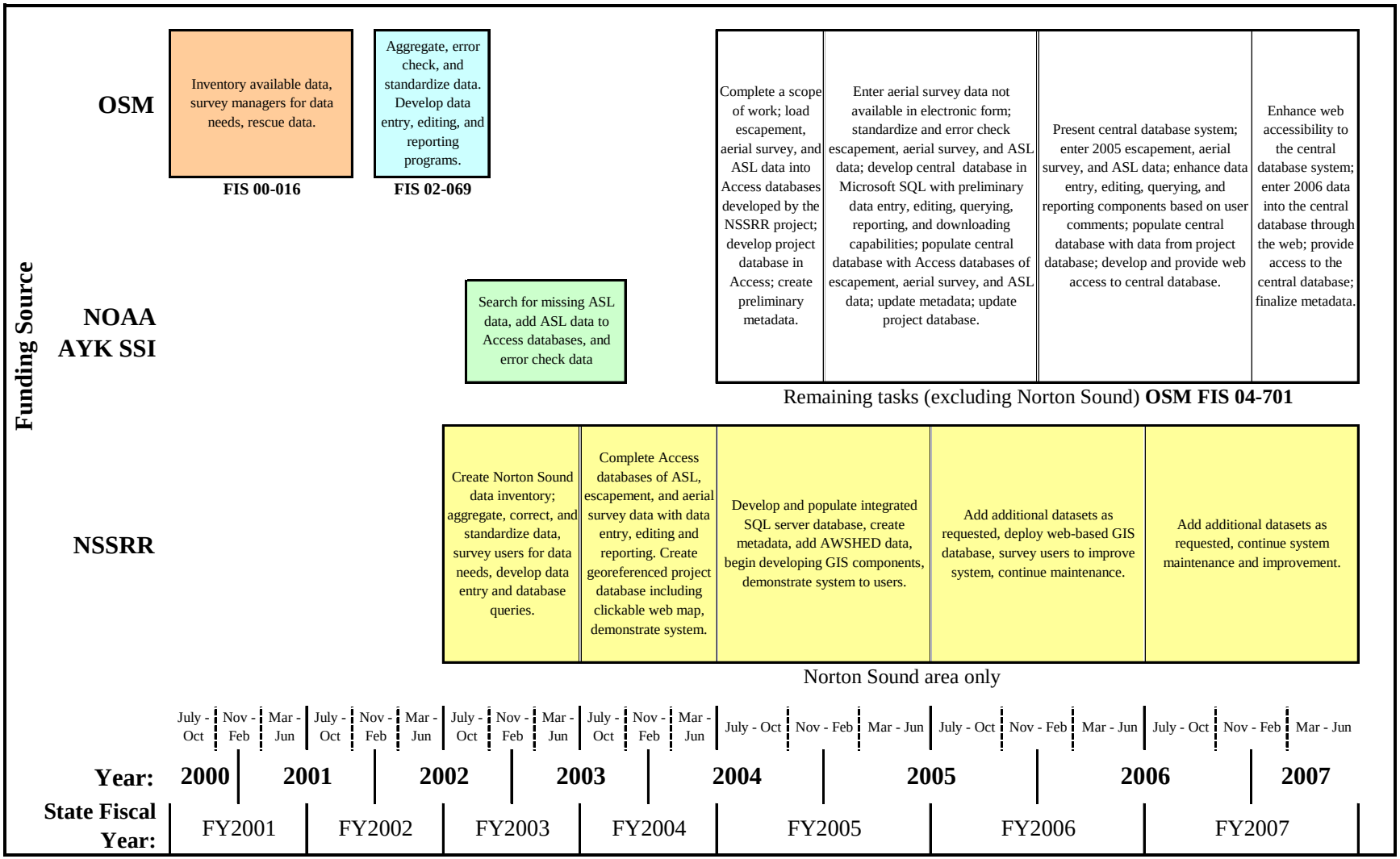

Figure 2. Project timeline for development of the AYK database management system. (OSM= USFWS Office of Subsistence Management; AYK SSI= AYK Sustainable Salmon Initiative, NOAA=National Oceanic and Atmospheric Administration, NSSRR= Norton Sound Salmon Research and Restoration, FY is state fiscal year).

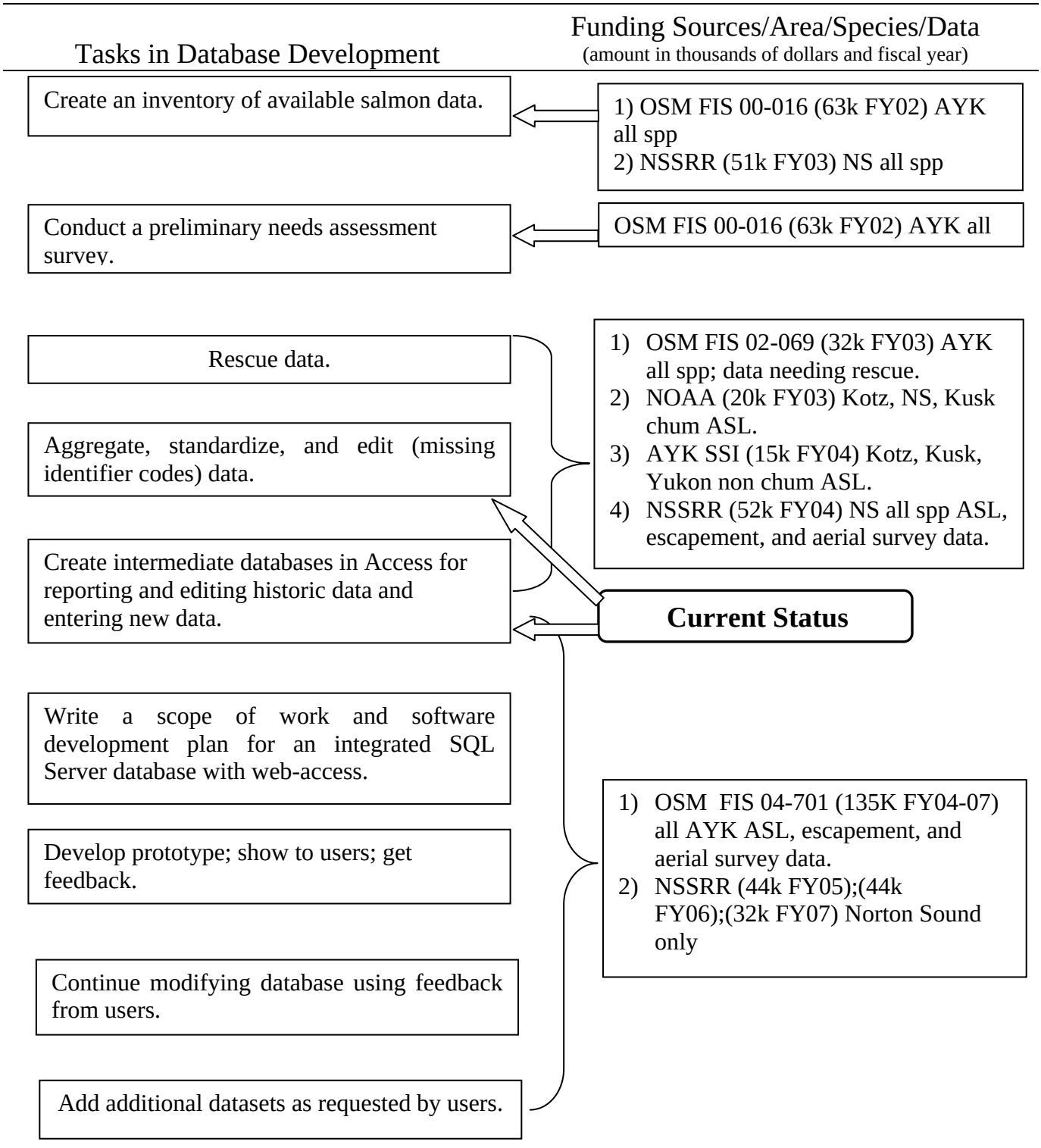

Figure 3. Tasks for development of the AYK salmon database management system including funding sources, amounts by fiscal year, and area. Not included are state general fund support of an Analyst/programmer IV, Analyst/programmer III, and a Research Analyst I ($150k annually).