

**FISH/SHELLFISH STUDY NUMBER 30**

Study Title:Database Management

Lead Agency:ADF&G

**PROJECT JUSTIFICATION**

FS-30 addresses the need to catalogue and maintain the principal electronic copies of raw data collected by the ADF&G Natural Resource Damage Assessment & Restoration (NRDA) Fish/Shellfish and Subtidal projects.

Assessment of injuries, successful restoration, and ongoing monitoring efforts ultimately are grounded in the data sets generated by NRDA studies.

Because of the fundamental role these data play in determining the effectiveness of any restoration program, it is important that care be taken to adequately document, archive, and maintain these principal electronic data sets.

In addition, the ADF&G personnel associated with FS-30 are directly responsible for the maintenance of critical historical fisheries databases referenced by many NRDA projects. This connection provides several added benefits to the NRDA effort, including direct access to historical data, technical expertise, and the use of the ADF&G Commercial Fisheries Wide Area Network (WAN) for electronic correspondence and transfer of data.

Principal Copy of  
Electronic Data Sets

- NRDA:  
FS-1, 2, 3, 4A, 11,  
13, 27, 28,  
ST-5, (FS-5, ST-2AB,  
ST-6).
- Historical Data:  
Commercial fisheries harvest  
Fisheries escapement data
- Restoration Projects:  
R-53, 58, 59, 60ABC, 105, and

113, (R-90, 106).

Documentation (DOCS) Database Management System (DBMS)

FS-30 supports 17 NRDA and restoration projects, with ancillary support to an additional 6 projects. Including historical data sets, the current amount of raw electronic data involved is estimated to be between 1,500,000,000 and 1,800,000,000 bytes. Tracking this volume of information requires significant time and effort; leaving it to Principal Investigators (PIs) would severely limit their ability to focus on project work. Ignoring data management entirely would ultimately lead to the loss of millions of dollars in data collection effort as projects complete or personnel transition to other projects. In brief, this project:

- Catalogues, archives, and maintains the principal copy of raw electronic data sets for FS-1, 2, 3, 4A, 11, 13, 27, 28, ST-5, (FS-5, ST-2AB, ST-6).
- Facilitates direct access by PIs to historical fisheries data sets essential to NRDA studies. Historical data includes commercial fisheries catch and escapement figures.
- Provides data processing and technical support for PIs and NRDA functions, including the use of ADF&G Commercial Fisheries WAN.
- Proposes to unify the data catalogues and maintenance of principal data sets for continuing ADF&G fisheries assessment, restoration, monitoring projects. This should facilitate sharing raw data between agencies and the ability to provide this information to the public.
- Proposes to catalogue, archive, and maintain the principal electronic data sets for R-53, 58, 59, 60ABC, 105, 113, (R-90 and 106).

Demonstration of the success of restoration effort depends directly on measurable results. Any restoration assertion is ultimately linked to principal assessment, restoration, and historical data sets.



For this reason, it is important that principal data sets of NRDA studies be documented, archived, and maintained.

## OBJECTIVES AND PRINCIPAL COMPONENTS

### Synthesis

Most of the original NRDA projects are targeted at the species level. Restoration projects which are targeted at the ecosystem level require synthesis of broadly disparate electronic data sets.

In addition, future restoration studies at the species level will depend heavily on access to established assessment and historical data.

FS-30 documents the content of existing NRDA and relevant historical data sets, making the sharing of data between projects possible. (This documentation covers all types of data relevant to a given study, and is not limited to chemical or GIS information.)

### Data Management

It is important that a principal copy of raw data be identified and archived. When data is shared between projects, multiple copies often diverge; any resulting conflicts must be resolvable through reference to a recognized true copy of the data.

FS-30 is responsible for storing true copies of raw data, and documenting changes made through ongoing quality control. Inherent in this responsibility is need to limit access to authorized parties. FS-30 also implements regular back-ups of electronic data sets.

### Requests for Information

When NRDA data are released by the Trustees, the public demand for this information may overwhelm PIs and support staff, thereby limiting their ability to focus on restoration work. By providing a central catalogue of raw data, and a recognized true copy of data sets, the extra work resulting from information requests should be reduced.

### Confidentiality

The dissemination of certain data integral to NRDA projects is limited by Alaska statute. Personnel associated with FS-30 are charged with tracking this information, and ensuring that state law is not violated.

## Technical Support

FS-30 is staffed by data processing and computer professionals. By its nature, this project has close contact with PIs and is aware of their technical needs. FS-30 assists PIs and their affiliates with technical issues, and thereby allows these individuals to focus on restoration work.

The ADF&G section charged with FS-30 also supports a statewide computer network (separately funded and maintained by ADF&G), which greatly facilitates sharing of information between PIs, Managers, and the Trustee Council. Current and future proposals, their support, and on-going requests for additional information rely heavily on this network.

## General

FS-30 is responsible for constructing a cost effective database management system (DBMS) to readily retrieve and order selected data from original data in electronic form according to user specified criteria of time, space, and other variables. The DBMS should be constructed to meet the following criteria, in order of priority:

1. Completeness of contents
2. Speed of retrieval
3. Ease of use in assembling primary data into datasets for further analysis by other software

## Specific Objectives

1. Continue maintenance of the secure repository for identified FS and ST studies NRDA and Restoration Project Data.

The data generated from studies relating to the Exxon Valdez oil spill are an important resource for the State of Alaska, the Federal Government, and the scientific community. Besides providing information for improved management of natural resources, these data will play a vital role in determining the success of ongoing restoration and enhancement projects.

The data will also serve an important role in subsequent legal actions related to the Exxon Valdez oil spill. Due to the data's potential role in the settlement of legal disputes, it is important that the conclusions derived from raw data be reproducible. When dealing with large raw data sets, reproducibility can only be ensured if a single repository

of the data is acknowledged. Without a recognized (single) repository, proliferation of updates and changes in the data across multiple copies may lead to contradictory conclusions.

2. Protect project data from accidental loss.

The datasets from NRDA and Restoration Projects should be adequately protected from inadvertent loss. Placing a study's principal copy of electronic data on a database server with regularly scheduled backup procedures should reduce the responsibility of the principal investigators in this important task.

3. Provide easy access to designated individuals and agencies.

It is essential that principal investigators have ready access to raw data. Research efforts should not be limited by access to data.

4. Limit unauthorized access.

The data storage facility must provide mechanisms for adequate security. Only designated individuals should have access to the data obtained from NRDA and Restoration projects.

5. Establish procedures for sharing data between studies and agencies.

It is recognized that the collective data of the NRDA studies may lead to conclusions which were not anticipated on a study by study basis. The data should be stored in such a fashion that it is possible to test hypotheses which span multiple studies. To this end, the data from the individual studies must be catalogued and stored in a way that facilitates sharing between studies and agencies.

6. Catalogue NRDA data and future Restoration Project data.

A complete catalogue of Assessment and Restoration Project data should include both general and detailed descriptions of the data. General descriptions should allow an independent party to determine the content and potential relevance of a dataset; detailed descriptions are essential for incorporating data into further studies.

7. Expedite information requests.

It is anticipated that future legal action on the Exxon Valdez oil spill may place heavy demands on PIs and staff, both in the

form of standard requests for information, as well as through the deposition process. While only the PIs and individuals associated with a specific study are qualified to evaluate and conduct analyses of data, a properly constructed repository of data and clearly defined procedures for accessing raw data should ease the burden of information requests and the demands on individuals involved in deposition.

8. Establish procedures for maintaining the repository data.

Clearly identified procedures will be implemented for maintaining information in the data repository. Such procedures should address the means for adding, deleting, and modifying data in the datasets, and should provide suitable documentation of relevant maintenance activities.

9. Describe the WAN database, and its implications.

A central repository of the data is envisioned. This repository should contain the current principal copy of electronic data for any given study. For performance reasons, it may be necessary to distribute portions of this database to local offices; procedures must be elaborated for synchronizing distributed copies of the datasets.

10. Provide both text and Graphical User Interface (GUI) access to historical data bases catch, escapement etc.

Develop direct access to important historical databases, including the capability for customized record selection, summary, and formatting. PIs can, from personal computers attached to the ADF&G Commercial Fisheries WAN, select data subsets using their own criteria, custom summarize data to 10 levels, then electronically transfer their new dataset to their location for use in their study environment (R:BASE, SAS, etc.) Records on a detailed level will be available, selected and sorted to the user's specifications.

#### STUDY METHODOLOGY

Access to historic databases in support of NRDA studies will be provided through an interface capable of providing summary and detail records sorted in a variety of output formats according to user specifications. The system will be accessible by authorized IBM-compatible personal computers on the ADF&G Commercial Fisheries WAN. It will be made available through a linked system of Local Area Networks (LANs) covering offices in Kodiak, Anchorage, Cordova and Juneau. The interface allowing



non-programmer access to the database will be developed for text and GUI platforms.

The NRDA study databases will be cataloged and stored in a central secure repository. Access to these data will be available to authorized staff. Documentation will include descriptions of each data set, covering aspects of physical layout, fields contained, purpose of data set, and author. This documentation should facilitate sharing data between PIs and agencies. Also, subsequent requests by the public will be facilitated by this documentation.

The original scope of data for FS-30 was commercial species from Prince William Sound, Kodiak, Cook Inlet, and Chignik areas. After discussions with assessment and restoration researchers we have changed the priority and type of observations to be incorporated. They are, in order of priority:

1. NRDA project data of global interest.
2. Commercial fisheries catch and effort data by area, species, and gear type.
3. Salmon escapement data, including aerial survey counts, stream counts, weir counts, and sonar counts.
4. Pre-emergent and egg density counts.
5. Biological data including age composition, size, sex, growth, and stock composition.
6. Groundfish and shellfish survey data.

This project will make use of an ADF&G statewide database network infrastructure being separately developed with State of Alaska general funds. Exxon Valdez oil spill settlement monies are not used to develop this network.

#### NRDA Project Data Sets

1. A secure database/file server system has been installed. (Objectives 1, 2, 3, 4).

A database server has been installed in the Region II ADFG office. Separate areas for each of the NRDA studies have been established. Procedures are being developed for establishing accounts, granting access, and ensuring appropriate backup of the datasets.

2. The documentation of NRDA (FS,ST) datasets is proceeding in two phases. (Objectives 5, 6, 7).

In the first phase, a general description of the datasets for a given NRDA study are completed. Included for each study are a qualitative description of the dataset, an estimate of the size of the dataset, the working format of the dataset, the individual responsible for the content of the data, the primary fields represented, and an estimate of the dataset's completeness and an estimate of the extent to which the data has been verified. In addition, primary investigators and their associates are identified as part of this general documentation process.

In the second phase, detailed descriptions of the data are elaborated. In addition to a textual description for each data field, the following data are defined at the field level: type, size, key status (must exist, must be unique), data validation rules, lookup tables (foreign keys), null values, value justification in the field, and leading fill characters. Synonyms for the fields are included where appropriate and known. Record definitions are defined as aggregates of the field definitions.

3. Procedures for data maintenance are under development. (Objectives 1, 8, 9).

Procedures are being developed for maintaining data in the repository. The repository holds the current principal copy of raw data for any given study. Procedures for reporting suspect data, modifying and updating datasets, and logging versions are under development. Performance of the WAN must be monitored before the procedures for data maintenance can be finalized.

#### Commercial Fisheries Historical Data

1. Programs have been written to analyze historic harvest data for errors. To date over 3.5 million records from spill affected areas have been searched.
2. Original documents have been obtained for incorrect records and corrections applied to the database.
3. Documentation has been written and assembled for changes made to the historic database (samples attached).
4. The technical card documenting codes has been revised and will be printed and distributed soon.

5. The detailed project plan for developing the historic commercial catch database has been substantially revised, now with an emphasis on NRDA direct access to detailed and summary data, and output formats in ASCII, spreadsheet, and R:BASE formats. (See attachments).
6. Purchase and development of a batch processor (separately funded) is under way and the interim detail data made available to NRDA researchers could be replaced by late spring.
7. The lead programmer is working closely with Commercial Fisheries networking staff to ensure that access to the wide area network is available and compatible with Oil Spill division administration and NRDA projects.
8. The Anchorage office is now connected to the department's wide area network. Cordova is scheduled to be connected by the end of November. The NRDA PIs, their affiliates, and Oil Spill Division staff now communicate and exchange documents via electronic mail.

#### DELIVERABLES

The primary deliverables for FS-30 include:

- Documentation of principal electronic data sets for selected NRDA FS/ST studies and future restoration projects. This documentation includes general description of data set content, import/export mechanisms facilitating data sharing between projects and agencies, and detailed data element definitions.
- Archives of principal electronic data sets, and modification logs to principal data.
- Support documentation (in electronic form) for selected NRDA FS/ST studies and future restoration projects.
- Software systems providing direct access to selected historical fisheries data sets by designated PIs and their affiliates.

#### SCHEDULE AND TIMELINES

The work of FS-30 is tied directly to the progress of NRDA FS, ST, and Restoration projects. Data collected by studies that FS-30 supports is keyed and subjected to quality control measures by the principal investigators of the specific FS, ST, or Restoration

Project. After QC is completed, a principal copy is made available to FS-30 for inclusion in the data repository. Concurrent with QC efforts are data documentation procedures which support the principal data sets.

Historical fisheries catch data is currently available to PIs through the Commercial Fisheries Fish Ticket System. New user interfaces will be in place for direct access to this data by the end of the state's fiscal year. Additional work on the Fish Ticket system will be completed May 1, 1992. (This additional work, which develops the existing system according to user requests, is separately funded through ADF&G's Commercial Fisheries budget).

#### FS-30 Database Management:

- Catalogues, archives, and maintains the principal copy of raw electronic data sets for FS-1, 2, 3, 4A, 11, 13, 27, 28, ST-5, (FS-5, ST-2AB, ST-6).
- Facilitates direct access by PIs to historical fisheries data sets essential to NRDA studies. Historical data includes fisheries catch and escapement figures.
- Proposes to unify the data catalogues and maintenance of principal data sets for continuing ADF&G fisheries assessment / restoration / monitoring projects. This should facilitate sharing raw data between agencies and providing this information to the public.
- Proposes to catalogue, archive, and maintain principal electronic data sets for R-53, 58, 59, 60ABC, 105, and 113, (R-90, and 106).
- Provides data processing and technical support for PIs and NRDA functions, including the use of ADF&G Commercial Fisheries WAN (wide area computer network).

#### ACRONYMS AND ABBREVIATIONS

DBMS - Database Management System  
EVOS - Exxon Valdez Oil Spill  
FS - Finfish / Shellfish (also, some Subtidal Studies)  
FT - Fish Tickets  
GUI - Graphic User Interface  
NRDA - Natural Resource Damage Assessment  
PI - Principal Investigator

SCENARIOS

FS-30 is unique among NRDA projects, in that it participates in the responsibility for maintaining principal copies of raw electronic data from other NRDA studies. The unique role FS-30 plays in NRDA efforts is exemplified by the following possible scenarios.

- 1.) Different copies of a particular data set are used leading to different conclusions. The validity of the data is questioned, and the entire data set is deemed unusable.

FS-30 defines and archives the principal copy of data, and ensures that any changes made to this data are appropriately catalogued.

- 2.) A given data set is inadequately documented, and units are omitted. For example, is a given value in fathoms, meters, feet, or yards?

FS-30 directly addresses and eliminates this problem through the data catalogue it maintains.

- 3.) A PI retires or moves on to other work. A recognized copy of raw data may not exist. The value of any existing primary raw data is questionable without adequate support documentation.

FS-30 archives and documents existing data sets, thereby facilitating transition of personnel.

- 4.) The raw data from a given study proves vital to a number of external public agencies and/or private concerns. The PI is overwhelmed with requests for information, leading to a compromise of effort on current restoration activity.

FS-30 can provide a first point of inquiry regarding raw data, including both general and detailed descriptions of principal data sets. Most initial information should be obtainable without direct interaction with the PI.

- 5.) Because a federal or state agency is not familiar with the details of data from an existing NRDA study, effort is expended to re-obtain information. Alternatively, a project is never proposed, because the prospective PI is unaware of existing data obtained through the efforts of other studies.

FS-30's catalogue of data sets should provide a general description of what is currently available, and what the possibilities are for future restoration.

BUDGET (\$K)

Salaries	\$154.0
Travel	6.9
Contractual	10.4
Supplies	4.6
Equipment	<u>2.8</u>
Subtotal	\$178.7
General Administration	<u>23.8</u>
Total	\$202.5