



## **NOAA Technical Memorandum NMFS F/NWC-18**

### **A Description of the Resource Survey Data-Base System of the Northwest and Alaska Fisheries Center, 1981**

Ralph J. Mintel  
and  
Gary B. Smith

November 1981

U.S. DEPARTMENT OF COMMERCE  
National Oceanic and Atmospheric Administration  
National Marine Fisheries Service

This TM series is used for documentation and timely communication of preliminary results, interim reports, or special Purpose information, and have not received complete formal review, editorial control, or detailed editing.

**This document is available to the public through:**

National Technical Information Service  
U.S. Department of Commerce  
5285 Port Royal Road  
Springfield, VA 22161

A DESCRIPTION OF THE RESOURCE SURVEY DATA-BASE SYSTEM  
OF THE NORTHWEST AND ALASKA FISHERIES CENTER, 1981

by

Ralph. J. Mintel and Gary B. Smith

Resource Assessment and Conservation Engineering Division  
Northwest and Alaska Fisheries Center  
National Marine Fisheries Service  
National Oceanic and Atmospheric Administration  
2725 Montlake Boulevard East  
Seattle, Washington 98112

November 1981

**THIS PAGE INTENTIONALLY LEFT BLANK**

## ABSTRACT

This report describes the computer data-base system that has been developed at the Northwest and Alaska Fisheries Center for handling the requirements of multiple, large-scale, groundfish trawl surveys. In addition to describing the organization and content of the system including data, software, and documentation, new concepts and approaches are discussed that improve the system's reliability, flexibility, and value as an information base.

The total volume of fisheries survey data included in the system is presently 1,215,873 data records, representing the results from approximately 30 yr of resource survey investigations by the National Marine Fisheries Service (and its predecessor agencies) in the Pacific Northwest and Alaska regions. The data are organized in the form of three major area data bases, include results from 201 vessel-cruises and 20,520 trawl samples, and are used as a shared resource by the organization.

The software, or computer programs, developed to link the survey data base with the users of the data include 576 source program (main, subroutine, and other source) files primarily written in FORTRAN. Principal functions include analyses, map plotting, general utility, data management, data editing and checking, and subroutine applications. Two objectives that have guided software development for the system have been generalizing software solutions to specific problems and then institutionalizing them. In addition, standards have been set to improve software reliability and decrease maintenance costs.

The system has an extensive library of computer-based documentation, approximately 14,700 lines total volume, that includes directories to the data and software resources, records of data characteristics, descriptions of the computer programs, other reference materials, and a news file.

Important new capabilities have resulted from the development of general utility programs that provide mechanisms for extending uses of the system's data and software. Of these, the implementation of relational concepts probably has the most significant implications. Data may be selected from the data base, joined with other data by cross-relation, manipulated, and then routed into the system's (or other) analytical and reporting programs. These capabilities provide versatility for broad-range applications, desirable adaptability to change, and open opportunities for interfacing with other major data bases in the future.

## TABLE OF CONTENTS

	<u>Page</u>
INTRODUCTION .....	1
Fisheries Resource Surveys .....	1
Interfacing with the Future .....	3
Purpose of Document .....	4
SYSTEM OBJECTIVES .....	7
Short-term .....	7
Long-term .....	8
Integrated Design .....	8
Data Control .....	9
Analytical Control .....	9
Data as a Shared Resource .....	9
Data Archives and Security .....	10
SYSTEM ENVIRONMENT .....	13
Computer Hardware .....	13
Users .....	13
Work Characteristics .....	14
SYSTEM STRUCTURE .....	15
Documentation Library .....	15
Content .....	15
Program Documentation .....	15
ADP Codebook .....	16
Data Record Formats .....	16
Data-Base Master Index .....	17
Species Code Dictionary .....	18
RACE News File .....	18
Data Resources .....	19
Major Data Bases .....	19
Data Types .....	19
Data Structure .....	21
Physical Organization .....	23
Standards .....	25
Software Resources .....	30
Major Program Groups .....	30
Data Files Related to Software .....	33
Standards .....	33
SYSTEM FUNCTION .....	39
Data Flow .....	39
Data Base Management .....	40

	<u>Page</u>
Data Base Retrieval .....	40
Security Control .....	49
SYSTEM ADMINISTRATION AND PERSONNEL .....	51
SYSTEM ACCESS .....	53
RACE Division Users .....	53
Other NWAFC Users .....	54
Others .....	54
ACKNOWLEDGMENTS .....	55
REFERENCES .....	57
APPENDICES	
Appendix A. Program Documentation Example .....	59
Appendix B. Program Documentation Index .....	65
Appendix C. ADP Codebook Excerpt .....	71
Appendix D. Data Record Formats Excerpt .....	75
Appendix E. Data-Base Master Index .....	77
Appendix F. Species Code Dictionary Excerpt .....	83
Appendix G. RACE News File Excerpt .....	85
Appendix H. DMSIII, the Relational Data Base Management Utility . .	87
Appendix I. Analytical Programs .....	93
Appendix J. Other Programs and Subroutines .....	99
Appendix K. Software Standards .....	107

## INTRODUCTION

### Fisheries Resource Surveys

The survey data-base system described in this report has been developed in response to the needs of fisheries resource assessment research. In particular, the report describes the system that has been developed at the Northwest and Alaska Fisheries Center (NWAFC) to handle the requirements of multiple, large-scale, groundfish trawl surveys conducted by the Resource Assessment and Conservation Engineering (RACE) Division in the Pacific Northwest and Alaska regions.

At the NWAFC, some parts of the RACE survey data-base system are also used for other survey functions. For example, the discrete trawl sampling data collected during hydroacoustic surveys of pelagic fish populations are added to the survey data-bases, although the hydroacoustic data are recorded and analyzed using a separate, portable computer system. Similarly, specialized data resulting from observations of crab and shrimp during RACE trawl surveys are managed and stored separately at the NWAFC's Kodiak Laboratory in Kodiak, Alaska.

In addition to the data that have resulted directly from resource surveys conducted by the RACE Division, extensive collections of trawl survey data from other sources have also been added to the RACE survey data-base system during the past 6 years. These collections have included published and unpublished cruise results from the International Pacific Halibut Commission, Alaska Department of Fish and Game, Far Seas Fisheries Research Laboratory (Shimizu, Japan), Fisheries Research Board of Canada, and Oregon Department of Fish and Wildlife. Fisheries survey data resulting from cooperative foreign research cruises by Japan, Poland, the USSR, Canada, and Republic of Korea have also been included.

Major types of groundfish trawl surveys include: (1) exploratory fishing, or resource appraisal surveys, designed to provide information on the composition, distribution, and abundance of biological resources and potential yields to fisheries; (2) monitoring surveys, intended to measure year-to-year and longer-term changes; and (3) research surveys, studying more specific research questions and problems. Among all of these, the kinds of data collected are relatively standard: information describing the cruise, the type of platform, time, sampling location, and sampling gear; taxonomic composition of the catch, weight, and number; sizes and ages; and more specialized observations that may include reproductive condition, stomach contents, and individual length-weight factors.

Trawl surveys are generally carried out on a very large scale, with extensive types of ecological sampling that result in voluminous data. For example, a demersal trawl survey conducted by the RACE Division in the eastern Bering Sea in 1979 resulted in a collection of 682 samples from a study area of 648,000 km<sup>2</sup>; 15,222 catch items for which biological identifications, weights, and number were recorded; 226,800 measurements of fish lengths; and 6,260 age determinations. In 1980, a trawl survey of fish populations along the Pacific Coast included 646 samples, 6,537 catch records, 43,220 length-frequency measurements, and 5,080 age determinations. Since several surveys of similar scale are conducted by the RACE Division each year, the total data flow is approximately 2,000-3,000 trawl samples (with associated information) per year.

The large-scale nature of these stock assessment surveys leads to unusual demands for organization of the field work, control and analysis of the data, and for providing service and control to multiple groups over successive years.

### Interfacing with the Future

The NWAFC resource survey data-base system has evolved from exploratory fishing and resource assessment investigations, including those by predecessor agencies, in the Pacific Northwest and Alaska regions from about 1940 to date and from the contributions of many individuals.

In the early 1940's, the U.S. Fish and Wildlife Service began exploratory fishing surveys off the Pacific Coast of the United States and Canada. These activities were later organized to form Exploratory Fishing and Gear Research Bases in Seattle, Washington, and Juneau, Alaska. In the 1950's and 1960's numerous exploratory fishing cruises were conducted from southern Oregon waters to the Chukchi Sea. Classic studies resulting from this period include those of Ellson et al. (1950), Pruter and Alverson (1962), Ronholt (1963), and Alverson et al. (1964).

Until 1975, resource surveys were relatively single-purpose and usually involved collections of less than 100 trawl samples per survey. Then in the mid-1970's, in response to needs to assess biological resources of the U.S. continental shelf and their vulnerability to oil development, and new responsibilities assigned under the Magnuson Fishery Conservation and Management Act of 1976, the scale and complexity of resource surveys underwent a rapid development. Surveys were bigger, 300-700 trawl samples per study, and total data volumes were very large; there was more emphasis upon multispecies assessments and studies of ecological relationships; there was need to improve timely reporting and data quality; more investigators became involved; and significantly, there emerged interest and need for analysis and comparisons of multiple surveys.

These trends have continued. New demands have included an increasing emphasis upon research and cross-relational uses of the survey data;

integration of all survey data to create a standardized, shared resource for broad-range applications by multiple users; establishment of a more stable software environment; and improvement in the user-system interface.

At the Center level, recent trends are inclining toward cross-relating different types of data and data-bases. For example, fisheries survey data might be combined or cross-related with U.S. foreign fisheries observer data to conduct a special analysis; marine mammals sighting data might be compared to a physical oceanographic data base; and foreign fisheries observer data might be cross-related with foreign catch statistics to make more detailed predictions of incidental catch rates.

Even at the level of the different major research centers of the National Marine Fisheries Service, there seem to be opportunities for standardization of stock assessment methods and perhaps sharing data and software resources.

In summary, the more recent pressures have been directed to improve the ability of survey data and software to accommodate change in its applications to fisheries and environmental management. This has been one of the most rapidly-developing areas of resource survey methodology in the past 5 years.

#### Purpose of Document

The RACE Division has recently completed a conversion of its data and software to a new Burroughs<sup>1/</sup> computer system. The system was installed at the NWAFC in its Seattle and Alaska facilities. Conversion activities included transferring data files and reorganizing them into a data-base system, converting and improving computer programs, preparing new documentaion, standardizing and cleaning up the data to conform to the new system, and educating users.

---

<sup>1/</sup> Reference to trade names does not imply endorsement by the National Marine Fisheries Service, NOAA.

The purpose of this document is, at this significant milestone, to:

1. Describe the contents of the new survey data-base system including data, software, and documentation;
2. Explain concepts, goals, and relationships of the new large-scale system;
3. Demonstrate approaches that have been taken to improve system quality, reliability, and management; and
4. Serve as a guide for potential users describing the available resources and system capabilities.

**THIS PAGE INTENTIONALLY LEFT BLANK**

## SYSTEM OBJECTIVES

The objectives for a resource survey data and computer software system must respond to both short and long-term goals of the stock assessment activities that it serves. However, since there are inherent conflicts between the suitable responses from different perspectives of time, this is an area of continuing challenge to those persons responsible for effective research management.

## Short-term

Short-term objectives, over the lifetime of a single stock assessment project, are usually dominated by interests of the project leader. These interests might include many or all of the following: to respond to some immediate organizational need(s), such as new information required for fishery management; to conduct the project work in a practical, efficient manner; to complete the project within the planned time schedule, budget, and standards of quality; and for personal/group accomplishment and recognition. In the environment of the RACE Division, most stock assessment projects have a maximum lifetime of one year, and project leaders are frequently responsible for managing 1-3 new major projects per year.

Short-term demands and the work load have led to the development and use of an established data processing and analytical system into which the project leaders can fit their various work. For example, many if not most of the initial types of information needed from stock assessment surveys are now fully anticipated and routine. These requirements have been met by developing general programs, shared among users, that perform standard analyses and produce standard reports. This aspect of the survey data-base system responds to short-term objectives and support needed for routine operations.

In the short-term, measures of success that the project leader applies to his analytical support system include convenience, flexibility, whether the required analytical tools, i.e., computer programs, are available, reliability., cost, and whether or not all steps can be adequately controlled and reviewed. Accuracy is frequently expressed as a concern, although its assurance actually requires attention through all stages of the information flow from data collection, to analysis and reporting.

#### Long-term

Long-term objectives become important after an organization has conducted stock assessment activities over a period of several years, and when there is an involvement of multiple investigators. General interests are to ensure that all of the work fits into a coherent plan; that investments of resources made by the organization are developed to the fullest extent possible and protected; to support and encourage the growth of the research itself, and the personal development of staff members; and to allow the organization to comfortably adapt to future change.

#### Integrated Design

Despite the usual immediacy and autonomy of short-term projects, over the long-term the various projects must add up and fit together in a coherent manner. This requires designing a system that is, and will continue to be, responsive to both user and organizational needs; that will have stability; that provides flexibility, such as responsiveness to ad hoc requests; yet that readily accommodates change and evolution. Central management and data independence (from the application programs) seem to be two keys to an effective long-term approach.

## Data Control

When resource survey activities extend over several years and involve multiple investigators, rules and standards become important to ensure the validity, consistency, and accuracy of the total data base. Definitions of particular data elements must be set and maintained, and rules must be made for codes and coding conventions. These rules and standards characteristically penetrate through all phases of an assessment project, and are perhaps best expressed and controlled through the development of a generalized field operations manual. This provides front-end control. Additionally, standards must be set and policed regarding the acceptable data quality, including both compliance to coding conventions and the error rate.

## Analytical Control

Two conflicting realities facing project leaders are the increasing scale and complexity of resource survey research, and the different levels of expertise offered by the persons available to do the work. As a result, an important aspect of long-term system design is to provide different levels of control and guidance: it is desirable to make the computer programs used to support routine operations simple, highly controlled, i.e., "canned", and to develop ample accompanying documentation; other research will require far more versatility, spontaneity, and background experience.

## Data as a Shared Resource

As resource survey work increases and becomes more complex, and as more investigators become involved, it becomes more advantageous to structure the organization's data so as to constitute a shared resource for broad-range

applications. This requires data control, as described in the previous section, and an organizational commitment to create a coherent data-base system. Two results might then be expected: (1) the data become increasingly viewed as an independent resource in themselves, an information base that can be used in a versatile manner for numerous types of information requests; and (2) by separating the data from the computer programs that use them, the data-base software is significantly insulated from effects of change.

Another important aspect of long-term analytical control is software reliability and the stability of the data-base software complex. To achieve these goals, an environment must be established where changes are planned, programming standards are set, policies are developed for acceptance testing of newly acquired software, the software system is insulated from effects of change, and adequate documentation is demanded.

#### Data Archives and Security

An important long-term objective, sometimes forgotten after the completion of individual projects, is ensuring that research data are not lost, tampered with, or misused. Fisheries survey data represent a substantial investment: on the average, one trawl sample represents field costs of about \$2,000; by extension, the data resulting from (and representing) a 600 trawl sample survey would have an investment value of \$1.2 million: and the total survey data held by the RACE Division, approximately 20,000 trawl samples, would represent an investment value of \$40 million.

In addition to collection costs, data from past surveys have intangible value as historical records. For example, an important type of information for fisheries management is the record of "virgin" stock conditions before the start of significant fishing. In other cases, resource survey data may be

valuable for providing similar descriptions of baseline conditions prior to additional types of important events--such as periods of different ocean climate, increased pollution exposures, or other man-made alterations.

Besides protecting against loss, another important aspect of data security over the long-term is preventing unauthorized modifications.

**THIS PAGE INTENTIONALLY LEFT BLANK**

## SYSTEM ENVIRONMENT

## Computer Hardware

Computer hardware is needed for operating a survey data-base system, but the fundamentals of the data-base (structure, standards, data management, and operation) are machine-independent. Regardless of the particular hardware configuration, documentation of the software and data is needed. A certain data organization such as haul records, catch records, and species code dictionary records is needed. Guidelines, or standards, for data and software quality control are needed. And someone must be assigned to manage and operate the system. These fundamentals transcend the particular hardware configuration and must be addressed in any system.

The NWAFC survey data-base system, originally developed on a mini-computer (32 kilo-characters of memory, card reader and punch, 2 million characters of disk storage, a plotter, FORTRAN compiler, and sort/merge utility), has been converted to a large-scale Burroughs 7800 system which includes multi-million character memory, printer, plotter, magnetic tape units, remote terminal support, and 170 million characters of disk storage for about 50 users in the RACE Division. Additional disk resources on the system are used for the operating system, by other divisions at the NWAFC, and by the computer vendor. Users also have access to statistical packages including "BMD," the University of California Biomedical Computer Programs (Dixon and Brown 1979); "SPSS," the Statistical Package for the Social Sciences (Nie et al. 1975); and "Minitab" (Ryan et al. 1981).

## Users

The primary users of the survey data-base system number about 25 in the groundfish assessment task unit, with another 25 occasional users of the

software and data throughout the rest of the RACE Division. Users in other divisions at NWAFC have access to the data-base and software. Users in Alaska at the Center's Auke Bay Laboratory and Kodiak Laboratory may access the survey data-base system via dial-up remote terminals, or via remote batch jobs.

#### Work Characteristics

The work characteristics have an important bearing on the nature of a data-base system in an organization. Some distinguishing characteristics of the resource assessment survey work at NWAFC include multiple groups conducting similar types of surveys in different geographical regions. Similar data processing problems are encountered by the multiple groups. This suggests that the data organization and software should be common to the multiple groups. High data volume suggests the development of software and methods that need to be more streamlined than those for low data volume. Current and -past projects requiring the analysis of numerous historical surveys suggest the standardization of data formats throughout the historical data. Increasing requests for cross-functional uses of the data also dictate the standardization of the data for effective and efficient access.

In summary, in developing a system, an assessment must be made of who the users of the system will be. This is important because it is a relatively small chore to write a computer program for only one person to use. It is quite another matter to develop systems for 25 or 50 users: in such cases, diagnostics, ease of use, documentation, and trace capabilities in the software often assume an importance that overshadows the program's original function. Similarly, the organization's work characteristics need to be considered in the design of a system. Finally, it is a straightforward technical task to elaborate the general system objectives into specific steps that can be accomplished on a particular hardware configuration.

## SYSTEM STRUCTURE

## Documentation Library

## Content

The documentation for the survey data-base system is made up of various components. These are:

1. Program documentation,
2. ADP codebook (ADP = Automatic Data Processing),
3. Data record formats,
4. Data-base master index,
5. Species code dictionary, and
6. RACE news file.

All of the documentation is maintained on disk files and is always available to users. Copies of the documentation may be listed on a terminal or written on the computer's printer.

Computer-based documentation, a tradition started on the old mini-computer at NWAFC, is a necessity for ease of maintenance, up to date documentation, availability of documentation, and ease of dissemination.

## Program Documentation

Program documentation describes the function of each program, how to execute the program, what file assignments are needed, and what parameters may be supplied for the run. Each program document is on a separate file, but the documentation for all programs may be printed with a single command, e.g.,

```
WRITE (PACE0360)DOC/=.
```

<sup>2/</sup>

---

<sup>2/</sup> Note: This is not intended to be a specific reference manual on the use of The Burroughs computer system. The computer commands shown here are for illustration only, as the particular invocation of the command may be different, depending on such things as a person's computer account number ("usercode," in Burroughs terminology), and default disk assignments.

The documentation for a single program might be listed on a terminal with a command, e.g.,

```
LIST (RACE0360)DOC/BIOMASS.
```

An excerpt from the program documentation is in Appendix A.

In some cases, closely related programs are documented in a single documentation file. Currently, there are over 120 entries in the program documentation library.

An index to the program documentation is kept on a disk file. This index shows the program title and a few words, one line, describing the program's function. The program documentation index is in Appendix B. The index will be printed when one requests a printed copy of all program documentation, but just the index may be printed with a command,

```
WRITE (RACE0360)DOC/A/INDEX.
```

#### ADP Codebook

The ADP Codebook is supplementary documentation describing the formats for field data collection forms, coding conventions for field data, vessel codes and names, gear codes and gear descriptions, and other tables of codes and meanings. This ADP Codebook reflects the current coding standards and meanings, and is used in conjunction with field data collection as well as with historical data retrieval and interpretation.

An excerpt of the ADP Codebook is in Appendix C. The ADP Codebook may be listed with a command,

```
WRITE (RACE0360)ADPCODEBOOK.
```

#### Data Record Formats

Inside the computer system most of the data base records are stored in binary, one computer word per item. The documentation of these record formats

describes the record type, the word number, each data item, and other information. Besides the standard data-base record formats (haul, catch, length-frequency), other internal or frequently referenced file formats are described.

This document would be referenced by anyone who would be selecting data from the master data base, as well as by any programmers. An excerpt of the formats is in Appendix D.

The formats may be listed with the command,

```
WRITE (RACE0360)DOC/A/FORMATS.
```

#### Data-Base Master Index

The Data-Base Master Index, one file per major area data-base, contains one record per unique vessel-cruise number combination. It is a summary of the data contained in the data base. The dates of start and finish, ranges in latitude and longitude, the nation(s) conducting the survey, and an abbreviated survey name are shown for each vessel-cruise. Also shown is a record count for each of the major file types in the data-base.

The master index, generated by a program which summarizes the data-base, is listed by running a specific program, e.g.,

```
RUN (RACE0010)LIST/MASTER/INDEX;,
FILE 1=(RACE0210)WC/BINE/MASTER/INDEX.
```

An example of the master index is shown in Appendix E.

The master index file will soon be a functional file, in addition to being informational, as programs for rapid retrieval of data by vessel and cruise numbers will first reference the data-base master index file.

It will also be used as a management tool for tracking the progress of processing new survey data for inclusion in the master data-bases.

## Species Code Dictionary

The Species Code Dictionary contains s-digit species codes, scientific names, and common names for the marine animal species that occur in the data-base. At present time, approximately 1200 species and higher taxonomic groups are included in the dictionary, and their codes have been organized and assigned on a systematic basis. As well as being a documentation file, the species code dictionary is a functional file since it is referenced by all application programs that report species names with other data.

The Species Code Dictionary may be printed with the command,

```
WRITE (RACE0010)SPECIES/DICT.
```

The dictionary is periodically issued in printed form, bound, and sorted numerically, alphabetically, and divided into lists of fish and invertebrates.

Although the coding system used in the Species Code Dictionary is unique to the NWAFC, the 5-digit species codes can be related or converted to the 12-digit taxonomic codes used as national standards by the National Oceanographic Data Center (National Oceanographic Data Center 1981) by using simple parallel code files.

An excerpt of the Species Code Dictionary is shown in Appendix F.

## RACE News File

Announcements of program problems, new software, software changes, meetings, and other information are made via a RACE News File. The most current announcement is at the beginning of the file, and the oldest information is at the end of the file.

The RACE News may be listed on a terminal with the command,

```
LIST (RACE0010)NEWS.
```

An excerpt of the RACE News File is shown in Appendix G.

## Data Resources

## Major Data Bases

The resource survey data are organized into three major area data bases reflecting the usual references to the data and the functional organization of RACE Division work:

1. Bering Sea and Aleutian Islands,
2. Gulf of Alaska, and
3. Pacific West Coast.

Each data base is identical in form to the other data bases, and each data base operates with the same software as the other data bases. A subset of a data base is a data base that can be manipulated with the same software as the original data base. Similarly, a combination of data bases is a data base.<sup>3/</sup>

Table 1 is a summary of the content and total volume of the three area data bases held by the RACE Division. In addition, the more detailed contents of the three area data bases at present time are shown by vessel-cruise and data type in the Data-Base Master Index examples of Appendix E.

## Data Types

Each major area data base has five primary file types that represent the field data:

1. Haul (station-specific data such as date, time, latitude, longitude, depth, sampling gear, temperature, gear performance);
- 2 . Position (a newer file type containing more details on position, e.g., latitude and longitude at the start and end of net tows);

---

<sup>3/</sup> This is an important concept from the system design standpoint. In many other systems, one program is not able to access two identical data bases without at least recompilation of the program. Furthermore, the creation of subsets or supersets of a data base might require a special program or programs, and extensive computer time.

Table 1.--Survey data volume: Summary of approximate record counts for each of the major area data bases by primary file type.

Area data base	Number of vessel-cruises	Haul records	Position records	Catch records	Length-frequency records	Specimen records	Total records
Bering Sea	67	8,775	3,003	132,519	432,554	71,835	648,753
Gulf of Alaska	102	8,557	2,891	108,200	192,931	93,140	405,821
Pacific Coast	32	3,188	2,071	33,413	62,704	59,891	161,299
<b>TOTAL</b>	<b>201</b>	<b>20,520</b>	<b>7,965</b>	<b>274,132</b>	<b>688,189</b>	<b>224,866</b>	<b>1,215,873</b>

3. Catch (species code, weight, number);
4. Length-frequency (species code, sex, length, frequency); and
5. Specimen (species code, sex, age-length, length-weight, reproductive condition).

#### Data Structure

The conceptual structure of the survey data base is illustrated in Figure 1. The data base is physically separated into the three major area data bases. For each major area there is a master index file containing one record per vessel-cruise. Ellipses in the illustration indicate that multiple records may be present for a given file type.

Subordinate to a master index record are all of the haul records for that vessel-cruise. Parallel to the haul records are the position records, which have additional details regarding the starting and ending latitudes, longitudes, and loran-C readings. The position records are optional, and are not necessarily present for all of the historical data.

For each haul record there may be multiple catch records, one per vessel-cruise-haul-species showing the catch weight and number. For hauls in which nothing was caught, there are no catch records or other subordinate records, but the haul records would be present.

Subordinate to the catch data are the length-frequency records, one record per vessel-cruise-haul-species-sex-length.

Specimen data records are usually one record per observation of length-age, length-weight, or length-maturity. Some specimen records have combinations of observations such as length, age, and weight all on the same record. From the standpoint of structure, specimen data does not fit in only one place. It may be subordinate to the length-frequency data, e.g., if the catch was subsampled



for the length-frequency data, and the length-frequency subsample was in turn subsampled for the specimen observations. In other cases the specimen data are not necessarily subsamples of the length-frequency data, and hence are subordinate directly to the catch records.

Particularly on the historical age data, some age-length specimen observations were not identified by haul number when the field collections were made as pooled samples. Specimen observations of this type can only be subordinate to a master index record, although a stratum code field on the specimen record allows classification finer than by just vessel and cruise numbers. Specimen observations with haul numbers, however, may be retrieved on the basis of criteria other than the original statistical design for collecting and pooling samples.

The dashed lines in Figure 1 show the possible structural relationships of the specimen data to the other data types.

#### Physical Organization

A traditional approach to data organization would be a physical hierarchy, e.g., with variable length records, haul record 1 followed by the catch records for haul record 1, followed by haul record 2; length-frequency records would be appropriately embedded in the catch data. In fact, some formal data base management system packages require such an embedded hierarchical physical organization of the data.

There are a number of disadvantages to such an organization. First, the historical specimen data would not necessarily fit in the structure, e.g., in some cases, it would be subordinate to the length-frequency data, other times subordinate to catch data, and sometimes subordinate only to a master index record. Next, programs that only needed to process a single record

type I e.g., catch data, would have to account for the other record types, if only to differentiate them from the desired data. And if in a data retrieval problem it is necessary to examine a subordinate record prior to selecting a superior record, the program must save the superior records, examine the subordinate records, and then process the superior records. Although such a programming scheme might be implemented for resource survey data, the programming effort would rapidly escalate if more data types were to be added to the data hierarchy.

As an alternative to such a physical hierarchy, the survey data are physically organized into separate files:

1. Master index file,.
2. Haul file,
3. Position file,
4. Catch file,
5. Length-frequency file, and
6. Specimen file.

Advantages of such an organization include the ability to include all of the historical specimen data: the ability to handle optional data, such as position records; simplicity--programs for a single type of data need only process that record type; and extensibility--additional data types can be, and have been, added to the structure without changing existing programs.

One problem induced, though, by the physically separate file types is that of joining related information from two separate files. That is accomplished by using the DMSIII relational utility, developed at NWAFC, which can be called as a subroutine by programs, or it can be executed as a stand-alone main program.

For example, to calculate catch per unit effort, the catch weight is needed from the catch record, and the effort (either duration or distance of net tow) is needed from the haul record. The DMSIII utility performs this associating function to produce a hybrid file containing catch weight and effort. This hybrid file then becomes the data source for the actual calculation, the dividing of catch weight by effort. This process is illustrated in Figure 2.

The relational utility allows the associating of data from two or more different files that have some data elements in common. As such, given any record in the survey data base, any corresponding data from any, even the same, file type may be produced by the relational utility.

Furthermore, the relational utility allows associating information with files outside the immediate domain of the data base, provided that the two sets of data have some data elements in common. The relational utility is the "glue" that binds the data base together, and it is the mechanism for associating internal data with external data. A more complete description of the relational utility is in Appendix H.

Figure 3 shows how data in separate files might be associated via the relational utility. The arrows indicate selecting a record and then associating the corresponding records in the file type at the end of the arrow. The data items listed on the arrows are the data items that are contained in both files.

#### Standards

An area of paramount importance in establishing a data base is that of standardization of the data. Easily overlooked, perhaps because of more immediate needs, standards are crucial to the longer-term uses of the data

## CPUE PROGRAM

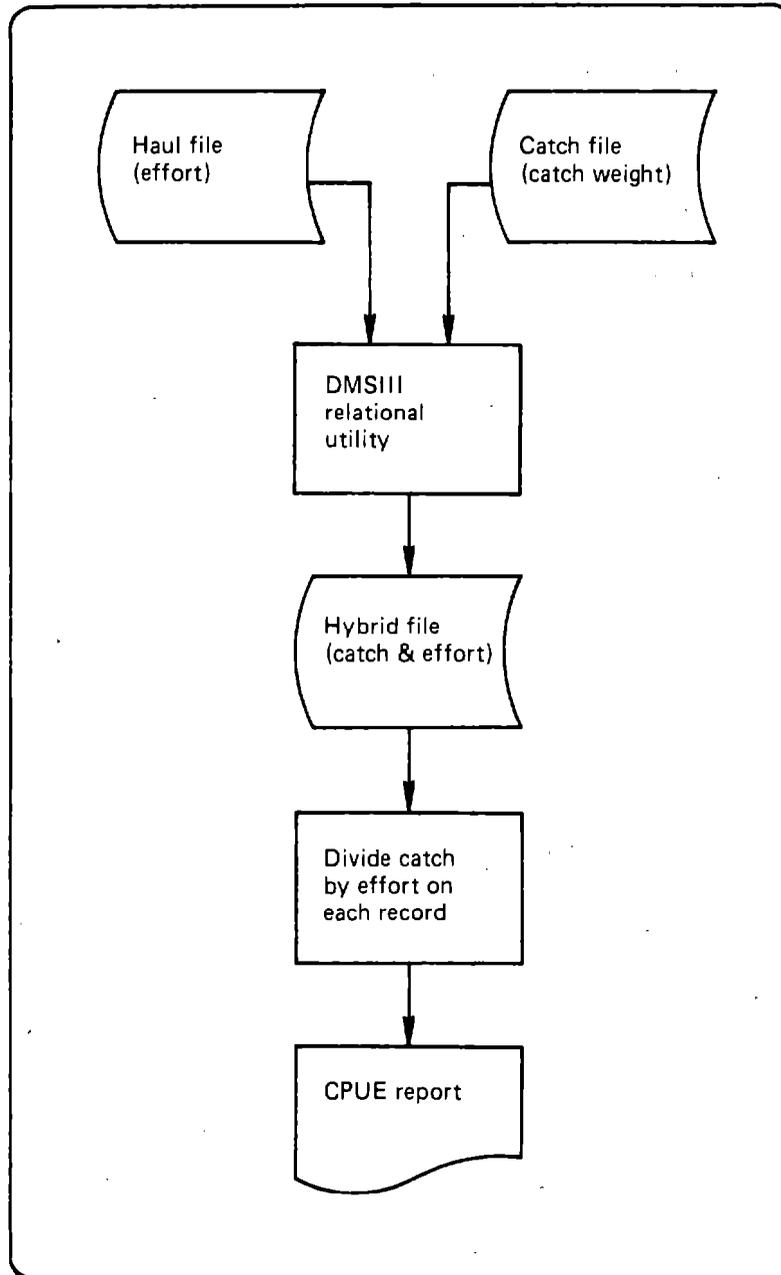


Figure 2. --Catch per unit effort (CPUE) calculation via the relational utility. Haul records and catch records have vessel, cruise, and haul numbers on both files. Hybrid records have catch weight and effort on each record.

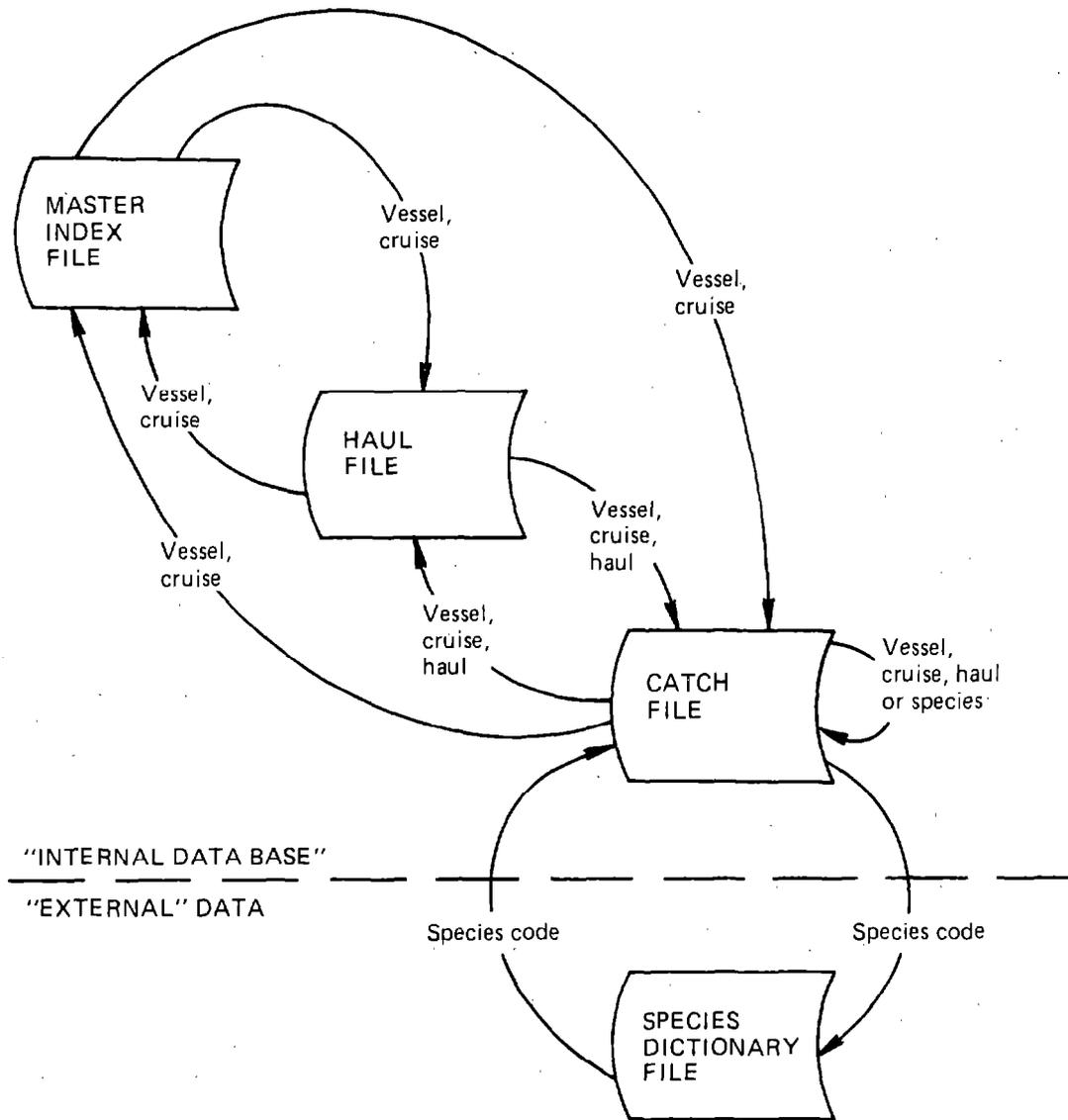


Figure 3.--Relating data in separate files. Arrows show examples of how records might be selected from one file type and then associated with corresponding records of another file type. Items listed on the arrows are contained in both files.

base. Without standards, all of the standardization issues/problems are encountered for each retrieval of data from the data base rather than just once as data are entered into the data base.

In fact, a lack of standardization may prevent the establishing of a data base, may render a data base useless, or make retrieval so time-consuming and costly (for having to repeatedly work the problems of standards) that the data base becomes ineffective.

Areas of data standardization treated in the establishing of the resource survey data base included media standards, format standards, measurement standards, taxonomic standards, and quality control standards.

Media standardization was the first area of data standards addressed in the data base. Historical data existed on decks of punched cards with embedded rubber bands, paper clips, and penciled notes. Considerable effort was devoted to interpretation of the groupings of data and the special notes. Other data was on magnetic tape from other agencies. Some data were in published and unpublished journals and reports that required transcribing into a machine-readable form.

Format standardization means that all of the records for a particular kind of data need to be in the same format. This area of standardization was especially necessary because much of the data originated in various organizations, each with a unique data format. Format standardization is important even when dealing with one organization's internal data, because formats usually change over time, and historical data are not always converted into the new format.

Measurement standards assure that all of the values for a particular data item are in the same units of measurement, e.g., that all catch weights are in pounds, to the nearest tenth of a pound in decimal places.

Taxonomic standards need to be applied -to biological identifications in the data base to ensure that the nomenclature is valid, clear, and consistent with systematic revisions.

Coding standards establish that all instances of a particular code, e.g., vessel code, have the same meaning, e.g., vessel name, and that there is one and only one code for a particular meaning, e.g., for the NOAA ship Miller Freeman that there is only one code, "21".

Different coding conventions have been used at NWAFC over the past years, and it has been a sizeable chore just to establish a unique vessel code table. The standardization of coding is not complete yet (which may indicate that it is a continuing process) since, for example, coding conventions differ in the separate functional areas.

Data quality control standards are necessary to assure that the data are correct and complete. A number of data editing programs are used to check the data, beginning with data entry of field data on programmed data loggers. Checking continues on the data throughout its progression to inclusion in the data base. Examples of editing follow.

1. Range checks on values:
  - (a) logical limits, e.g., no latitude minutes over 60 or latitude degrees over 90;
  - (b) reasonable limits, e.g., a fish length exceeding 1200 mm which in some cases is acceptable, and in other cases not;
2. Duplicate records check, e.g., two catch records for the same vessel-cruise-haul-species;
3. Missing data fields, e.g., missing species code or missing weight;
4. Missing records, e.g., no parent haul record for a subordinate catch record;

5. Relationship of data to other data and if comparisons are reasonable, e.g., average weight per fish in a-trawl sample exceeding by two standard deviations the average weight per fish for that species in other samples; and
6. Inventory of data values.

### Software Resources

#### Major Program Groups

The software, or computer programs, link the survey data base with the users of the data. This is accomplished with various categories of programs:

1. Analytical programs,
2. Map plotting programs,
3. General utility programs,
4. Data management programs,
5. Data editing and checking programs,
6. Other programs, and
7. Subroutine library programs.

A list of the main programs by functional category is in Appendix B.

The analytical programs are used for the standard processing of the resource survey data. Relatively standard analyses and products include using program CPUE to compute reports of catch per unit effort, i.e., relative density, for sets of individual samples; using program RANK, to compute values of frequency of occurrence and mean catch per unit effort, i.e., mean relative density, for all species and other catch items within a set of samples, and list them in descending sequence; using program BIOMASS and the area-swept method (Alverson and Pereyra 1969) to compute the population weight and number for a species from a set of samples taken within an area surveyed, and

to compute the size composition of the apparent population, i.e., distribution of population number among length intervals; and using program AGE/LENGTH, to determine mean lengths-at-age and the distribution of population number among age classes. Additional descriptions of the analytical programs are in Appendix I.

A map plotting program, MAP, and associated plot work file programs are used to present survey results on a geographical map with optional coast-line features. Various analytical programs "feed" a plot work file to the map program, which superimposes the results on a computer-drawn map. Results that can be plotted in this manner include simple haul data, e.g., station, numbers, sampling depths, bottom water temperatures, using program STATION; values of species catch per unit effort using program CPUE; values of median length, or sex ratio, using program LLF/40; and values of the catch per unit effort of particular size or age classes of a species using program LIMIT/CPUE. Details of the map plotting programs are included in Appendix J.

The general utility programs are the building blocks of a system, and are perhaps the first capabilities that should be developed. With the exception of three programs (SELECT/SPECIES, INVENTORY/SPECIES, and GROUP/SPECIES), the general utility programs used in the RACE software system are independent of the nature, content, or subject matter of a data base system. These utilities could be used in a business, administrative, or some other scientific system totally unrelated to fisheries resource assessment work. The utilities are not, however, machine-independent or independent of the operating system. Among these utilities are the programs SELECT, DMSIII, COMPACT, SORT, and CONVERT/METRICS. These particular utilities, and some others, exist not only as stand-alone main programs, but also as subroutines so that these capabilities may be harnessed in other application programs. For example, the BIOMASS program

has 13 calls to DMSIII, 3 calls to CONVERT/METRICS, 11 calls to COMPACT, and 9 calls to SORT.

The general utilities, particularly DMSIII and SELECT, constitute a mechanism for extending the use of the data base beyond the range of the existing software. For example, via SELECT and DMSIII, data may be extracted from the data base and routed into other analytical and statistical packages such as "SPSS," the Statistical Package for the Social Sciences (Nie et al. 1975).

The data management programs are used to maintain the data base files and to keep the data in proper order. The SORT/BINE and SORT data management programs sort the survey haul, position, catch, length-frequency, and specimen data in order, primarily by vessel, cruise, and haul numbers, and species code.

The other function of the data management programs is to support changes to the data items. Although there are some specific file update programs for changing certain data items, there are no general master file update programs for the data base. Instead, when a change needs to be made to a data item, the computer vendor's text editor package, CANDE, is used. Then the character (sometimes called "EBCIDC") files are converted to binary by the MAKE/BINE data management programs. Some checking of the data is performed by the MAKE/BINE programs. The binary files are expected by all of the analytical programs, which enforces another kind of standardization. Binary files can be transformed back into character versions for inspection and updates by the MAKE programs.

Numerous data editing and checking programs are used to detect erroneous or missing data.

Other programs are used seldom, or only by a single person, rather than by the entire organization. These include the species code dictionary programs and the coastline construction programs.

In addition to the main programs, there are about one hundred subroutines that are in a library. Some of these subroutines are versions of the general utility programs, mentioned before. Other routines are for latitude and longitude conversion, graphics, and for scanning parameters. As new application programs are written, the portions of them that can be used in other future programs are written as subroutines and included in the subroutine library.

#### Data Files Related to Software

In addition to the survey data files and the software, there are a few other data files that are part of the system. These include the species code dictionary, which is accessed by many application programs that report species names since only the species codes are carried in the catch, length-frequency, and specimen data records. Similarly, a gear code dictionary carries sampling dimensions (mean net width and height, where available) for the various gear codes. A coastline file, accessed by the MAP program, is an extract of the world Data Bank-II coastline with digitized coastline, rivers, islands, lakes, and state and national boundaries. For rapid access to the coastline data, there is a coastline index file.

A summary of the software and related files is presented in Table 2.

#### Standards

Just as data standards are essential for a data-base system, so are programming standards important for the software. Two primary objectives for software development for the RACE survey data-base system are that software is generalized and institutionalized.

Generalizing software means that programs are developed with a view toward general solutions to specific problems, the specific problems being those encountered in fisheries resource assessment work. Generalizing the software

Table 2.--Summary of software and related file sizes.

Files	Number of files	Number of characters of disk space (millions)	Notes
Source programs <sup>1/</sup>	576	8.2	135 Main programs. 98,104 lines of code. Multiple files per main program.
Object programs <sup>2/</sup>	271	7.3	1 file per main program, 1 file per subroutine library entry, and other object modules.
Coastline file	2	6.5	528,529 points. Index file is 2,167 records. Includes Pacific coast of Central and North America, equator through Alaska; and the Pacific coast of Asia, Siberia through Hong Kong.
Miscellaneous files	66	1.7	Species code dictionary, gear code dictionary, news file, and obsolete source programs.
Documentation files	135	1.3	14,700 lines of documentation.

<sup>1/</sup> Primarily FORTRAN and some ALGOL.

<sup>2/</sup> Machine language code.

implies that the same, or similar, problems in the organization can be solved with the same software.

Software that is institutionalized is made available to the organization. This means: (1) a library of object programs must be available at all times, (2) the programs must be documented so that people will know how to use the software, (3) the software must be maintainable so that errors can be corrected and that capabilities can be extended, and (4) users are informed of changes to the software.

Details of the software standards are given in Appendix K. The standards are somewhat informal and flexible in that not all of the programs meet all of the standards. Also, the programming staff is small--2 persons--and as such the standards are less rigorously policed than they might be in a larger group. Nonetheless, the standards express how programs are written for the group.

Illustrative of one of the standards is Table 3 which shows the volume of comment statements in some of the programs. The emphasis is on placing the programming documentation in the source program itself by using comment statements rather than trying to maintain separate documents such as flowcharts; noncomputer-based documents are seldom maintained. A separate user's document, computer-based, is maintained though in addition to the source program file.

Institutionalizing and generalizing the software is a marked departure from traditions of assigning programmers directly to individual investigators. It also means more costly software development, initially, than the "quick and dirty" approach. The more costly software development is recognized in the literature by Brooks (1979:4-6), who differentiates software products into four categories:

1. A program that can be run only by its author;
2. A programming product--a generalized, tested, documented, maintainable, and extendable product;

Table 3.--Comment statements in source programs.<sup>1/</sup>

Program file name	Lines	Total number of comment lines	Percentage comments <sup>2/</sup>	Number of nonblank comments	Percentage nonblank comments
BIOMASS	1,646	517	31	355	22
RANK	1,348	431	32	252	19
SUB/CROSSG	1,055	527	50	384	36
AGE/LENGTH	2,287	850	37	481	21
SUB/LORAN	1,189	503	42	329	28
SUB/SELECT	1,237	486	39	311	25
CPUE	742	239	32	115	15
DLG/01	828	341	41	197	24
MAP/SUB/CTLCRD	740	257	35	115	16
SUB/BINEX	<u>232</u>	<u>100</u>	<u>43</u>	<u>57</u>	<u>25</u>
Total	11,304	4,251	38	2,596	23

1/ This sample is not necessarily statistically random. Rather, these are a few of the larger or frequently used routines that illustrate the emphasis placed on comments in the source programs.

2/ Not all comment statements have text on them. The substantive comment statements are indicated in the right-most two columns in the table. Blank comment lines are often used for spacing, itself a valid documentation technique.

3. A programming system--a collection of interacting, related, and coordinated programs for accomplishing large tasks; and
4. A programming systems product--a programming system that is generalized, tested, documented, maintainable, and extendable.

His estimates are that a programming product is at least three times as costly as a program (that does the same function), and that a programming systems product is nine times as costly as a program with the same function.

The RACE software might best be characterized as a programming product in the terminology of Brooks. The return on the investment in generalizing and institutionalizing the software, though, is the solving of a problem once rather than repeated, diverse efforts.

**THIS PAGE INTENTIONALLY LEFT BLANK**

## SYSTEM FUNCTION

## Data Flow

Control of the survey data flow begins with the entering of the field data in programmable data loggers aboard ship. Data recorded on on-deck sampling forms and length-frequency strip charts are transcribed into the terminal via prompting and editing programs. (Crab specimen data are entered directly on the terminal via sound-powered headsets without intermediate on-deck forms.) Printed and magnetic cassette tape copies of the data are produced by the terminal.

The data loggers aboard ship perform first-level editing, such as range checks, assure necessary completeness, and assure a consistent format for the field data as well as provide a machine-readable output needed for timely subsequent survey analytical results.

At the end of each leg of a cruise, the data tapes are transmitted into the computer at NWAFC by the data manager, who also checks (via checking programs) for, and corrects, any transmission errors.

Other survey or historical data may be entered into the system by having the appropriate forms transcribed (keypunching or key-to-tape service) and loading the data onto disk files. Transformation into the standard data formats is often necessary for external or historical data, but reformatting programs and the computer system's text editor facilitate this process.

The scientific party responsible for the survey or the analysis of the data then performs the editing, checking, and correcting of the data. Numerous data editing programs are used in this phase of the data flow. The corrected data files are submitted to the data manager, who reviews and rechecks the data and adds the data to the master data base files while the scientific party continues with their analysis of the survey data.

Subsequently, the data may be retrieved from the master data base by the end user of the data, with or without the consultation or assistance of the data manager. The survey data flow is illustrated in Figure 4.

#### Data Base Management

Contrary to some uninformed belief, data base management is not automatic in this survey data-base system nor in any data-base system. Some aspects of data base management may be performed with software (as with the extensive inventory of data editing and checking programs, and general utilities such as the SELECT and DMSIII programs), but an essential ingredient for data base management is the assigning of responsibilities to a specific person for carrying out the proper data base management procedures. For the RACE survey data-base system, that person is the data manager.

Besides responsibilities for data flow, the data manager also keeps and assigns codes, e.g., vessel codes and gear codes, and disseminates system documentation. Also, the data manager provides advice and assistance to others who use the system.

#### Data Base Retrieval

A data-base system is designed for certain kinds of retrieval (or for general types or classes of data retrieval). The approach taken to a data-base system and its design determine the kinds of retrievals that are possible and also dictate the necessary methodology.

Sometimes existing applications in an organization are rewritten with a data base management system package without extending the functions of the applications. Martin (1976:322) described such a phenomenon as only proving that data-base systems are more expensive than the previous systems. Such implementations are commonly, and mistakenly, called data-base systems merely

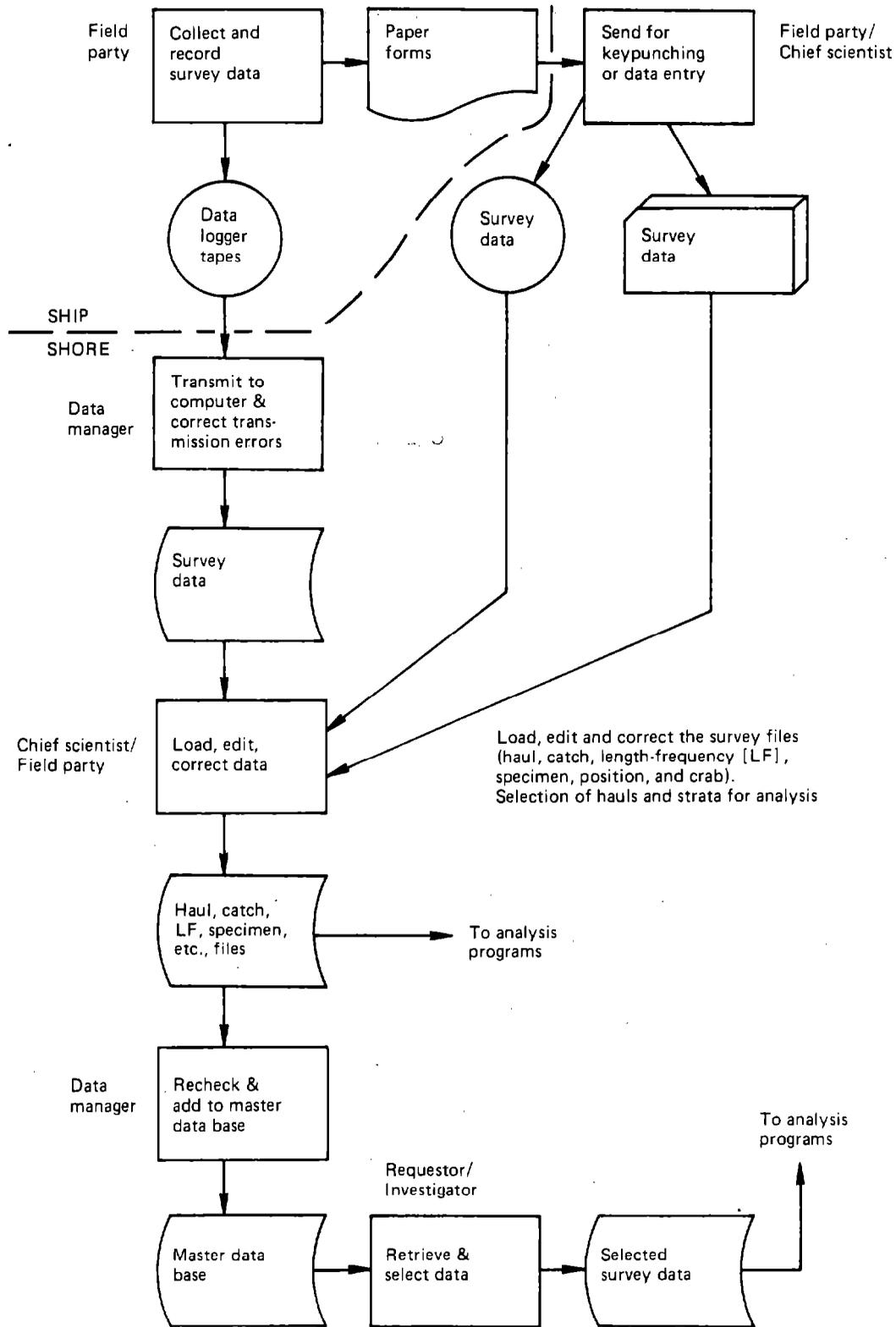


Figure 4. --Survey data flow and responsibilities.

because a vendor's data base management system package is used in implementing the new applications; they are truly not data-base systems because of the lack of integration of data across functional boundaries. Hence, data retrieval possibilities are limited to the scope of the original applications.

Other data-base systems are designed for applications that are peripheral to the mission of the organization; that is, they are not used for accomplishing the normal work of the organization. Rather, such data-base systems are supplementary to, or layered on top of, other systems, but removed nonetheless from the mission-oriented mainstream work processes.

In contrast to such data base approaches, the RACE survey data-base system is, first, a mechanism for accomplishing the normal work of the organization. Second, it is a mechanism for integrating the data resources across functional boundaries. Data retrieval techniques for each of these aspects of the data-base system are presented below;

The RACE survey data-base system is designed to support the resource assessment work on a survey by survey basis. A general strategy to follow in using the system for analyzing data for a single survey is to select the desired records (haul, position, catch, length-frequency, and specimen) onto separate files, and route only those desired records into the analytical programs. Frequently, an investigator will want to assign strata codes to haul records prior to running the analytical programs that aggregate the results by statistical subdivisions.

Many analytical programs that read a haul file and a subordinate file, e.g., catch or length-frequency, automatically select the subordinate data corresponding to the haul records. Thus, it is often unnecessary to select, for example, a subset of a catch file after selecting a subset of a haul

file. Program DMSIII can be used to select a subset of records that correspond to another subset of records, if that is desired or necessary.

In addition to the normal work of an organization, a data base is the mechanism for cross-functional uses of the data resources. An organizing principle for the RACE survey data base was the integration of the data across multiple surveys, as documented in a formal conversion plan (Intel 1980) prior to converting the system onto the new computer at NWAFC. Establishing a data base that truly integrates data across functional boundaries makes possible new uses of the data resources and can even change the nature of the work that an organization does, such as a change from single survey analysis to multisurvey analyses, and interdisciplinary approaches to, and uses of, the organization's data.

With integration of the multiple surveys and the features of the relational data base concept as implemented at NWAFC, data base retrieval can begin with any field on any type of record. Associated data can be retrieved, even if it is of a different record type, and hybrid data records can be produced for subsequent analyses.

Five examples of data base retrieval problems were prepared in conjunction with the RACE conversion plan. The new (now current) system was exercised with these five examples to verify the viability of the data-base system and to demonstrate its capabilities. These examples are presented in Figures 5-9 to illustrate only some of the retrievals that are now possible with the RACE survey data-base system.

Navigation of the data base, the moving from problem to solution via the numerous programs and other steps, is not automatic, however. A data retrieval problem usually involves many steps, and there are usually many paths to the

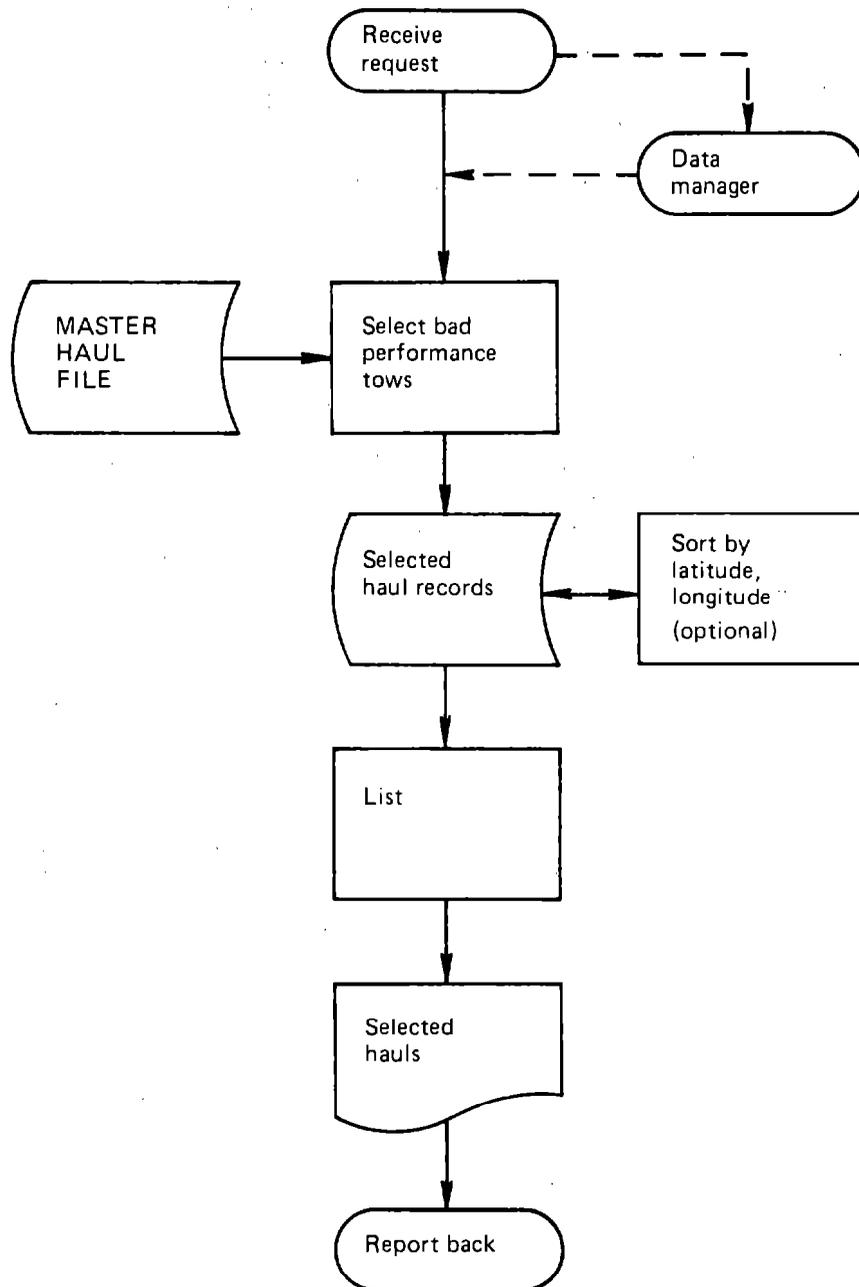


Figure 5.--Data base retrieval example 1: A commercial fisherman wants to know all places in the Gulf of Alaska where we have ripped trawl nets during fisheries surveys so that he, and others, can avoid them.

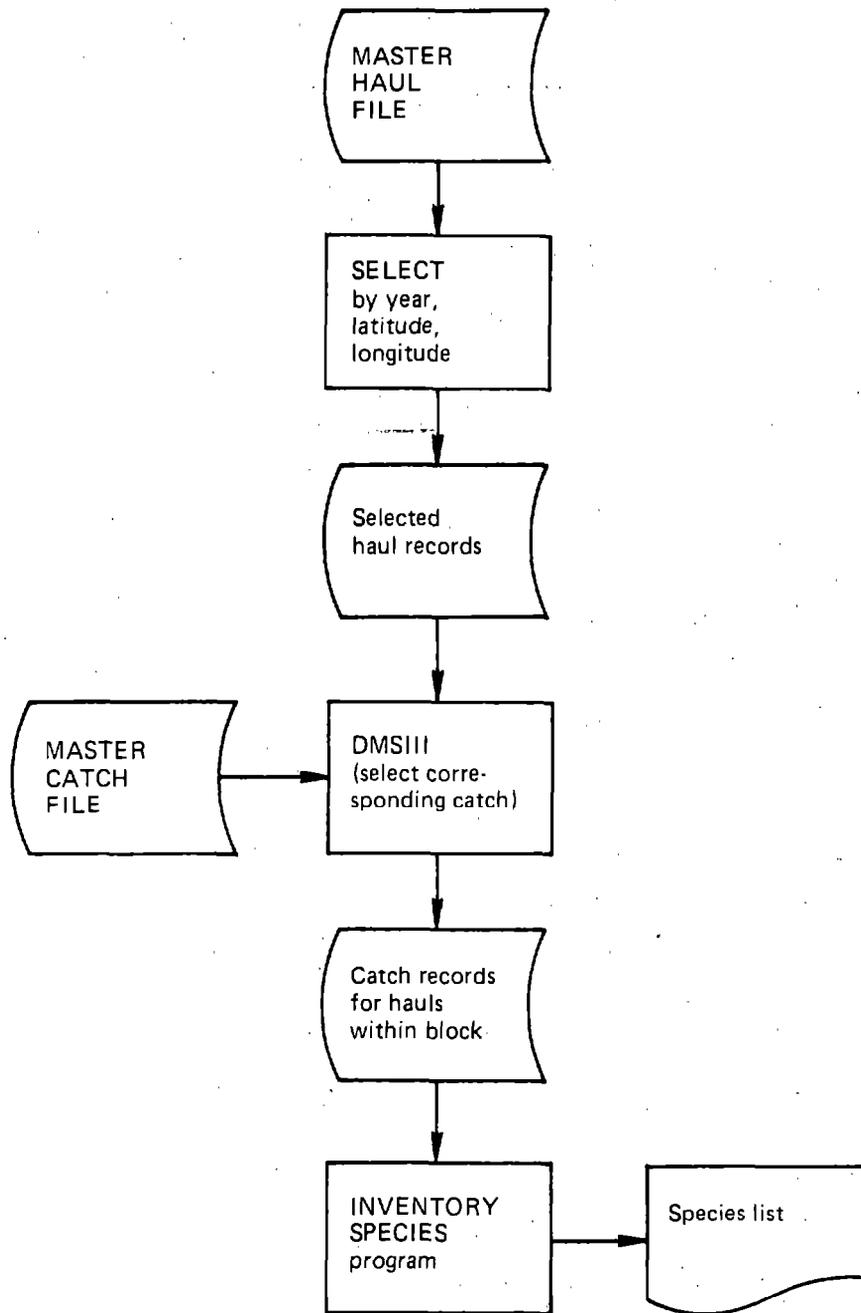


Figure 6. --Data base retrieval example 2: What fish and invertebrate species have been encountered since 1975 in a rectangular block bounded by latitudes N1 N2, and by longitudes W1, W2?

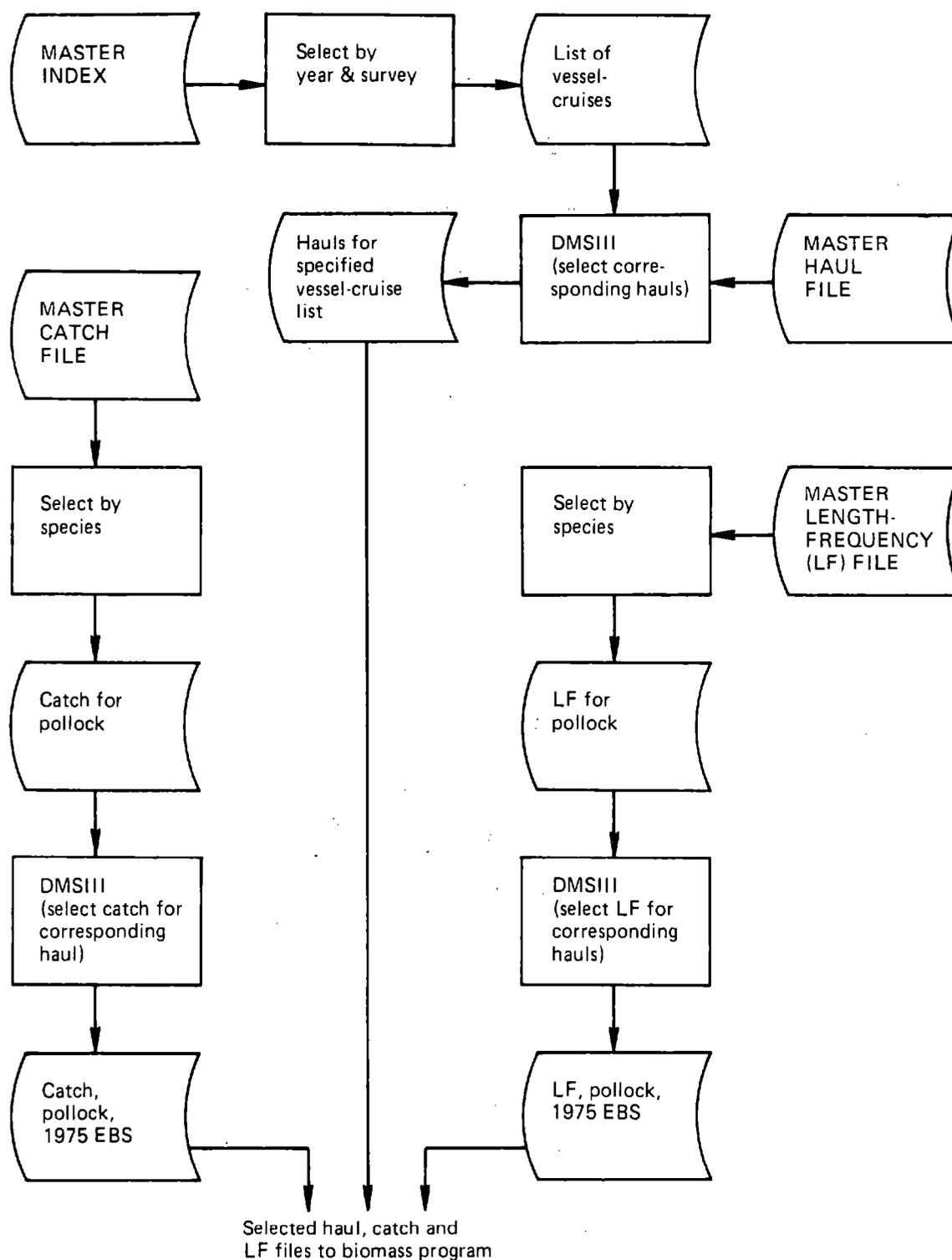


Figure 7.--Data base retrieval example 3: Repeat the estimates of walleye pollock, *Theragra chalcogramma*, population biomass and length composition for the 1975 eastern Bering Sea (EBS) survey.

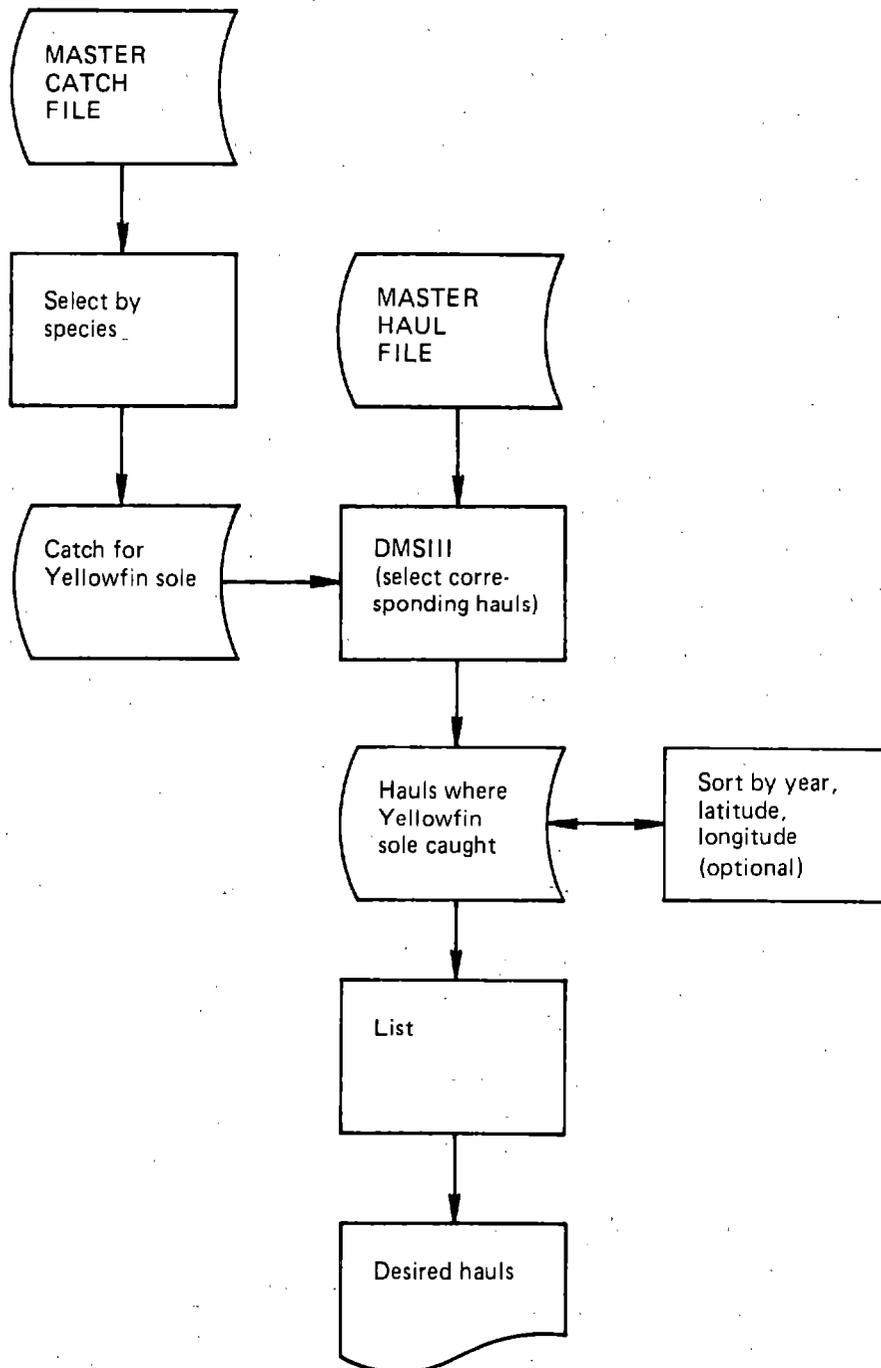


Figure 8. --Data base retrieval example 4: Where have all trawl samples been located in which yellowfin sole, Limanda aspera; have been caught in the Bering Sea?

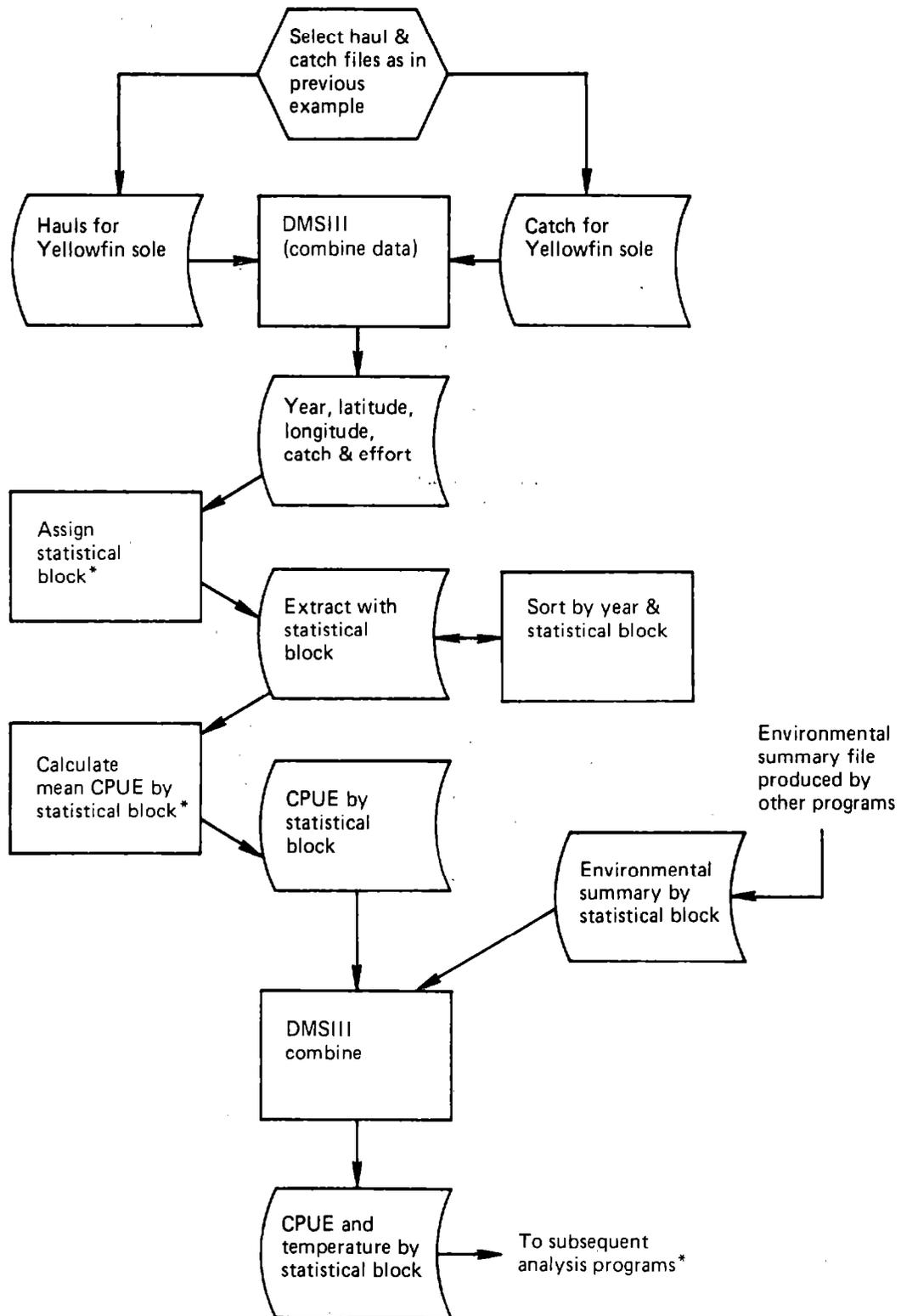


Figure 9. --Data base retrieval example 5: We want to examine the mean catch per unit effort (CPUE) of yellowfin sole in the Bering Sea versus mean sea temperature (or some other environmental data from an external, e.g., oceanographic summary file) by year, by 1/2 degree latitude and 1/2 degree longitude statistical blocks. Asterisks indicate steps requiring new programs.

same result. Hence, consultation with the data manager is recommended for successful and efficient use of the system.

#### Security Control

Security, an important consideration in any system, was also addressed in the RACE conversion plan. The data security and access privileges as defined in that plan are shown in Table 4.

Individual computer users are responsible for the data files on their account number(s), for all transactions against their files, and for backups of their data files. The data manager is responsible for the survey data base, and is the only person who can alter or delete data from the master data base files. Other users have read-only access privileges to the master data base files.

The computer operating system has commands for granting or prohibiting access to data (and program) files. Furthermore, selective privileges can be extended to certain account numbers while allowing other, or no, privileges to all others. Generally, though, file security is either "private" (only the owning account number may modify the file), or "public in" (meaning that any user on the computer may read, but not alter, the file).

Additional support is provided by the computer center staff who create daily backups of disk files that are accessed each day. Usually these daily backups are available for a maximum of two weeks, after which they are purged.

Table 4.--Summary of data security and access privileges for the Resource Assessment and Conservation Engineering Division (RACE) survey data-base system (from Mintel 1980).

File type	Who creates it	Who can change it	Who can read it	Who is responsible for backups
Source programs	RACE programmers	RACE programmers	RACE programmers	Manager of programming
Bound application programs	RACE programmers	RACE programmers	Any users at Center	Manager of programming
Individual survey files	Survey staff and RACE data manager	Survey staff and RACE data manager	Any RACE user	Survey staff
Multiple survey data bases	RACE data manager	RACE data manager only	Any users at Center	RACE data manager
System-wide files (species dictionary, ADP codebook, documentation)	RACE data manager	RACE data manager	Any users at Center	RACE data manager

## SYSTEM ADMINISTRATION AND PERSONNEL

A number of personnel roles were crucial to the development of the RACE survey data-base system. Such roles might be filled similarly in the development of other data-base systems of a similar magnitude.

First, and perhaps paramount, management must recognize and desire the capabilities that a data-base system can provide for an organization. This recognition may often be expressed in terms of personnel resource allocations in direct support of a data-base system.

Next, an overall administrative role is necessary for defining the various other technical roles and responsibilities for a data-base system. Then, the planning and design of a data-base system need to be carried out and documented. The overall administrative role, planning, and design at NWAFC were performed by one person who also served as one of the two programmers for the project.

The programming roles also involved documentation, some training of others, and consultation with users.

A data manager's role is crucial, too. First, there are enormous conversion responsibilities for transferring the data into the new system, standardizing the data, checking, and resolving the numerous data questions that arise. Also, the data manager trains and assists others in the organization in the use of the new system. Two persons were assigned to these conversion tasks for the RACE data-base system.

Of these two persons, one continues in the ongoing role of data manager after the initial conversion effort. Numerous continuing responsibilities include maintaining the master data base files and ADP code tables, e.g., vessel code and gear code tables, maintaining security backup copies of the data base files, quality control and checking new data added to the data

base, disseminating documentation and assistance to the users of the system, and serving as a first-level reference for questions about the system.

Recognition of these various roles is important in the development of a data-base system. A successful data-base system is truly a group effort and a group result that cannot be accomplished by any single person.

## SYSTEM ACCESS

## RACE Division Users

The survey data-base system is available for use on-line, 24 hours per day, 7 days a week. Any person with a current NWAFC computer account number may access and use the documentation library and major programs. And subject to certain authorization controls, data may be retrieved from the major area data bases or accessed in the form of some other user's copied, or unique, temporary files. New data may also be provided by the user or entered interactively with some programs, e.g., program NORM/SEP.

Computer services at the NWAFC are provided by the Office of Fisheries Information Systems (OFIS), telephone (206) 442-1541. Questions regarding computing services, telephone numbers for accessing the computer, account numbers, and any service problems should be directed to OFIS.

RACE Division users are advised to work in close cooperation with the Division's data manager, Ken Weinberg, telephone (206) 442-7796. The data manager should be able to provide the most current information regarding the status of particular data, information regarding any current problems affecting uses of the survey data-base system, information regarding any undocumented special knowledge required for use and interpretation of data, and to recommend if any potential conflicts might result from preliminary or shared uses of data.

Examples of problems where undocumented special knowledge might be required for the correct interpretation of survey data include cases where unique and unrecorded coding, e.g., strata codes in haul and specimen data, remain in the data; cases where unique and unrecorded biasing factors, e.g., artifacts of sampling design, vessel fishing powers, any variances from normal sampling gear performance and selectivity, were present in the implementation of the field survey; any unreliability in the biological identifications; and any

exceptions to normal standards of measurement. To explore or resolve questions such as these, it will probably be necessary to contact the persons most familiar with the data of interest.

An important aspect of sharing data is showing courtesy and recognition for the work of others. Most of the Division's survey data have been established in the data-base system as a result of the interest and work of several research groups and many individuals. Success in sharing data, then, requires an atmosphere of responsibility and trust: good communication of plans/intentions between individuals and groups, avoidance/resolution of conflicting goals, a spirit of working together, and full acknowledgment of contributions.

#### Other NWAFC Users

Other NWAFC personnel who would like to use the survey data-base system are free to do so, but are advised to first contact the RACE data manager for current information and assistance. NWAFC users working from locations other than Seattle may want to make arrangements with the OFIS for delivery of their computer outputs by mail, if this service has not already been requested.

#### Others

Other persons, not affiliated with the NWAFC, may submit requests to the RACE Division for information searches and summaries from the survey data-base system. For these services or further information, please contact Dr. Murray L. Hayes, Director, Resource Assessment and Conservation Engineering Division, 2725 Montlake Boulevard East, Seattle, WA 98112.

## ACKNOWLEDGMENTS

We would like to thank the many persons who have supported the development of the present NWAFC resource survey data-base system. Individuals who significantly contributed to implementation of the system are Mike McPhail, program conversion and applications programming; Ken Weinberg and Gene Feldman, data management; Karen Halliday, data-base standardization; and Bruce Gibbs, for proposing the cross-relational concept.

**THIS PAGE INTENTIONALLY LEFT BLANK**

## REFERENCES

- ALVERSON, D. L., A. T. PRUTER, and L. L. RONHOLT.  
1964. A study of demersal fishes and fisheries of the northeastern Pacific Ocean. H. R. MacMillan Lectures in Fisheries, Inst. Fish., Univ. Brit. Columbia, Vancouver, B.C., 190 p.
- ALVERSON, D. L., and W. T. PEREYRA.  
1969. Demersal fish explorations in the northeastern Pacific Ocean--an evaluation of exploratory fishing methods and analytical approaches to stock size and yield forecasts. J. Fish. Res. Board Can. 26: 1985-2001.
- BLASGEN, M. W., M. M. ASTRAHAN, D. D. CHAMBERLIN, J. N. GRAY, W. F. KING, B. G. LINDSAY, R/ A. LORIE, J. W. MEHL, T. G. PRICE, G. R. PUTZOLU, M. SCHKOLNICK, P. G. SELINGER, D. R. SLUTZ, H. R. STRONG, I. L. TRAIGER, B. W. WADE, and R. A. YOST.  
1981. System R: an architectural, overview. IBM Syst. J. 20(1): 41-62.
- BROOKS, F. P., Jr.  
1979. The mythical man-month. Addison-Wesley, Menlo Park, Calif., 195 p.
- DIXON, W. J., and M. B. BROWN.  
1979. BMDP-79: biomedical computer programs p-series. Univ. Calif. Press, Berkeley, Calif., 880 p.
- ELLSON, J. G., D. E. POWELL, and H. H. HILDEBRAND.  
1950. Exploratory fishing expeditions to the northern Bering Sea in June and July, 1949.' U. S. Fish. Wildl. Serv.; Fish. Leaflet. 369, 56 p.
- FAGER, E. W.  
1957. Determination and analysis of recurrent groups. Ecology 38: 586-595.
- MACDONALD, P. D. M.  
1980. A FORTRAN program for analyzing distribution mixtures. McMaster Univ., Hamilton, Ont., Dep. Math. Sci., Stat. Tech. Rep. BO-ST-1, 74 p.
- MARTIN, J.  
1976. Principles of data-base management. Prentice-Hall, Inc., Englewood Cliffs, N.J., 352 p.
- MINTEL, R. J.  
1980. RACE conversion plans. January 1980 version. Unpubl. manuscr., 32 p. Northwest and Alaska Fish. Cent., Natl. Mar. Fish. Serv., NOAA, 2725 Montlake Blvd. E., Seattle, Wash. 98112.
- NATIONAL OCEANOGRAPHIC DATA CENTER.  
1981. NODC taxonomic code, 3rd edition. U.S. Dep. Commer., Natl. Oceanic Atmos. Admin., Natl. Oceanogr. Data Cent., Washington, D.C., 554 p.

- NIE, N. H., C. H. HULL, J. G. JENKINS, K. STEINBRENNER, and D. H. BENT.  
1975. SPSS: statistical package for the social sciences. McGraw-Hill,  
N.Y., 675 p.
- PRUTER, A. T., and D. L. ALVERSON.  
1962. Abundance, distribution, and growth of flounders in the southeastern  
Chukchi Sea. J. Cons. Int. Explor. Mer 27(1): 81-99.
- RONHOLT, L. L.  
1963. Distribution and relative abundance of commercially important  
pandalid shrimps in the northeastern Pacific Ocean. U.S. Fish.  
Wildl. Serv., Spec. Sci. Rep. Fish. 449, 28 p.
- RYAN, T. A., Jr., B. L. JOINER, and B. F. RYAN.  
1981. Mini tab reference manual. Pa. State Univ., University Park, Stat.  
Dep., 154 p.
- SANDBERG, G.  
1981. A primer on relational data base concepts. IBM Syst. J. 20(1):  
23-40.
- STUCKI, L. G.  
1981. Software tools. In Sources and selected readings from a profes-  
sional development seminar, Seattle, Washington, 18 April 1981,  
p. 1-25. Assoc. Comput. Mach., Puget Sound Chapter, Seattle,  
Wash.



## PARAMETER DEFAULTS:

PLOT HEIGHT= 9.0 INCHES  
 PLOT WIDTH= 12.0 INCHES  
 HEMISPHERE = NORTH  
 HEMISPHERE = WEST  
 TITLE LETTER SIZE=0.10 INCHES  
 STATION LETTER SIZE= 0.10 INCHES  
 BORDER LETTER SIZE=0.10 INCHES  
 BOUNDRIES TOP,BOTTOM,LEFT,RIGHT= IF OMITTED THE PROGRAM  
 WILL SCAN THE DATA AND DETERMINE  
 THE BOUNDS,ROUNDED TO THE NEXT  
 10 MINUTES,AUTOMATICALLY

WITHOUT TICK PARAMETER NO TICKS WILL BE DRAWN  
 MAJOR TICK SIZE 0.14 INCHES  
 MINOR TICK SIZE 0.02 INCHES  
 BORDER LABELS= THE FOUR CORNERS WILL BE LABELED

ONLY COLUMNS 1-72 MAY BE USED FOR PARAMETERS.  
 PARAMETERS MUST BEGIN IN COLUMN 1.  
 ONLY ONE PARAMETER MAY APPEAR ON A LINE.  
 PARAMETERS MAY NOT CROSS LINES.

## REGISTRATION MARKS

REGISTRATION MARKS ARE DRAWN BY THE PLOTTER AND SERVE AS THE  
 KEY INDICATOR FOR A MECHANICALLY SUCCESSFUL DRAWING.  
 THE MARK IS COMPOSED OF TWO INVERTED RIGHT ANGLES. WHEN  
 THESE TWO ANGLES SHARE A COMMON VERTEX (APPEAR AS A PLUS "+" SIGN)  
 THAN THE PLOT HAS BEEN DRAWN SUCCESSFULLY--EXCEPT IF GLICHES  
 WERE TRANSMITTED. HOWEVER IF THERE ARE ERRORS IN THE USER'S  
 PARAMETER DESIGNATIONS THAN THE RIGHT ANGLES WILL NOT BE  
 JOINED TOGETHER. IN THIS CASE THE REGISTRATION MARKS WILL BE  
 SEPARATED. REGISTRATION MARKS CAN BE FOUND IN THE LOWER  
 RIGHT -HAND CORNER OF THE PLOT.

## PROGRAM LIMITATIONS

- 1)EITHER ALL FOUR CORNERS OF THE PLOT MUST BE SPECIFIED OR NONE.
- 2)MAPS FOR THE SOUTHERN HEMISPHERE ARE UNTESTED.
- 3)MAPS WHICH SPAN ACROSS 0 DEGREES LATITUDE ARE UNTESTED.
- 4)MAPS MUST NOT SPAN MORE THAN HALF THE EARTH IN WIDTH.

## PROGRAM MAP EXAMPLES

## WFL EXAMPLE 1:

```

?BEGIN JOB PLOT;
  QUEUE = 60;
  FETCH = "PLEASE REFER TO JOB INSTRUCTIONS";
  INSTRUCTION 1
    PLEASE USE SMALL (11") PAPER FOR THIS PLOT.
    START PEN ON RIGHT SIDE WITH A 1" MARGIN.
    BALL POINT PEN.
  
```

(ANY SIZE PAPER AND ANY PEN WILL BE O.K.)  
 JOHN DOE ROOM 768E 442-4545  
 THANK YOU.)  
 RUN (RACE0010)OBJECT/MAP ON RACE;  
 FILE 5=MAP/PARMS;  
 ?END JOB.

PARAMETER FILE FOR WFL EXAMPLE 1:  
 (FILE)MAP/PARMS

SMOOTH 30  
 COAST  
 SMOOTH 20  
 NATION  
 HIGH 9.0  
 BOTTOM 5000  
 TOP 6300  
 LEFT 15000  
 RIGHT 13500  
 TICKS  
 LABELS

WFL EXAMPLE 2:

?BEGIN JOB PLOT;  
 QUEUE = 60;  
 FETCH = "PLEASE REFER TO INSTRUCTION BLOCK";  
 INSTRUCTION 1  
 PLS USE LARGE PAPER FOR THIS PLOT.  
 USE BALL POINT PEN STARTING ON RIGHTSIDE  
 WITH 3 INCH MARGIN  
 IF PROBLEMS OCCUR CALL JOHN DOE  
 2-4545 THANKS;  
 RUN(RACE0010)OBJECT/MAP;FILE 1=PWOK ON NWAFC;  
 FILE 2=PTITLE ON NWAFC;FILE 5(READER);  
 DATA  
 SMOOTH 20.  
 COAST  
 SMOOTH 10  
 RIVER  
 HIGH 18.62  
 WIDE 28.35  
 TOP 6100  
 BOTTOM 5400  
 LEFT 17600  
 RIGHT 15600  
 TYPET 0.10  
 TYPES 0.05  
 TYPEB 0.10  
 FANCY  
 TICKX 13500,"W",60,10,0.10,0.05  
 TICKY 5400,"N",60,10,0.10,0.05  
 LABELX 13500,"W",300  
 LABELY 5400,"N",450  
 PLACES 1  
 ?END JOB.

THIS RUN UTILIZES FILES CREATED BY THE CPUE PROGRAM  
 CALLED PWOK AND PTITLE LOCATED ON NWAFC PACK.

THE MAP DRAWN WILL HAVE A FANCY 3-LINE BORDER WITH MAJOR LONGITUDINAL TICK MARKS PLACED EVERY DEGREE AND MEASURING 0.10 INCHES LONG. MINOR LONGITUDINAL TICK MARKS WILL BE PLACED EVERY 10 MINUTES AND MEASURE 0.05 INCHES LONG. LATITUDE TICK MARKS WILL BE SPACED AND SIZED AS THE LONGITUDINAL TICKS WERE, BEGINNING AT 5400 NORTH. LONGITUDE AND LATITUDE LABELS WILL BE DRAWN EVERY 5 AND 6.4 DEGREES RESPECTIVELY. CPU# NUMBERS WILL BE PLOTTED TO THE FIRST DECIMAL PLACE.

#### Q U E S T I O N S   &   A N S W E R S

Q: WHAT IF THERE IS NO PLOT CREATED?

A: RUN THE PROGRAM THRU CANDE USING THE NOPLOT OPTION AND HAVE THE OUTPUT DIRECTED TO YOUR TERMINAL FILE 6(REMOTE). ANALYZE YOUR LISTING FOR ERRORS.

Q: HOW DO YOU GET A PLOT WITHOUT COASTLINE BUT WITH STATIONS?

A: DO NOT INCLUDE THE "COAST" PARAMETER.

FEATURES STILL TO BE TESTED AS OF 12/22/80:  
1) AUTOMATIC SCAN OF DATA FOR MAP BOUNDS

## PROGRAM GROWFIT.

GROWFIT IS DESIGNED TO FIT EITHER THE VON BERTALANFFY FUNCTION OR THE GENERALIZED GROWTH FUNCTION TO SIZE-AGE OR WEIGHT-AGE DATA.

THIS PROGRAM IS A MODIFICATION OF A PROGRAM WRITTEN BY DAVID SOMERTON CENTER FOR QUANTITATIVE SCIENCE IN FISHERIES, FORESTRY AND WILDLIFE. UNIVERSITY OF WASHINGTON, FEBRUARY 1977.

## THE VON BERTALANFFY FUNCTION

$$L = LINF * (1. - EXP(-K * (T - TO)))$$

IS A THREE PARAMETER MODEL WHICH HAS AN ASYMPTOTIC UPPER BOUND AND NO FINITE INFLECTION POINT. IT IS USED TO FIT LENGTH-AGE DATA.

## THE GENERALIZED GROWTH FUNCTION

$$W = WINF * (1. - EXP(-K * (T - TO))) ** DELTA$$

IS A FOUR PARAMETER MODEL WHICH HAS AN ASYMPTOTIC UPPER AND LOWER BOUND AND AN INFLECTION POINT. IT IS USED TO FIT WEIGHT-AGE DATA OR LENGTH-AGE DATA WHICH DISPLAY AN INFLECTION.

AN OPTION IS AVAILABLE FOR CALCULATING WEIGHTING FACTORS FOR FITTING WEIGHT OR LENGTH MEASUREMENTS WHOSE VARIABILITY INCREASES WITH AGE.

## PRINTED OUTPUT...

- 1) THE PROGRAM NAME, INPUT FILE NAMES, AND PROGRAM PARAMETERS.
- 2) COUNTS OF THE NUMBER OF RECORDS READ AND NUMBER OF RECORDS SELECTED.
- 3) THE REPORT TITLE AND SPECIES CODE, NAME, AND SEX.
- 4) THE NUMBER OF DATA VALUES, THE INITIAL AND FINAL PARAMETER ESTIMATES, AND THE RESIDUAL SUM OF SQUARES.
- 5) THE PARAMETER VARIANCE-COVARIANCE MATRIX.

## LIMITATIONS

3000 OBSERVATIONS - THE PROGRAM WILL TERMINATE IF THIS LIMIT IS EXCEEDED.

## FILE DEFINITIONS

FILE 1=INPUT BINARY SPECIMEN FILE(REQUIRED)  
FILE 5=INPUT PARAMETER FILE (REQUIRED)  
FILE 6=PRINTOUT

## PARAMETER SYNTAX

< PARAMETER > ::= < SPECIES PARAMETER > (REQUIRED)  
                   < INITIAL PARAMETER > (REQUIRED)  
                   < TITLE PARAMETER > (OPTIONAL)  
                   < SEX PARAMETER > (REQUIRED)  
                   < FIT PARAMETER > (REQUIRED)  
                   < WEIGHTING FACTOR PARAMETER > (REQUIRED)  
                   < COMMENT PARAMETER > (OPTIONAL)

< SPECIES PARAMETER > ::= SPECIES < SPECIES CODE >  
                           SPECIES = < SPECIES CODE >

< INITIAL PARAMETER > ::= INITIAL < L INF >,< K >,< TO > (VON BERTALANFFY)  
                           INITIAL < L INF >,< K >,< TO >,< DELTA > (GEN.GROWTH)

< TITLE PARAMETER > ::= TITLE < REPORT TITLE >  
 TITLE = < REPORT TITLE >

< SEX PARAMETER > ::= SEX < GENDER >  
 SEX = < GENDER >

< GENDER > ::= MALE  
 FEMALE  
 UNSEXED  
 OTHER  
 ALL

< FIT PARAMETER > ::= FIT AGE-LENGTH  
 FIT AGE-WEIGHT

< WEIGHTING FACTOR PARAMETER > ::= WEIGHTED  
 NOT WEIGHTED  
 UNWEIGHTED

< COMMENT PARAMETER > ::= COMMENT < COMMENT TEXT >  
 % < COMMENT TEXT >

PARAMETER DEFAULTS:

- 1) <SEX PARAMETER>: "SEX = ALL"
- 2) <FIT PARAMETER>: "FIT AGE-LENGTH"
- 3) <WEIGHTING FACTOR PARAMETER>: "UNWEIGHTED"

NOTES...

- 1) AT MOST ONE OF EACH TYPE OF PARAMETER SHOULD BE ENTERED.  
 IF MORE THAN ONE PARAMETER OF A GIVEN TYPE IS ENTERED THEN  
 THE LAST ONE ENTERED WILL BE USED.
- 2) IF THERE ARE ANY PARAMETER SYNTAX ERRORS THEN THE PROGRAM  
 WILL TERMINATE AFTER ALL PARAMETERS HAVE BEEN READ.

PARAMETERS ARE FREE FORM, WITH THE FOLLOWING EXCEPTIONS:

- 1) ONLY COLUMNS 1-72 MAY BE USED FOR PARAMETERS.
- 2) ONLY ONE PARAMETER MAY APPEAR ON A LINE.
- 3) A PARAMETER MUST BEGIN AND END ON THE SAME LINE.

COMPUTATIONAL METHODS...

-----  
 THE GROWTH FUNCTIONS ARE FIT USING THE INSL NONLINEAR LEAST SQUARES  
 ROUTINE ZXSSQ. THIS ROUTINE STARTS WITH THE INITIAL PARAMETER  
 ESTIMATES AND ITERATIVELY IMPROVES THEM UNTIL A MINIMUM IN  
 THE RESIDUAL SUM OF SQUARES IS FOUND. IF THE INITIAL PARAMETER  
 ESTIMATES ARE QUITE CLOSE TO THE TRUE VALUES, CONVERGENCE IS DIRECT  
 AND RAPID. IF THE INITIAL ESTIMATES ARE RATHER POOR--GREATER THAN AN  
 ORDER OF MAGNITUDE AWAY FROM THE TRUE VALUES--THEN CONVERGENCE MIGHT  
 NOT BE ACHIEVED. IF THIS OCCURS THE PROGRAM WILL PRINT AN ERROR MESSAGE  
 AND THE BEST ESTIMATES FOUND TO THAT POINT. THESE BEST ESTIMATES  
 SHOULD BE USED AS INITIAL ESTIMATES IN A SECOND RUN.

WHEN THE VARIABILITY OF WEIGHT OR SIZE MEASUREMENTS INCREASES WITH AGE,  
 WEIGHTING IS NECESSARY TO PRODUCE PARAMETER ESTIMATES WITH MINIMUM VARIANCE.  
 WEIGHTING FACTORS ARE ESTIMATED BY A TWO-PART PROCESS.  
 FIRST, A GROWTH FUNCTION IS FIT TO THE DATA WITH NO WEIGHTING.

APPENDIX B

Program Documentation Index

ANALYTICAL PROGRAMS

\*\*\*\*\*

AGE/LENGTH (2/20/1981)  
 LIST/AGE/LENGTH (1/27/81)  
 AGE/LENGTH/PLOT (7/28/81)  
 FPC (5/11/81)  
 LLF/35 (4/1/81)  
 LLF/40 (12/1/80)  
 CPUE (12/1/80)  
 NORM/SEP (4/1/81)  
 PLOT/LF (2/19/81)  
 RANK (2/20/81)  
 TAXA (12/1/80)  
 CATCH/FISHING/POWER (12/1/80)  
 LIMIT/CPUE (5/11/81)  
 LENGTH/WEIGHT/PLOT (1/23/81)  
 LENGTH/WEIGHT/REGRESSION (3/19/81)  
 LENGTH/SUMMARY (1/12/81)  
 BIOMASS (2/13/80)  
 LIST/SIZECOMP (1/12/81)  
 CLUSPAK (4/16/81)  
 GROWFIT (3/30/81)  
 REGROUP (4/8/81)  
 MATURITY/LENGTH (1/27/81)  
 LIST/MATURITY/LENGTH (2/9/81)  
 STATION (3/16/81)

AGE-LENGTH KEY AND POPULATION ESTIMATES BY LENGTH AND AGE CLASS  
 PRINTS OUTPUT FROM AGE-LENGTH PROGRAM  
 VON BERTALANFFY AGE-LENGTH PLOT  
 DETERMINES IF FISHING POWER COEFFICIENTS NEED TO BE APPLIED TO CATCH  
 LENGTH-FREQUENCY SUMMARY BY SPECIES, SEX AND HAUL  
 LENGTH COMPOSITION BY HAUL  
 CATCH PER UNIT EFFORT BY HAUL AND SPECIES  
 SEPARATES NORMALLY DISTRIBUTED COMPONENTS FROM A GROUP  
 PLOTS POPULATION ESTIMATES BY LENGTH  
 RANKS SPECIES BY FREQUENCY OR CPUE IN SETS OF HAUL SAMPLES  
 COMPUTES NUMBER OF SPECIES TAKEN PER STATION  
 DIVIDES CATCH WEIGHT AND NUMBER BY FISHING POWER COEFFICIENT  
 GENERATES CPUE PLOT WORK FILE BY SEX, SIZE AND AGE RANGES  
 LENGTH/WEIGHT PLOT  
 LINEAR LEAST SQUARES REGRESSION OF WEIGHT ON LENGTH  
 GENERATES UNWEIGHTED SIZE COMPOSITION FILE FROM LENGTH FILE  
 COMPUTES POPULATION BIOMASS, NUMBER, AND LENGTH COMPOSITION  
 LISTS CONTENTS OF A BINARY SIZE-COMPOSITION FILE  
 CLUSTER ANALYSIS PACKAGE  
 FITS GROWTH CURVES TO LENGTH-AGE OR WEIGHT-AGE DATA FROM A BINARY SPECIMEN FILE  
 A SERIES OF PROGRAMS FOR ANALYSIS OF SPECIES ASSOCIATIONS  
 PRINTS MATURITY/LENGTH KEY AND ESTIMATES POPULATIONS BY LENGTH AND MAT. CLASS  
 PRINTS OUTPUT FROM MATURITY/LENGTH PROGRAM  
 SELECTS HAUL DATA FIELDS FOR PLOTTING & GENERATES PLOT WORK FILE

MAP PLOTTING PROGRAMS

\*\*\*

MAP (1/19/81)  
 LIST/PLOTWORK (4/14/81)  
 UNDER/PLOT (4/9/81)

DRAWS MERCATOR PROJECTION MAP & SUPERIMPOSES PLOT WORK FILE DATA  
 LISTS THE CONTENTS OF A BINARY PLOTWORK FILE  
 ELIMINATES OVERPLOTTING - RESOLVES MULTIPLE STATIONS IN A GRID

COASTLINE CONSTRUCTION PROGRAMS

\*\*\*\*\*

COAST/MAKE/BINE (2/19/81)  
 COAST/MAKE/INDEX (2/19/81)  
 COAST/CHECK/INDEX (2/19/81)  
 COAST/SELECT/SEGMENTS (2/19/81)  
 COAST/COMBINE/INDEX (2/19/81)  
 COAST/PRINTER/PLOT (2/19/81)

CONVERTS CHARACTER WDB-2 COASTLINE FILE INTO BINARY  
 CREATES WDB-2 BINARY INDEX FILE  
 CHECKS VALIDITY OF COMBINED COASTLINE INDEX FILES  
 WDB-2 COASTLINE FILE WITH SEGMENTS FOUND IN INDEX FILE  
 UPDATE A COMBINED COASTLINE INDEX FILE  
 PLOTS COASTLINE FEATURES

SPECIES/EDIT (12/1/80)  
SPECIES/SORT (12/1/80)

DOCUMENTATION (5/7/81)

PRINT/SCALE/LABELS (5/13/81)  
GEODETC/TD/LORAN (4/22/81)

SELECT (12/1/80)  
DMSIII (2/25/81)  
INVENTORY (2/12/81)  
CONVERT/METRICS (2/12/81)  
COMPACT (1/23/81)  
REFORMAT (12/1/80)  
LONGER (12/1/80)  
RGEN (3/1/81)  
UTAH/SORT (4/13/81)  
SORT/BINE (2/12/81)  
MAKE (6/24/81)  
MAKE/BINE (6/24/81)  
SELECT/SPECIES (6/24/81)  
INVENTORY/SPECIES (12/1/80)  
GROUP/SPECIES (12/1/80)

SPECIES CODE DICTIONARY PROGRAMS

\*\*\*\*\*

GENERATES SPECIES CODE DICTIONARY, INDEX AND BINARY FILES  
SORTS SPECIES CODE DICTIONARY

MISCELLANEOUS

\*\*\*\*\*

PRINTS A CHARACTER FILE SPACED FOR LETTER SIZE PAPER

CRUISE PREPARATION

\*\*\*\*\*

PRINTS LABELS TO PUT IN OTOLITH VIALS OR SCALE ENVELOPES  
CONVERTS LATITUDES AND LONGITUDES TO LORAN COORDINATES

GENERAL UTILITY PROGRAMS

\*\*\*\*\*

SELECTED BINARY RECORDS ARE WRITTEN TO A NEW BINARY FILE  
RELATIONAL DATA BASE MANAGEMENT UTILITY  
PRINTS VALUES & FREQUENCIES OCCURRING IN A BINARY FILE  
METRIC CONVERSIONS OR ARITHMETIC ON ANY FIELDS OF A BINARY FILE  
REDUCES A SORTED BINARY FILE ON DUPLICATE KEY FIELDS  
REFORMATS CHARACTER FILES  
ADDS WORDS ONTO AN EXISTING BINARY FILE  
REPORT GENERATOR FOR BINARY FILES. (REFM0010 PROGRAM)  
INTERACTIVE. SORT A CHARACTER FILE. (NMML1000 PROGRAM)  
SORTS A BINARY FILE  
CONVERTS A CHARACTER FILE INTO A BINARY FILE  
CONVERTS A BINARY FILE INTO A CHARACTER FILE  
SELECT DATABASE RECORDS BY SPECIES CODE.  
LISTS THE SPECIES FOUND IN A BINARY FILE  
REASSIGNS SPECIES CODES TO GROUP CODES ON CATCH, L-F, OR SPECIMEN DATA

DATA MANAGEMENT PROGRAMS

\*\*\*\* \*\*\*\*\*

LIST/HAUL (12/1/80)	PRINTS BINARY HAUL FILE
LIST/CATCH (12/1/80)	PRINTS BINARY CATCH FILE
LIST/LENGTH (12/1/80)	PRINTS BINARY LENGTH FILE
LIST/MASTER/INDEX (6/24/81)	DATABASE MASTER INDEX FILE: DATES, POSITIONS, NATION, SURVEY, RECORD COUNTS
LIST/STATION (4/22/81)	LISTS A BINARY STATION FILE WITH LORAN RATES AND READINGS
LIST/POSITION (12/1/80)	PRINTS BINARY POSITION FILE
LIST/SPECIMEN (12/1/80)	PRINTS BINARY SPECIMEN FILE
MAKE/BINE/HAUL (12/1/80)	CONVERTS EBCDIC HAUL FILE TO BINARY
MAKE/BINE/CATCH (12/1/80)	CONVERTS EBCDIC CATCH FILE TO BINARY
MAKE/BINE/SQUID (4/22/81)	MAKES A BINARY FILE FROM AN EBCDIC FILE WITH SQUID DATA
MAKE/BINE/LENGTH (12/1/80)	CONVERTS EBCDIC LENGTH FILE TO BINARY
MAKE/BINE/STATION (4/22/81)	MAKES A BINARY FILE OF STATION DATA
MAKE/BINE/POSITION (12/1/80)	CONVERTS EBCDIC POSITION FILE TO BINARY
MAKE/BINE/SPECIMEN (12/1/80)	CONVERTS EBCDIC SPECIMEN FILE TO BINARY
MAKE/HAUL (12/1/80)	CONVERTS BINARY HAUL FILE TO EBCDIC
MAKE/CATCH (12/1/80)	CONVERTS BINARY CATCH FILE TO EBCDIC
MAKE/LENGTH (12/1/80)	CONVERTS BINARY LENGTH FILE TO EBCDIC
MAKE/POSITION (12/1/80)	CONVERTS BINARY POSITION FILE TO EBCDIC
MAKE/SPECIMEN (12/1/80)	CONVERTS BINARY SPECIMEN FILE TO EBCDIC
SORT/BINE/HAUL (5/21/81)	SORTS BINARY HAUL FILE BY VESSEL, CRUISE, AND HAUL
SORT/BINE/CATCH (5/21/81)	SORTS BINARY CATCH FILE BY VESSEL, CRUISE, HAUL, AND SPECIES
SORT/BINE/LENGTH (5/21/81)	SORTS BINARY LENGTH FILE BY VESSEL, CRUISE, HAUL, SPECIES, SEX, & LENGTH
SORT/BINE/POSITION (5/21/81)	SORTS BINARY POSITION FILE BY VESSEL, CRUISE, AND HAUL
SORT/BINE/SPECIMEN (5/21/81)	SORTS BINARY SPECIMEN FILE BY VESSEL, CRUISE, HAUL, SPECIES, SEX, LENGTH
SORT/HAUL (5/21/81)	SORTS CHARACTER HAUL FILE BY VESSEL, CRUISE, AND HAUL
SORT/CATCH (5/21/81)	SORTS CHARACTER CATCH FILE BY VESSEL, CRUISE, HAUL, AND SPECIES
SORT/LENGTH (5/21/81)	SORTS CHARACTER LENGTH FILE BY VESSEL, CRUISE, HAUL, SPECIES, SEX, LENGTH
SORT/POSITION (5/21/81)	SORTS CHARACTER POSITION FILE BY VESSEL, CRUISE, AND HAUL
SORT/SPECIMEN (5/21/81)	SORTS CHARACTER SPECIMEN FILE BY VESSEL, CRUISE, HAUL, SPECIES, SEX, LENGTH
UPDATE/MASTER/INDEX/NAMES (8/25/81)	UPDATE NAMES IN DATABASE MASTER INDEX FILE
GENERATE/MASTER/INDEX (6/24/81)	GENERATE OR UPDATE DATABASE MASTER INDEX FILE

DATA CHECKING, EDITING, AND LISTING PROGRAMS

\*\*\*\* \*\*\*\*\*

DLG/01 (12/1/80)  
 DLG/09 (12/1/80)  
 DLG/10 (12/1/80)  
 DLG/90 (12/1/80)  
 DLG/91 (12/1/80)  
 DLG/CRAB/CHECK/RANGE (5/13/81)  
 DLG/CRAB/COMPARE (5/13/81)  
 DLG/HAUL (12/1/80)  
 DLG/SORT (12/1/80)  
 DLG/CATCH (12/1/80)  
 DLG/LENGTH (12/1/80)  
 DLG/POSITION (12/1/80)  
 DLH/05 (5/13/81)  
 HCA/10 (12/1/80)  
 LHC/15 (3/26/81)  
 LHC/96 (12/1/80)  
 LLF/27 (12/1/80)  
 PHC/15 (12/1/80)  
 WGT/10 (12/1/80)  
 AUDIT/LF (3/30/81)  
 AUDIT/HAUL (3/30/81)  
 AUDIT/CATCH (3/30/81)  
 AUDIT/POSITION (3/30/81)  
 AUDIT/SPECIMEN (3/30/81)  
 LORAN/AND/DISTANCE (6/24/81)  
 MERGE/ADULT/AND/JUVENILE (5/13/81)  
 CONVERT/DEPTHS (2/12/81)  
 POSITION/MOVE/HAULDF (6/24/81)  
 POSITION/MOVE/DFP2HAUL (6/24/81)  
 POSITION/PICK/DF (6/24/81)  
 POSITION/PICK/LATLONG (6/24/81)  
 DUPLICATE/LENGTH (12/1/80)

CHECK DATALOGGER OUTPUT FOR FORMAT & TRANSMISSION ERRORS  
 EDIT DATA LOGGER SPECIES NAMES  
 EDIT DATA LOGGER SPECIES CODES  
 DATALOGGER PERFORMANCE SUMMARY  
 DATALOGGER PERFORMANCE SUMMARY  
 FLAGS UNREASONABLE DATA IN CRAB DATA LOGGER RECORDS  
 COMPARES CRAB COUNTS ON DATA LOGGER FILE AGAINST BINARY CATCH  
 CREATES DATA LOGGER HAUL FILE  
 SORTS DATA LOGGER FILES  
 CREATES DATA LOGGER CATCH FILE  
 CREATES DATA LOGGER LENGTH FILE  
 CREATES DATA LOGGER POSITION FILE  
 CHECKS CRAB DATA LOGGER RECORDS FOR FORMAT ERRORS  
 PRINTS DUPLICATE BINARY CATCH RECORDS  
 HAUL & CATCH RECORD CROSS REFERENCE  
 CHECKS SPEED & DISTANCE BETWEEN HAULS  
 LENGTH & CATCH FILE CROSS REFERENCE  
 PRINTS HAUL & CATCH FILES BY HAUL  
 AVERAGE WEIGHT PER FISH  
 QUANTITATIVELY SUMMARIZES BINARY LENGTH FILE BY VESSEL AND CRUISE  
 QUANTITATIVELY SUMMARIZES BINARY HAUL FILE BY VESSEL AND CRUISE  
 QUANTITATIVELY SUMMARIZES BINARY CATCH FILE BY VESSEL AND CRUISE  
 QUANTITATIVELY SUMMARIZES BINARY POSITION FILE BY VESSEL AND CRUISE  
 QUANTITATIVELY SUMMARIZES BINARY SPECIMEN FILE BY VESSEL AND CRUISE  
 COMPUTES LATITUDES, LONGITUDES, DISTANCE TOWED FROM LORAN READINGS  
 COMBINES CATCH & L-F DATA FOR TWO SPECIES CODES INTO ONE  
 CONVERTS DEPTH METRICS ON HAUL RECORDS  
 PLACES CAPTAIN'S ESTIMATES OF DISTANCE TOWED ONTO POSITION FILE  
 STARTING POSITIONS MOVED ONTO HAUL FILE  
 INTERACTIVELY SELECTS DF FROM POSITION FILE FOR THE HAUL FILE  
 INTERACTIVELY SELECTS LAT & LONGS FROM POSITION FILE FOR HAUL FILE  
 PRINTS DUPLICATE LENGTH RECORDS

**THIS PAGE INTENTIONALLY LEFT BLANK**

## APPENDIX C

## ADP Codebook Excerpt

**TITLE** ADP CODE BOOK - AUGUST 20, 1981  
RESOURCE ASSESSMENT AND CONSERVATION ENGINEERING DIVISION

THE DATA CODES PRESENTED IN THIS MANUAL ARE A CURRENT SUMMARY OF CODES LISTED IN THE RACE DIVISION'S CENTRAL DATA PROCESSING SYSTEM CODE FILE. THIS VERSION REPLACES ALL PREVIOUS DATA CODE LISTS.

THE PURPOSE OF THESE CODES IS FOR CODING DESCRIPTIVE INFORMATION FOR USE ON THE RACE DIVISION'S HAUL-POSITION FORMS, SPECIES CATCH FORMS, LENGTH-FREQUENCY FORMS, SPECIMEN FORMS, PREDATOR-PREY FORMS, AND TAG RELEASE FORMS.

THE SOURCE OF THESE CODE LISTINGS IS NOW A SINGLE DISK FILE MAINTAINED AT THE NWAFC, SEATTLE. NEW ASSIGNMENTS OF DATA CODES MUST BE CONSIDERED PROVISIONAL UNTIL THE NEW CODES HAVE BEEN CHECKED WITH, AND INCORPORATED WITHIN, THE CENTRAL CODE FILE IN SEATTLE. SUBMIT PROPOSED CODES TO THE DATA PROCESSING MANAGER, RACE, NWAFC.

TABLE OF CONTENTS	PAGE
*****	****
HAUL-POSITION FORM . . . . .	3
SPECIES CATCH FORM . . . . .	7
LENGTH-FREQUENCY FORM . . . . .	8
SPECIMEN FORM . . . . .	9
PREDATOR-PREY FORM . . . . .	11
TAG RELEASE FORM . . . . .	13
SUPPLEMENTARY TABLES	
VESSEL CODE . . . . .	15
GEAR CODE . . . . .	17
GEAR ACCESSORIES CODE . . . . .	22
MATURITY CODE . . . . .	24
BOTTOM TYPE CODE . . . . .	26
LORAN RATES . . . . .	30
TAGGING CODE . . . . .	34

## HAUL-POSITION FORM DP-001 77-06-18

## POSITION CARD

COLUMNS	DESCRIPTION
*****	*****
1-2	VESSEL CODE SEE VESSEL CODE TABLE
3-5	CRUISE NUMBER
6-8	HAUL NUMBER
9-10	YEAR
11-12	MONTH
13-14	DAY
15-16	APPROXIMATE STARTING POSITION DEGREES LATITUDE
17-19	MINUTES LATITUDE, IN TENTHS OF A MINUTE. THERE IS AN IMPLIED DECIMAL POINT BETWEEN COL. 18 AND 19.
20-22	DEGREES LONGITUDE
23-25	MINUTES LONGITUDE, IN TENTHS OF A MINUTE. THERE IS AN IMPLIED DECIMAL POINT BETWEEN COL. 24 AND 25.
26	HEMISPHERE 'W' = WEST, 'E' = EAST.
NOTE - IF YOU HAVE ACCURATE BEGINNING AND ENDING LORAN POSITIONS, FILL IN ONLY THE APPROXIMATE STARTING POSITION (TO WITHIN 5 NAUTICAL MILES, IF POSSIBLE) AND DO NOT CODE THE ENDING APPROXIMATE POSITION.	
IF YOU HAVE DETERMINED POSITION BY MEANS OTHER THAN THE LORAN READINGS, E.G., FROM A RADAR FIX, THEN ENTER THE STARTING AND ENDING POSITION TO THE NEAREST TENTH OF A MINUTE.	
27-28	ENDING POSITION ENDING DEGREES LATITUDE
29-31	ENDING MINUTES LATITUDE, IN TENTHS OF A MINUTE. THERE IS AN IMPLIED DECIMAL POINT BETWEEN COLUMNS 30 AND 31.
32-34	ENDING POSITION DEGREES LONGITUDE
35-37	ENDING POSITION, MINUTES LONGITUDE, IN TENTHS OF A MINUTE. THERE IS AN IMPLIED DECIMAL POINT BETWEEN COL. 36 AND 37.
38	ENDING POSITION, LONGITUDE HEMISPHERE.

## SPECIES CATCH FORM OP-002 77-06-17

COLUMNS *****	CONTENTS *****
1-2	VESSEL (SEE VESSEL CODE TABLE)
3-5	CRUISE
6-8	HAUL
63-67	SPECIES CODE (SEE SPECIES CODE BOOK)
68-73	WEIGHT TO NEAREST TENTH OF A POUND. THERE IS AN IMPLIED DECIMAL POINT BETWEEN COLUMNS 72 AND 73.  IF YOU HAVE A CATCH, BUT IT IS LESS THAN ONE-TENTH OF A POUND, ENTER IT ON THE FORMS AS ONE-TENTH OF A POUND (I.E., ENTER A '1' IN COLUMN 73).
74-79	NUMBER OF INDIVIDUALS
80	'1' = RECORD CODE PRINTED ON FORM FOR SPECIES CATCH RECORDS.

## LENGTH-FREQUENCY FORM DP-003 77-06-18

COLUMNS	DESCRIPTION
*****	*****
1-2	VESSEL (SEE VESSEL CODE TABLE)
3-5	CRUISE
6-8	HAUL
9-14	TOTAL WEIGHT - FOR THIS SPECIES IN THIS HAUL - TO THE NEAREST TENTH OF A POUND. (THIS SHOULD MATCH THE WEIGHT IN COL. 68-73 ON THE SPECIES CATCH FORM.) THERE IS AN IMPLIED DECIMAL POINT BETWEEN COL. 13 AND 14.
15-19	SAMPLE WEIGHT - TO THE NEAREST TENTH OF A POUND. THERE IS AN IMPLIED DECIMAL POINT BETWEEN COL. 18 AND 19.
40-44	SPECIES CODE (SEE SPECIES CODE BOOK)
47	SEX *1*=MALE *2*=FEMALE *3*=UNDETERMINED
49-52	LENGTH IN MM.
54-56	FREQUENCY (MUST BE AT LEAST 001)
78	SUBSAMPLE TYPE *1*=RANDOMLY SELECTED. *2*=STRATIFIED BY SIZE *3*=STRATIFIED RANDOM
79	LENGTH TYPE BLANK, *0* OR *1* = FORK LENGTH MEASUREMENT FROM TIP OF SNOUT TO FORK OF TAIL. *2*=MIDEYE TO FORK OF TAIL. *3*=TIP OF SNOUT TO HYPURAL PLATE *4*=MIDEYE TO HYPURAL PLATE. *5*=TOTAL LENGTH (EXTREMITY TO EXTREMITY) *6*=SNOUT TO SECOND DORSAL (E.G., RATFISH). *7*=LENGTH OF CARAPACE FROM BACK OF RIGHT EYE SOCKET TO END OF CARAPACE. *8*=WIDTH OF CARAPACE. *9*=HEAD LENGTH (SNOUT TIP TO POSTERIOR OPERCULAR MARGIN) *11*=SNOUT TO ANUS (E.G. RATTAILS) *12*=MANTLE LENGTH (E.G. SQUID)
80	RECORD CODE *3*=LENGTH-FREQUENCY RECORD, PRINTED ON FORMS.

## APPENDIX D

## Data Record Formats Excerpt

HAUL RECORD FORMAT		DIMENSION HREC(35)	EQUIVALENCE/HAUL	
BINARY WORD	MNEMONIC	FIELD	EXTERNAL FORMAT	COLUMNS
*****	*****	*****	*****	*****
1	INACT	ACTIVE/INACTIVE FLAG. *-1=INACTIVE HAUL, BLANK=ACTIVE HAUL.	C1	1
2	IVES	VESSEL	I3	2-4
3	ICRS	CRUISE	I4	5-8
4	IHAL	HAUL	I4	9-12
5	IYEAR	YEAR	I3	13-15
6	IMON	MONTH	I2	16-17
7	IDAY	DAY	I2	18-19
8	RLAT	LATITUDE, +DDMM.MM NORTH HEMISPHERE IS BLANK OR + SIGN, SOUTH HEMISPHERE IS SIGNED -. INTERNAL FORMAT IS DEGREES AND FRACTIONS THEREOF. EXAMPLE. 56 DEGREES 45 MINUTES NORTH IS +5845.00 EXTERNALLY, AND +58.75 INTERNALLY.	F8.2	20-27
9	RLON	LONGITUDE, +DDMM.MM EXTERNAL REPRESENTATION... WEST HEMISPHERE IS SIGNED + OR BLANK. EAST HEMISPHERE IS SIGNED -. ON INTERNAL REPRESENTATION, POSITIONS ARE IN DEGREES AND FRACTIONS THEREOF, WITH POINTS IN EAST HEMISPHERE REPRESENTED BY NUMBERS GREATER THAN 180.  EXAMPLE. 176 DEGREES 15 MINUTES WEST IS REPRESENTED EXTERNALLY AS +17615.00, AND INTERNALLY AS +176.25 178 DEGREES, 30 MINUTES EAST IS REPRESENTED EXTERNALLY AS -17830.00, AND INTERNALLY AS +181.50	F9.2	28-36
10	IGRDEP	GEAR DEPTH, METERS	I4	37-40
11	IBMDEP	BOTTOM DEPTH, METERS	I4	41-44
12	IFST	FISHING START TIME (ALSO KNOWN AS EQUILIBRIUM TIME, TO THE NEAREST HOUR).	I3	45-47
13	DUR	DURATION OF HAUL IN HOURS, AND FRACTIONS THEREOF.	F5.2	48-52
14	DIST	DISTANCE FISHED, NAUTICAL MILES, MM.MM	F6.2	53-58
15	IHLTYP	HAUL TYPE	I3	59-61
16	ISTR	STRATUM,	I3	62-64
17	ITRDEP	TRACE DEPTH	I4	65-68
18	TMPSRF	SURFACE TEMPERATURE	F5.1	69-73
19	TMPCR	GEAR TEMPERATURE NOTE - MISSING TEMPERATURES ARE REPRESENTED EXTERNALLY AS EITHER BLANKS OR -9. INTERNALLY, MISSING TEMPERATURES ARE REPRESENTED AS -9.	F5.1	74-78
20	ICLOUD	CLOUD COVER NOTE - MISSING CLOUD COVER IS RECORDED EXTERNALLY AS EITHER BLANKS OR -9., AND INTERNALLY AS -9.	I3	79-81
21	IBTYPE	BOTTOM TYPE CODE	I3	82-84
22	IWIRE	WIRE CUT, METERS ALSO KNOWN AS 'SCOPE'	I4	85-88

23	IRPM	RPM	I4	89-92
24	IGEAR	GEAR CODE	I4	93-96
25	IDOOR	ACCESSORIES CODE	I3	97-99
26	IDYNAM	DYNAMOMETER CODE (-9=MISSING/BLANK FIELD)	I2	100-101
27	IPERF	PERFORMANCE CODE	I2	102-103
28	IECHO	ECHOSOUNDER CODE (-9=MISSING/BLANK FIELD) (BLANK COLUMN)	I2 1X	104-105 106
29	VAR1	VARIABLE FIELD 1	C5	107-111
30	VAR2	VARIABLE FIELD 2	C4	112-115

THE FOLLOWING FIELDS ARE CALCULATED AND SET BY THE PROGRAM THAT CONVERTS EBCDIC HAUL RECORDS TO BINARY. THESE FIELDS WILL BE VISIBLE ON OUTPUT HAUL EBCDIC HAUL RECORDS WHEN CONVERTING BINARY BACK TO EBCDIC.

31	ISBLAT	STATISTICAL BLOCK LATITUDE, FOR 1/2 (LAT) BY 1 (LONG) DEGREE STATISTICAL BLOCKS. (550 IS 55N, 553 IS 550 30+M N.) (172 IS FOR 172D XXM W.) (180+ FOR E. HEMISPHERE)	I4	116-119
32	ISBLON	STATISTICAL BLOCK LONGITUDE	I3	120-122
33	INPFC	INPFC STRATUM CODE A FUNCTION OF LAT, LONG.	I4	123-126
34	IDZONE	DEPTH ZONE 0-99 IS ZONE 100, 100-199 IS 200, ETC.	I4	127-130
35	IQTR	QUARTER (=F(MONTH))	I2	131-132

CATCH RECORD FORMAT DIMENSION CREC(6) EQUIVALENCE/CATCH

BINARY WORD	MNEMONIC	FIELD	EXTERNAL FORMAT	COLUMNS
*****	*****	*****	*****	*****
1	IVES	VESSEL	I3	1-3
2	ICRS	CRUISE	I4	4-7
3	IHAL	HAUL	I4	8-11
4	ISPEC	SPECIES CODE, RACE	I6	12-17
5	WEIGHT	WEIGHT WWW,WWW.W MAX.	F8.1	18-25
6	INUM	NUMBERS	I7	26-32

LENGTH-FREQUENCY RECORD FORMAT DIMENSION FREC(9) EQUIVALENCE/LF

BINARY WORD	MNEMONIC	FIELD	EXTERNAL FORMAT	COLUMNS
*****	*****	*****	*****	*****
1	IVES	VESSEL	I3	1-3
2	ICRS	CRUISE	I4	4-7
3	IHAL	HAUL	I4	8-11
4	ISPEC	SPECIES CODE, RACE	I6	12-17
5	ILEN	LENGTH IN MM (SHRIMP SPECIES IN .1MM INCREMENTS)	I6	18-23
6	IFREQ	FREQUENCY	I5	24-28
7	ISEX	SEX CODE (1=M, 2=F, 3=UN, 4=SHRIMP CODE)	I2	29-30
8	ISAMTP	SAMPLE TYPE CODE	I2	31-32
9	ILENTP	LENGTH TYPE CODE	I2	33-34

APPENDIX E

Data-Base Master Index

Table E-1.--West Coast master index.

VESSEL	CRUISE	DATE RANGE		LATITUDE RANGE		LONGITUDE RANGE		NATION	SURVEY NAME	HAULS	POSITN	CATCH	LENGTH	SPECMN	
		YYMMDD	YYMMDD	DDMM.MM	DDMM.MM	DDMM.MM	DDMM.MM								
4	715	71	9 5	7110 3	4415.00	4610.00	12359.00	12445.00	US	ODF&WFLATF	96	0	1,089	0	0
4	724	72	9 2	72 913	4245.00	4423.00	12409.00	12450.00	US	ODF&WFLATF	59	0	684	0	0
4	735	73	9 7	7310 4	4419.00	4611.00	12359.00	12445.00	US	ODF&WFLATF	102	0	1,441	0	0
2	742	74	4 5	74 5 9	3406.00	4125.00	11924.00	12427.00	US	PACIFHAKE	60	0	808	753	4,683
4	749	74	9 7	7410 2	4245.00	4449.00	12410.00	12453.00	US	ODF&WFLATF	93	0	1,091	0	0
2	752	75	3 29	75 5 1	3359.00	4138.00	11840.00	12429.00	US	PACIFHAKE	69	0	790	1,410	2,093
2	754	75	9 4	751024	3657.00	4955.00	12212.00	12725.00	US	PACIFHAKE	115	0	1,746	4,075	1,968
620	751	75	9 4	751012	4145.00	4931.00	12403.00	12712.00	US	PACIFHAKE	121	0	1,713	2,907	2,089
4	754	75	9 12	7510 2	4617.00	4821.00	12414.00	12517.00	US	PACIFHAKE	82	0	889	102	444
2	763	76	8 3	76 923	3618.00	5131.00	12155.00	12939.00	US	76RF PILOT	100	0	1,376	2,969	2,402
22	762	76	8 7	76 926	3616.00	5134.00	12156.00	13003.00	US	76RF PILOT	77	0	378	1,420	423
19	762	76	8 12	76 923	3615.00	5135.00	12159.00	12956.00	US	76RF PILOT	27	0	100	559	35
4	771	77	7 4	77 927	3403.00	4829.00	11918.00	12542.00	US	77 HAKE/RF	288	288	3,741	7,605	4,699
22	771	77	7 13	77 926	3427.00	4826.00	12042.00	12539.00	US	77 HAKE/RF	237	237	3,221	7,139	4,691
21	772	77	7 20	77 928	3406.00	4945.00	11940.00	12722.00	US/PL	HK/RF H/A	116	0	693	3,558	4,650
25	775	77	7 29	77 818	3950.00	4339.00	12400.00	12445.00	US	77 HAKE/RF	85	85	834	2,052	1,411
723	772	77	8 15	77 915	3904.00	4813.00	12347.00	12614.00	POL	HK/RFH/A	65	65	1,014	1,557	3,200
620	772	77	8 30	77 921	4523.00	4619.00	12408.00	12446.00	US	77 HAKE/RF	76	76	1,002	2,137	792
14	791	79	3 22	79 423	4656.00	4823.00	12448.00	12542.00	US	79POP COOP	51	51	717	1,273	1,068
6	791	79	4 18	79 427	4616.00	4819.00	12423.00	12544.00	US	79POP COOP	67	67	852	1,505	3,225
8	791	79	4 18	79 427	4428.00	4520.00	12414.00	12452.00	US	79POP COOP	61	61	672	1,041	1,091
7	791	79	4 20	79 5 2	4524.00	4620.00	12421.00	12448.00	US	79POP COOP	63	63	528	1,524	1,384
2	792	79	8 8	79 921	4327.00	4754.00	12422.00	12523.00	US	BKCODINDEX	100	100	243	2,253	3,386
21	794	79	8 29	7910 3	4751.00	4912.00	12402.00	12646.00	US	HERRING	34	34	146	290	800
5	801	80	7 12	80 927	3647.93	4933.28	12152.26	12713.56	US	80 HAKE/RF	293	293	3,072	6,407	4,171
19	801	80	7 12	80 926	3649.00	4942.90	12206.30	12727.10	US	80 HAKE/RF	318	318	3,184	4,622	5,068
21	803	80	7 14	80 9 8	3659.00	4919.10	12224.50	12707.90	US	HK/RF H/A	77	77	367	1,601	1,979
2	802	80	8 6	80 926	4322.60	4753.30	12421.90	12522.50	US	BKCODINDEX	102	102	246	2,136	1,995
36	801	80	8 13	80 821	4000.00	4000.00	12430.00	12430.00	US	RFISHCOMM.	27	27	241	326	253
35	801	80	9 9	80 913	4700.00	4700.00	12500.00	12500.00	US	RFISHCOMM.	8	8	39	152	358
1	802	80	10 22	801119	3235.90	3812.20	11924.20	12337.50	US	BKCODINDEX	93	93	224	1,195	1,533
1	812	81	4 12	81 424	4258.00	4442.10	12437.00	12456.00	US	WIDRFBHAV	26	26	272	136	0

32=INDEX RECORDS

Table E-2. --Gulf of Alaska master index.

VESSEL	CRUISE	DATE RANGE		LATITUDE		LONGITUDE		RANGE	NATION	SURVEY NAME	HAULS	POSITN	CATCH	LENGTH	SPECMN		
		YYMMDD	YYMMDD	DDMM.MM	DDMM.MM	DDDDMM.MM	DDDDMM.MM										
2	15	53	310	53	4	7	5925.00	5955.00	13937.00	14015.00	US	EXPL YAKUT	79	0	649	0	0
2	16	54	226	54	4	6	5930.00	6107.00	14540.00	14837.00	US	EXPL PR.WM	120	0	758	0	0
2	20	54	713	54	9	8	5913.00	6107.00	14452.00	14826.00	US	EXPL PR.WM	178	0	1,918	0	0
620	2	57	718	57	9	29	5346.00	5530.00	16019.00	16630.00	US	EXPL SHUM	61	0	725	0	0
620	3	57	9	6	57	930	5344.00	5548.00	15948.00	16631.00	US	EXPL SHUM	36	0	437	0	0
2	39	58	722	68	7	31	5655.00	5936.00	15021.00	15455.00	US	EXPL KODK	109	0	1,493	0	0
2	44	59	1014	59	11	11	5922.00	6102.00	14552.00	15013.00	US	EXPL KENAI	101	0	1,097	0	0
32	60	1	60	9	9	6010	5445.00	5852.00	13348.00	13808.00	US	EXPL SE	37	0	315	0	0
603	61	61	511	61	7	29	5454.00	5854.00	15124.00	15830.00	USCAN	IPHC KODK	124	0	1,580	742	0
630	61	61	511	61	7	29	5439.00	5905.00	15044.00	15850.00	USCAN	IPHC KODK	140	0	1,593	1,364	0
604	61	61	513	61	7	29	5339.00	5856.00	15112.00	16501.00	USCAN	IPHC KODK	186	0	1,768	47	0
620	61	61	6	4	61	914	5751.00	6001.00	13645.00	15032.00	US	GRFSH YAKT	207	0	1,771	0	0
630	61	61	819	61	11	27	5425.00	5812.00	15125.00	16031.00	USCAN	IPHC KODK	63	0	823	425	0
603	61	61	823	61	11	18	5606.00	5854.00	15100.00	15500.00	USCAN	IPHC KODK	64	0	866	88	0
604	61	61	828	61	11	27	5338.00	5816.00	15128.00	16503.00	USCAN	IPHC KODK	67	0	778	92	0
2	52	61	914	61	10	11	5916.00	6018.00	14445.00	14852.00	US	GRFSH PR.WM	53	0	819	430	0
620	61	61	917	61	9	25	5849.00	5912.00	15203.00	15311.00	US	EXPL COOK	25	0	277	0	0
604	62	62	2	2	62	420	5338.00	5846.00	14853.00	16502.00	USCAN	IPHC KODK	49	0	512	114	0
630	62	62	2	3	62	420	5339.00	5844.00	14630.00	16450.00	USCAN	IPHC KODK	45	0	513	336	0
603	62	62	215	62	4	20	5348.00	5845.00	14843.00	16430.00	USCAN	IPHC KODK	33	0	339	47	0
2	54	62	423	62	5	31	5720.00	6014.00	14458.00	15038.00	US	GRFSH PR.WM	82	0	947	50	0
606	62	62	622	62	8	27	5735.00	6020.00	14545.00	15028.00	USCAN	IPHC KENAI	108	0	1,148	515	0
630	62	62	622	62	8	27	5758.00	6016.00	13715.00	14619.00	USCAN	IPHC KENAI	83	0	992	87	0
603	62	62	625	62	8	27	5742.00	6003.00	14245.00	15030.00	USCAN	IPHC KENAI	67	0	649	161	0
33	62	62	711	62	8	15	5711.00	5847.00	14833.00	15203.00	US	EXPL KENAI	63	0	633	0	0
33	62	62	821	62	10	2	5857.00	6047.00	14416.00	15029.00	US	EXPL PR.WM	92	0	1,040	0	0
603	62	62	910	62	11	26	5651.00	5945.00	14015.00	15145.00	USCAN	IPHC KENAI	52	0	575	320	0
606	62	62	911	62	12	8	5740.00	6019.00	14020.00	15129.00	USCAN	IPHC KENAI	63	0	682	157	0
630	62	62	919	62	10	19	5816.00	5926.00	13717.00	14010.00	USCAN	IPHC KENAI	28	0	356	14	0
603	63	63	124	63	3	19	5551.00	6015.00	14215.00	15530.00	USCAN	IPHC KODK	29	0	236	15	0
606	63	63	124	63	3	18	5755.00	6015.00	13748.00	15337.00	USCAN	IPHC KODK	48	0	476	66	0
630	63	63	3	1	63	311	5540.00	5803.00	15144.00	15530.00	USCAN	IPHC KODK	9	0	96	26	0
34	63	63	520	63	6	11	5812.00	6001.00	13426.00	14433.00	US	EXPL FW-YK	85	0	192	0	0
33	63	63	712	63	9	10	5540.00	6016.00	14652.00	15541.00	US	EXPL PR.WM	229	0	2,848	0	0
9	63	63	713	63	8	1	5410.00	5951.00	13135.00	14830.00	CANDA	FRBC RFISH	49	49	513	281	0
9	63	63	826	63	9	15	5554.00	5819.00	14850.00	15432.00	CANDA	FRBC RFISH	25	25	315	356	0
38	64	64	616	64	9	15	5333.00	5839.00	15050.00	17008.00	US	EXPL KODK	308	0	4,287	0	0
9	64	64	8	2	64	816	5307.00	5551.00	15457.00	16706.00	CANDA	FRBC RFISH	32	32	377	132	0
9	65	65	214	65	2	28	5616.00	5930.00	14027.00	15255.00	CANDA	FRBC RFISH	21	21	244	270	0
9	65	65	828	65	9	7	5441.00	5752.00	13400.00	13653.00	CANDA	FRBC RFISH	29	29	421	258	0
9	66	66	825	66	9	14	5116.00	5649.00	12854.00	13558.00	CANDA	FRBC RFISH	39	39	584	626	0
34	67	67	711	67	8	27	5846.00	6005.00	14820.00	15035.00	US	EXPL KENAI	79	0	997	0	0
9	67	67	925	67	9	28	5556.00	5619.00	13505.00	13529.00	CANDA	FRBC RFISH	11	11	128	178	0
34	68	68	7	3	68	926	5638.00	5841.00	15118.00	15625.00	US	EXPL KODK	79	0	1,102	0	0
9	70	70	3	9	70	516	5402.00	5938.00	13328.00	14234.00	CANDA	FRBC RFISH	63	63	878	912	0
15	70	70	830	70	10	10	5820.00	6107.00	14609.00	15224.00	US	SHRMP KODK	107	0	1,078	0	0
14	71	71	417	71	5	24	5634.00	5718.00	15130.00	15329.00	US	SHRMP KODK	61	0	669	0	0
4	71	71	6	9	71	720	5629.00	5739.00	15152.00	15349.00	US	SHRMP KODK	184	0	1,114	0	0
2	72	72	510	72	5	31	5628.00	5710.00	15242.00	15350.00	US	SHRMP KODK	60	0	827	0	0
2	72	72	722	72	8	17	5627.00	5757.00	15123.00	15500.00	US	GRFSH	62	0	630	1,273	199
14	72	72	825	72	9	28	5406.00	5552.00	15855.00	16239.00	US	SHRMP SHUM	101	0	1,583	0	0
2	73	73	510	73	6	5	5648.00	5846.00	15127.00	15521.00	US	GRFSH	45	0	622	2,529	0

Table E-2. --Gulf of Alaska master index (cont'd).

VESSEL	CRUISE	DATE RANGE		LATITUDE RANGE		LONGITUDE RANGE		NATION	SURVEY NAME	HAULS	POSITN	CATCH	LENGTH	SPECMN			
		YYMMDD	YYMMDD	DDMM.MM	DDMM.MM	DDDMM.MM	DDDMM.MM										
4	732	73	523	73	614	5634.00	5700.00	15258.00	15348.00	US	SHRMP KODK	52	0	614	0	0	
2	734	73	826	7310	7	5614.00	5849.00	15002.00	15642.00	US	GRFSH	82	0	1,075	5,158	270	
14	735	73	829	731023		5423.00	5624.00	15712.00	16241.00	US	SHRMP SHUM	142	0	2,442	0	0	
14	741	74	415	74	522	5633.00	5705.00	15301.00	15347.00	US	SHRMP KODK	40	0	486	0	0	
2	744	74	717	74	8	5302.00	5430.00	16214.00	16752.00	US	GRFSH	60	0	722	2,671	0	
14	743	74	9	1	741027	5422.00	5634.00	15737.00	16306.00	US	SHRMP SHUM	177	0	2,590	0	0	
14	751	75	4	1	75	522	5748.00	5839.00	15002.00	15233.00	US	SHRMP KODK	58	0	868	0	0
16	751	75	5	3	75	8	5910.00	6017.00	14010.00	14748.00	US	GRFSH YAKT	148	0	3,763	15,828	9,110
2	753	75	626	75	8	5510.00	5958.00	14752.00	15717.00	US	GRFSH	98	0	1,282	5,570	0	
14	753	75	9	7	751031	5344.00	5634.00	15739.00	16631.00	US	SHRMP SHUM	167	0	2,845	0	0	
2	762	76	410	76	524	5444.00	5926.00	13057.00	13957.00	US	GRFSH INSD	87	0	1,470	4,993	1,232	
14	763	76	9	8	761028	5427.00	5634.00	15738.00	16317.00	US	SHRMP SHUM	156	0	3,031	0	0	
21	771	77	126	77	3	5528.00	5828.00	14902.00	15631.00	US	GRFSH KODK	156	156	1,902	10,758	3,161	
723	772	77	7	4	77	8	5438.00	5957.00	13350.00	15051.00	POL	POLLKRF	71	71	591	2,448	2,973
2	773	77	727	77	8	5453.00	5739.00	13355.00	13628.00	US	GRFSH	27	0	284	633	471	
14	774	77	821	77	923	5427.00	5634.00	15735.00	16312.00	US	SHRMP SHUM	146	0	2,581	0	0	
21	781	78	321	78	324	5737.00	5813.00	15001.00	15142.00	US	GRFSH MARMT	28	28	216	1,623	0	
21	783	78	327	78	411	5603.99	5806.05	14938.67	15609.60	US	GRFSH KODK	55	55	796	4,294	10,386	
14	781	78	413	78	5	5940.00	6052.00	14551.00	14825.00	US	GRFSH PRMH	70	0	1,343	4,768	662	
2	783	78	610	78	811	5524.00	5940.00	13454.00	14252.00	US	BKCCDINDEX	80	80	270	961	1,646	
30	781	78	725	78	6	5461.00	5615.00	13059.00	13458.00	US	RFISH SE	39	39	471	302	0	
24	781	78	727	78	921	5407.00	5933.00	13150.00	15503.00	US	RF SE&KODK	99	99	1,121	3,631	2,465	
27	782	78	727	78	921	5414.00	5918.00	13106.00	15540.00	US	RF SE&KODK	105	105	1,185	3,008	2,115	
14	783	78	825	781016		5330.00	5633.00	15736.00	16712.00	US	SHRMP	171	0	3,620	0	0	
21	782	78	910	78	917	5639.00	5705.00	15232.00	15318.00	US	POLLK H/AC	2	0	8	0	402	
21	785	78	925	781116		5537.00	6011.00	14440.00	15544.00	US	GRFSH KODK	63	63	930	1,826	2,238	
21	791	79	120	79	413	5534.00	5807.00	13428.00	15529.00	US	GRFSH&RF SH	103	0	1,518	7,547	7,004	
24	791	79	526	79	830	5341.00	5954.00	13455.00	16509.00	US	GF&C.OH.RF	215	215	2,772	8,194	7,976	
2	792	79	6	3	79	715	5433.00	5751.00	13249.00	13703.00	US	BKCCDINDEX	97	97	291	1,775	1,673
21	801	80	220	80	413	5510.54	5810.94	15226.67	15640.81	US	GRFSH KODK	203	204	2,718	13,922	11,583	
2	802	80	6	6	80	710	5432.50	5751.20	13246.70	13703.50	US	BKCCDINDEX	94	94	274	1,890	1,736
3	801	80	616	80	8	5353.94	6018.18	13547.55	16430.06	US	IPHC JVHAL	208	208	3,026	11,141	1,774	
44	801	80	7	3	80	815	5114.20	5236.20	17253.40	17323.00	US JPN	GRFSH LEUT	129	129	1,599	2,810	2,522
552	801	80	7	6	8011	4	5119.22	5624.30	16506.82	17043.91	US JPN	GRFSH LEUT	217	217	3,891	34,233	1,151
31	802	80	730	80	820	5218.16	5452.35	16513.68	17053.72	US	GRFSH	89	89	816	5,064	550	
14	803	80	812	80	916	5325.58	5535.20	15849.40	16731.60	US	SHRMP	92	92	1,837	5,070	553	
42	801	80	823	80	916	5502.53	5742.69	15531.39	16209.98	US	ADFG SHRMP	131	131	2,111	2,643	110	
43	801	80	823	80	9	5643.80	5812.18	15209.70	15455.00	US	ADFG SHRMP	76	76	866	1,495	109	
41	801	80	9	9	80	923	5723.62	5945.03	13951.54	15356.20	US	ADFG SHRMP	35	35	561	1,169	45
21	804	8011	9	801120		5440.72	5817.00	13329.53	13544.30	US	RFISH SE	42	42	506	1,421	58	
21	812	81	3	2	81	4	5535.20	5807.00	15404.20	15612.20	US	POLLK H/AC	54	54	288	2,545	0
19	811	81	413	81	5	5348.80	5740.50	15416.90	16357.80	US	WGULFBKCD	120	120	1,410	6,208	732	
46	811	81	519	81	528	5641.80	5804.10	13426.60	14454.10	US	BKCCDINDEX	32	32	91	686	482	
3	811	81	611	81	628	5353.94	5723.00	15220.50	16430.94	USCAN	IPHC JVHAL	91	91	1,045	4,735	0	

Table E-3. --Bering Sea water index.

VESSEL	CRUISE	DATE RANGE				LATITUDE RANGE				LONGITUDE RANGE				NATION	SURVEY NAME	HAULS	POSITN	CATCH	LENGTH	SPECMN
		YYMMDD	YYMMDD	DDMM.HH	DDMM.HH	DDMM.HH	DDMM.HH	DDMM.HH	DDMM.HH											
2	43	59	8	6	59	830	6520.00	6917.00	16348.00	16853.00	US	CHUKCHI/NS	59	0	412	0	0	0		
630	631	63	519	63	812	5430.00	5830.00	15900.00	16845.00	US/CN	IPHC HALIB	99	0	1,107	0	0	0			
620	651	65	612	65	827	5445.00	5915.00	15830.00	16915.00	US/CN	IPHC HALIB	204	0	1,914	0	0	0			
501	661	66	518	66	716	5119.00	5852.00	16017.00	17926.00	JAPAN	JFA SURVEY	134	0	1,366	2,775	0	0			
630	661	66	6	2	66	819	5445.00	6000.00	15800.00	17200.00	US/CN	IPHC HALIB	105	0	865	0	0	0		
631	671	67	531	67	729	5445.00	6230.00	15845.00	17000.00	US/CN	IPHC HALIB	147	0	1,117	0	0	0			
502	671	67	7	2	67	9	5114.00	6102.00	16013.00	16445.00	JAPAN	JFA SURVEY	106	0	1,023	3,059	0	0		
631	681	68	6	1	68	710	5441.00	6100.00	15915.00	17400.00	US/CN	IPHC HALIB	83	0	683	0	0	0		
503	681	68	621	68	720	5508.00	6258.00	16015.00	17215.00	JAPAN	JFA SURVEY	148	0	1,335	4,980	0	0			
632	691	69	6	1	69	625	5435.00	5845.00	15845.00	16515.00	US/CN	IPHC HALIB	62	0	532	0	0	0		
504	691	69	630	69	9	6	5115.00	6157.00	15946.00	17242.00	JAPAN	JFA SURVEY	224	0	2,524	8,162	0	0		
626	701	70	531	70	629	5444.00	5845.00	15915.00	16515.00	US/CN	IPHC HALIB	104	0	860	0	0	0			
505	701	70	7	2	70	8	2	5509.00	6123.00	16015.00	17214.00	JAPAN	JFA SURVEY	124	0	1,381	7,042	0	0	
506	711	71	530	71	630	5437.00	5952.00	16014.00	17545.00	JAPAN	JFA SURVEY	152	0	1,717	9,599	0	0			
626	711	71	6	6	71	8	9	5351.00	5904.00	15915.00	17400.00	US/CN	IPHC HALIB	147	0	1,327	0	0	0	
14	714	71	723	71	831	5440.00	5841.00	16010.00	17015.00	US	CRAB/GRFSH	52	0	1,162	2,041	1,458	0			
14	722	72	528	72	724	5438.00	5240.00	15936.00	16852.00	US	CRAB/GRFSH	103	0	2,203	7,500	1,466	0			
627	721	72	6	6	72	627	5322.00	5800.00	16000.00	16750.00	US/CN	IPHC HALIB	69	0	496	0	0	0		
507	731	73	5	2	73	624	5445.00	6510.00	15830.00	17930.00	JAPAN	JFA SURVEY	117	0	1,306	6,335	0	0		
620	731	73	618	73	629	5440.00	5800.00	15919.00	16500.00	US/CN	IPHC HALIB	38	0	307	0	0	0			
16	733	73	626	73	718	5440.00	5740.00	16435.00	17132.00	US	CRAB/GRFSH	63	0	1,349	3,259	1,010	0			
14	734	73	7	9	73	812	5439.00	5800.00	15858.00	16550.00	US	CRAB/GRFSH	94	0	2,087	2,427	291	0		
508	741	74	5	5	74	611	5436.00	5907.00	16216.00	17453.00	JAPAN	JFA SURVEY	64	0	958	1,458	0	0		
620	741	74	6	2	74	610	5439.00	5802.00	15918.00	16500.00	US/CN	IPHC HALIB	44	0	360	0	0	0		
14	742	74	614	74	0	1	5438.00	5740.00	16135.00	17232.00	US	CRAB/GRFSH	101	0	2,287	7,507	1,366	0		
17	743	74	623	74	724	5437.00	5843.00	15816.00	17134.00	US	CRAB/GRFSH	97	0	2,072	4,963	702	0			
508	751	75	5	4	75	7	3	5437.00	6005.00	16148.00	17817.00	JAPAN	JFA SURVEY	123	0	1,298	2,926	0	0	
14	752	75	6	1	75	810	5440.00	5801.00	15820.00	17240.00	US	CRAB/GRFSH	155	0	3,209	10,518	1,437	0		
620	751	75	6	2	75	614	5431.00	5743.00	16033.00	16600.00	US/CN	IPHC HALIB	53	0	354	0	0	0		
17	751	75	8	7	75	930	5437.00	6141.00	15804.00	17829.00	US	EBS SURVEY	224	0	3,075	12,649	4,543	0		
19	751	75	8	7	75	930	5437.00	6153.00	15801.00	17845.00	US	EBS SURVEY	211	0	4,128	6,370	3,261	0		
21	751	75	818	75	1020	5437.00	6023.00	15937.00	17156.00	US	EBS SURVEY	219	0	6,024	16,703	4,446	0			
21	761	76	4	1	76	531	5436.00	5846.00	15952.00	17433.00	US	CRAB/GRFSH	117	0	3,118	9,337	4,661	0		
17	761	76	421	76	613	5431.00	5909.00	15806.00	17546.00	US	CRAB/GRFSH	161	0	2,647	8,746	3,086	0			
19	761	76	421	76	620	5451.00	5950.00	15909.00	17444.00	US	CRAB/GRFSH	219	0	3,831	12,698	5,815	0			
508	761	76	513	76	714	5436.00	6007.00	16146.00	17815.00	JAPAN	JFA SURVEY	104	0	984	3,515	0	0			
14	762	76	529	76	8	9	5440.00	5820.00	15821.00	17234.00	US	CRAB/GRFSH	166	0	4,123	6,466	1,587	0		
620	761	76	6	4	76	619	5410.00	5805.00	15917.00	16557.00	US/CN	IPHC HALIB	77	0	379	0	0	0		
21	762	76	9	2	76	10	9	6304.00	6818.00	16115.00	17151.00	US	CHUKCHI/NS	268	0	4,896	11,998	4,387	0	
21	763	76	9	3	76	10	9	6304.00	6811.00	16118.00	16912.00	US	CHK. GLNET	33	0	58	142	237	0	
14	770	77	525	77	6	1	5524.00	5540.00	16332.00	16354.00	US	COMP TOWS	22	0	457	962	0	0		
14	773	77	6	4	77	8	5	5440.00	5841.00	15819.00	17259.00	US	CRAB/GRFSH	173	0	3,784	8,311	2,031	0	
620	771	77	6	6	77	619	5432.00	5801.00	15831.00	16603.00	US/CN	IPHC HALIB	48	0	551	1,727	566	0		
20	771	77	710	77	8	8	5356.00	5845.00	15748.00	16638.00	US	NHFS CLAM	230	0	998	4,317	459	0		
26	771	77	8	2	77	831	5505.00	5815.00	15905.00	17138.00	US	ICHTHPLKTN	75	0	148	571	0	0		
21	780	78	211	78	3	9	5435.00	5957.00	16546.00	17712.00	US	HERRING	41	41	462	1,654	728	0		
14	782	78	520	78	7	7	5432.00	5740.00	15903.00	17051.00	US	CRAB/GRFSH	114	114	2,861	9,848	1,010	0		
620	781	78	6	6	78	817	5238.00	6018.00	13049.00	16617.00	US/CN	IPHC HALIB	236	0	1,842	0	0	0		
27	781	78	618	78	7	6	5500.00	6015.00	16304.00	17516.00	US	SHRIMP/PLK	58	58	758	2,647	189	0		
28	781	78	618	78	816	5454.00	6100.00	15857.00	17815.00	US	CRAB/GRFSH	202	202	4,897	18,166	2,941	0			
509	781	78	618	78	716	5219.00	6051.00	16737.00	17141.00	JAPAN	HYD/AC PLK	78	0	144	530	0	0			
29	781	78	7	4	78	8	8	0.00	5746.83	0.00	18147.18	US	NHFS CLAM	487	467	1,467	3,618	0	0	

Table E-3. --Bering Sea master index (cont'd).

VESSEL	CRUISE	DATE RANGE		LATITUDE RANGE		LONGITUDE RANGE		NATION	SURVEY NAME	HAULS	POSITN	CATCH	LENGTH	SPECMN
		YYMMDD	YYMMDD	DDMM.MM	DDMM.MM	DDMM.MM	DDMM.MM							
28	791 79 522 79 819	5449.00	6339.00	15759.00	17835.00	US	USJPN COOP	339	339	9,174	34,928	5,542		
14	792 79 525 79 824	5459.00	5800.00	15857.00	17225.00	US	USJPN COOP	165	165	3,009	14,019	2,963		
21	793 79 610 79 722	5220.00	6059.00	16630.00	17839.90	US	HYD/AC PLK	35	35	85	1,117	4,588		
512	791 79 610 79 728	5419.00	5954.00	16021.00	17657.00	JAPAN	USJPN COOP	229	229	4,168	36,583	2,968		
513	791 79 629 79 8 2	5700.00	6320.00	16619.00	17928.00	JAPAN	USJPN COOP	155	155	2,348	26,236	0		
12	792 79 7 7 79 822	5420.00	6128.00	16243.00	17922.00	US	USJPN COOP	178	178	3,038	10,437	1,458		
21	792 79 726 79 8 5	6310.00	6446.00	16132.00	16935.00	US	CHUKCHI/NS	118	118	3,342	3,534	940		
21	794 79 829 7910 3	4751.00	4912.00	12402.00	12646.00	US	HERRING	0	0	0	0	0		
21	800 80 125 80 213	5433.00	5940.00	16547.00	17805.00	US	HERRING	17	17	90	395	0		
31	801 80 512 80 730	5459.04	6139.79	15817.84	17843.12	US	GRFSH	259	259	5,863	22,842	2,640		
14	802 80 522 80 7 9	5440.30	5940.41	16242.27	17736.13	US	CRAB/GRFSH	127	127	2,636	9,463	2,761		
21	811 81 211 81 224	5406.98	5700.70	16410.56	17096.29	US	GRFSH	70	70	984	6,990	0		
1	813 81 522 81 8 3	5441.36	6138.00	16054.75	17904.06	US	CRAB/GRFSH	230	230	4,668	25,873	0		
37	811 81 522 81 720	5440.57	5940.06	15821.38	17036.39	US	CRAB/GRFSH	179	179	3,671	14,611	0		

66=INDEX RECORDS

## APPENDIX F

## Species Code Dictionary Excerpt

1	20810	CHILARA TAYLORI
2	20810	SPOTTED CUSK-EEL
1	20820	OTOPHIDIUM (OPHIDIUM) SCRIPPSAE
2	20820	BASKETWEAVE CUSK-EEL
2	20841	POMFRET UNIDENT
1	20842	BRAMA JAPONICA
2	20842	PACIFIC POMFRET
1	20844	TARACTES ASPER
2	20844	ROUGH POMFRET
2	20900	BROTULA UNIDENT
1	20910	DICROLENE FILAMENTOSA
1	20920	CATAETYX SP
1	20921	CATAETYX RUBRIROSTRIS
1	20930	BROSMOPHYCIS MARGINATA
2	20930	RED BROTLA
1	20952	TRACHURUS SYMMETRICUS
2	20952	JACK MACKEREL
1	20953	SERIOLA DORSALIS
2	20953	YELLOWTAIL
1	20955	CARANGIDAE
2	20990	CETOMIMID UNIDENT
1	21000	CHAULIODONTIDAE
2	21000	VIPERFISH UNIDENT
1	21010	CHAULIODUS MACDUNI
2	21010	PACIFIC VIPERFISH
1	21110	CLUPEA HARENGUS PALLASI
2	21110	PACIFIC HERRING
1	21120	ALOSA SAPIDISSIMA
2	21120	AMERICAN SHAD
1	21130	SARDINOPS SAGAX
2	21130	PACIFIC SARDINE
1	21200	MACROURIDAE
2	21200	GRENADIER UNIDENT
1	21201	NEZUMIA STELGIIDOLEPIS
2	21201	CALIFORNIA GRENADIER
1	21210	CORYPHAENOIDES SP
1	21220	CORYPHAENOIDES ACROLEPIS
2	21220	PACIFIC GRENADIER
1	21230	ALBATROSSIA (CORYPHAENOIDES) PECTORALIS
2	21230	GIANT GRENADIER
1	21232	CORYPHAENOIDES CINEREUS
1	21233	CORYPHAENOIDES SERRULA
1	21234	CORYPHAENOIDES SPINULOSUS
1	21235	CORYPHAENOIDES ARMATUS
1	21238	CORYPHAENOIDES FILIFERA
2	21238	FILAMENTED GRENADIER
1	21239	CORYPHAENOIDES LONGIFILIS
1	21300	COTTIDAE
2	21300	SCULPIN UNIDENT

**THIS PAGE INTENTIONALLY LEFT BLANK**

## APPENDIX G

## RACE News File Excerpt

```
=====
T O R A C E   U S E R S
=====
```

SEPT. 24, 1981

-----

BIOMASS PROGRAM

IN ALLOWING SYNONYMS FOR THE EFFORT PARAMETER, I GOOFED. SYNONYMS FOR 'HR' (DURATION EFFORT) ARE NOT ALLOWED. SYNONYMS FOR THE OTHER METRICS (NM, KM, KG, LB, MT) ARE ACCEPTABLE.

BIOMASS RUNS BETWEEN SEPT. 22 AND TODAY THAT SPECIFIED 'HR' WOULD HAVE FLAGGED THE 'HR' PARAMETER WITH A WARNING AND WOULD HAVE USED THE DEFAULT 'NM' EFFORT. IF YOU RESUBMIT SUCH RUNS, THE 'HR' PARAMETER SHOULD NOW BE RECOGNIZED AND ACCEPTED.

SEPT. 22, 1981

-----

BIOMASS PROGRAM

IF THE WEIGHT PARAMETER HAPPENED TO HAVE EXTRA CHARACTERS ON THE END, E.G., "LBS" OR "LB." INSTEAD OF "LB" AS SPECIFIED IN THE DOCUMENTATION, AND YOU SPECIFIED POPULATION ESTIMATE CALCULATION METHOD 2 (USING LENGTH-WEIGHT RELATIONSHIPS), THE PROGRAM WOULD NOT PROPERLY RECOGNIZE THAT YOU MEANT "LB" (AND SIMILARLY FOR "KG" AND "MT") WHEN CALCULATING POPULATION ESTIMATES.

ALTHOUGH THE BIOMASS WAS CALCULATED PROPERLY, THE POPULATION ESTIMATE CALCULATION METHOD ACTUALLY USED WOULD HAVE BEEN 3, BECAUSE THE PROGRAM DID NOT RECOGNIZE THE MISSPELLED PARAMETER WHEN TRYING TO USE THE LENGTH-WEIGHT RELATIONSHIPS.

NO INCORRECT INFORMATION WAS PRINTED BY THE BIOMASS PROGRAM, HOWEVER. IN THIS CASE, METHOD 3 WOULD HAVE BEEN USED, AND THE REPORT WOULD HAVE SHOWN METHOD 3. YOU REALLY MAY HAVE WANTED METHOD 2, THOUGH, AND DID NOT GET IT.

THEREFORE, THE PROGRAM HAS BEEN CHANGED TO ACCEPT SYNONYMS FOR THE WEIGHT AND EFFORT METRICS ON THE PARAMETER FILE.

METRIC	ACCEPTABLE SYNONYMS
*****	*****
LB	LB LB. LBS LBS.
KG	KG KG. KGS KGS.
MT	MT MT. MTS MTS.
NM	NM NM. NMS NMS.
KM	KM KM. KMS KMS.

SEPTEMBER 4, 1981

-----

PROGRAMS LLF/35 AND GENERATE/MASTER/INDEX CHANGED.

LLF/35 WAS CHANGED TODAY TO CORRECT A PROBLEM WITH AN INFINITE LOOP WHEN PROCESSING THE PARAMETERS.

GENERATE/MASTER/INDEX WAS CHANGED YESTERDAY TO CORRECT  
A PROBLEM WHEN UPDATING ONLY SOME, RATHER THAN ALL, OF THE  
POSSIBLE FILE TYPES.

AUGUST 4, 1981  
-----

TERMINAL EVALUATION REQUESTED

FORMS FOR EVALUATING THE NEW RACE REMOTE COMPUTER TERMINALS  
ARE AVAILABLE IN ROOM 168-W. ALL RACE COMPUTER USERS SHOULD  
FILL IN AN EVALUATION FORM SO THAT IT CAN BE DETERMINED  
WHICH MAKES AND MODELS OF TERMINALS WILL BE PURCHASED  
AT THE END OF OUR 6-MONTH LEASE CONTRACT.

ON OR ABOUT SEPTEMBER 1, THE TALLY PRINTER AND THE  
HP-2621-P CRT/PRINTER TERMINALS AT MONTLAKE WILL BE SWAPPED  
WITH THE TI KSR-820 AND THE ADM-31 CRT TERMINALS AT  
SAND POINT. THEREFORE, PLEASE COMPLETE YOUR EVALUATION  
FORMS BEFORE THEN FOR THESE UNITS.

IF YOU HAVE ANY QUESTIONS ABOUT TERMINAL EVALUATION,  
PLEASE CALL KEN WEINBERG.

JULY 29, 1981  
-----

```
*****
****                               ****
**** TAPE DRIVE EQUIPMENT         ****
**** MALFUNCTION                   ****
****                               ****
*****
```

TAPES THAT YOU WROTE LAST FRIDAY, JULY 24, 1981,  
MAY BE UNREADABLE DUE TO AN EQUIPMENT MALFUNCTION.

THEREFORE, IF YOU WROTE ANY TAPES TOWARD THE END OF  
LAST WEEK (JULY 23-24), PLEASE CHECK THOSE TAPES AS SOON  
AS POSSIBLE BY COPYING THE FILES FROM TAPE ONTO DISK.

IF THERE ARE ANY TAPE ERRORS, IT MAY BE NECESSARY TO  
ATTEMPT TO RECREATE PRIOR FILES.

PLEASE COMPLETE YOUR CHECKING OF YOUR TAPES BEFORE FRIDAY,  
JULY 31, BECAUSE RESTORATION OF PRIOR FILES MAY BE IMPOSSIBLE  
AFTER THEN.

(IF YOUR TAPE IS READABLE, PLEASE REMOVE ANY UNNEEDED FILES  
FROM DISK.)

IF YOU HAVE ANY QUESTIONS, PLEASE CALL RALPH MINTEL,  
442-7214 OR 442-7796.

## APPENDIX H

## DMSIII, the Relational Data Base Management Utility

Although some relational data base management system packages such as System 1022 have been available commercially for some years, other relational systems are still in a developmental stage (Blasgen et al. 1981). The newness, though, of relational data base management systems should not prevent an organization from implementing the relational concepts, as has been accomplished with DMSIII at the NWAFC.

In a relational system, the data are stored in separate files--the literature uses the term "tables," rather than files (Sandberg 1981)--one file per data type. Data from two files may be associated if each of the two files has some data items in common with the other file (the "join" function). The two other functions of a relational system are "selection" (record selection or selection of rows of data) and "projection" (a selection of columns of data).

At NWAFC the selection function is performed by a separate program, SELECT. The input to the program is a data file, and the output file is a subset of the input file. The "projection" and "join" functions of a relational system are performed by the DMSIII relational utility program.

DMSIII has two input data files, called, for reference purposes, the master and slave files. There can be up to three output logical data files from DMSIII. These logical output files are for:

- 1 . Matches--which occur when a master record matches a slave record on the data items declared to be common in both files;
2. Master faults--which occur when a master record has no corresponding slave record; and
- 3; Slave faults--which occur when a slave record has no corresponding master record (the mirror image of master faults).

The three logical output files may be on separate physical files, or they may be combined into one or two physical files. For example, the matched records could be directed to one output file, and the master and slave faults could be combined onto a second output file.

Two additional files are used by DMSIII. A parameter file specifies the details required for the run, and the printed output file, which can be suppressed, shows record counts and other control information. Figure H-1 shows the files used in DMSIII.

The parameter file specifies the details for the program. First, if DMSIII is being executed as a stand-alone program rather than as a subroutine, the physical output files are assigned for the three possible logical output files. (If DMSIII is executed as a subroutine, the file assignments are in the calling parameters.) Next, the key fields, the data items common to the two files, are defined as well as the ascending or descending order for each key field.

DMSIII by default will sort the master and slave files in the order specified by the key field definitions. Optional parameters allow bypassing both or either of the sorts if the user is certain that the input files are already in the proper order.

Parameters also define the nature of the relationship between the master and slave records. The relationship may be one-to-one, many-to-one, or one-to-many, but not many-to-many in this version of DMSIII. A one-to-one relationship means that one master record may match at most one slave record. Many-to-one means that multiple master records may match a given slave record. Similarly, a one-to-many relationship means that a given master record may match many slave records. The matching of records between master and slave files occurs on the basis of the defined key fields common to both files.

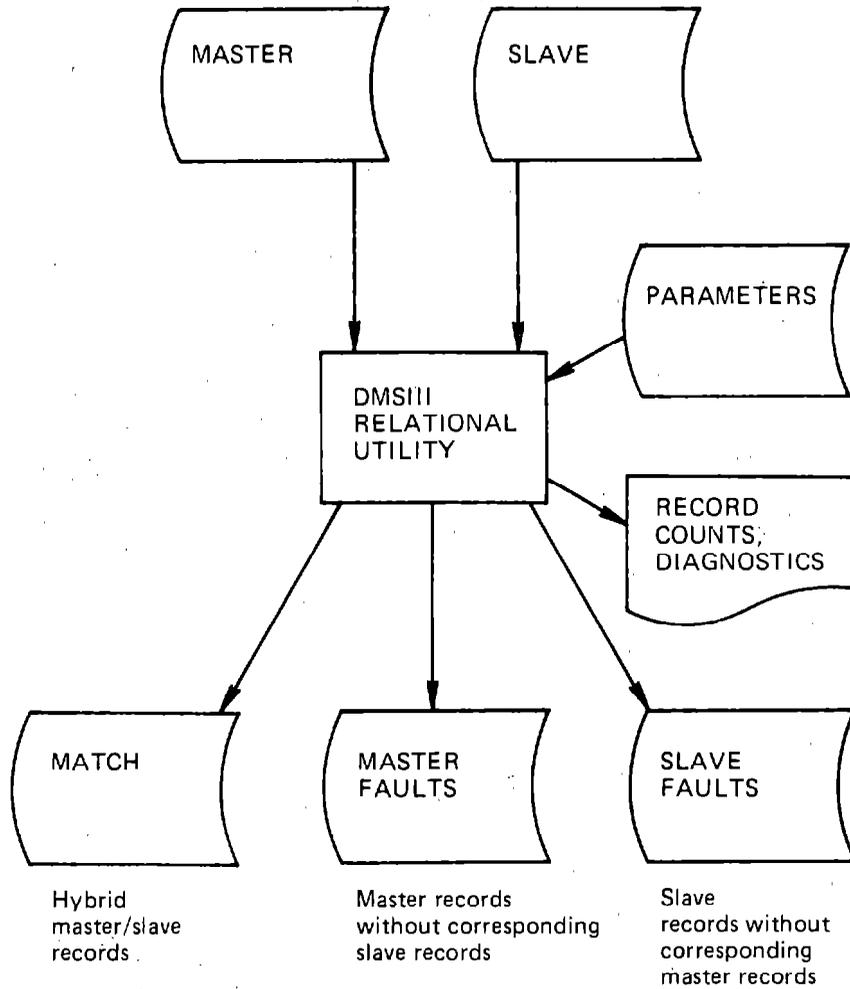


Figure H-1. --DMSIII relational utility files.

In the survey data, the haul-position record relationship is one-to-one; haul-catch is one-to-many since there may be many species caught in a single net tow; and the catch-species dictionary relationship is many-to-one over an entire survey or collection of hauls because a single species may occur in many tows.

The output record formats for the matches, master faults, and slave faults are defined with parameters. The matched output records may be any hybrid of the master and slave data items (the "join" and "projection" functions). The master and slave fault records may have data from their respective source records ("projection" function).

Also, record type codes, indicating whether the record is a match, master fault, or a slave fault, may be inserted in the output records. Zeroes may be inserted also for formatting purposes, e.g., to assure a certain size of output record.

Some uses of DMSIII are illustrated in the following application examples.

1. Match haul and catch records.

Suppose that a haul file contains vessel, cruise, haul, and effort, and that a catch file contains vessel, cruise, haul, species, and weight. A catch per unit effort report, derived from the information contained in both the haul and catch files, is desired. Although a specific program could be written to match the haul and catch records, DMSIII can be used to produce a hybrid file containing, e.g., vessel-cruise-haul-effort-species-weight. The parameter file for DMSIII would specify what to do with haul records that have no corresponding catch records, catch records that have no corresponding haul records, and whether or not multiple catch records per haul record are allowed.

2. Determine missing data.

Given a haul and catch file as before, it is desired to find catch records that do not have corresponding haul records. In this case, the haul file could be declared the master file, the catch file the slave. The matched output would be ignored. The slave fault records, i.e., catch records without haul records, could be written on an output file and routed to subsequent programs.

3. Master file update.

Sequential master file update. can be accomplished by treating the transaction file as the slave file. The transactions might have been edited prior to running DMSIII, and may be applied to the master by DMSIII, or by a subsequent program that is reading a master/transaction hybrid record produced by DMSIII.

4. Data selection from a list.

Given a catch file and a list of species codes on another file, it is desired to select only those catch records that match the list of species codes. The master file could be the catch file, and the slave file the list of species codes. The matched output file would be defined to be catch records. Master faults and slave faults would be ignored.

5. Multiple DMSIII runs for data retrieval.

Multiple invocations of DMSIII might be used to satisfy a data retrieval request. For example, we might want all of the haul records where any one of a list of species was caught. First, we would run DMSIII, as in example 4 above, to produce selected catch records. These catch records would become input to a second run of DMSIII along with a haul file. The output from

this second run of DMSIII would be just those haul records that correspond to the selected catch records.

The first use of DMSIII could have been accomplished with the "SELECT" program; the second use of DMSIII, however, was for matching records across files.

6. Increase/reduce the size of records, or rearrange words in a record.

Admittedly, this may be an unusual application of DMSIII. Some of these functions may be more economically accomplished with other utility programs. Nevertheless, a file can be matched with itself or a null file to produce new (longer, shorter, or rearranged) records.

7. Select corresponding records from another file.

Given a haul file, matching catch records could be extracted from a large catch file. In this case, the haul file would be the master, the large catch file the slave. The match output would be only the catch data items (the hybrid output usually has data from both master and slave records, but it does not have to include data from both), and the relationship between the haul and catch data would be one-to-many.

8. Select records without corresponding records on another file.

Given haul and catch files, it is desired to select haul records for which there are no corresponding catch records, as is the case for net tows in which nothing was caught. The haul file is the master file, the catch file is the slave file. Matches and slave faults are ignored. The master faults, i.e., haul records without catch, are produced by DMSIII.

## APPENDIX I

## Analytical Programs

Appendix I contains additional descriptions of the general analytical programs used in the NWAFC survey data-base system.

1. STATION

Program STATION is used to select data from a binary haul file for plotting onto a geographical map. Any data element (word number) of the haul file may be specified, e.g., haul number, month, sea surface temperature, gear performance code, then the data for that element are selected from each haul record and written onto a plot work file.

STATION is very commonly used for both editing data and analyses. Examples of editing applications include using STATION to provide a record of cruise track, check for errors in the recorded sampling positions, and verify correct assignments of geographical strata codes. Plots of sampling gear performance made using STATION can also be a convenient record of locations of bottom substrates unfavorable for trawling. Typical analytical uses are to produce maps of bottom depth, sea surface temperature, or bottom water temperature.

## 2. FPC

Program FPC compares the relative fishing power of two vessels or two gear types, and determines whether standardization factors need to be applied to the catch data of one or more species.

3. CATCH/FISHING/POWER

If fishing power correction factors or other standardizations do need to be applied to a set of catch data, this is done using program CATCH/FISHING/POWER which applies a divisor to the binary catch file.

#### 4. CPUE

one of the most frequently used analytical programs, CPUE is used to plot station values of catch per unit effort, i.e., relative density, for a species onto a geographical map. Input information includes binary haul and catch files, and species code. Values of catch per unit effort are then computed and written onto a plot work file.

The most typical use of program CPUE is to produce maps of animal densities--graphics products that can subsequently be contoured, photographed, or used as overlay sheets.

#### 5. LIMIT/CPUE

Whereas program CPUE is used to compute station values of catch per unit effort (relative density) for a species, program LIMIT/CPUE can be used to proportion and limit the values output from program CPUE to specified sex, length, or age classes. For example, programs CPUE and LIMIT/CPUE could be used in combination to produce maps of the relative density distributions of 1-yr-old individuals of a particular species, individuals in the size range 12-20 cm, or only females.

Of course, values for subgroups can only be determined for sampling locations at which length-frequency data were collected.

#### 6. RANK

Program RANK summarizes the frequency of occurrence and mean catch per unit effort of all species and other catch items in a set of samples, and lists them in descending order; confidence intervals for the mean estimates are also reported. The input data (binary haul file) may be grouped by any number of statistical subdivisions or strata, and factors (such as geographic areas) may be applied to obtain weighted overall estimates.

Typical applications of RANK are to summarize the species list recorded by a survey and to provide evaluations of the relative importance of the various species in terms of frequency and apparent abundance.

#### 7. BIOMASS

Program BIOMASS uses the area-swept method to expand estimates of the mean catch per unit effort for a species, within an area covered by a trawl survey, to estimates of population weight and number. If length-frequency data are available, then estimates are also made of population length composition, i.e., the distribution of population number among length intervals.

Population length-frequency estimates may also be written by BIOMASS onto temporary binary size composition files for use by other programs. Analytical programs that are closely related in the usual sequence of analyses are LIST/SIZECOMP, which lists the contents of binary size composition files, and PLOT/LF, which is used to produce x-y plots of the length-frequency distributions.

#### 8. AGE/LENGTH

Program AGE/LENGTH summarizes an age-length key for a given species from a set of specimen data and, optionally, applies the key against the data in a binary size composition file (from program BIOMASS) to compute the distribution of population number among age classes. Mean lengths-at-age are also reported.

Age-length data may be selected from the specimen file by sex. Similar to program BIOMASS, output files may be generated by AGE/LENGTH for subsequent use by other programs. For example, program LIST/AGE/LENGTH lists the age-length key files and age-length population files that may be written by AGE/LENGTH. Program GROWFIT can be used to fit either of two growth models

to age-length or age-weight data. Program AGE/LENGTH/PLOT may be used to produce an x-y plot of mean lengths-at-age and fitted age-length curve.

9. NORM/SEP

Program NORM/SEP is an interactive program used for analyzing distribution mixtures, developed from MacDonald (1980), that provides an alternative to program AGE/LENGTH for distinguishing age-class populations from length-frequency data. NORM/SEP is used for analysis of length-frequency distributions to distinguish and classify modes, and to determine integrals of the various distribution components. Data are input either as a binary size composition file (from program BIOMASS) or entered interactively.

10. LLF/40

Program LLF/40 is used to compute and, optionally, plot geographic maps of the sex ratio or length of a species at individual sampling stations. Lengths may be reported as the minimum, median, or maximum length recorded at each station, or the total length range.

11. LENGTH/WEIGHT/REGRESSION and LENGTH/WEIGHT/PLOT

Relationships are frequently needed to convert body lengths to body weights, and vice versa. Program LENGTH/WEIGHT/REGRESSION selects data for length and weight from a binary specimen file, and performs a linear least-squares regression of  $\log(\text{weight})$  on  $\log(\text{length})$ . Program LENGTH/WEIGHT/PLOT can then be used to produce an x-y plot of the length-weight data points and fitted length-weight curve.

12. MATURITY/LENGTH

Program MATURITY/LENGTH' functions similarly to program AGE/LENGTH, although it summarizes length-maturity observations (actually, coding of

gonad or reproductive conditions) instead of age-length data. A table is formed, for each length class, of the number of occurrences of each gonad condition within a set of samples.

13. TAXA

As a tool for examining zoogeographic patterns, program TAXA is used to compute and plot values of the number of biological taxa within a specified group, i.e., species code range, that were recorded at individual stations. For example, TAXA could be used to produce a geographic map of the number of flatfish species recorded at each sampling location of a survey.

14. REGROUP Package

REGROUP is a package of six programs used to conduct various steps of recurrent group analysis (Fager 1957). Recurrent group analysis is a statistical method used to summarize relationships between species and to group taxa on the basis of their co-occurrences. Program REGROUP/GRP10 selects data for the analysis from binary haul and catch files. Program REGROUP/GRP15 counts and lists species in the selected data. Program REGROUP/REGROUP determines joint occurrences, indices of affinity, and species groupings. Program REGROUP/CONNEX determines the number of connections between groups. Program REGROUP/STATION lists the stations at which each group was present. Program REGROUP/GRPLT is used to plot maps of the locations of group occurrences.

15. CLUSPAK Package

Somewhat similar to the REGROUP package, CLUSPAK is a package of four programs used to perform numerical classification ("cluster") analyses of species associations. CLUSPAK, however, enables selection of species data by size-specific categories. Program CLUSPAK/PREP selects data for the analysis from binary haul, catch, and length-frequency files. The selected

data are then converted to a binary, i.e., present/absent, haul x entity matrix using program CLUSPAK/BIMAT. Program CLUSPAK/COEF is then used to convert the data matrix to a matrix of entity x entity similarity coefficients. The actual clustering of entities is subsequently performed using program BMDP1M of the University of California Biomedical Computer Programs (Dixon and Brown 1979). Finally, program CLUSPAK/GROUP/SELECT is used to plot maps of the locations where associated entities occurred.

## APPENDIX J

## Other Programs and Subroutines

This section provides further descriptions of other computer programs that have been found to be particularly useful.

## 1. MAP

The approach taken for presentation of results in a map form has been to have a single main program for drawing maps. Numerous options allow the user to specify the size of the map, the resolution, and kind of coastline features to be included. Other data may be superimposed on the map by means of a plot work file. Such plot work files are produced by various analytical programs, e.g., CPUE, STATION, TAXA, and LLF/40.

By isolating the map program from other application programs, programming is simplified and graphical presentation capabilities are easily extended to many programs--i.e., programs simply write a plot work file. Each record on a plot work file has latitude, longitude, station symbol, and a station value. The station symbol and station value are plotted by MAP at the specified latitude and longitude. Station values may represent sample numbers, catch per unit effort values for a species, bottom depths, temperatures, vessel codes, or almost anything that is contained in, or can be derived from, the data base.

A supplementary program lists the plot work file. This is a valuable aid in troubleshooting questionable plots.

Another program, UNDER/PLOT, is used to resolve overplotting of stations. For example, plotting station results where the stations are close to each other may result in an unreadable plot. Sometimes, in such cases, the user may manually remove stations from the appropriate files and rerun the analytical programs and redraw the map.

Where appropriate, though, UNDER/PLOT can be used to impose a conceptual grid on the plot work file and resolve multiple stations within a grid block to a single station. Parameters for UNDER/PLOT specify the grid sizes and boundaries and the resolution action desired, such as average, minimum, maximum, or count. The result is a new plot work file with a maximum of one record per grid block. The new plot work file can then be routed into the MAP program.

## 2. SELECT

The SELECT program selects a subset of a binary input file onto an output file. A parameter file specifies the selection criteria which are stated as a logical expression. The expression may contain a few or many terms. The relational operators less than, less than or equal to, equal to, not equal to, greater than or equal to, and greater than (LT, LE, EQ, NE, GE, GT) are allowed, as are the logical operators AND, OR, and NOT. Parentheses may be used to specify the precedence of evaluation of the logical expression.

For each input data record, the logical expression is evaluated. If it is "true," the record is selected onto the output file. If it is "false," the record is not selected. The data items are referenced by their word numbers (position) in the records. For example-, to select catch records for walleye pollock, Theragra chalcogramma, the selection parameters would be

```
"W4 EQ 21740"
```

because the species code is in word 4 of the catch records, and 21740 is the code for walleye pollock.

Character data may be selected by enclosing character constants within single or double quotation marks, e.g.,

```
W17 EQ 'USA'
```

but often specific knowledge of character fields regarding size is necessary for successful selection.

Tautologies are not recognized as syntax errors. They are evaluated as stated previously. For example,

(W1 LE 53) OR (W1 GT 53)

will cause all records to be selected, while

(W1 LT 53) AND (W1 GT 53)

will cause no records to be selected, but a "null" file is created nonetheless.

(A null file is a file without any records on it. When it is read, the end-of-file branch is immediately taken.)

The operands in the selection criteria can be either data items in the records (word numbers) or constants. For example,

(W1 GT W15)

is a legitimate statement for selecting records in which word 1 is greater than word 15.

### 3. INVENTORY

Program INVENTORY literally takes an inventory of all values that occur for a field throughout an entire binary file. The output is a report showing the values sorted into ascending sequence, their frequencies of occurrence, and a histogram of those frequencies. Report headings are assigned via the parameters for the program.

This report is an invaluable aid for finding the out of range or suspect values in a data file. The INVENTORY program itself calls other utilities, SORT/BINE and COMPACT.

### 4. CONVERT/METRICS

CONVERT/METRICS may be used for converting data from one metric to another, such as pounds to kilograms, for specified fields throughout a binary file.

This utility is extremely useful because such conversions are frequently needed. Furthermore, many of the analytical programs are independent of the metrics. When analytical results are desired in a number of different metrics, this is accomplished by converting the data on the appropriate file (usually a catch file) via CONVERT/METRICS and rerunning the analytical programs.

A parameter file for CONVERT/METRICS specifies what operations to do and upon what data fields to perform them. Other functions and arithmetic involving any of the data items on a file can be performed. For example, two fields may be added, then multiplied by a constant, truncated to an integer value, and placed in a third field on each record. The operations are performed uniformly throughout the file on the specified fields.

Another use for CONVERT/METRICS is to adjust the haul numbers for a survey. Sometimes two hauls have been assigned the same number, such as can easily happen when one haul is the last haul of one leg of a survey, and the other haul is the first haul on the second leg. In such a case, the file can be physically separated into two parts, and then CONVERT/METRICS can be run to add a constant of one to the haul number field on the second file.

#### 5. COMPACT

Another frequently used building block for other programs, COMPACT summarizes a sorted binary file. One record is written for each sequence change on the key fields in the input binary file. A parameter file for COMPACT defines the key fields and the format and content of the output summary records. The output records may contain any fields from the input records as well as sums of fields, a count of how many records were present in the group with identical key field values, or a pointer to the first record in each control sequence group. COMPACT is a key routine that is used in virtually any program that summarizes data.

6. REFORMAT

A utility used on character files, REFORMAT moves columns on a file. Columns may be added onto each record, and columns may be set to a constant value propagated throughout the file. A parameter file specifies the transformations. REFORMAT is often used to convert data from an existing format to a standard format.

7. LONGER

A utility for binary files, LONGER (a misnomer, because "wider" would be more appropriate) appends words to each record of a binary file. A parameter file specifies how many words to append to each record.

8. RGEN

A report generator for binary files, RGEN was developed by another division at NWAFC. Supplied on a parameter file are the key fields, the fields to summarize, and FORTRAN format statements for the report page headings and summary lines. RGEN sorts the input file to the specified key field sequence, and prints one line per sequence change on key field values.

9. UTAH/SORT

A sort utility for character files, UTAH/SORT was obtained from another agency and installed by another division at NWAFC.

10. SORT/BINE

SORT/BINE is a utility for sorting binary files. A parameter file specifies the key fields and whether to sort in ascending or descending sequence on each field. This and other sort utility programs do not actually sort the data files directly. Rather, they set up compare procedures according to the parameters and then call the computer vendor's sort/merge routine. Nonetheless, a convenient and expedient user interface is provided by SORT/BINE and other programs.

11. MAKE and MAKE/BINE

MAKE and MAKE/BINE are used to convert binary files to character files and vice versa. A parameter file specifies how many columns or words to use for the output records, an optional list of variables to write, and a FORTRAN format statement.

12. SELECT/SPECIES

Actually three main programs, one per data type for catch, length-frequency, and specimen files, the SELECT/SPECIES programs are used for selecting records by species codes, species code ranges, or a list of species codes/ranges. Although selection could be accomplished by using the SELECT program, the SELECT/SPECIES programs execute in about one-fourth the time of the SELECT program. As such, they are preferred over the SELECT program, especially when retrieving data from the master data base files by species code. These programs are, of course, unique to the survey data record formats.

13. INVENTORY/SPECIES

Another utility unique to the survey data is INVENTORY/SPECIES. The input is a binary catch, length-frequency, or specimen file, or any other file with species codes. A parameter file specifies the type of data being read. The output is a list of the species codes that occur in the file and their species dictionary names. Any species codes that are not in the dictionary are listed as exceptions.

14. GROUP/SPECIES

Frequently analytical results are desired not only for individual species, but also for larger taxonomic groups such as all flatfish, all rockfish, all fish, or all crabs. Rather than program the aggregation logic into each analytical program, a formidable task, species codes are reassigned on a copy

of the binary catch, length-frequency, or specimen files by program GROUP/SPECIES. A parameter file specifies the species code groupings as well as the type of data.

The output file from GROUP/SPECIES, which has multiple records per new species group code, is frequently routed into the COMPACT program to combine the multiple records per species group code into a single record per new group species code (per vessel-cruise-haul). Caution must be exercised, though, when combining catch records if some have incomplete data for the number of individuals. The resulting file with new species group codes is then routed into the desired analytical programs.

**THIS PAGE INTENTIONALLY LEFT BLANK**

## APPENDIX K

## Software Standards

The standards described in this section express how computer programs have been written for the survey data-base system and were implemented to improve the management and reliability of the system's software.

1. Programs are generalized and not written for just one person. Emphasis is placed on longer-term use or wider application.
2. Software is institutionalized. The object programs are available at all times for execution, much as, for example, the FORTRAN compiler is available.
3. Documentation: The program, when entered in the library, is accompanied by a document that explains, briefly, the function of the program, how to execute the program, what file assignments are needed, and the parameters that must be supplied for the run. The documentation also reflects a history of program modifications.
4. Source programs are generally not available to the user of the software. One reason for this is to control software redundancy, i.e., eliminate multiple versions of the same program.
5. The assumption is that a program is being written for maintenance, and will have to be understood and modified later.
6. Program/software development is differentiated from program use (execution). That is, programmers are responsible for writing the software, but persons in the host organization's other functional areas are responsible for using the software.

7. The modification philosophy is to add an option to a program, test it, and document it, rather than repeatedly modifying a source program for individual runs.
8. Data record association for two or more files is via the DMSIII relational utility routine, rather than by using (writing) record matching logic in the application programs.
9. File numbering conventions are: 1 = input data file; 2 = output data file; 5 = input parameter file; 6 = printed output file. Although there are variations from the standard, especially in programs with multiple input or output files, the standard is a good mnemonic device for reassigning files to various devices, e.g., to remote terminals, instead of disk or printer.
10. Parameters supplied by the user (for selecting options for the particular execution of the program) are echoed (printed) and diagnosed by the program.
11. Parameters generally may be in any order, although in a few programs order is important.
12. Parameter format is a keyword, usually starting in column 1, followed by a free-format, i.e., not column-number specific, value, or list of values.
13. File titles for the input and output data files are printed on the output. Since the programs are written for generic file types, and since the specific file titles can be assigned at program execution time, the printing of the specific file titles by the program is of great benefit when one is troubleshooting a run. Although the printing of file titles is dependent upon the operating system (non-standard FORTRAN), this is accomplished by calling a

subroutine to print the title, and having the system-dependent statements in the one subroutine called by all other programs.

14. Not enough can be said about commenting the program. The 2,000-line (or even 20-line) program without explanatory comments is not doing anyone any favors, especially when the time comes to answer questions about the functioning of the program or to modify the program.

15. Comments in parameter files: Many of the application programs make calls on utility subroutines, each of which in turn requires a parameter file. Comments on these parameter files document items such as intermediate record formats and sort keys.

16. FORTRAN COMMON declarations are placed in a separate source program file. This file is inserted in the main program and subprograms at compilation time by using the library "INCLUDE" feature (most compilers have such a feature). As a result we never have a problem of COMMON mismatches between the main program and the subprograms.

17. COMMON items are documented with comment statements explaining each item.

18. Statement numbers are in ascending sequence from top to bottom of the program listing.

19. Format statements, contrary to what some programming style manuals dictate, are placed directly below the first read/write statement in which they are referenced. Although there was a historical (early 1960's) compilation speed or object program size advantage in placing all of the format statements at the beginning or end of a program, that simply is not the case now. The emphasis is on quickly associating the format statement with the input/output list,

because when a change is made in the input/output list, a change usually must be made in the format statement, too.

20. Spacing and parentheses are used liberally for legibility, appearance, and understandability. For example,  $A=B+C*D**E$  is written as  $A = B + (C * (D**E))$ .

21. Items in format statements are separated by spaces for legibility, as (lx, 13, F4.2, 315) rather than (lX,I3,F4.2,315).

22. Variable names, and other items, are not split between column 72 of one statement and column 7 of the continuation statement. Rather, the list is terminated prior to column 72 and continued so that names are not split between lines.

23. Record counts are usually printed at the end of a program to show how many records were actually read and written. These are invaluable aids to anyone who is troubleshooting a run.

24. Standard data file record formats (for haul, catch, length-frequency, and specimen records) are inserted in the source program with the library "INCLUDE" feature. All references to particular variables are via the standard mnemonics defined in the included segment.

25. Debug options, frequently inserted for testing a program, are left in the program and may be selected by including a "DEBUG" option on the parameter file. When selected, the debug option prints a trace of certain variables during the execution of the program.

26. Some programs always write a trace file (similar to a debug option request, described above). The trace file is a temporary disk file that usually disappears

at the end of the job. By reassigning the trace file with the job control language, the trace may be routed to the printer, a terminal, or to a permanent disk file.

27. Batch plotting programs, i.e., non-interactive graphics, usually have a "NOPLOT" option which allows exercising the program except for the plot calls. This allows rapid testing of a program (except for the plotting functions) without the turnaround delay associated with graphics.-

28. For testing of plot routines, a substitute set of subroutines is used in place of the vendor's plot subroutines. The substitute set merely prints the parameter values on the printer and returns.

29. Test coverage analysis: A recently installed tool, "NODAL," is used to provide a report and analysis of the tested and untested portions of a program. The test coverage report identifies all statements in a FORTRAN program that have and have not been executed, down to each possible branch on each "IF" statement. This is an invaluable aid in planning subsequent test runs for a program. Some research findings, such as described by Stucki (1981), have shown 40% of the "IF" statement branch paths to have been untested in programs in a so-called production status.