

This paper not to be cited without prior reference to the authors*)

International Council for
the Exploration of the Sea

C.M. 1974/D:5 (Supplement)
Statistics Committee
Ref: Demersal Fish (N) and
(S) Cttees, Pelagic
Fish (N) and (S) Cttees

Report of the Meeting of the Working Group
on the Eventual Establishment of an ICES ADP
System for Fisheries Statistics

APPENDIX III

Trial Run of System c) Using North Sea Herring Data

by

Hans Lassen

Keld Laumann

*) Danmarks Fiskeri- og
Havundersøgelser
Charlottenlund Slot
2920 Charlottenlund
Denmark

Objectives and Requirements of the Trial Run

The objectives of the trial run were stated so as to establish the databasis necessary for the Herring Working Group to carry out their prognoses e.g. the catch in numbers of each age group. This should be done for 1972 only. The method to follow was also specified as a verbatim flow diagram. Note was to be taken of the resources required for the trial run as a probable guide to assessments of the costs.

The data requirements of the trial run were given as

- a Catch broken down by rectangle and month wherever possible, the rest of catch data broken down as finely as possible.
- b Numbers per kg broken down as finely as possible (see a).
- c Percentage of spring spawners in the samples.
- d Age distributions from samples taken at random from the catch.

Datamaterial

During the ADP meeting the datamaterial which should be included in the databank was specified and punching forms were prepared. The catch data were extracted from the ICES Statistical Newsletters (Herring and Mackerel) 1972 and from the Bulletin Statistique 1972. The biological data, numbers per kg and age-distributions were extracted partly from Statistical Newsletters and some material had been sent to the Statistician of ICES to be used for this exercise.

Design of System

We decided to design the system using a standard programme package and we chose that which has been used by the Danish laboratory for some years, the OSIRIS package.

We started with an analysis of the facilities required by the herring experts and from that list we found which of the operations could be handled by the OSIRIS system. The left over operations are the programming required for the trial run.

Analysis of the flow-diagram submitted by herring experts led us to the following list of required operations:

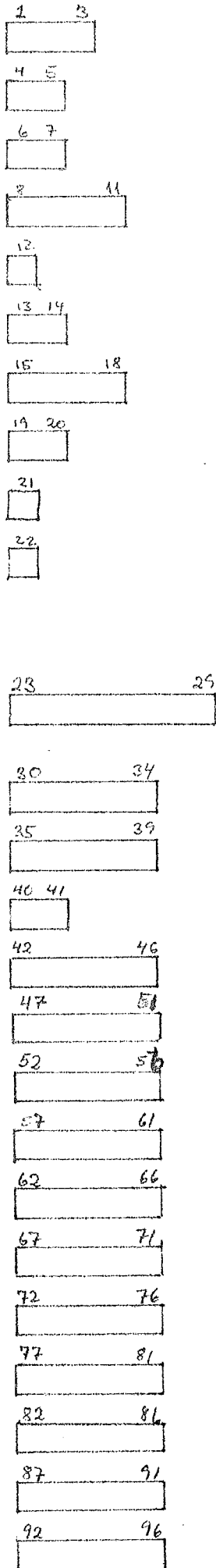
1. Display of a file.
2. Simple calculations within catch and biological data.
3. Extracting data with specified properties.
4. Tables of catch, biological material split on geographical areas time, gear etc.
5. Creating files including all data where at least a specified breakdown is available.
6. Split a total according to subtotal which have a finer structure either in time area etc.
7. Completing information from neighbouring months.

Conclusions are as follows

- a. OSIRIS will be most efficiently used if all information is collected into one file.
- b. OSIRIS will under a be able to handle points 1-4 of the above list with a single exception, i.e. tables of age distributions.

We then continued to draw up flow diagrams and a file definition for the system.

File Definition



Identifier

Country

Year

Month

Division

Subdivision

Species Area

Rectangle

Gear

Utility of Catch

Source of Sample

Data

Catch (metric tons fresh round weight)

Number counted and weighed

Weight of above number

Percentage of spring spawners

Number of fish aged

Number of fish 0-group

-do- I-group

-do- II-group

-do- III-group

-do- IV-group

-do- V-group

-do- VI-group

-do- VII-group

-do- VIII-group

-do- VIII+ -group

Keywords to be used as input.

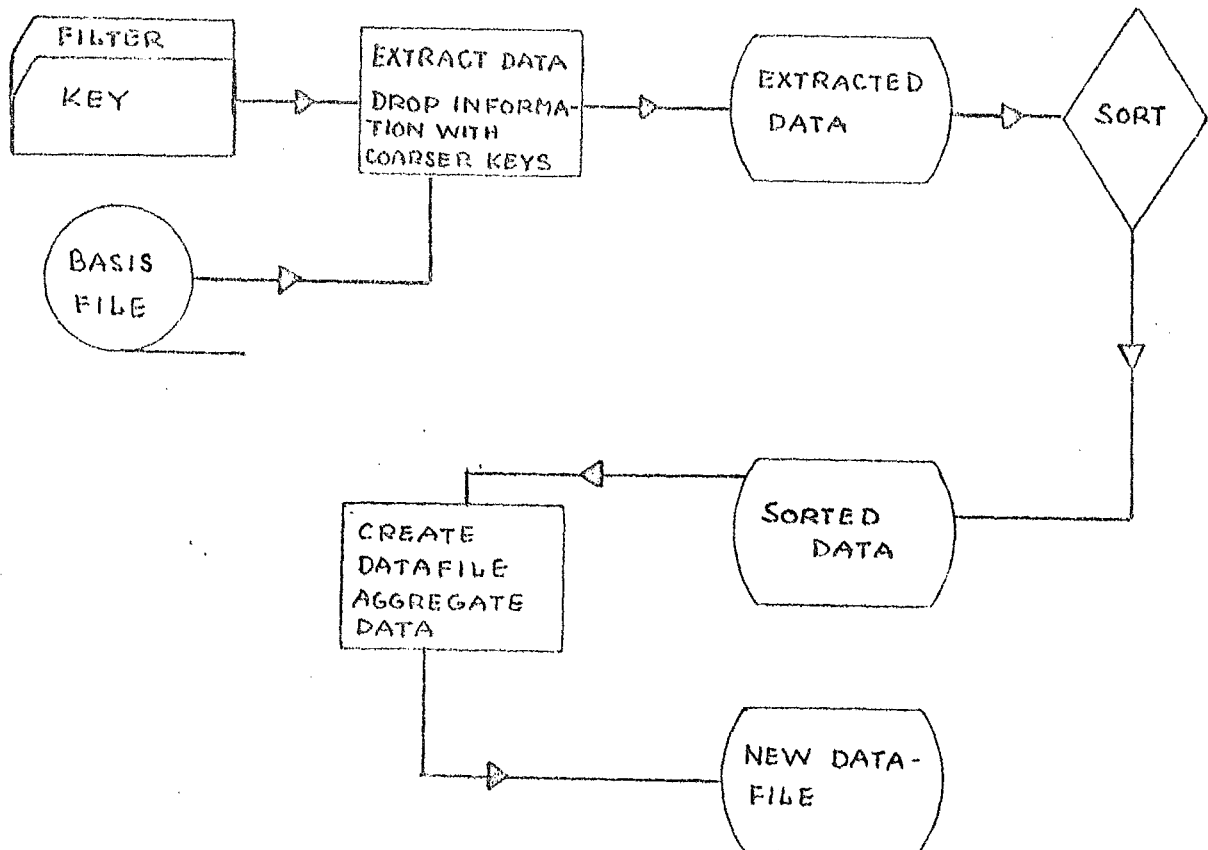
	<u>Variable</u>	<u>Keyword</u>
Identifier		
	Country	COUN
	Year	YEAR
	Month	MNTH
	Division	DIVS
	Subdivision	SBDV
	Species Area	SPAR
	Rectangle	RECT
	Gear	GEAR
	Utilization	UTIL
	Source of samples	SOUR
Data		
	Catch	CTCH
	Number weighed	NUMW
	Weight	WGHT
	Percentage of spring	SPRS
	Number aged	NUMA
	Ages	AGES

5. Creating files including data with limited breakdown.

This operation will do the following

1. It filters the basis file for all information which has a key coarser than the key specified.
2. The records are delivered on disc and sorted.
3. The sorted file is aggregated.
 1. Catch, only the information with coarsed key is retained.
 2. The biological information is summed and delivered on a disc file. The data is completed by taking data from neighbouring months.

Flow diagram



CREATE continued

The format of the key input is

```

1           8 9   12 13   16 17   20 21
CREATE      [ ID1 ] , [ ID2 ] , [   ]

```

The ID is the identifier variable which is to be retained in the identifier. Variables not listed will be summed over.

6. Split a total

This operation will do the following:

1. It takes as input two keys; one must be a refinement of the other.
2. It filters the basis file for all information which has a key coarser than the coarser input key. Records which do not fulfil this requirement are left out, the other records are transferred to disc.
3. The data are sorted, the month being the finest criterion.
4. The sorted files is aggregated on the coarser key, a new file is created.
5. The sorted file is aggregated on the finer key, a new file is created. Data is completed.
6. The following calculation is performed.

$$\frac{\text{Data var. (coarser key)} \times \text{Data var. (finer key)}}{\sum \text{Data var (finer key)}}$$

Finer
key

the result is delivered in Data var (finer key). This result is printed and put on disc in the resulting files.

The data-var is a data variable, say catch, which is going to be split.

FORMAT

```

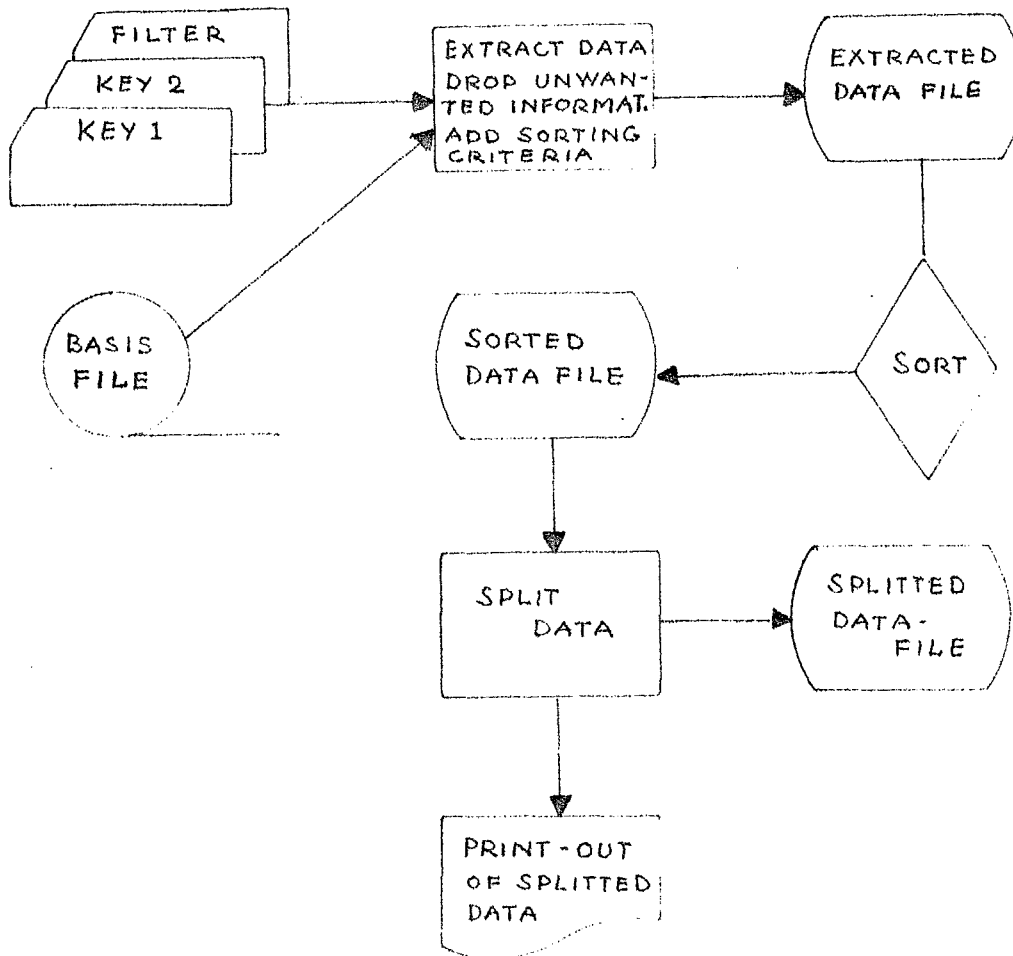
1           8 9   12 13   16 17   20 21
SPLIT      DATA VAR  ID1   ,           ,

```

The data var is the data variable which is going to be split.

The ID variables are the part of the identifier to be retained in the key.

Flow-diagram



7. Completing information from neighbouring month

In the case where biological information is lacking for a certain area in a certain month but catch has been recorded, we need some facility to complete the data. The herring experts stipulated that the neighbouring month should be used.

We have not generalized this facility but have used this way of completing data as the only one available within the system. It should be obvious, that other ways of completing are needed for the general system.

We have defined that both SPLIT and CREATE will automatically complete data. If it is not possible, the run will be aborted.

Appendix III

Description of Work and Resources Used.

Punching.

The punching was done by the Service Hydrographique of ICES directly from the prepared punching forms. No screening took place. The total number of cards are about 1000. Cost of punching is at present ca. 0.50 D.kr./card.

Costs of punching	500 D.kr.
-------------------	-----------

Data vetting.

The vetting procedure was extended to establish the automatic logical check required by the final system. This includes two computer programmes, one for the catch data, one for the biological data. These checks showed some difficulties in the data. The problems were mainly of the type that two or more identifiers were identical, but the data given were different. In some cases the statistical Newsletters gave an indication of what had happened f.ex. landings of vessels from foreign countries had been included in different ways but in some cases we had to leave the data given by the Working Group. The basic point of this task is that even in the very favourable situation given to the Working Group, data vetting requires special care and is time consuming.

The costs were

Man costs	1 man-month
Machine-time for establishing the programmes	200 D.kr.
Run of check - 6 runs	60 D.kr.

Programming

The analysis of the system given on p.2 defined 7 specified tasks which the system should be able to carry out. The first 4 of these could be handled by the OSIRIS system while point 5, 6 and 7 required special programmes. These jobs could be handled by only one programme which is not finished at the present moment. The status is that apart from final testing on real data the programme is completed.

The Danish data were not prepared on punch forms during the meeting in June but data have been extracted directly from the Danish files resident on magnetic tapes. This showed very few problems and this way of submitting data is strongly recommended if the system is put into production.

The data file in the system had to be treated using a special programme which had been written.

The total resources used were

Man costs	1½ man-month
Computer costs	1300 D.kr.
Magnetic tape, paper etc.	75 D.kr.

Estimated costs for Establishing the System c.

As a special task under the trial run a budget for establishing system was set up. We have prepared this under the following assumptions

- a ICES will do the ADP itself having the expertise available within the staff.
- b The computer costs are taken as hire of blocktime at the Danish market.
- c Data will be available in computer readable form e.g. magnetic tapes which should be supplied by the national agencies. Punching may be carried out at extra costs either by ICES itself or at a suitable firm.
- d Programmes will use higher language FORTRAN, ALGOL, COBOL or PL/1 resulting in reduced establishing costs and slightly increased running costs.

Establishing cost Budget

Man power

1 systems analyst academic level
1/2 secretary

Material

1 terminal	D kr. 35.000
1 punching equipment	25.000
1 modem + telephone	6.000
magnetic tape, paper etc.	<u>4.000</u>
total excl. tax	<u>D kr. 70.000</u>

Running costs

Estimated costs per year

computer time *)	D kr. 65.000
telephone	10.000
technical service	4.000
storage media, paper etc.	4.000
bus between computer and ICES	<u>3.000</u>
	<u>D kr. 86.000</u>

*) includes establishing data files and run out of standard tables for the Bulletin Statistique and Statistical Newsletters, plus necessary security back-ups etc.

Working groups and working parties are expected not to use more than 3500 Dkr. per meeting bearing in mind that a minicomputer Nova 1200 is available according to agreement with the Danish Institute.

Time schedule for Establishing system C

Under identical assumptions as the budget we find it realistic that the system can be setup and programmed by one man during one year. This applies to the basic system with limited analyses package. This package can be worked out in coming years.

General Remarks

With regard to the amount of data handled by ICES, the present requirements are only moderate and do not enforce the ADP-technique, but future needs, bearing quota regulations in mind, will probably drastically increase both the speed at which data should be made available and the amount of data. The present plan will allow a very quick reporting almost without increasing costs.

The availability of data on computers will increase possibilities of analyses to the working group and much tedious and time consuming work may be avoided.

The design of the system for the trial run may be used directly for other species provided that no additional data are to go into the system. Such data are likely to be length and age/length distributions. That is for the general system the design has to be redone on significant points.

The authors would like to stress that most of the data required are already available on computer readable forms at the national agencies and only therefore it is possible to establish the system with very limited manpower.

Status at 27/9 1974

	state	month man costs	Dkr. computers	Dkr. others
Data punch and vetting	finished	1	260	500
Establishing data file	almost finished	1	1300	75
Analysis programmes	under way	1/2		
Final run	not started			