Data Processing and Management at IRMA: Overview and On-going Activities

G. Garofalo*, A. De Santi, G. Norrito and S. Cusumano

Abstract

This note describes the main features of the Software for the Exploratory Analysis of Trawling Information in the Mediterranean (SeaTrim) which is currently under development at CNR-IAMC. It is an advanced software designed for the handling of data collected during experimental bottom trawl surveys. Its main component provides a flow of integrated routines to accomplish the exploratory analysis of the data, both globally and spatially oriented. SeaTrim outputs range from abundance indexes to biological features, such as length–frequency structure, sex-ratio, maturation stages, size at maturity. The software is designed to import any existing data base by converting it to a standardized format, and to provide a consistent modular architecture that can be tuned for regional data-management needs.

In view of the establishment of harmonized data-collection protocols and common analytical procedures within the framework of the FAO MedSudMed Regional Project (MedSudMed, 2003), the package may provide an excellent tool for easy data handling and exploratory analysis, especially on the shared stocks.

Introduction

CNR-IAMC has been carrying out bottom trawl surveys in the Strait of Sicily since 1985, using a traditional Italian commercial gear (GRUND project; Relini, 2000), and since 1994, using an *ad hoc* designed gear with high vertical opening (MEDITS project; Anon., 2000; Abelló *et al*., 2002). At present, the MEDITS historical data-series comprises nine consecutive spring surveys, whereas the GRUND series consists of 26 campaigns. This latter series is discontinuous in time, however (surveys were not performed in some years), and presents a seasonal heterogeneity (only the autumn season was regularly covered).

Since 1984, researchers at IRMA have been devoting much effort to improving the quality and timeliness of survey data collection, processing and analysis. Over the years, different software tools have been designed for, and applied to, the data and information management, often under the pressure of rapid technological advances. To date, however, there is neither a unique data base nor a common analytical environment to serve the demand of the different scientific programmes. Moreover, what has become increasingly important in recent years, and will become even more crucial in the future, is putting emphasis on the possibility to share data, analytical methods and assessment of results with other scientific communities in the framework of joint projects. This process includes applying harmonized codes to different data bases to increase the amount, the uses and hence the value of the data collected. Foreseeing these priorities, scientists of CNR-IAMC are now developing a new comprehensive computerized data base for the analysis of data from bottom trawl surveys. The package, called SeaTrim (Software for the Exploratory Analysis of Trawling Information in the Mediterranean), provides a standard data-base structure as well as the possibility to import or customize regional data bases. In both cases, procedures for the input and validation of the data are available. Starting from raw catch and biological (length measurements, etc.) data, a set

* MaLiRA-Group, CNR-IAMC, Via L. Vaccara 61, 91026 Mazara TP (Italy); Tel.: 0923-948966; Fax: 0923-906634; germana.garofalo@irma.pa.cnr.it

of integrated routines performs exploratory analysis and estimates biological and population dynamics parameters. Basic cartographic functions are also provided to allow a geographical representation of the population features. Preliminary analysis and results provided by SeaTrim can be used under other software (FISAT, ArcView, etc.) to accomplish the assessment of the demersal stocks.

In this note, after a brief overview of the data-management systems currently used at CNR-IAMC, the architecture and the main components of SeaTrim will be described.

Overview of existing software packages

TrawlBase

TrawlBase was developed in the early 1990s (Garofalo *et al.,* 1991; 1993) and was the first PC-based data-base management system of the GRUND programme. It was designed to manage three different types of data files (Figure 1): (a) haul files; (b) catch files; (c) species files. The "haul file" is structured to contain all the technical and operational information regarding the execution of the haul, such as geographical coordinates, duration, speed and depth. The "catch file" is designed to record the total number and weight of each species caught. Finally, the "species file" contains the biological parameters of the target species. The analytical capabilities of the system are limited to the calculation of the abundance indices, by species, both in weight (kg/km^2) and number (N/km^2).

New Trawl

The package called "New Trawl" was developed to improve the management of the data in the GRUND data base. Its purpose is to provide routines both for checking the data quality-controlling and producing a number of predefined outputs. In other words, it allows a quick look at the data and a preliminary exploratory analysis.

In particular, the software allows the automatic drawing of the graphs and the export (on Excel spreadsheets) of the following results of analysis:

- number and weight per haul
- abundance indices per haul
- maturation stages
- length-distribution by maturation stage
- number of specimens by sex
- sex ratio by size
- length–weight relationship.

Moreover, routines have been developed for exporting the data to the FISAT package according to the length–frequency (LFQ) and catch-probability (PBT) formats.

Finally, the software provides statistical and summarization routines which allow an assessment of the consistency of the data in the data base.

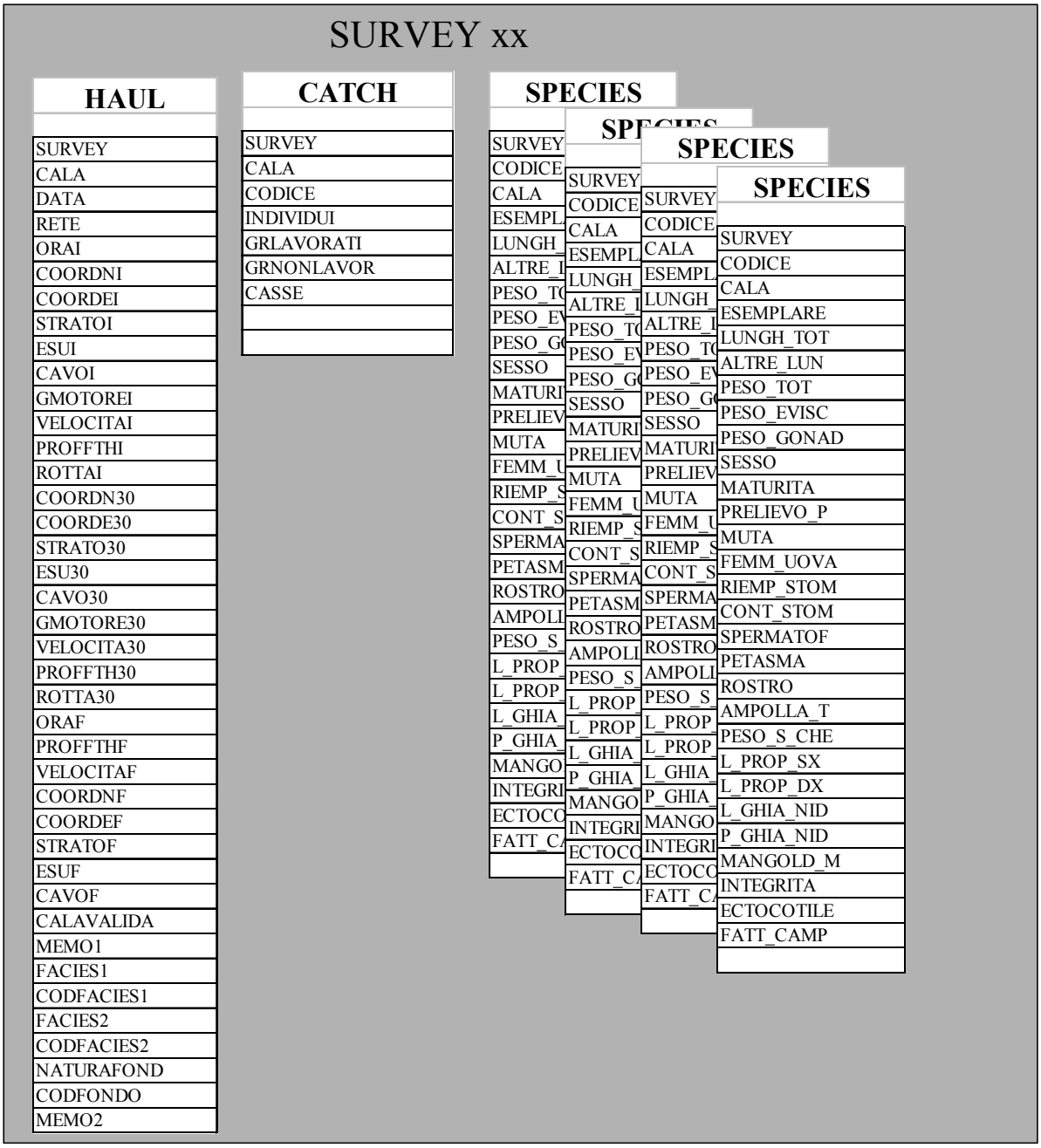

Figure 1. File structure of TrawlBase

MEDITS data base

The MEDITS Programme does not have its own data-base management system. It requires data to be recorded according to three standard exchange formats (text TA, TB, TC files; Anon., 2001), including normalized coding (Table 1) of: (a) characteristics of each haul; (b) total catch per haul; (c) biological parameters of the samples.

Observations on the target species comprise the count of individuals, length–frequency distribution, sex (including sexual maturation stage) and total weight. For all the other species of commercial interest (fishes, crustaceans and molluscs), the total number and total weight are recorded for each haul.

Two specific software tools have been developed to manage the data. The first one, CheckMed (Souplet, 1996a), has been written to allow an automatic checking of the data, the other one, IndMed (Souplet, 1996b), performs simple standard analyses of the data. These analyses include the production of biomass and abundance indices (in kg/km² and in N°/km²), as well as length–frequency distribution for each of the reference species and each of the depth strata.

SeaTrim

SeaTrim is a software for the management and standard preliminary analysis of trawl survey data. The package originates from the experience gained during the SAMED experiment (Stock Assessment in the Mediterranean; SAMED, 2002). Based on MEDITS data, scientists of the different countries involved worked together to study the abundance, biological features and age-structure of the most important target species; the scientists also performed equilibrium and non-equilibrium assessments. Indeed, the SeaTrim software has the potentiality to work on a standardized data base (the international MEDITS format has been selected as the reference) and be also employed with flexibility to export data sets for estimating the age-structure of the main demersal resources and for developing predictive capability aimed at the management of the fishery resources.

The architecture of SeaTrim, synthesized in the flow chart in Figure 2, comprises four different parts:

- data acquisition
- project definition
- elaborations supporting the analysis
- utility functions

Data acquisition – Besides the acquisition of data by importing text-format files (TA, TB, TC, standard MEDITS), which are supposedly error-free, it is also possible to create a new data base, with four information fields: survey, haul, catch data and laboratory measurements. The fields' only requirement is to be compatible with the structure of SeaTrim analytical procedures.

Data acquisition is an interactive process, following the successful completion of a series of formal checks; these checks, besides the formal aspects, produce reports and graphs which should be employed to verify the correctness of the data directly inputted into SeaTrim.

Having validated the data, an *ad hoc* data base is created for the specific survey; there is no limit to the number of these data bases (a.k.a. "sources"); in this way, a single analytical package allows the management of many different choices which may arise from the various interests of the research teams.

Project definition – The management of the projects is centered on a data base of reference, holding information on: data-base sources (with filtering options); geographical units; target species; macrostrata of interest. Moreover, an outfit data base stores pre-defined geographical areas, their coordinates and the proportion of each depth stratum; the user may define his/her own study area, for *ad hoc* projects.

Elaborations supporting the analysis – Every time a project is chosen, SeaTrim creates (following the previous definitions) an *ad hoc* processing data base which, together with a technical "query"-operating system, allows analyses of, and export procedures for: raw haul and catch data; text TA,

TB, TC files (standard format for data exchange); Excel files of the detailed exploratory analyses performed using SeaTrim "Job" procedures.

The exploratory analysis, in particular, comprises nine "Jobs":

- Job.0 – Descriptive features of the geographical units encompassed by the project
- Job.1 – Spatial distribution
- Job.2 – Abundance
- Job.3 – Length structures as box plots
- Job.4 – Sex ratio
- Job.5 – Maturity stages and size at sexual maturity
- Job.6 – Length-structure stability
- Job.7 – Length–frequency distributions
- Job.8 – Young of the year.

Side-lined functions - Among the outfit functions, the most relevant ones are: the utilities to assign a unique code to each haul; and the users' management of the data sets.

SeaTrim has been designed to operate both in a computer network and on a single desktop. The operative system is Windows 98 or succeeding versions, whereas the data bases are structured in the MDB (Access 97) format, although the application does not require the corresponding software. As a matter of fact, also Excel was employed in SeaTrim, but to manage the outputs (as *.xls files) and as a shuttle for the quick transfer of data. Consequently, it is necessary to install the corresponding Microsoft programme (Excel).

Table 1. MEDITS data-exchange formats (ASCII).

Haul characteristics (file TA)	**Catch per haul (file TB)**	**Biological parameters (file TC)**
Country	Country	Country
Vessel	Vessel	Vessel
Gear	Year	Year
Rigging	Haul number	Haul number
Doors	Codend closing device	Codend closing device
Year	Part of the codend	Part of the codend
Month	Faunistic category	Faunistic category
Day	Species code (Rubbish)	Species code (Rubbish)
Haul number	Total weight in the haul	Length-class code
Codend closing device	Total number in the haul	Fraction weight
Shooting time	Females (number)	Sub-sample weight
Shooting quadrant	Males (number)	Sex
Shooting latitude	Unsexed (number)	Number of measured individuals
Shooting longitude		Length-class
Shooting depth		Maturation stage
Haul time		Number of individuals in that class
Haul quadrant		
Haul latitude		
Haul longitude		
Haul depth		
Haul duration		
Validity code		
Course (rectilinear or not)		
Species reporting code		
Distance		
Vertical opening		
Wing opening		
Bridle length		
Warp length		
Warp diameter		
Hydrological station number		
Observations		

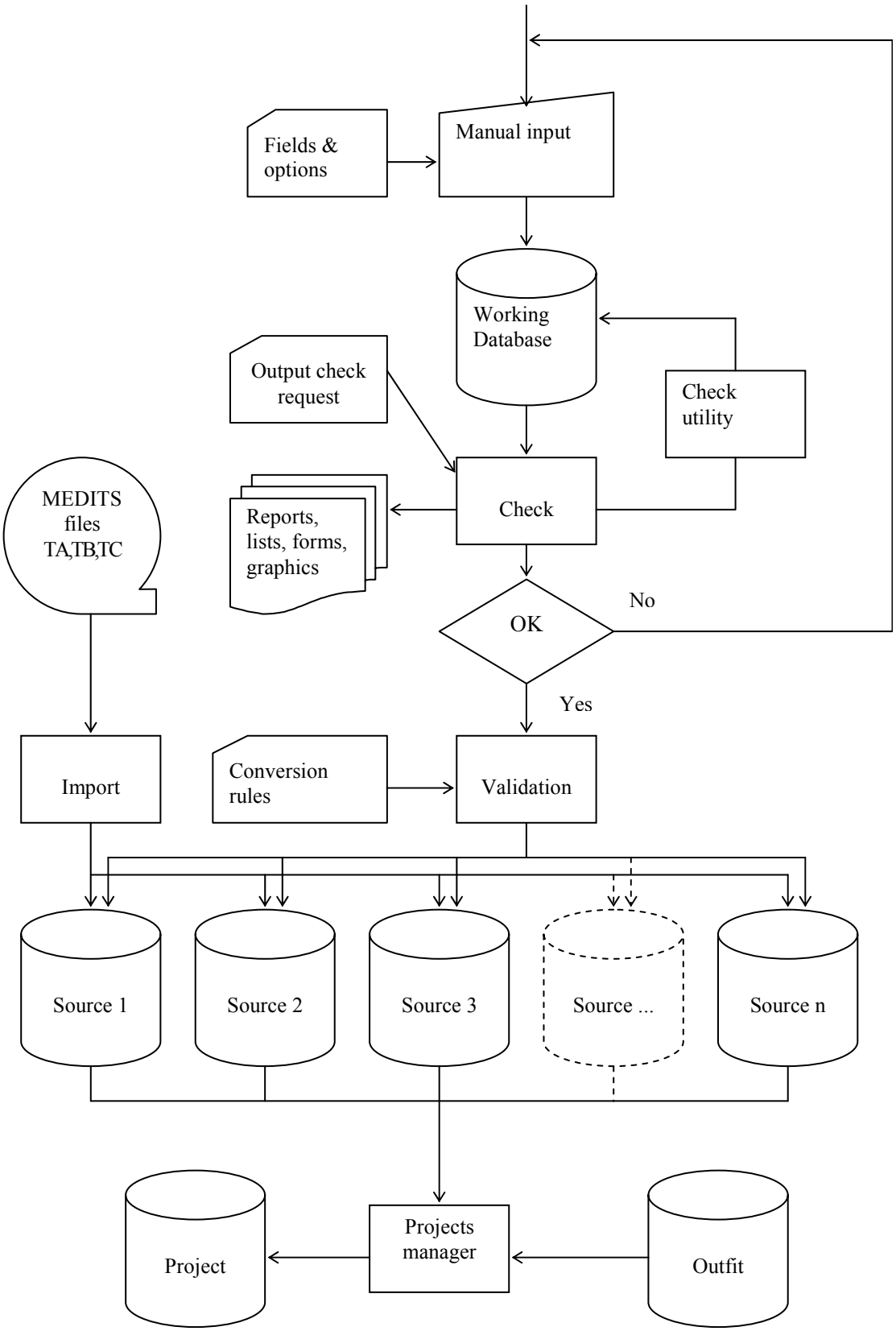

Figure 2. Flow chart representation of the **SeaTrim** software

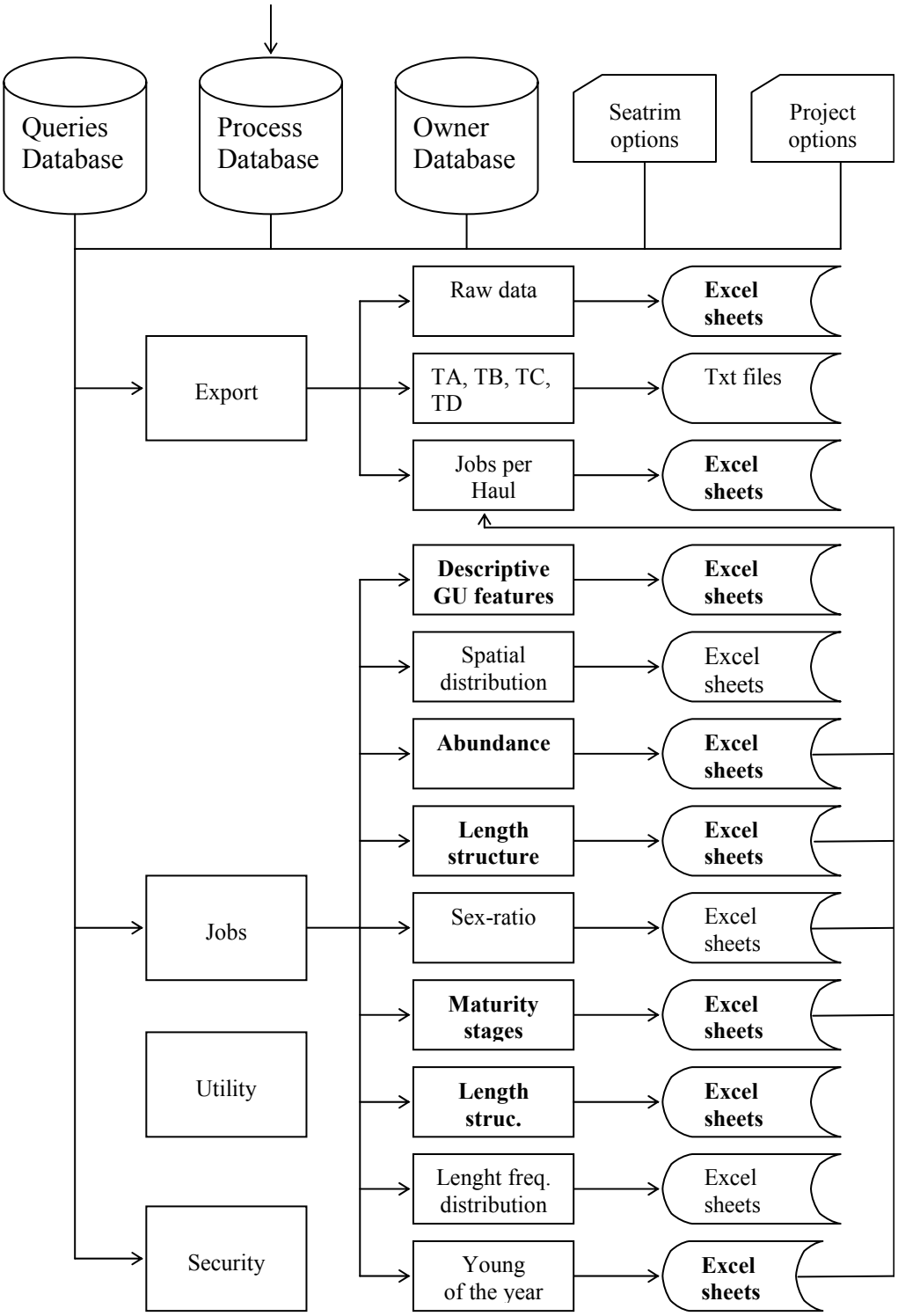

Figure 2 (continued). Flow chart representation of the **SeaTrim** software

Conclusions

Today, the concerned research communities have a growing need for collaboration and the sharing of information and data to carry out common research projects. This process includes applying harmonized codes to the data bases to allow easier summarization and use of data across all partner jurisdictions.

We believe that the development and use of SeaTrim could contribute to the promotion of scientific cooperation among research institutions involved in the assessment and monitoring of the fisheries resources through experimental trawl surveys. By standardizing and validating the storage and exploratory analysis of data, SeaTrim is expected to: increase the value and quality of the data collected; provide easy and timely access to information by a wider range of users; and increase the amount and uses of the data collected. In addition, it will reduce the labour and cost involved in storing, managing and analysing the data. Ultimately, by keeping the data in a standard format, the potential for wide data exchange and improved scientific communication is being exploited.

References

Abelló, P., Bertrand, J.A., Gil de Sola, L., Papaconstantinou, C., Relini, G., Souplet, A. 2002. Mediterranean Marine Demersal Resources: the MEDITS International Trawl Survey(1994–1999). *Scientia Marina*, 66(2): 1-280.

Anon. 2000. Programma nazionale raccolta dati alieutici ex Reg. CE 1543/2000, nell'ambito del programma nazionale di ricerca per la pesca e l'acquacoltura (PNR-P/A) 2000/2002. Ministero Politiche Agricole e Forestali, DG Pesca e Acquacoltura, Roma, Italia: 88pp.

Anon. 2001 International bottom trawl surveys in the Mediterranean (MEDITS). Volume 2: Instruction manual. Ifremer Ed., Séte (France): 57pp.

Garofalo, G., Cannizaro, L., Levi, D. 1991. A first note on the organization of the data-base of the program T.R.A.W.L. N.T.R.-I.T.P.P. n° 36: 14pp.

Garofalo, G., Cannizzaro, L., Levi, D. 1993. TrawlBase, a configurable database for trawl surveys data. ICES CM 1993/D:62, 5pp. (mimeo).

MedSudMed, 2003. Report of the first meeting of the MedSudMed Coordination Committee (Rome, 19-20/09/2002). FAO-MiPAF Assessment and Monitoring of the Fishery Resources and the Ecosystems in the Straits of Sicily. GCP/RER/010/ITA/MSM-TD-01. *MedSudMed Technical Documents, 1*: 57pp.

Relini, G. 2000. Demersal trawl surveys in Italian seas: a short review. *Actes de Colloques* Ifremer, 26: 76-93.

SAMED, 2002. Stock Assessment in the MEDiterranean. Final report to the European Commission, DG XIV, Project 99/047 (CD-Rom).

Souplet, A. 1996a. Checkmed: a software to check the MEDITS data files. In : Bertrand J. *et al*., 1996. Campagne internationale de chalutage démersal en Méditerranée (MEDITS). Campagne 1995. Rapport final. Vol. I. Rapport de contrat CEE-IFREMER-IEO-SIBM-NCMR (MED/93/020, 018, 006, 004): 21-26.

Souplet, A. 1996b Calculation of abundance indices and length frequencies in the MEDITS survey. In : Bertrand J. *et al*., 1996. Campagne internationale de chalutage démersal en Méditerranée (MEDITS). Campagne 1995. Rapport final Vol. III. Rapport de contrat CEE-IFREMER-IEO-SIBM-NCMR (MED/93/020, 018, 006, 004): 5-9.