Col. Vol. Sci. Pap. ICCAT, 55(1): 318-330 (2003)

ICCAT RELATIONAL DATABASE SYSTEM: CURRENT STATUS AND FUTURE DEVELOPMENT

C. Palma[1] and P. Kebe[1]

SUMMARY

The ICCAT relational database system (ICCAT-RDB) is the new framework developed to manage statistical and general purpose information at the ICCAT Secretariat. This document provides a description of the advancements made during last year (September 2001 to August 2002) focusing on subjects such as database status, progress made, and future strategy that will guide the implementation in the short run. Statistical databases are described in greater depth than the general purpose databases, which are outlined briefly. Special emphasis is placed on the numerous mechanisms that can effectively improve the integration of data submitted and increase the quality of statistical information. Technical wording has been avoided as much as possible.

RÉSUMÉ

Le système de base de données relationnelles (ICCAT-RDB) est le nouveau cadre mis au point pour gérer au Secrétariat de l'ICCAT l'information statistique et l'information à des fins générales. Le présent document fournit une description des avancées réalisées au cours de l'année antérieure (septembre 2001 à août 2002), en mettant l'accent sur des thèmes tels que l'état de la base de données, les progrès réalisés, et la future stratégie qui orientera la mise en œuvre dans le court terme. Les bases de données statistiques sont décrites avec plus de détails que les bases de données destinées à des fins générales, lesquelles font l'objet d'une brève description. L'accent est particulièrement mis sur les nombreux mécanismes qui peuvent effectivement améliorer l'intégration des données soumises et améliorer la qualité de l'information statistique. Le jargon technique a été évité autant que possible.

RESUMEN

El sistema de base de datos relacional de ICCAT (ICCAT-RBD) es el nuevo marco desarrollado para gestionar la información general y estadística en la Secretaría de ICCAT. Este documento proporciona una descripción de los avances realizados durante el último año (septiembre de 2001 hasta agosto de 2002), y se centra en temas como el estado de la base de datos, los progresos realizados y la estrategia futura que guiará la implementación a corto plazo. Se describen las bases de datos estadísticas con mayor profundidad que las bases de datos generales, que son objeto de una breve presentación. Se ha puesto un énfasis especial en los numerosos mecanismos que pueden mejorar de forma eficaz la integración de los datos presentados y que pueden mejorar la calidad de la información estadística. Se ha evitado la terminología técnica en la medida de lo posible.

KEYWORDS

Statistics, Database, Quality control, Validation

[1]ICCAT, Corazón de María 8, 28002 Madrid, Spain. E-mail: info@iccat.es

1. INTRODUCTION

The work plan presented at the 2000 Sub-committee on Statistics (Palma 2001 - SCRS/2000/043) development of the ICCAT-RDB. Since then, various adjustments to the original work plan have been made (Anon, 2002 - SCRS/2001/007), (Palma, 2002 - SCRS/2001/40) with the aim of including important suggestions made by the Sub-committee on Statistics (SCStat).

The sustained progress made during last two years has led to a database system with a reasonable degree of functionality. A substantial proportion of the scientific requests is now fulfilled by the new system. Despite the progress made, the system is far from its optimal performance and the development must continue according to the work plan established.

The major part of the work foreseen for this second year was accomplished with some minor problems. However, two circumstances delayed the development of the substitution and raising procedures used for the specific catch-at-size estimates (CAS). The heavy schedule of ICCAT meetings for 2002 in conjunction with the need to have the old and the new system working in parallel had side effects for this particular task (especially due to its complexity). The old system was used for CAS estimates needed for the assessments in 2002.

ICCAT-RDB has attained a level of maturity where important issues need to be discussed and decisions need to be made to improve the overall quality of statistical information. These issues involve the creation of an efficient data exchange protocol, normalization of information based on international standards, creation of hierarchical classifications (gears, fleets, geographic areas), adoption of standard rules on data estimation processes, *etc.* The success of the ICCAT-RDB depends directly on these issues.

One of the major aims of this document is to present to the 2002 meting of the Sub-committee on Statistics guidelines and proposals that can increase the quality of statistics used in the assessments.

2. DESIGN ADJUSTMENTS

2.1 Conceptual model

Client/server based systems are built so that a database on a central computer (server) is shared among multiple users, each of them accessing simultaneously the server side through an application (client).

ICCAT-RDB adopts the above architecture and makes use of the latest relational database advancements. It is a relational database system with a two-tier architecture, where the server side is implemented in Microsoft SQL Enterprise Server 2000, and the client side is made up of several programs, each having a specialized purpose (generic input, automatic input, simple output, specialized output, validation, *etc*.).

Besides its two-tier design, ICCAT-RDB has also been implemented to support a multi-tier design. Over the last year, several tests have been made using a three-tier architecture (with an internet browser as a client, a web server as the middle-tier and MS-SQL Server) to produce dynamic HTML pages as a result of database query requests. The major benefit of this approach lies in the fact that the client can be any internet browser (no need to develop client front-ends) the web server being responsible for managing all the queries and output functionality) which allows the possibility of web publishing (locally or on internet).

As the ICCAT-RDB design was intended to be in constant evolution, the web based dynamic publication of information should be understood as a long term goal, and in consequence, a strategic development issue has been focused on maintaining this flexibility to permit a low cost migration to a multi-tier architecture.

2.2 Structural review

The ICCAT-RDB has two distinct types of database groups: statistical and general purpose databases. Last year, various structural changes were made (see Palma, 2002 for comparison) in order to incorporate more flexible and efficient cross database operability. The databases now include:

Statistical group:

- task1: task 1 information
- cef: catch and effort statistics (including sharks)
- size: size samples reported
- cas: catch-at-size (CAS) reported and estimated
- catdis: catches estimates distributed on 5x5 squares
- trade: bluefin trade (BSD) and other ICCAT species traded externally
- vessels: vessel registration
- tag: tag/release information for tuna and tuna like species

General purpose group:

- biblio: ICCAT bibliographic information
- contacts: contact management
- meetings: meeting management
- mailreg: mail registration (entry / exit)
- recres: recommendations / resolutions management

The above structure is closer to what we understand as the most effective for ICCAT needs. However, the structure must be left open for additional changes so that additional databases can be incorporated and re-arrangements can be made.

2.3 Status, progress made and future planning

During this second year of development great improvements have been made. ICCAT-RDB is now able to provide a large set of querying facilities for the major databases. Some databases are in a more advanced stage of maturity than others. Below is a short description of the status of each database status (with an approximate functionality indicator), the progress made during last year and future guidelines.

Task1

This is fully operational. Major input/output routines and validation procedures were implemented. Specific querying facilities can now be implemented on demand. The "zeros" revision proposed at the last SC-Stat has been completed. More than 50% of the "zeros" figures " < 1 MT" have been updated to the real catch figures. Also, progress has been made on the historical updates tracking process. At present, the system can track backwards the last 4/5 updates made over time. Going beyond this level of historical updates will be a time consuming process and must be understood as a long term task.

cef

This is currently operating at 90% of its potential. All the information available at the Secretariat has already been integrated. Major input/output routines have been implemented. Synchronization between the data itself and the "*cef*" catalogue has been completed for 95% of the cases. This important task will allow the database to have the catalogue published in real time. Validation and quality control are the major ongoing tasks.

One of the most important changes made to this database was the integration of "shark" statistics, only for those cases where shark by-catch was submitted as catch-and-effort statistics. This process has just been started and will be finished during the next year.

size

This is operating at 70% of its potential. All the information available is already integrated. Input routines are under development and major output routines have been implemented. The synchronization between the data itself and the "size" catalogue (data record publications) has just been started and must be finished before the validation and quality control phase. This process will be finished during the next year.

Various adjustments have been made to the database in order to incorporate attributes of individual fish. Various observed parameters per individual (biometric, sex, maturity stages, etc) can now be stored. The little information available at the Secretariat with these characteristics will be incorporated over the next year. The database has already incorporated a subset of tables for storing biological parameters.

cas

This is now at 70% of its potential. Its structure is similar to the "size" database. It has been created mainly to store Secretariat catch-at-size estimates and CAS reported by nation. A great portion of the functionality of "size" is here replicated. Ar present, the information stored only covers the CAS of swordfish, bluefin and bigeye (last assessments made). One peculiarity of this database is its built-in potential to store CAS estimates using various methodologies and/or in different time periods. However, this characteristic is preliminary and under refinement. Various other automation routines (*e.g.* a chart plotting tool that allows to plot multidimensional matrices of histograms directly in Microsoft Word) are under development and testing.

One important task that has just been started is the catch-at-size (CAS) estimation procedures and related issues (substitution rules and raising methodology). At present, the Secretariat plans to build up a temporary database that will collect for each species, as far as possible, historical substitutions and raising processes made over the years. In parallel, various routines for CAS estimates will be under development and testing. This is the most important implementation task for the next year.

Catch-at-age estimates (CAA) can also be included here, if slicing algorithms used by ICCAT are integrated into the ICCAT-RDB engine.

catdis

This is in the design and experimental phase. Its main purpose is to serve as a container for CAS estimates distributed on a 5x5 square grid (output obtained by crossing "task1" and "cef" databases) . Considering that this is mainly an working database, the development of a program that will allow this database to be exported integrally to a stand alone database (e.g., MS ACCESS) for publication purposes is foreseen.

trade

For trade data associated with bluefin statistical document (BSD) the database is operating at 90% of its potential. Changes have been made to incorporate the new conversion factors used to estimate round weight (SCRS/2002/010). The information is up to date. This database has also been adjusted to store BSD document based data. The little information with these characteristics (only USA trade of 1999 and 2000) has been integrated accordingly.

Trade data for other species is simply on independent tables given the heterogeneity of information available. In consequence, minimum querying facilities are provided.

vessels

This is in the design phase. The Secretariat is currently compiling and treating the available information (EXCEL files) for integration in the near future. There are still some doubts in relation to the final design structure, which depends mainly on two issues: vessel attributes to be collected, and degree of actualization. The FAO High Seas Vessel Authorization Record database (HSVAR) is under development, and can be used as a guide for the ICCAT database structure. The second issue is most problematic. The ICCAT Secretariat does not have, currently, human resources to manage high levels of updating for such a system.

tag

This is at 50% of its potential. A large portion of the tagging information available at the Secretariat is already in the database. Thus far, the major portion of data is on individual tables which complicates the task of creating global querying facilities. Input routines also need to be developed. Integrating all the sub-datasets and refinements into the relational structure are tasks to be implemented in the near future. Validation and quality control procedures are the next steps forward.

General purpose databases

Given the lower priority of this database group, only some have been subject to development. The bibliographic database (biblio) was designed on MS SQL Server 2000. All the others are still in the relational structures developed last year (MS ACCESS stand alone databases).

The "biblio" database already includes the ASFA references associated with ICCAT (nearly 2000 records), but is still incomplete as it does not have all the related ICCAT publications and papers. Input routines are under development and output routines are already been designed and are in the testing phase. The "contacts" database did not undergo any major improvements. The "meetings" manager database is in a similar situation to "contacts". The "mailreg" database has all the information up to date. As statistical databases require information stored on "mailreg" (reference numbers associated with the statistical data submitted) there is a need to migrate it to the server and to develop the input routines in the short term. The "recres" database is its design phase.

2.4 Proposals for database enhancement

There has been a dramatic increase in the amount of data processed by the Secretariat over the last decade. While this cannot be exactly quantified, a simple exercise allows the statistical workload evolution over the years to be deduced qualitatively**. Figure 1** shows the total number of records available by year in the "size" database and distributed according to the respective processing period. As can be noted, almost 50% of size data available was processed in the last 3 years. This high figure, is distorted since it includes the large revision made to the surface fleet of European Community. The increasing trend in the workload is a consequence of the scientists' needs for more accurate and detailed information. Hence, more efficient approaches for organizing information must be considered.

Database enhancement covers both its structural optimization and improvements at data quality level through the adoption of more efficient definitions and functional associations between objects and events.

2.4.1 General purpose

Refers to proposals that apply to all the databases (ICCAT-RDB system).

Creation of ICCAT Data exchange protocol

As recognized by SC-Stat, the most problematic issue that the ICCAT Secretariat handles every year is processing the wide variety of file formats submitted. In 2002, the Secretariat processed almost 1000 electronic files (preliminary and final versions) with a total of around 70 MB. This has several implications, namely delays in data integration, possible increase in hand typing errors, impossibility of automating input procedures and delays in validation and quality control work, all of which lead to a workload beyond the Secretariat human resources capacity. A minimum of 70% of the available time of the ICCAT Secretariat Statistical department is spent on this task. In consequence, there is an imperative need to optimize this process.

The optimal solution requires the creation (and application) of an ICCAT data exchange protocol. Its main objective is to govern efficiently data submissions and information requested.

The protocol proposal must be a written document and should contain, for each data type (Task 1, catch and effort, size, catch-at-size, vessel lists, *etc*.):

data submitted:

- submission rules (formats allowed, ICCAT standard codes)
- standard submission forms
- deadlines and tolerance limits
- data validation rules for acceptance
- rules applied to datasets revised (updates)
- Minimum limits of data granularity (aggregation on time, area, etc.)
- Secretariat confirmation actions (reception and data processing)

data requested:

- List of major output information available
- Level of confidentiality associated with each data aggregation level
- Secretariat answer time limits for each requested data type group
- Rules applied to information requests other than standard outputs

Advice on this draft scheme of the data exchange protocol is welcomed.

Create built-in algorithms for Catch-at-age estimates

At present, CAA estimations are made using several stand-alone programs developed in FORTRAN. This facility can also be recreated on ICCAT-RDB engine, improving the automation level and reducing the Secretariat response time. Nevertheless, the recreation of the slicing algorithms (or other methodology used) imply the migration of FORTRAN code to the SQL and C++. Moreover, it implies the compilation of the growth equations (and inherent application criteria) for all the tuna species that require these type of estimations. If this purpose is considered, SCSat should provide guidance that could allow this task.

Revision of ICCAT classification codes

Classifications based on hierarchical trees are the most practical way of classifying objects or events. ICCAT codes (gears, vessel categories, countries/parties, geographic areas, fleets, *etc*.) lack a hierarchical categorization. Major difficulties arise when there is a need for additional codes and associations among them.

The authors recognize that hierarchical classifications are needed and that they should be, as far as possible, based on international standard classifications. They also recognize that a major revision of ICCAT standard codes can have a great impact on the national frameworks for managing fisheries statistics.

In consequence, this subject should be treated with care and a provisional study should be carried out by the Secretariat, which will present a draft proposal of hierarchical codes classification of the objects in question at the next SC-Stat meeting.

ICCAT fleets classification

ICCAT country codes have intrinsically defined various elements. For example, EC-Italy has other codes that also define geographic areas for Italy. Other countries have intrinsically defined sub EEZ's (*e.g.*, EC-Azores, EC-Madeira, EC-Canary). Others also define charter vessel flags (*e.g.*, Brazil: national fleet and foreign fleets: BRA-USA, BRA-JPN, etc.).

This classification is inefficient when grouping outputs at more aggregated levels (by country for instance), and especially when revised datasets needs to substitute the old ones in the database.

The ICCAT Secretariat is building up a list of national fleets (combination of nation, flag, fishing area and gear) in order to propose the creation of a flexible hierarchical classification of the form:

Parties > Flags > Nations > Area > gear

Table 1 shows a draft structure to build up the list of National fleets (with some examples).

2.4.2 Database related

Task 1

Task 1 updates made during Species group meetings.

Updates to the task 1 database made during species group meetings is a problematic issue, especially when species group meetings coincide. The Secretariat cannot handle efficiently all the updates and thus errors can be made with the traditional mode of collecting updates (compilation of all stand alone papers with changed figures).

This can be solved by using a standard form that records all the updates (and associated links with respective old values) made during a species group. Adopting this approach reduces updating errors, Secretariat response time, and makes it possible to track backwards the updates made over time. **Form 1** in the Appendix is a proposal for a standard form for this purpose. This approach was tested at the albacore and tropical species group this year (where one person, delegated by species group rapporteur is in charge of collecting all the updates and passes the final table to the Secretariat). It is for the SCStat to decide the best way of implementing this facility.

Reclassify ambiguous areas in Task 1.

For some species, task 1 has a certain number of records classified by geographic regions which do not reflect their stock units. When catches are grouped into stock units this information is joined in a group called "unclassified Stock area". **Table 2** shows the fraction of the catches with this ambiguity. For instance, yellowfin tuna (with management units Atlantic East and Atlantic west) has around 2% of the total catches in areas which cannot be classifed as east or west. Only a few species have information with ambiguous geographic areas. These ambiguities should be eliminated by the respective species groups or by a proposal of the Secretariat to each species group.

Reclassify unclassified gears in Task 1.

Efforts should also be made to revise task 1 information with unclassified gears, especially for the most important ICCAT species. **Table 3** shows the portion of the catches that each species has associated with unclassified gears.

Reclassify unclassified tuna (TUN and MIX) unclassified gears in Task 1.

In overall, mixed tuna catches represent around 2% of all the tuna and tuna like species in task 1 database. However, analyzing it by year (Figure 3) it can be noted that, some years have relatively large mixed tuna (e.g., 1972 and 1992 with more than 15000 MT). This information is just lost in terms of assessments. Measures should be taken to reduce this amount to more insignificant values.

Catch and effort (cef)

Eliminating duplicate effort.

The catch and effort (CEF) database has an inherent weakness resulting from the way in which data was reported over time. A reasonable amount of information is normally reported on a species oriented basis. That is, for a certain fleet in a certain period of time, catch and effort is reported for species "a", then for species "b" and so on. When the fleet in question is not mono-specific (more than one species is caught), which is the case of the majority of the tuna and tuna-like species fleets, effort is theoretically repeated for each species. This leads to an over-reporting of effort. **Figure 2** shows the distribution of CEF observations associated to the number of species reported. It can be shown that around 30% of cases have been reported as having mono-specific catches. Another deduction that can be made is that more than 75% have only three associated species, which could indicate the absence of by-catch species (tuna or sharks).

This matter does not have a short term solution for much of the historical data (as noted above, when available, sharks by-catch is included here). However, simple changes to the reporting structure can be made in order to avoid this problem in future submissions. The solution is to orient the CEF data on an effort basis ,*i.e.* report the catches of all targeted and by-catch species (tuna and sharks) for each measure of effort, according to the current standard form for CEF.

Effort categorization

Two distinct effort categories can be found on "cef" database: gear independent and gear dependent. **Table 4** shows the number of records distributed by effort measures (ICCAT codes) on "cef" database. It can be verified that each category has around 50% of the information. Currently, the catch and effort database has only one measure of effort.

Quality of data can be greatly improved if two effort measures are collected in the near future. A re-definition of effort measures according to the above categories is a consequent requirement.

Another related issue is the fact that nearly 4% of CEF information has no effort associated. This CEF data is superfluous and should not be allowed in the future by imposing quality control rules.

Minimum limits to group time and area.

A small portion of CEF data (mainly historical series) is grouped by year or quarter (less than 5%). The major portion is monthly-based. Future CEF data reporting should, therefore, maintain at least this level of time period detail. Efforts should also be made to revise the information reported on a yearly or quarterly basis.

Geographically, around 95% of available CEF information is reported in 1x1 or 5x5 square grids (65% and 30% respectively). The remaining 5% is classified by various other grids (10x10, 10x20, 10x10, 10x20 grids) and ICCAT sampling areas. Less than 1% does not have any kind of geographical classification.

Effort should be made to maintain this level of geographic detail (e.g., establishing the 5x5 grid as the minimum geographic grid for CEF data).

Size and CAS

Re-enforce submission of datasets

Both datasets, size sampling information and CAS estimates obtained by national scientists, are mandatory submissions. A considerable amount of size data submitted does not cover both datasets. In particular, size sampling is submitted without the respective CAS dataset. Once again, the SC-Stat should re-enforce this request.

Minimum time and geographic limits.

Table 5 shows the overall relative distribution of size sampling (size) with two attributes: "Time period" and "Kind of square". More than 90% of the "size" data is monthly-based. Monthly samples should be adopted has the minimum time period allowed for data submissions (validation rule).

With respect to geographic areas, more than 70% of the information is classified in a 5x5 square grid or finer area. It is quite a good indicator of quality. The remaining size information has ICCAT areas associated (15%) or larger grids (12%). Less than 1% of the information has no area associated.

Geographic grids larger than 5x5 squares are hard to handle (especially rectangles), particularly when they need to be grouped inside the standard (species based) sampling areas of ICCAT. They should be avoided in the future. Using 5x5 squares can be used as a minimum geographic limit.

Other databases

Deeper analysis is required to pinpoint possible weaknesses.

REFERENCES

PALMA, C. 2001 Strategy towards the ICCAT Relational Database Management System (ICCAT-RDB). Palma. C. *ICCAT Collective Volume of Scientific Papers* Vol LII pp 1875-1894: SCRS/2000/043.

ANON. 2000. Detailed Report: Intersessional Meeting of the Subcommittee on Statistics. *CollectiveVolume of Scientific Papers*. Vol. LI. Pages ; 2119-2169. ICCAT, Madrid.

ANON. 2002 Meeting of the Ad hoc Committee for advice on the implementation of the ICCATrelational database system (ICCAT-RDB). *ICCAT Collective Volume of Scientific Papers* Vol LIV pp 1693-1709 (SCRS/01/007).

PALMA. C. 2002. Current status of ICCAT relational database management system (ICCAT-RDB). *ICCAT Collective Volume of Scientific Papers* Vol LIV pp 1710-1712. (SCRS/01/040).

Table 1. Table structure to compile the list of National fleets. Some examples are shown.

Fisheries						
Fleets			Gears			
FLAG	**COUNTRY**	**base_port/region**	**GEAR/cat**	**GEAR TYPE**	**CHARACTERISTIC**	**TARGET SPECIES**
BRAZIL	*BRAZIL*	*RIO*	*LONGLINE*	*SURFACE*	*JAPANESE-LL*	*ALB*
BRAZIL	*U.S.A.*	*NATAL*				
BRAZIL	*JAPAN*	*SANTOS*				
BRAZIL	*URUGUAY*	*STA CATALINA*				
BRAZIL	*SPAIN*	*ITAJAI*				
EC-FRANCE	*EC-FRANCE*	*TROPICAL*				
EC-FRANCE	*EC-FRANCE*	*G DE GASC.*				
EC-FRANCE	*EC-FRANCE*	*MEDI*				
EC-ITALIA	*EC-ITALIA*	*TYRRHENIAN*				
EC-ITALIA	*EC-ITALIA*	*ADRIATIC*				
EC-ITALIA	*EC-ITALIA*	*S.SICILY*				
EC-ITALIA	*EC-ITALIA*	*LIGURIAN*				
EC-PORTUGAL	*EC-PORTUGAL*	*AZORES*				
EC-PORTUGAL	*EC-PORTUGAL*	*MADEIRA*				
EC-PORTUGAL	*EC-PORTUGAL*	*MAINLAND*				
EC-SPAIN	*EC-SPAIN*	*CANARY*				
EC-SPAIN	*EC-SPAIN*	*MALAGA*				
EC-SPAIN	*EC-SPAIN*	*SANTANDER*				
EC-SPAIN	*EC-SPAIN*	*BASQUE*				
EC-SPAIN	*EC-SPAIN*	*TROPICAL*				

Table 2. Task 1 specific catches (MT) with inconsistent statistical areas, if Stock units are applied for each species (Total refers to total specific catches of all Task 1 information).

			Inconsistent statistical areas			
Species	Stock units	TOTAL	ATL	ETRO	WEST	%
ALB	AT.N and AT.S	3362138	8045.5			0.239%
BIL	AT.N and AT.S	3554	449.2		19.0	13.175%
BUM	AT.N and AT.S	148102	3579.9	0.2		2.417%
SAI	AT.E and AT.W	116745	492.0			0.421%
SKJ	AT.E and AT.W	4204993	1443.0			0.034%
SWO	AT.N, AT.S and MED	1197637	20.8	0.0		0.002%
WHM	AT.N and AT.S	72211	558.6	0.2		0.774%
YFT	AT.E and AT.W	5196019	112610.8			2.167%

Table 3. Task 1 specific catches (MT) with unclassified gear (UNCL) and respective proportion of the total (Total refers to total specific catches of all Task 1).

Species	TOTAL	with gear	UNCL gear	% UNCL
ALB	3362138	3261894	100244	2.98%
BET	2802339	2799897	2441	0.09%
BFT	1490568	1380955	109613	7.35%
BIL	3554	2419	1135	31.93%
BLF	88041	70636	17405	19.77%
BLM	866	848	18	2.08%
BON	1482614	619922	862693	58.19%
BOP	39063	38725	338	0.87%
BRS	256730	236133	20597	8.02%
BUM	148102	143358	4744	3.20%
CER	15139	15138	2	0.01%
FRI	658210	584328	73883	11.22%
KGM	437198	313872	123326	28.21%
KGX	51051	15797	35254	69.06%
LTA	536004	467577	68427	12.77%
MAW	89592	86063	3529	3.94%
MIX	3894	3208	685	17.60%
SAI	116745	110317	6428	5.51%
SBF	60703	60692	11	0.02%
SKJ	4204993	4191664	13329	0.32%
SPF	6212	6210	2	0.03%
SSM	508443	206841	301602	59.32%
SWO	1197637	1098038	99599	8.32%
TUN	367332	212152	155181	42.25%
WAH	47598	31290	16308	34.26%
WHM	72211	71368	843	1.17%
YFT	5196019	5121509	74510	1.43%

Table 4. Number of records per Kind of effort category, and effort type categorization.

Effort category	Kind Effort	N. Recs	% effort	% of category
gear dependent	N.POLE-D	122	0.0	51.8
	LINE.DAYS	20	0.0	
	NO.HOOKS	167691	50.1	
	NO.LINES	325	0.1	
	NO.SETS	4393	1.3	
	NO.TRAPS	188	0.1	
	NO.POLES	20	0.0	
	TRAP D.	529	0.2	
gear independent	D.AT SEA	5825	1.7	43.9
	D.FISH.	134218	40.1	
	FISH.HOUR	4608	1.4	
	NO.BOATS	1020	0.3	
	NO.TRIPS	784	0.2	
	SUC.D.FI	491	0.1	
	SUC.SETS	61	0.0	
?	**NO DATA**	**14381**	**4.3**	**4.3**
Total		334676	100.0	100.0

Table 5. Overall relative distribution of size sampling (size) with two attributes: Time period and Kind of square.

Time Period	%
Monthly	92.7
Semester	0.0
Trimester	6.5
Yearly	0.8

Kind Square	%
1x1	40.3
5x5	32.6
10x10, 10x20, 20x20,5x10	11.9
ICCAT area	14.8
No square	0.4

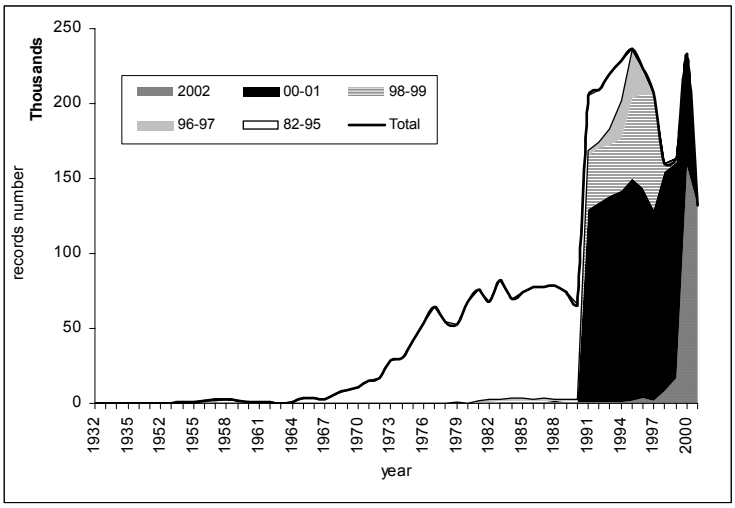

Fig.1. Database "size" (as of Aug 2002): Total number of records available by year, and distributed according to data processing period.

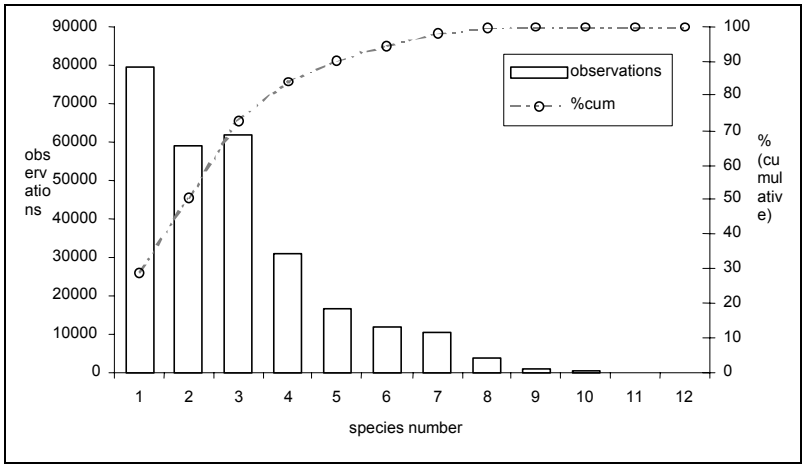

Fig.2. Catch and effort database (cas). Distribution of catch and effort observations according to the number of species reported.

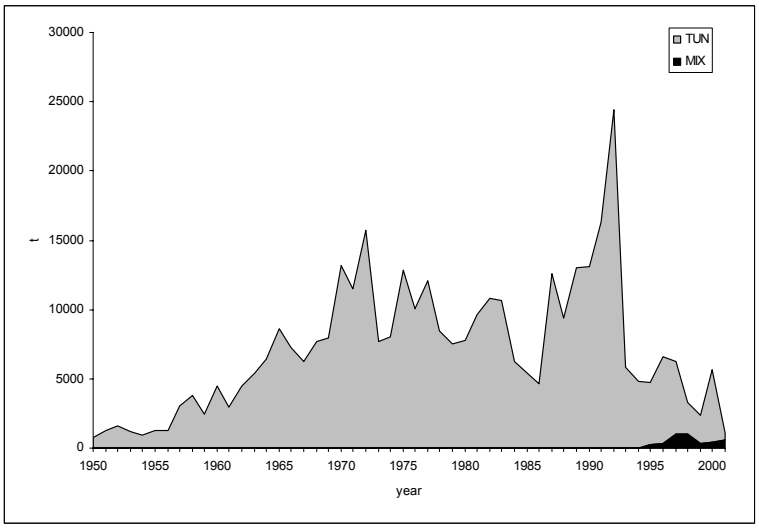

Fig.3. Mixed tuna total catches between 1950 and 2001.