

AN UPDATE OF THE ICCAT GBYP DATA MINING AND DATA RECOVERY ACTIVITIES

Antonio Di Natale¹

SUMMARY

The Atlantic-wide research programme on bluefin tuna, conventionally GBYP, among several objectives, has the duty to improve the knowledge of bluefin tuna biology, ecology and ethology. The results of the first years of the data recovery and data mining activity are here presented. The GBYP has been able to recover a considerable amount of historical and recent data sets, which concern most of the gears and many fishing grounds. The data related to fishing gears used by vessels cover the years from 1903 to 2010, while the data related to tuna traps constitutes a very long historical series, from the year 1512 to 2009, constituting the largest time series among all RFMOs. Data were also recovered from farmed fish. Most of the data concern Task II (length, weight, effort), but catch data are also in high number. The data were all cross-checked against the ICCAT bluefin tuna data base, and then individually quality checked. A high amount of trade and market data was also recovered and validated in the last year. This report includes a general overview of the various data sets.

RÉSUMÉ

Le Programme de recherche sur le thon rouge englobant tout l'Atlantique, dénommé conventionnellement « GBYP », a parmi plusieurs objectifs la mission d'améliorer les connaissances sur la biologie, l'écologie et l'éthologie du thon rouge. Le présent document fournit les résultats des premières années d'activités de récupération des données et d'exploration des données. Le GBYP a pu récupérer un volume considérable de jeux de données historiques et récentes qui se rapportent à la plupart des engins et à de nombreuses zones de pêche. Les données relatives aux engins de pêche utilisés par les navires couvrent les années allant de 1903 à 2010, alors que les données relatives aux madragues thonières constituent une série historique très longue, partant de 1512 à 2009, ce qui constitue la plus longue série temporelle de toutes les ORGP. Des données relatives aux poissons en cages ont également été récupérées. La plupart des données appartiennent à la Tâche II (longueur, poids, effort), mais il existe aussi un grand volume de données de capture. Les données ont toutes été vérifiées par croisement par rapport à la base de données de l'ICCAT sur le thon rouge et la qualité de chaque donnée a été individuellement vérifiée. Un grand nombre de données de marché et commerciales ont aussi été récupérées et validées au cours de la dernière année. Le présent rapport inclut un aperçu général des divers jeux de données.

RESUMEN

El Programa de investigación de atún rojo para todo el Atlántico, denominado GBYP, tiene, entre otros objetivos, la tarea de mejorar los conocimientos de la biología, la ecología y la etología del atún rojo. Se presentan los resultados de los tres primeros años de actividades de recuperación y minería de datos. El GBYP pudo recuperar una cantidad considerable de conjuntos de datos históricos y recientes, que afectan a la mayoría de los artes y a muchos caladeros. Los datos relacionados con los artes pesqueros utilizados por los buques cubren los años desde 1903 a 2010, mientras que los datos relacionados con las almadras constituyen una serie histórica muy larga, desde el año 1512 hasta 2009, lo que constituye la serie temporal más larga de todas las OROP. También se recuperaron datos de peces engordados. La mayoría de los datos se refieren a la Tarea II (talla, peso, esfuerzo) pero también hay una gran cantidad de datos de captura. Los datos fueron verificados con la base de datos de atún rojo de ICCAT y posteriormente se comprobó su calidad individualmente. Una cantidad muy alta de datos comerciales y de mercado se han recuperado y validado en este último año. Este informe incluye una perspectiva general de los diversos conjuntos de datos.

KEYWORDS

*Bluefin tuna, Large pelagic species, ICCAT, Data collection,
Data recovery, Data analyses, Mediterranean Sea, Atlantic Ocean*

¹ ICCAT, GBYP, c/Corazón de María, 8, 6a. 28002 Madrid, Spain.

1. Introduction

Bluefin tuna data used in the assessment were officially classified as “unreliable” by the SCRS in most of the reports over the last decade and, for this reason, data mining and data recovery was set by the Commission as one among the first priorities of this programme.

As usual, the first preliminary activity was conducted at the ICCAT Secretariat. An updated analysis of the ICCAT data base on bluefin tuna was carried out, with the purpose to identify the most relevant gaps in the data series which are potentially useful for the stock assessment, taking into account the data already collected under GBYP Phase 1; this gap analysis was provided by GBYP to the SCRS Scientists and National statistical correspondents to help them in detecting the lacking data.

2. Objectives of the data mining and data recovery

The objective of data recovery and data mining activities is to fill the many gaps existing in several data series currently present in the ICCAT data base, concerning both recent and historical data, which causes a large amount of substitutions in the assessment process, increasing uncertainties. At the same time, data mining activities should provide reliable data series, longer than those currently available, recovering data from many sources, including archives having difficulties for the access. This activity will allow for a better understanding of the long-time catch series by gear, improving the data available for the assessment and possibly for replacing substitutions used for data gaps.

For Phase 3, the GBYP Steering Committee limited the data mining only to an exploratory work to be done for the Ottoman archives and, if this was not be possible, to a further data recovery for historical trap data. The GBYP Steering Committee excluded again any possibility for recovering more recent data from other fisheries.

3. Data recovered in Phase 1, 2 and 3

In total, the ICCAT GBYP data recovery and data mining activities in Phase 1, Phase 2 and Phase 3 were carried out by issuing 11 Calls for Tenders and 18 contracts. The very high amount of data recovered in the first three Phases was detailed by Di Natale *et al.*, (2014). The data collected in the first three Phases were fully checked and validated, they have been fully discussed by the SCRS, the Subcomstat and the BFT Species Group, and now they are finally available for the ICCAT bluefin tuna data base.

4. Activities in Phase 4

The objective which was set for data recovery and data mining in Phase 4 has been accomplished, even if formally the report will be provided in September, after the 2014 SCRS meeting. Partial reports, up to September 2013, were provided by Justel-Rubio *et al.* (2014b), Di Natale *et al.* (2014a), Justel-Rubio *et al.* (2014a), Örenc *et al.* (2014) and Di Natale *et al.* (2014b). Following the recommendations of the Steering Committee, the objective for Phase 4 was set again focused on the Ottoman data, because additional data from the Eastern Mediterranean Sea, the Marmara Sea and the Black Sea are considered of basic importance for understanding the evolution of both the bluefin tuna population and fisheries. Due to the very peculiar difficulties included in this data mining activity, it was decided to have a meeting with the team of scientists who conducted the first part of the exploratory work in Phase 3 before issuing a Call for Tenders. The meeting took place in Istanbul in April 2013, also attended by Dr. Saadet Karakulak, and all problems and possibilities were deeply analysed. As a result of this important operational meeting, it was decided to suspend the data mining activities in Turkey, until real possibilities of finding historical data will be detected. The frank and very professional cooperation of the Turkish team was very useful for adopting the right decision.

Then, after taking into account the initial data mining and data recovery objectives set by the Commission when the GBYP was established and the additional and new information provided to GBYP by some scientists, concerning the opportunities for recovering or mining various data sets, it was issued the ICCAT-GBYP Call for Tenders 06/2013, targeting detailed fishery and size frequency data from NW Atlantic and historical genetic data from the Eastern Mediterranean. Two bids were received but then one offer was not finalised for local difficulties of the tender. As a matter of fact, the objectives were re-conducted to the original one and a contract was issued for recovering ancient genetic data from samples collected at the beginning of the XX century in the central-eastern Mediterranean and in the Marmara Sea. The samples were recovered and the genetic analyses will be presented at the SCRS in September 2014.

Furthermore, following the recommendation of the GBYP Steering Committee and the SCRS BFT Species Group, GBYP carried out a difficult work for setting a team able to analyse the many trade and auctions data provided as a gift in kind to GBYP in Phase 3. Finally, after many discussions and negotiations, a contract was provided to an external expert who analysed in details all the many data sets, following very carefully all the instructions `provided by the TORs and the result of this huge work was presented at the SCRS BFT Data Preparatory Meeting in May (Mielgo Bregazzi, 2015). Two important sets of data were selected as “reliable”, while a third set would need additional efforts, particularly for cross-checking the data with the BCD. The SCRS BFTDPM had acknowledged the quantity and quality of the work carried out on the trade, auction and market data, setting further provision for their official incorporation in the SCRS data base.

In addition to the activities already planned, many historical data on tuna traps were donated to Ph.D. Jean-Marc Fromentin (IFREMER) to ICCAT GBYP. These data were provided on an excel file, having 10 spreadsheets. The data cover the period 1525-2000; they include 25 traps from several countries, concerning 1,569,939 bluefin and an undefined amount of tons. Several of these data, obtained during the ICCAT GBYP data mining activities, have been already included in the ICCAT BFT data base and, for this reason, the “new” data need to be carefully checked against the official data base. This work, which was not planned in Phase 4, will be carried out in Phase 5 if a data analyst support will be made available for GBYP.

4.1 Bluefin tuna fishery data analyses

As reported on the above point 3, all fishery and size data recovered by ICCAT GBYP in the first Phases have been deeply quality checked, cross-controlled against the ICCAT BFT data base and analysed, as requested by the GBYP Steering Committee. The result of this intense and complex work were provided to the ICCAT Bluefin tuna Meeting on Biological Parameters Review (Tenerife, May 2013) (Anon., 2014), specifically charged by SCRS to provide recommendations for the use of data recovered by GBYP. The analyses and the overview of the data have been included in Di Natale *et al.* (2014a). The summary results showing the total number of data recovered are on **Tables 1 and 2**. The numbers on the two tables do not include the data recently donated to ICCAT GBYP, mentioned in the last paragraph of the previous point 4.

Many GBYP data sets were directly used by the various small WG established during this Meeting. The final recommendations by the Tenerife Meeting (see the final Report on http://www.iccat.int/Documents/Meetings/Docs/2013-BFT_BIO_ENG.pdf) are the followings:

- a) For the Task II size data the Group considered that the methods used to validate those data have been appropriate and agreed to incorporate these data to the ICCAT data bases.
- b) As regards Task II catch and effort series that fill gaps in ICCAT current data base, once the quality checking is passed, be incorporated in the ICCAT data base.
- c) Regarding Task II catch and effort series recovered under the GBYP that overlap, scientists from the involved CPCs will work in collaboration with the Secretariat in order to solve the problem. Those corrections will be submitted to the next BFT Species Group for approval by the SCRS.
- d) Regarding Task II catch and effort series that overlap with those already in ICCAT database, scientists from the involved CPCs will work in collaboration with the Secretariat in order to solve the problem. Those corrections will be submitted to the next BFT Species Group for approval by the SCRS.

The action recommended in point a) has been already successfully completed and then the data will be finally officially incorporated in the ICCAT BFT Data Base. For action recommended in point b), overlapping and conflicts were fully identified and finally resolved, all other catch and effort data will be directly incorporated in the ICCAT BFT Data Base. For the actions recommended in points c) and d), the GBYP and the Secretariat worked with the Statistical Correspondents and the national scientists of each CPC concerned; overlapping and conflicts were fully identified and finally resolved (as communicated to the SCRS Data Preparatory Meeting on May 2014). Catch/effort data for the period prior to 1950 were added to the ICCAT BFT data base.

The analysis work carried out by GBYP concerned also Task I data recovered by the Programme. These data were cross-checked against the data already existing in the ICCAT Task I BFT data base and the results are in Justel-Rubio (2014). A few conflicts were noticed and these were further discussed according to the procedures in place. All other GBYP catch data including those for the period prior to 1950 were added to the ICCAT BFT Task I data base.

5. Trade, auction and marked data validation

One of the objective set for Phase 4 was the validation of the several data sets including millions of trade, auction and market data, which were provide as in-kind donation to GBYP in previous Phases. These data were deeply discussed by the SCRS BFT Species Group and a strategy was recommended for validating them. The GBYP Steering Committee elaborated very detailed terms of reference that were largely discussed for about three months with the interested parties. Besides all many efforts and commitments, the contradictory validation was not possible and, in agreement with the Steering Committee, the validation was carried out by only an external expert.

The first overview of the analyses was presented to ICCAT on 27 April 2014, some comments were provided to the expert and then the draft report was reviewed, taking all comments into account. The final report about this validation work (Mielgo Bregazzi, 2015) was presented to the SCRS Data Preparatory meeting in May 2014 (the data recovered are included in **Table 1** and **Table 2** and showed in **Figure 1**).

Many data were considered fully reliable and then fully validated, while other were selected and identified for additional validation against statistical documents and BCDs, a work which needs much additional time and efforts, with the strict cooperation of the CPCs concerned, national experts and the ICCAT Secretariat.

The SCRS BFT Data Preparatory meeting “recommended the creation of a group of experts (to be established by the Group rapporteurs, the Secretariat and the GBYP coordinator during the September species group meeting) to review and fully validate the trade data compiled and presented in Mielgo Bregazzi (2015) for use in the 2015 stock assessment. The Group acknowledged the important work of preliminary validation carried out by the external expert contracted by GBYP”.

6. Conclusions

These first four Phases of ICCAT-GBYP activities confirmed both the good opportunities to recover and make available many bluefin tuna data sets and the high importance of this work for improving our understanding of bluefin tuna fisheries.

The data mining concerning ancient data is now showing some limits, because finding additional data sets may imply very considerable efforts, either in terms of funds or in time required for carrying out the mining in ancient archives. Furthermore, some promising archives (like the Ottoman one) showed that data were very limited, even if important scientific information can be always recovered.

An important activity which can be done in future Phases, following also the recommendation of the ICCAT GBYP mid-term Review (Fonteneau *et al.*, 2014), is to recover all the available data sets which have been published by many authors or which are included in old books and reports, which are not in the ICCAT data base. This important but not easy work will be able to further improve the data base, finally making available for everybody data sitting in various libraries, which are quite often very difficult to access.

The data recovery activity, as also indicated by the opinion of both managers and senior scientists (see Di Natale *et al.*, 2014), has many additional opportunities and could also provide several additional recent data sets that might better explain some aspects of the bluefin tuna fisheries in more recent times.

As clearly stated by the Commission, when ICCAT GBYP was launched, the use of bluefin tuna data recovered under GBYP activities must be limited to scientific uses, excluding any possibility of using these data for any compliance issue.

The procedures for incorporating all GBYP data sets in the ICCAT data base should be speeded-up, of course excluding any detrimental effect on the necessary quality controls and taking into account all the established ICCAT procedures.

Bibliography

- Anonymous, 2014, Report of the 2013 Bluefin tuna meeting on biological parameters. Tenerife, Spain, May 7 to 13, 2013. Collect. Vol. Sci. Pap. ICCAT: 70(1): 1-159.
- Di Natale A., Idrissi M., Justel-Rubio A., 2013, BFT catch and size historical data recovered under the Atlantic-wide research programme for bluefin tuna (ICCAT-GBYP Phase 1 and 2) Preliminary report. Collect. Vol. Sci. Pap. ICCAT, 69(2): 816-856.
- Di Natale A., Idrissi M., Justel Rubio A., 2014a, Bluefin catch and size historical data recovered under the Atlantic-wide research programme for bluefin tuna (ICCAT-GBYP phases 1 - 3). Collect. Vol. Sci. Pap. ICCAT, 70(1): 241-248.
- Di Natale A., Idrissi M., 2014b, ICCAT GBYP Atlantic-wide Research Programme for Bluefin Tuna 2013. GBYP Coordination detailed activity report for Phase 3 (last part) and Phase 4 (first part). Collect. Vol. Sci. Pap. ICCAT, 70(2): 459-498.
- Fonteneau A., Suzuki Z., Payne A.I.L., 2014, Mid-term review of the ICCAT Atlantic-wide research Programme on Bluefin Tuna. Collect. Vol. Sci. Pap. ICCAT, 70(2): 565-584.
- Justel-Rubio A., Ortiz M., 2013. Review and preliminary analyses of size frequency samples of bluefin tuna (*Thunnus thynnus*) 1952-2010. Collect. Vol. Sci. Pap. ICCAT, 69(2): 297-330.
- Justel-Rubio A., Ortiz M., Parrilla A., Idrissi M., Di Natale A., 2014a, Preliminary review of ICCAT bluefin tuna conventional tagging database. Collect. Vol. Sci. Pap. ICCAT, 70(2): 299-320.
- Justel-Rubio A., Ortiz M., Palma C., Gallego J.L., Di Natale A., Idrissi M., 2014b. Preliminary Evaluation of the Total Catch Removals Eastern Bluefin tuna. A comparison of the GBYP and ICCAT Task I databases. Collect. Vol. Sci. Pap. ICCAT, 70(2): 518-536.
- Mielgo Bregazzi R., 2015, Weight/Size structure of Atlantic Bluefin Tuna fished and/or ranched in the Mediterranean and Northeast Atlantic during the period 1995 to 2014 as revealed by trade, market & corporate biometric data. Collect. Vol. Sci. Pap. ICCAT, 71. *In this Volume*.
- Örenc A.F., Ünver M., Düzcü L., Di Natale A., 2014. Tentative bluefin tuna data recovery from the Ottoman Archives, the Maritime Museum Archives and the Archives of the Istanbul Municipality. Collect. Vol. Sci. Pap. ICCAT, 70(2): 447-458.

Table 1. Total data recovered by GBYP in Phase 1, Phase 2, Phase 3 and the first part of Phase 4. The additional trap data provided in Phase 4 are still to be checked and were not included.

TOTAL PHASES 1 to 4	origin	1st Total	Total data
# Records	OG	87,761	509,620
	TP	30,923	
	TAMD	311,415	
	FARM	49,354	
	GEN	172	
	DTBV	29,995	
BFT (no.)	OG	34,753	26,377,340
	TP	23,247,666	
	TAMD	825,485	
	FARM	49,354	
	GEN	172	
	DTBV	2,219,910	
BFT (t)	OG	114,596	1,191,312
	TP	744,227	
	TAMD	80,408	
	FARM	474	
	DTBV	251,607	
	#Fish sampled	OG	94,932
TP		7,610	
TAMD		825,485	
FARM		49,354	
GEN		152	
DTBV		2,219,910	
Legenda: OG = Other Gears; TP = Trap; TAMD = Trade, Auction and Market Data; FARM = Farmed tunas; GEN= Genetic; DTBV = Data To Be Validated			
Note: TAMD data include 29,995 records, 2,219,910 bft (no.) and 251,607 t to be further checked and validated.			

Table 2. Total data recovered by GBYP in Phase 1, Phase 2, Phase 3 and the first part of Phase 4 by century (<1500-1900) and by decade (1900 onwards) (TP = Traps; OG = Other gears; TAMD* = Trade, Auction and Market data, provisional; FARM = data provided by farms; GEN = Historical genetic samples; DTBV = trade, auction and marked data to be further validated; TBA = validated data to be attributed by year at the moment of their incorporation in the ICCAT BFT data base).

	year	<1500	1500	1600	1700	1800	1900	1910	1920	1930	1940	1950	1960	1970	1980	1990	2000	2010	TBA	DTBV
# Records	OG						9	10	87	11.509	15.616	29.992	17.946	6.201	1.781	1.174	3.210	236		
	TP		252	171	211	6.100	3.005	4.353	6.705	2.301	1.021	1.040	2.032	184	777	1.221	1.548		3	3
	TAMD*																		311.415	
	FARM															851	18.492	30.021		
	HGEN	20							60	60		2			30					
	DTBV																			
BFT (no.)	OG													107	70	9.937	21.559	3.080		
	TP		3.978.087	1.292.782	425.335	4.472.749	1.613.889	1.883.967	2.971.129	2.013.583	1.787.209	1.566.956	614.611	51.510	178.743	204.806	186.199		6.111	6.111
	TAMD*																		825.485	
	FARM															851	18.492	30.021		
	HGEN	20							60	60		2			30					
	DTBV																			
BFT (t)	OG						44	163	601	2.497	6.057	29.059	14.842	24.461	17.880	17.086	1.704	203		
	TP					141.907	40.327	70.723	75.579	83.592	86.204	111.417	71.842	11.981	8.755	19.568	15.306	711		
	TAMD*																		80.408	
	FARM															207	268			
	DTBV																			
#Fish sampled	OG											18.614	18.548	9.053	804	18.569	28.000	1.344		
	TP							153	170								2.225	5.062		
	TAMD*																		825.485	
	FARM															851	18.492	30.021		
	HGEN	20						60	60		2				10					
	DTBV																			

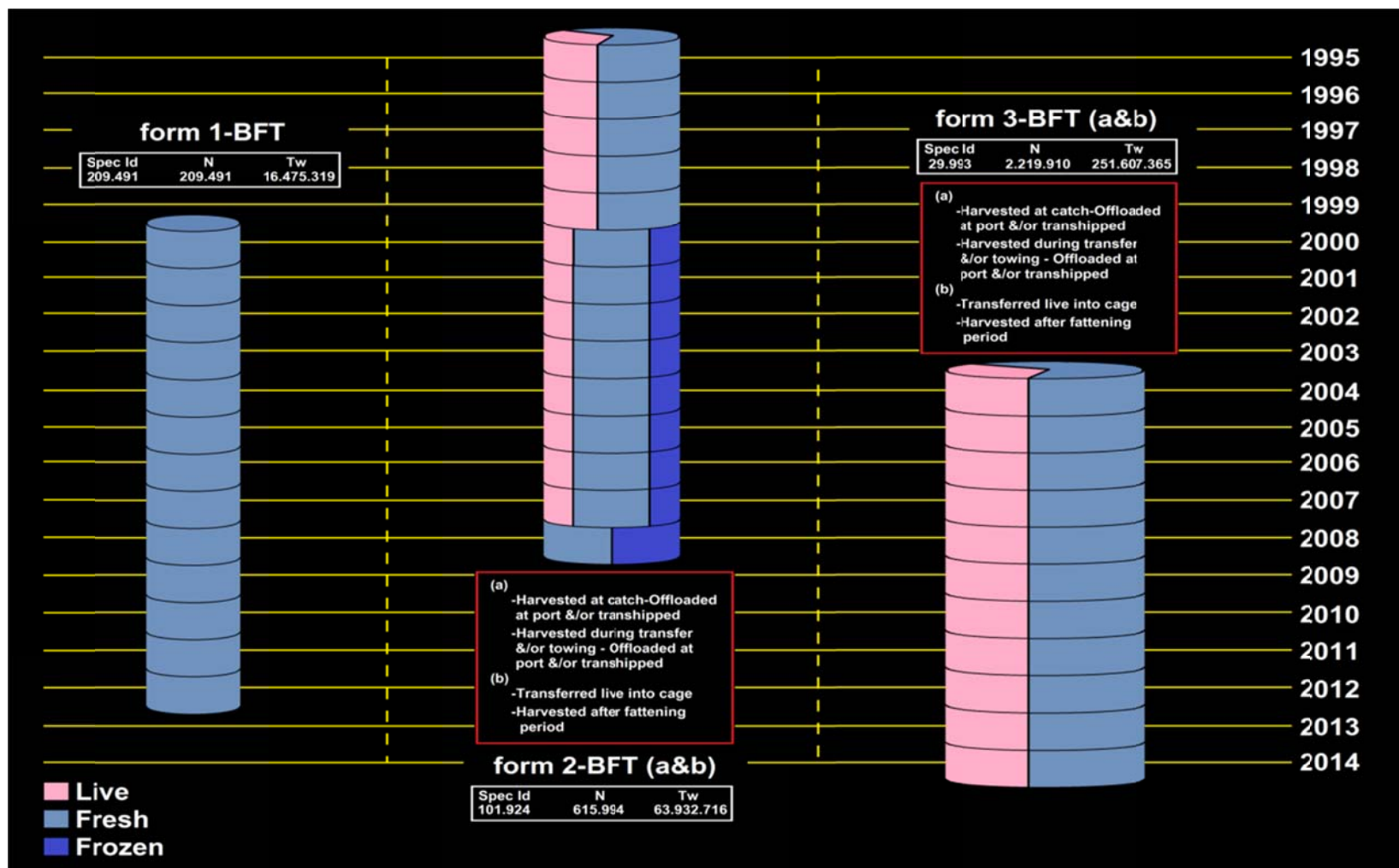


Figure 1. Chronology and structure of trade, market and auction data for the period 1995 to 2014, as they have been reviewed and validated. Form1 and form2 were have been fully validated by the external expert, while the data included in form3 need further checks and validation analyses.