Open biological and ecological data for tuna and tuna-like species: The time has come!

Emmanuel Chassot^{1,2,*}, Julien Barde³, Laurent Floch⁴, Laurent Ibanez², Rodney Govinden², Nathalie Bodin^{1,2}

¹ Institut de Recherche pour le Développement, SFA, BP570, Victoria, Seychelles

² Seychelles Fishing Authority, PO BOX 449, Victoria, Seychelles

³ Institut de Recherche pour le Développement, Indian Ocean Commission, Ebène, Mauritius

⁴ Institut de Recherche pour le Développement, Sète, France

*Corresponding author: <u>emmanuel.chassot@ird.fr</u>

Abstract

We describe the methodological approach developed to collect and manage a large range of biological and ecological data for large pelagic species sampled throughout several routine monitoring and research programs conducted in the Indian Ocean in the last decades. To cover the large habitats of occurrence of tuna and tuna-like species, collaboration with fishermen and processing factories was instrumental in fish sampling. We propose a minimum set of standard variables to be collected to describe and keep track of the fishing environment. In industrial fisheries, information from logbooks and traceability of tuna and tuna-like species through the storage process is key to fully represent the spatio-temporal uncertainty associated with the capture of each fish. Morphometric measurements and data retrieved from specific analysis are hosted in a relational database in a way that allows for any type of measure to be added without modifying the database design and structure. Standard code lists and metadata are used to describe the data managed in the database, which facilitates their visibility and diffusion. We argue that the use of standard, generic, well described sampling and analytical protocols combined with standard code lists is key for sharing and merging data sets at the scales of interest of tuna and tuna-like species. The comprehensive description of the methods used to collect and produce biological and ecological data requires interactions with computer scientists to make the most of data that are costly and mostly financed by public funds.

Keywords: metadata, open data, tuna fisheries, Indian Ocean

1-Introduction

Ecological and biological data collected on marine populations are managed by the different administrative and scientific institutes, Universities, and Non-Governmental Organizations (NGOs) that conduct the monitoring and research programmes. The cost of the data includes the logistics, materials, and human resources employed during the design of the studies, the acquisition of funding, the collection of the samples, their preparation and analysis with methods that can be very sophisticated in some cases (e.g. nuclear magnetic resonance, mass spectroscopy), and ultimately the management of the data and remaining samples when available. Data acquisition is therefore a long process with culminates with the publication of scientific results in reports and peer-review articles and can promote careers and support the recognition of an individual or research group. For this reason, ecological and biological data are difficult to access and their sharing is generally restricted to short-term collaborations within a specific project. Even when not used, data are generally not released or shared as they are expected to be used in the future. Another key limiting factor for data release is the cost in time associated with the description of the data necessary for the (potential) user and the lack of knowledge on metadata methods. Consequently, it is not only difficult to access published data in fisheries ecology but having a good knowledge of the sets of ecological and biological data existing in a given marine region is currently intractable.

Sharing ecological and biological data is however key to promote the broad, integrative analyses that are required for ecosystem-based fisheries management. It may also be used to help predict the effects of global changes on large pelagic populations, and more globally on marine ecosystems considering tuna and tuna-like species as biodincators. The development of standards for metadata, data formats and access protocols has recently enhanced interoperability functions in information systems, supporting open access to ecological and biological data (Reichman et al. 2011). This is particularly relevant for large pelagic populations that cover areas of millions square kilometres and are exploited by several fishing nations. A large range of data on biological and ecological traits of tuna and tuna-like species are collected everyday through routine monitoring programs and research projects globally. Data may be of numerical, continuous and categorical form, and include simple size/weight measurements of fish and its body components (e.g., soft and hard tissues), sex, descriptors of reproductive activity (maturity, fecundity), estimates of age and individual growth, and various more specific analyses of particular tissues tracing population structure, habitat use, diet habits and health condition. For instance, in fisheries monitoring, morphometric data are key to derive conversion equations to standardize size data, raise size samples to catch weight, and compute sampling coverage. In longline fisheries, conversions of length into weight has been shown to be critical for stock assessment as they can affect the information provided on gear selectivity (Langley et al. 2006, Minte-Vera et al. 2016). In purse seine fisheries, the species composition of the catch is generally derived from size sampling at landing after conversion of length into weight (Pallarés & Hallier 1997, Pianet 1999). Poor estimation of length-weight relationships can bias estimates of species composition, with potential effects on management (Marsac et al. 2017). The IOTC Secretariat manages a database of length, weight, and sex (IOTC 2005) which is however limited in spatio-temporal coverage and lacks visibility. Morphometric data are also essential for research purpose: combined with genetic and elemental analyses of hard/soft tissues, it helps monitoring changes in population structure and habitat (e.g. Chouvelon et al. 2017). Moreover, investigation of the trophic ecology and energy allocation strategy of fish through biochemical analyses (e.g., stable isotopes, lipids and fatty acids) or modelling gives insights on how species balance energy between growth and reproduction and how it may be affected by global changes may affect (Jusup et al. 2011, Zudaire et al. 2014, Sardenne, Bodin, et al. 2016).

In the present paper, we describe the methodological approach developed throughout several routine monitoring and research programs conducted on tuna in the Indian Ocean, including some data from the Indian Ocean Tuna Tagging Program (IOTTP), to showcase how biological and ecological data for tuna and tuna-like species can become open-access. The overarching objective of the study is to

promote discussions on the opportunities and challenges raised by data openness. First, we describe the main features of the fish sampling operations and how collaboration with fishermen is instrumental in retrieving information on the origin of each fish. We argue that confidentiality issues can easily be addressed and solved through removal of individual names of vessels or companies. Second, we provide the general principles we adopted to describe the different steps of the process conducting from fish collection to final data acquisition in a database built for hosting all biological and ecological data collected. We emphasize the need to use open, standard protocols for data collection and analysis, and describe each step of the process with metadata to promote full understanding and transparency of the data and foster collaboration through data sharing. Our approach allows for full traceability of the samples throughout their lifetime, including storage methods and location, to help in data interpretation in case of apparently aberrant data records, and support the development of a bank of samples for long-term comparative studies on large pelagic populations.

2- Fish sampling

More than 45,000 tuna and tuna-like species were sampled throughout several projects conducted in the Indian Ocean during 1987-2015 from eight species: albacore tuna (Thunnus alalunga; ALB), bigeye tuna (Thunnus obesus; BET), blue marlin (Makaira nigricans; BUM), frigate tuna (Auxis thazard; FRI), sailfish (Istiophorus platypterus; SFA), skipjack tuna (Katsuwonus pelamis; SKJ), swordfish (Xiphias gladius, SWO), and yellowfin tuna (Thunnus albacares; YFT) (Table 1). Species are defined following the standard ASFIS classification (Garibaldi & Busilacchi 2002). Sampling operations took place either at sea by observers onboard fishing vessels, or at land during unloading of the vessels at landing sites or at processing plants in Seychelles, Mauritius, South Africa, La Réunion, Thailand, Sri Lanka and Kenya. The sampling platform (i.e., field, laboratory or factory), fish status (i.e. fresh or frozen) and conditions of storage before processing (mode and duration) of the sampled fish were recorded as it may affect the morphometric measurements and the nature of the analyses of biochemical tracers that can be conducted thereafter (Bodin et al. 2014).

Year	ALB	BET	BUM	FRI	SFA	SKJ	SWO	YFT
1987	0	23	0	0	0	0	0	615
1988	0	240	0	0	0	0	0	707
1989	0	68	0	0	0	0	0	481
1990	0	625	0	0	0	0	0	899
1991	0	5	0	0	0	0	0	32
2003	0	0	0	0	0	2	0	316
2004	0	0	0	0	0	41	56	1199
2005	0	0	0	0	0	0	0	2413
2006	0	45	0	0	0	5	0	4542
2007	0	185	0	0	0	72	36	2299
2008	0	69	0	0	0	4	23	2986
2009	0	6	0	0	1	782	413	4286
2010	1	7	0	0	0	582	131	4611
2011	0	33	0	0	0	0	50	2895
2012	47	16	0	0	0	0	0	3653
2013	367	208	2	4	34	164	38	1851
2014	1437	430	0	0	0	983	55	1075
2015	337	558	3	0	0	857	37	1121

Table 1. Annual number of tuna and tuna-like species sampled in the Indian Ocean during 1987-2015

Information on the fishing environment was retrieved thanks to the collaboration of fishermen and staff at processing factories (Table 2). Access to skipper logbooks and well plans that describe the location of the fish storage in purse seiners, as well as information on the traceability of the fish processed at processing factories enabled linking fish sampled with fishing operations (Fig. 1).



Fig. 1: Spatial distribution of tuna and tuna-like fish sampled during 1987-2015. Average position is used when there is some uncertainty in

Uncertainty in the location arose when fish were sampled from purse seiners brine-freezing wells that can contain fish from multiple sets and at factories where only information tracing back to the fishing trip is generally recorded. In such cases, the range of fishing dates was recorded and information on spatial uncertainty was computed as the surface area (km²) inside the convex hull of the positions of the potential fishing sets (**Fig. 2**). In some cases, positions of fishing operations were not available and spatial uncertainty was computed as the surface area (km²) of the polygon linking the extreme fishing positions of the vessel during the trip collected from fishermen.



Fig. 2: Example of uncertainty associated with the catch of some fish sampled from brine-freezing wells of purse seiners and at the canning

For each fish, the fishing gear of the commercial and recreational fisheries that captured the fish was recorded as it provides qualitative information on fishing depth and size/age selectivity based on the FAO standard classification of fishing gears (FAO 2016). In addition, the type of school association (i.e. free-swimming school or school associated with a floating object) was recorded as it affects habitat use and trophic interactions (e.g. Jaquemet et al. 2011). No standard classification for the type of school association exists among tuna RFMOs yet but some initiatives have recently started to propose a standard classification (Gaertner et al. 2016).

Variable	Description
ocean_code	Acronym of the ocean where the fish was caught
gear_code	The gear that caught the fish
vessel_name	Name of the fishing vessel that caught the fish
vessel_code	Code of the fishing vessel that caught the fish
landing_date	Date of arrival of the fishing vessel in the port
landing_site	Place of arrival of the fishing vessel in the port
fishing_date	Date the fish was caught
fishing_date_min	Minimum date of fishing of the fish when the accurate date is not available
fishing_date_max	Maximum date of fishing of the fish when the accurate data is not available
activity_number	Sequential number of activity in the day
fishing_time	Time of fishing operation
latitude_deg_dec	Latitude of the location of capture
latitude_deg_dec_min	Minimum latitude of the location of capture when the accurate position is not available
latitude_deg_dec_max	Maximum latitude of the location of capture when the accurate position is not available
longitude_deg_dec	Longitude of the location of capture
longitude_deg_dec_min	Minimum longitude of the location of capture when the accurate position is not available
longitude_deg_dec_max	Maximum longitude of the location of capture when the accurate position is not available
school_type	Nature of the association of the fish when caught (DFAD: Drifting fish aggregating device;
	FSC: Free swimming school; MIX: Mix of log-associated and free-swimming school; AFAD:
	Anchored fish aggregating device)
well_number	The number of the freezing-well in the fishing vessel
well_position	The position of the tank in the fishing vessel (portside or starboard)
vessel_storage_mode	The mode of conservation of fish in the vessel (brine, freeze, chill)

Table 2. List of variables used to describe the fishing environment of the tuna and tuna-like species sampled in the Indian Ocean during 1987-2015.

For each fish, different morphometric measurements were taken on the body and its soft and hard tissues components (**Table 2**). The information is stored in the database (see **section 4**) in a table with a column 'measure_name' and a column 'measure_value' for each fish identifier, which provides allows any type of new measurement to be added to the database. Each measurement type is described in a table of metadata. Information on the tool used for measurement and associated precision is described. In addition, the sex (M: Male; F: Female; Ind: Indeterminate) and macroscopic maturity stage for females were derived from visual exam. Although the macroscopic maturity stage does not accurately describe the spawning oocyte development stage in tunas (Albaret 1977), it provides qualitative information on the spawning activity of the fish (i.e. developing vs. mature) and it was found useful to identify gonads to be further analysed for fecundity and/or biochemical analyses.

Notation	Definition	Description	Precision	Tool
FL	Fork length	Projected straight distance from the tip of the upper jaw (snout) to the shortest caudal ray (fork)	0.1 cm	Calliper
EFL	Eye fork length	Projected straight distance from the caudal margin of orbit to the fork of the tail	0.1 cm	Calliper
LJFL	Lower jaw fork length	Projected straight distance from the tip of the lower jaw to the shortest caudal ray (fork)	0.1 cm	Calliper
LD1	Predorsal length	Projected straight distance from the tip of the snout to the anterior base of the first dorsal fin	0.1 cm	Calliper
CFL	Curved fork length	Projected curved-body distance from the tip of the upper jaw (snout) to the shortest caudal ray (fork)	0.1 cm	Tape measure
CLJFL	Curved lower jaw fork length	Projected curved-body distance from the tip of the lower jaw to the shortest caudal ray (fork of the caudal fin)	0.1 cm	Tape measure
TG1	First thorax girth	Projected curved-thorax length just behind the first pectoral fins, i.e. where the fish height is greatest	0.1 cm	Tape measure
TG2	Second thorax girth	Projected curved-thorax length in front of the second dorsal fin and the anal fin	0.1 cm	Tape measure
W _T	Total fish weight	Weight of the whole fish	0.1 kg	Scale
W _{TG}	Gutted fish weight	Weight of the fish from which the gut (internal organs of the abdominal cavity) has been removed	0.1 kg	Scale
W _{TGG}	Gilled gutted fish weight	Weight of the fish from which (i) the gills have been removed, and (ii) the gut has been removed (whether or not the operculae or gill covers, dorsal, pelvic or anal fins have been removed)	0.1 kg	Scale
WTHGFT	Headed gutted finned tailed fish weight	Weight of the fish from which the gut has been removed and (i) the head and that portion of the body immediately forward of the anterior insertions of both pectoral fins have been removed, (ii) the gills removed, (iii) the tail removed at a point at or behind the caudal notch, (iv) the pectoral, dorsal, and anal fins have been removed	0.1 kg	Scale
W _G	Gonad weight	Weight of the two gonads	0.1 g	Scale
WL	Liver weight	Weight of the liver	0.1 g	Scale
\mathbf{W}_{SF}	Full stomach weight	Weight of the full stomach (stomach tissue and contents)	0.1 g	Scale
W _{SE}	Empty stomach weight	Weight of the stomach after removing the contents	0.1 g	Scale
W _{VR}	Rest of viscera weight	Weight of all viscera (i.e., heart, oesophagus, pylorus, intestine, pyloric caeca and anus) without liver and stomach	0.1 g	Scale

Table 2. Description of morphometric measurements taken on tuna and tuna-like species

3- Sample analysis

Hard and soft tissues from different organs of the tunas and billfishes were sampled according to the sampling protocols in use within each project. The conditions of sampling (i.e. fish state and quality, sampling tools, storage conditions, etc.) determine the subsequent analyses that can be carried out. For instance, microchemistry of otoliths or estimation of contaminant concentrations in fish tissues require specific tools and methods of sample handling that will not affect the chemical composition. For this reason, it is essential to (i) develop generic sampling protocols that will facilitate further studies not planned at the time of the project, (ii) comprehensively describe and make available the protocols used for data collection (e.g. <u>https://www.protocols.io/</u>), and (iii) use lab procedures that are internationally recognized (e.g. ISO/IEC 17025).



Fig. 3: Scheme illustrating the fish and tissue sampling procedure, the types of analysis conducted on the samples collected and their general purpose

We collected samples of liver, gonad, red muscle, white muscle as well as otoliths from fresh and defrosted fish. Analyses of lipid classes and fatty acids profiles, concentration of proteins and organic and metallic contaminants, and stable isotopes ratios of nitrogen and carbon were carried out and used for different studies on the trophic ecology, habitat use and fish quality of tuna and tuna-like species in the Indian Ocean (e.g. Zudaire et al. 2014, 2015, Sardenne, Bodin, et al. 2016, Grande et al. 2016, Sardenne, Chassot, et al. 2016, Bodin et al. 2017).

4. Data management and dissemination

4.1. Data management

We developed a PostgreSQL/PostGIS database to host all biological and ecological data collected throughout the research projects (**Fig. 4**). Information on all fields is described with a comprehensive data dictionary and a set of reference tables providing code lists for the fishing environment, the sample collection (tissue, sample position, operator, etc.), the descriptors of reproductive activity (atresia, post-ovulatory follicles, micro- and macro maturity stage), and the sample preparation (drying mode, extraction mode, packaging, analysis mode, storage mode, analysis lab, certified material). The database is conceived to host any type of morphometric measurement, biochemical tracer and biological descriptor. Values and generic information for each analysis (e.g. type of tracer, laboratory, operator, mode, date, mass, mass unit) are stored in linked tables. Additional tables enable to add specific features for each type of analysis (e.g. fraction type for fatty acids).

We used a free version of the Extract-Transform-Load software Talend to manage the data flow. The permanent acquisition of data from different projects, including new sampling of fish as well as results progressively obtained from analytical labs required to implement a data flow facilitating the reporting of the data with tabular files (.xls), their formatting with the open-source programming language R in generic formats, and their loading into a PostgreSQL/PostGIS database.



Fig. 4. Simplified structure of the database

4.2 Data dissemination

Ongoing work includes the description of the main features of the data sets (e.g. context of development, lineage, keywords, contacts, etc.) with the ISO norms and standards for environmental data (e.g. Dublin Core; <u>http://www.dublincore.org</u>). The next step will be to publish the metadata online through the open-source Geonetwork catalogue application (<u>http://www.geonetwork-opensource.org</u>). The publication of metadata online will provide better visibility to the data and is expected to foster collaborations with other tuna Commissions and research institutes. In a second step, we will develop the extraction from the data sources and preparation of the data and publish them through open data repositories such as Zenodo (<u>https://zenodo.org/</u>) and pangaea (<u>https://www.pangaea.de/</u>) which provide a unique DOI to each data set and a minimal set of metadata elements.

A final step for data dissemination will consist in connecting the data sets to the GBIF with the Integrated Published Toolkit (IPT) (Robertson et al. 2014). The IPT relies on Darwin Core and Ecological Metadata Language and enables to publish biodiversity data sets online and facilitate their merging with other data sets. By complying with these standards for both metadata and data, we will support interoperability with regional (WIOMSA, SAIAB, AfrOBIS, MadBIF) and global initiatives (GBIF, IOBIS, Encylclopedia of Life and IPBES: <u>http://www.ipbes.net/</u>).

Acknowledgements. We thank the many fishermen, observers, scientists, managers and staff at processing factories, and students who assisted with tuna and billfish collection and sample processing over the last decades. This work benefited from funds from the Seychelles Fishing Authority (SFA) and the French National Research Agency (ANR-11-JSV7-007-01; EMOTION research project).

References

- Albaret J-J (1977) La reproduction de l'albacore (*Thunnus albacares*) dans le Golfe de Guinée. Cah ORSTOM Sér Océan 15:389–419
- Bodin N, Lesperance D, Albert R, Hollanda S, Michaud P, Degroote M, Churlaud C, Bustamante P (2017) Trace elements in oceanic pelagic communities in the western Indian Ocean. Chemosphere 174:354–362
- Bodin N, Lucas V, Dewals P, Adeline M, Esparon J, Chassot E (2014) Effect of brine immersion freezing on the determination of ecological tracers in fish. Eur Food Res Technol 238:1057–1062
- Chouvelon T, Brach-Papa C, Auger D, Bodin N, Bruzac S, Crochet S, Degroote M, Hollanda SJ, Hubert C, Knoery J, Munschy C, Puech A, Rozuel E, Thomas B, West W, Bourjea J, Nikolic N (2017) Chemical contaminants (trace metals, persistent organic pollutants) in albacore tuna from western Indian and south-eastern Atlantic Oceans: Trophic influence and potential as tracers of populations. Sci Total Environ 596–597:481–495
- FAO (2016) Report of the twenty-fifth session of the Coordinating Working Party on Fishery Statistics. FAO, Rome, Italy
- Gaertner D, Ariz J, Bez N, Clermidy S, Moreno G, Murua H, Soto M, Marsac F (2016) Results achieved within the framework of the EU research project: Catch, Effort, and eCOsystem impacts of FAD-fishing (CECOFAD). In: IOTC Proceedings. Victoria, Seychelles, 11-13 Novembre 2016, p 32p
- Garibaldi L, Busilacchi S (2002) ASFIS List of Species for fishery statistics purposes. FAO, Roma, Italy
- Grande M, Murua H, Zudaire I, Arsenault-Pernet EJ, Pernet F, Bodin N (2016) Energy allocation strategy of skipjack tuna *Katsuwonus pelamis* during their reproductive cycle. J Fish Biol 89:2434–2448
- IOTC (2005) Biological data on tuna and tuna-like species gathered at the IOTC Secretariat: Status Report. In: IOTC Proceedings. IOTC, p 22p
- Jaquemet S, Potier M, Ménard F (2011) Do drifting and anchored Fish Aggregating Devices (FADs) similarly influence tuna feeding habits? A case study from the western Indian Ocean. Fish Res 107:283–290
- Jusup M, Klanjscek T, Matsuda H, Kooijman SALM (2011) A full lifecycle bioenergetic model for bluefin tuna. PLoS ONE 6:e21903
- Langley A, Okamoto H, Williams P, Miyabe N, Bigelow K (2006) A summary of the data available for the estimation of conversion factors (processed to whole fish weights) for yellowfin and bigeye tuna. In: Second regular session of the Scientific Committee of the Western-Central Pacific Fisheries Commission. WCPFC, p 21
- Marsac F, Baez J-C, Floch L, Fonteneau A (2017) Potential changes affecting species composition and tuna catch at size for purse seine fleets by using the new length-weight relationships for tropical tunas in the Indian Ocean. In: IOTC Proceedings. Victoria, Seychelles, 26-28 November 2017, p 11p

- Minte-Vera CV, Aires-da-Silva A, Satoh K, Maunder MN (2016) Changes in longline size-frequency data and their effects on the stock assessment models for yellowfin and bigeye tunas. In: Seventh meeting of the Scientific Advisory Committee of the Inter-American Tropical Tuna Commission, 09-13 May 2016, La Jolla, California, USA. IATTC, La Jolla, California (USA), p 22
- Pallarés P, Hallier J-P (1997) Analyse du schéma d'échantillonnage multispécifique des thonidés tropicaux. IEO/ORSTOM
- Pianet R (1999) Evolution du système de collecte et de traitement des données de la pêche thonière des senneurs européens et assimilés de 1981 à 1998. In: IOTC Proceedings.p 74–96
- Reichman OJ, Jones MB, Schildhauer MP (2011) Challenges and Opportunities of Open Data in Ecology. Science 331:703–705
- Robertson T, Döring M, Guralnick R, Bloom D, Wieczorek J, Braak K, Otegui J, Russell L, Desmet P (2014) The GBIF Integrated Publishing Toolkit: Facilitating the efficient publishing of biodiversity data on the internet. PLOS ONE 9:e102623
- Sardenne F, Bodin N, Chassot E, Amiel A, Fouché E, Degroote M, Hollanda S, Pethybridge H, Lebreton B, Guillou G, Ménard F (2016) Trophic niches of sympatric tropical tuna in the Western Indian Ocean inferred by stable isotopes and neutral fatty acids. Prog Oceanogr 146:75–88
- Sardenne F, Chassot E, Fouché E, Ménard F, Lucas V, Bodin N (2016) Are condition factors powerful proxies of energy content in wild tropical tunas? Ecol Indic 71:467–476
- Zudaire I, Murua H, Grande M, Go\~ni N, Potier M, Ménard F, Chassot E, Bodin N (2015) Variations in the diet and stable isotope ratios during the ovarian development of female yellowfin tuna (*Thunnus albacares*) in the Western Indian Ocean. Mar Biol 162:2363–2377
- Zudaire I, Murua H, Grande M, Pernet F, Bodin N (2014) Accumulation and mobilization of lipids in relation to reproduction of yellowfin tuna (*Thunnus albacares*) in the Western Indian Ocean. Fish Res 160:50–59