DATA TOOLBOX FOR FISHERIES: THE CASE OF TUNA FISHERIES

Paul Taconet[1], Emmanuel Chassot[2], Jérôme Guitton[3], Carlos Palma[4], Fabio Fiorellato[5], Enrico Anello[6], Julien Barde[1]

SUMMARY

Assessing the status of tuna and tuna-like populations for providing management advice requires the analysis of multiple data sets collected by the contracting parties and cooperating non-contracting parties of Tuna Regional Fisheries Management Organizations (tRFMOs) Conventions. Data on the magnitude and composition of landings, discards, and fishing effort are currently managed at basin scale by the Secretariats of the tRFMOs. Consequently, data formats and reference codes have evolved rather independently despite some links with the FAO Coordinating Working Party on Fishery Statistics. We have developed a global harmonized database for tuna fisheries data by collating the public domain datasets (total catch, monthly-spatially aggregated catch and effort, and catch at size) from IOTC, ICCAT, IATTC and WCPFC. The database – named SARDARA - currently covers the period 1919-2014 and is freely accessible online along with a set of open source codes (a « toolbox ») to handle the data, i.e. transform the data formats, load the standardized data into the database, and compute a suite of indicators (e.g. global maps of catch). The use of harmonized coding systems and standard nomenclatures is critical to simplify data exchange and dissemination, resulting in benefits for the scientific community and *in fine* for the conservation of healthy stocks. Our objective is to propose a core of services to format and exchange tuna fisheries data and indicators, and promote standards for metadata and data formats to facilitate the access to the data through web-based tools. Over the coming years, this toolbox will be enriched with additional contributions by the community of users through a collaborative web site. Among others, the expected benefits of the project are the promotion of communication towards tRFMOs and their member States as well as to the general public and the increase of transparency and accessibility to fisheries data sets, indicators, underlying codes, and related research in several fisheries domains from ecology to assessment.

KEYWORDS

Catch/effort, High seas fisheries, Fishery statistics

--

This project / work has received funding from the European Union's Horizon 2020 research and innovation programme under the BlueBRIDGE project (Grant agreement No 675680).

1 Institut de Recherche pour le Développement, UMR MARBEC, CRH, Avenue Jean Monnet, BP171, 34203 Sète cedex, FRANCE. Paul.Taconet@ird.fr ; Julien.Barde@ird.fr
2 IRD, Seychelles Fishing Authority, BP570, Victoria, SEYCHELLES. Emmanuel.Chassot@ird.fr
3 Agrocampus Ouest, UMR ESE 0985, 65 Rue de Saint-Brieuc, 35000 Rennes, FRANCE
4 ICCAT SECRETARIAT - Corazón de María, 8. 28002 Madrid, SPAIN
5 IOTC Secretariat, Le Chantier Mall (2nd floor), PO Box 1011, Victoria, SEYCHELLES. fabio.fiorellato@iotc.org
6 FAO, Viale delle Terme di Caracalla, 00153 Rome, ITALY

1. Introduction

Assessing the status of tuna and tuna-like populations for providing management advice requires the analysis of multiple data sets collected by the contracting parties and cooperating non-contracting parties (CPCs) of Tuna Regional Fisheries Management Organizations (tRFMOs) Conventions. In absence of fishery-independent data for most tuna fisheries over the world, stock assessment models mostly rely on commercial fisheries data that describe the magnitude and composition of landings, discards, and fishing effort. Such data are collected and processed by the CPCs through logbooks, landings and size-frequency samples and provided to the tRFMOs Secretariats following the rules (i.e. nature, formats, deadline) defined by the Conservation and Management Measures and Resolutions in force with each Commission. Datasets are stored and managed by the Secretariats to provide a holistic view of the tuna and tuna-like exploited populations and fisheries, and prepare the datasets for scientific analyses, including stock assessments. Hence, data formats and reference codes are currently managed at basin scale by each Secretariat. Consequently, they have evolved independently over time despite some common backgrounds and links with the Coordinating Working Party on Fishery Statistics of the Food and Agriculture Organization (FAO).

Annual time series of catch by species, gear, and flag collated by tRFMOs contribute to the global statistics compiled by FAO and appear as such in the FAO yearbook and bi-annual report entitled ‘The State of World Fisheries and Aquaculture’ (SOFIA). Tuna data are accessible through the Fishtat software and a dedicated web portal (http://www.fao.org/fishery/statistics/tuna-catches/query/en), this latter covering the period 1950-2010. In addition, some past projects (e.g. the FAO tuna atlas7 and the Agrocampus-IRD atlas8) have focused on the monthly spatially-aggregated data collated by tRFMOs to describe the extent of global tuna fisheries and provide an overview of the world tuna fisheries.

Building upon these projects, we have developed a global harmonized database for tuna fisheries data by collating the public domain datasets available from the International Commission for the Conservation of Atlantic Tunas (ICCAT), the Indian Ocean Tuna Commission (IOTC), the Inter-American Tropical Tuna Commission (IATTC) and the Western-Central Pacific Fisheries Commission (WCPFC). Our overarching objectives are to: (i) review the tuna datasets available from each tRFMO and propose coding systems and standard nomenclatures to facilitate their merging for analysis, (ii) give more visibility to the data and more transparency to the processing steps driving to the datasets used as inputs for assessment models, and (iii) provide tools to facilitate data extraction and visualization to anyone interested in tuna fisheries. In the present article, we describe the general methodological approach used for the collation and formatting of the data through the development of a database that hosts the public-domain data and a set of open source codes (a « toolbox ») to handle the data, i.e. convert the formats, load the standardized data into the database, generate global harmonized datasets, and compute a suite of indicators. We provide a few examples to illustrate the interest of the approach for comparing the data available between tRFMOs and describing the temporal evolution of tuna fisheries across oceans. In addition, to promote comparative analyses between oceans and stocks, we argue that the outputs of the project (i.e. database and toolbox) would be useful to foster participation and involvement of researchers from coastal countries to scientific analyses through enhanced understanding of the use of their data, as well as to improve communication of science to policy makers and to the general public.

2. Materials & Methods

2.1 Data sources

The most recent public domain datasets were collated from the Secretariats of IOTC, ICCAT, IATTC and WCPFC and loaded into the database. The datasets include total catches, spatially-aggregated catches and efforts, and catch-at-size derived from size-frequency data. Work is ongoing to include datasets from the Commission for the Conservation of Southern Bluefin Tuna. The temporal extent of data currently spans the period 1919-2014. Data were downloaded from the tRFMOs websites or provided by the Secretariats. In this section, we will address a quick overview of the data, i.e. their nature, meaning, and the main similarities and differences among tRFMOs.

7 http://www.fao.org/figis/geoserver/tunaatlas/
8 http://halieut.agrocampus-ouest.fr/sirs_sardara/

2.1.1 Total catches

The total catches are the reported catches by CPC, fishing gear, species and year. They shall represent 100% of the catches made by each country within the convention area of each tRFMO. These data are submitted on an annual basis to the tRFMOs by each CPC. They cover the tuna and tuna-like species and sometimes non-target species. Each catch value in the total catches dataset is linked to the following dimensions: (1) the time frame (one year), (2) the country or contracting party (flag), (3) the gear, and (4) the species or group of species in some cases. Depending on the tRFMO, the spatial resolution associated to a catch value can be the whole area of competence of the tRFMO (i.e. IATTC and WCPFC), the FAO major fishing area (i.e. IOTC), and ocean-specific sampling areas (i.e. ICCAT). It should be noted that the resolution of these dimensions may vary a lot across tRFMOs, e.g. a total of 122 species is reported in the total catches of the IOTC while the WCPFC database only includes 8 species. It is also noteworthy that total catches should include fish discarded-at-sea but it is rarely the case and catches actually represent landings in most cases.

2.1.2 Catch-and-effort

Catch-and-effort data are data aggregated over spatio-temporal strata that are collected by the CPCs or the tRFMOs in some cases. Generally, catch-and-effort data are defined over one month time period and 1° or 5° size square spatial resolution. Following ICCAT, catch and fishing effort statistics are defined as "the complete species (tuna, tuna like species and sharks) catch composition (in weight <kg> or/and in number of fish) obtained by a given amount of effort (absolute value) in a given stratification or detail level (stratum). T2CE are basically data obtained from sampling a portion of the individual fishing operations of a given fishery in a specified period of time." (ICCAT Task 2). Hence, geo-referenced catch data and associated effort can represent only part of the total catches. Some tRFMOs such as the IOTC however indicate in the resolutions that the data should be extrapolated to the total national monthly catches for each gear and that documents describing the extrapolation procedures should be made available on a routine basis.

Overall, catch-and-effort data usually come with the following dimensions: (1) the time frame (generally one month), (2) the spatial resolution (generally 1° or 5° size square), (3) the country or contracting party (i.e. flag), (4) the species (only for the catches), (5) the fishing gear, (6) the fishing operation mode for purse seine fisheries (i.e. associated with a floating object or in free swimming school), (7) the unit for catch and effort (e.g. catch may be expressed in metric tons or in number of fishes).

However, catch-and-effort data differ from one tRFMO to another. The diverging confidentiality policies across the tRFMOs may result in missing dimensions (e.g. absence of flag for WCPFC), or voluntary missing data (e.g. when single vessels can be identified). The dissemination mode, formats and number of files of the source data also vary a lot across tRFMOs.

2.1.3 Catch-at-size

Catch-at-size (CAS) matrices are derived from size-frequency data which are collected by the CPCs or through dedicated research projects (e.g. Stobberup & Geehan (2015)) following specific guidelines to obtain unbiased and representative size samples of the population of interest. In the Indian Ocean for instance, Resolution 15/02 of the IOTC stipulates that "sampling coverage shall be set to at least one fish measured by ton landed, by species and type of fishery, with samples being representative of all the periods and areas fished". Size-frequency data are generally available on a 5° grid area by month, gear and fishing mode but the spatial resolution for longline fisheries data can be coarser (i.e. 10° grid). CAS describe the overall size structure of a fish stock by raising the size-frequency data to the total catches. They are generally produced by the Secretariats for assessment purposes and concern solely the principal market tunas (yellowfin, skipjack, bigeye, and albacore) as well as swordfish for which the quality and coverage of the input data (total catch, catch-and-effort and size frequency) is considered to be adequate enough. They might rely on strong assumptions (e.g. use of proxy fleets or different levels of fleet/gear aggregations) when size measurements are lacking as for most small-scale fisheries. The format and timespan of the catch-at-size matrices varies between tRFMOs and data are not always available with spatial information. CAS are available on a 5 by 5 degree square grid for the periods 1969-2014 for the Atlantic Ocean and 1950-2015 for the Indian Ocean. For the Pacific Ocean, CAS are not geo-referenced and work is ongoing to include them in the database.

2.2 Constraints and technical design of the database

2.2.1 The constraints of transparency and reproducibility

The database has been implemented with open-source software (PostgreSQL and PostGIS). Through a collaboration between FAO, IRD and technology partners in the context of the BlueBRIDGE project, it has been ported and is currently hosted on the iMarine platform and accessible online. At this stage, SARDARA hosts public domain data available from the Secretariats (i.e. some data may have been removed from the files). In the eventuality that confidential data would be made available and stored, some access-control mechanism should be agreed with the original data providers.

Bringing transparency and reproducibility to the data and the processes is a major goal of the project. In this context, the database that hosts the data has been built with the following objectives:

- Express the same data by using different code lists, i.e. original tRFMOs code lists and standard code lists (see section 2.3);
- Store various processed levels of the data (primary data as provided by the tRFMOs, raised data as transformed by scientists, etc.) (see section 2.4);
- Keep track of the transformations applied to the data (i.e. transparent and reproducible workflow);
- Archive all historical and updated datasets;
- Store data with any type of spatial and temporal resolution.

In a first step, data are converted from their original formats to the database format with a set R-scripts. Data are then loaded into the database with a second set of R-scripts – these data transformation and load scripts are part of the toolbox and are also accessible online. The primary data (or “raw” data) are therefore stored and available in SARDARA. Third, the data are mapped with standard coding systems and standard nomenclatures within the database. To cope with transparency and reproducibility constraints, all the code lists used (i.e. tRFMOs and standard FAO code lists), the mapping, additional data for data processing (e.g. to convert catches from number to weight) and the processes used to transform the data are stored inside SARDARA. This method enables: (i) to use the whole set of data services with the data at any processed level (from the primary data to the transformed data) and expressed either with raw code lists or standard ones, and (ii) to be fully transparent regarding the mappings and processes applied to the raw data. **Figure 1** summarizes the data flow and shows how transparency and reproducibility constraints on data and treatments have been set up technically in the project.

2.2.2 Database design

SARDARA is built as a data warehouse designed with a star schema. In other words, it is composed of some « fact » tables referencing some « dimension » tables. The fact tables are the total catches, the geo-referenced catches, the geo-referenced efforts and the catch-at-size. The dimension tables are the sets of information associated to each catch/effort/CAS numerical value: the species, the gear, the flag, the area, the time frame, etc. Each line on a fact table is therefore a combination of information coming from the dimension tables, plus a numerical value (e.g. the value of the catch). The database is divided into schemas, which can be seen as folders. Each schema can have several tables or views. In our database, each dimension is represented as a schema. For a given dimension, the corresponding schema has: (i) all the raw code lists coming from all the tRFMOs, (ii) the standard FAO code list if any, and (iii) a table of the mapping between the raw code lists and the standard code list.

The fact tables (catches, efforts, catch at size) are stored under the «fact_tables » schema. Under this schema are stored the primary datasets priorly formatted to cope with the database's format.

The additional data used for further processing of the data are stored under the schema « conversion_factors ». It consists in conversion factors (for catch units and effort units – see section 2.4).

Finally, the « tunaatlas » schema holds the processed datasets – see section 2.4 for the details of the available tables. The tables of the schema « tunaatlas » are *views* of the tables of the database, which mean that they are entirely built from these tables – they do not use data coming from outside the database. To be in line with the transparency and reproducibility constraints, the processed datasets are built as database views, enabling anyone to see the exact query that has built the dataset.

2.3 Reference codes and formatting

Coding systems and nomenclatures used to describe the data (e.g. gears, flags, species) may differ according to tRFMOs. With the collaboration of the Secretariats, the code lists for the gears (and gear groups), species (and species groups) and countries of the ICCAT, IOTC, IATTC, and WCPFC were collated, stored into the database and partially mapped to standard code lists. Most tRFMOs codes have been mapped with the standard FAO code list, i.e. 95% for the species, 99% for the flags and 84% for the gears. Ongoing work is conducted to finalize and validate the mapping. The standard FAO code lists used are the ones recommended by the Coordinating Working Party on Fishery Statistics (CWP):

- For the species: the ASFIS List of Species for Fishery Statistics Purposes;
- For the gears: The International Standard Statistical Classification of Fishing Gear (ISSCFG);
- For the flags/countries: The UN. Standard country or area codes for statistical use. Note that the overseas territories (e.g. Ile de la Réunion for France) are presented as countries in this code list, which is compliant with the tRFMOs data.

An IRD-defined species group code list oriented to tuna fisheries (with 12 groups) has also been set-up, and tRFMOs species mapped to it. It should be noted that no standard combination {gear ; effort unit} has been recommended by the CWP, even though some examples of meaningful combinations of gears and associated efforts have been released. The use of standard effort units is discussed in the processes applied to the geo-referenced efforts (section 2.4.3) and the discussion (section 4).

The code lists (both the tRFMO and the FAO ones) and the preliminary mappings are available as tables in the database (under each dimension's name). They are also available on this link: http://mdst-macroes.ird.fr/BlueBridge/Tuna_Atlas/sardara_code_lists_and_mappings.zip. The main difficulties encountered for the mapping are:

- The absence of correspondence between code lists, for some tRFMOs own-defined codes that usually are aggregation of existing codes (e.g. flag “IDPH” - Indonesia and Philippines – for WCPFC; species “Otun” - other tuna – for ICCAT). These codes have hence been mapped with more aggregated code lists – i.e. group of species, group of gear. Concretely, it means that querying the data using the standard code lists at the finest resolution (species, flag, gear) will return “not mappable” values for these codes; however, at a more aggregated resolution (group of species, group of gear) it will return the associated standard code;
- The lack of description of some codes, which results in the impossibility of determining the exact correspondence in the standard FAO code lists.

2.4 Processed data sets

Total catches datasets formats are very similar among the different tRFMOs and a simple merging of the datasets is sufficient to get the total catches at a global scale. By contrast and although geo-referenced datasets are overall similar between tRFMOs, they differ to some extent, especially regarding the dimensions that are included, the processes applied to the data prior to dissemination, the units of catch and effort used, the number of input datasets, and the confidentiality policies (see section 2.1). Therefore, the simple merging of these datasets is not enough to get a pertinent overview of tuna fisheries at a global scale: some processes are required to harmonize the datasets.

This section describes the processes applied by IRD to the primary (or « raw ») geo-referenced catch datasets so as to produce a global dataset of geo-referenced catch data (so-called ‘tuna atlas’ dataset). The resulting dataset of the processing workflow is a table containing one single value of catch per stratum (i.e. year, month, area, species, gear, flag, type of school), expressed in weight, and raised whenever possible to the total catches. The raising factor for each stratum is also given. The intermediate tables of the workflow are available in the database, along with the queries that built them, enabling anyone to trace-back the full flow. A major objective of our approach is to allow any user to propose their own processing method at any step of the workflow to produce their own datasets. Below is the comprehensive list of the specificities of the raw dataset, and the transformations and corrections applied by IRD to obtain the global tuna atlas dataset:

- The units used to express the catches may vary between tRFMOs datasets. Catches are expressed in weight, or in number of fishes, or in both weights and numbers in the same stratum. Values expressed in weight were kept and numbers were converted into weight using simple conversion matrices (A. Fonteneau, *pers. com*). These conversion factors depend on the species, the gear, the year and the main geographical area (equatorial or tropical). They were computed from the Japanese and Taiwanese size-frequency data as well as from the Japanese total catches and catch-and-effort data. Some data might not be converted at all because no conversion factor exists for the stratum: those data were not kept in the final dataset.

- Geo-referenced catches were raised to the total catches for all tRFMOs. Depending on the availability of the flag dimension (currently not available for the geo-referenced catch-and-effort dataset from the Western-Central Pacific Ocean), the dimensions used for the raising are either {Flag, Species, Year, Gear} or {Species, Year, Gear}. Some catches cannot be raised because the combination {Flag, Species, Year, Gear} (resp. {Species, Year, Gear}) does exist in the geo-referenced catches but the same combination does not exist in the total catches. In this case, non-raised catch data were kept.
- For confidentiality policies, information on flag and school type for the geo-referenced catches is available in separate files for the eastern Pacific Ocean. For each stratum, the catch from the flag-detailed dataset was raised to the catch from the school type-detailed dataset to get an estimation of the catches by flag and school type in each stratum

3. Results

We illustrate here the interest of the approach through 3 simple examples of fisheries indicators that can be directly available from the database. First, the merging of the 4 data sets of total catches provides an overview of the importance of tuna fisheries with regards to other fisheries and aquaculture production. Here, time series of tuna catch by gear group show that surrounding nets (i.e. purse seine) now predominate the global tuna fisheries with >3 million t caught in the recent years while about 1,3 million t were caught with hooks and lines (i.e. longline) and about 0,25 million t with gillnets and entangling nets (**Fig. 2**). Similar charts can be easily displayed by flag and species. In particular, global catch data by species have been used to assess the status of tuna and tuna-like species at global scale based on the IUCN criteria (Collette et al. 2011). Patterns in annual time series of catch by species can also be useful to detect collapses in fish stocks, including tunas (Mullon et al. 2005).

Second, comparison between the total catches and cumulated catches available from catch-and-effort data can be useful to monitor the overall quality of data available for stock assessments. In the Indian Ocean, the major decline in fishing effort from industrial fisheries in the late 2000s in relation with piracy threat resulted in a major decrease in the amount of geo-referenced catch data available to the Secretariat (**Fig. 3**). In the recent years, no information is available for about 70% of the total tuna catches in the Indian Ocean, which affects the assessments and general understanding of the stock dynamics and movements. Spatial information on the fisheries, linked to the occurrence of artisanal fleets, appears to be more available for the other oceans, particularly in the Pacific Ocean (**Fig. 3**). It is noteworthy that the availability of catch-and-effort data does not preclude that the data is of good quality. The IOTC Secretariat developed an estimate of the quality of the data for each record (Unknown, Poor, Fair, Good) and the nature of its estimate (Original, Raised and if so, to what extend and with what process).

Third, the data collated can provide information on the changes in the types of tuna school association targeted by purse seiners. The massive increase in the use of artificial drifting fish aggregating devices (FADs) since the 1990s worldwide has deeply modified the size and species composition of the purse seine catch. The use of FADs has raised some concerns among NGOs and tRFMOs about their impacts on the amounts of bycatch, including sensitive species such as some sharks and turtles. Compiling data from the different tRFMOs provides a global picture of the extent of FAD-fishing in the world oceans, with more than 50% of purse seine tuna catch coming from schools caught in association with floating objects (**Fig. 4**).

4. Discussion

4.1 Processes applied to primary data

Our current processing steps involved some ancillary data (e.g. mean weights for Japanese longliners) and assumptions (e.g. choice of spatial strata) for producing the global monthly spatially-aggregated dataset. In particular, mean weights currently used for the conversion from number to weights for Japanese longliners have been assumed stable for several years and such conversions should be improved in the future. The issue of size structure of the catch in longline fisheries is particularly tricky (Geehan & Hoyle 2013) and recent work has been conducted in the Eastern Pacific Ocean to better understand the inconsistency observed in some datasets (Satoh et al. 2016). In the Indian Ocean, ongoing work is currently conducted by the IOTC Secretariat to better define the whole process of selection of proxy fleets and strata resulting in the production of average weights required for conversions of numbers into weights. Similarly, the availability of different effort units for similar gears (e.g. searching days or numbers of trips for purse seine) requires some conversion factors in order to include all the data when producing a global dataset of effort. Collaboration between tRFMOS and the FAO through the CWP would be useful to define standard effort units for each gear and standard protocols for harmonizing and raising efforts. The issue of selection of effort unit (e.g. trip vs. day vs. set) is particularly important for raising observations of bycatch at the scale of a fishery. Again, our methodological approach provides the possibility to include different processing criteria which would result in different datasets that could be compared and eventually give insight into the uncertainty linked to the choice of a particular method. Finally, our results indicate that there are currently some inconsistencies between total catch and catch-and-effort datasets (e.g. missing fleet due to *ad hoc* referential) which can result in some catch-and-effort data that are not raised.

4.2 Global database and toolbox for fisheries science and management

Coordination and collaboration between TRFMOs, as initiated through the Kobe process, are essential for improving tuna science as they share common scientific and methodological issues, from the collection and management of data to the production of scientific advice. Indeed, several features of tuna fisheries are common to all oceans, e.g. the strong decline in catch rates with the initial expansion of Japanese longline fisheries from the 1950s (Hampton et al. 2005, Polacheck 2006), the apparition of deep longline in the mid-1970s which generally results in a break in the time series of CPUEs available for stock assessments, the major decrease in mean weight of tunas caught by purse seiners in relation with the use of artificial fish aggregating devices in purse-seine fishing from the mid-1990s. The spatially-explicit fisheries data collected through the TRFMOs revealed particularly useful to analyse the patterns of tuna distribution in the open-oceans (Worm et al. 2005, Worm & Tittensor 2011, Reygondeau et al. 2012), to untangle the effects of environment and fishing on tuna dynamics (Rouyer et al. 2008), to describe the extent of their spatial distribution (Chassot et al. 2010), and to appreciate changes in fishing strategies over time (Fonteneau et al. 2013). Such studies were however generally restricted to some oceans or fisheries (e.g. Japanese longline) or required substantial initial work to collect and harmonise the data prior to their analysis. By contrast, global analyses relying on shared datasets benefit from the information provided by the multiplicity and range of case studies (e.g. meta-analyses) and can provide more general and robust conclusions about ecological and economic drivers of fisheries (Worm et al. 2009, Pons et al. 2016). Our approach is consistent with the project initiated with the RAM Legacy database on fish stock assessments (Ricard et al. 2012) and the general trend observed in many scientific domains such as genetics or oceanography (Poloczanska et al. 2008).

4.3 On the benefits of using standard data format and code lists for tuna stock management

For the purposes of the project, the process of mappings between code lists have been managed by IRD with the collaboration of the tRFMOs. A great part of our work has consisted in handling the datasets and the code lists – gather, format, transform, process (taking into account all the specificities), map with standard code lists. The use of harmonized coding systems, standard nomenclatures and common data formats is critical to simplify data exchange and dissemination, resulting in benefits for the scientific community, the managers and *in fine* for the conservation of healthy stocks. As an example, the adoption of a common format would result in the ability of using the data services developed directly with the source data, thus benefiting from a work that has already been achieved. In order to be fully efficient, the management of the mappings between the tRFMOs code lists and the standard FAO ones should be done by the tRFMOs themselves. The mappings should then be made publicly available in their latest revision (as well as for the historical mappings) either as downloadable files or as results of a remote service execution. Our database might then query these services and reflect changes within its tables. Additionally, and for the same efficiency reasons, the common standard to adopt for the exchange of information should also be discussed by the five tRFMOs within a dedicated working group, ideally led by FAO or through the inter-Kobe process. Our work could serve as recommendations within the discussions.

5. Conclusions

For all the reasons stated above, the assistance of tRFMOs and FAO including through the CWP, is needed to address a number of issues that cannot be solved at a single organization level. The assistance of FAO is very important to set up standards (i.e. code lists) that are clearly defined and described. FAO's long experience in data collection coming from various sources and data dissemination can be of great benefit for the project. As addressed several times, one critical issue of the project is the coherence and the consistency of the primary data and formats distributed by the tRFMOs. It is therefore essential to strengthen the collaboration with the Secretariats to facilitate the flow of the data. In particular, the assistance of the tRFMOs is critical to cope with the following issues in their datasets: (i) metadata are not always complete, in particular with regards to confidentiality rules; (ii) there is a wide variety of data formats, distribution mode and number of files used. This results in complex and fragile transformation scripts and queries to build global and coherent data sets. On this point, IOTC is working at providing access to its data through well-documented remote services; (iii) Heterogeneity and lack of description of the code lists used, resulting in some unmatchable codes with the FAO's ones; (iv) There is lack of visibility and recognition of data ownership in datasets and reports.

Acknowledgments

We are grateful to all staff and personnel involved in collection and management of tuna fisheries data over the world over the last decades. We sincerely thank Nick Vogel and Carolina Minte-Vera from IATTC and Peter Williams and John Hampton from SPC for great help with the datasets from the Pacific. We are also grateful to our FAO colleagues, particularly Fabio Carocci and Anton Ellenbroek, for their work and support to the project. We finally thank Alain and Viveca Fonteneau for initiating the SARDARA project as early as the 1990s, Olivier Maury for continuing the work through the Pelagic Fisheries Research Program of the JIMAR, the French ANR project REMIGE, and the IRD Tuna Observatory organization. PT acknowledges financial support from the the French National Research Institute for Sustainable Development (IRD), the French Ministry of Fisheries and Aquaculture and the International Seafood Sustainability Foundation.

References

Chassot E, Kaplan DM, Ortiz de Zarate V, Romanov E, Fonteneau A (2010) Comment on “Clarification regarding the distribution of bigeye tuna Thunnus obesus in the Atlantic Ocean, including British waters.” J Fish Biol 77:1449–1454

Collette BB, Carpenter KE, Polidoro BA, Juan-Jordá MJ, Boustany A, Die DJ, Elfes C, Fox W, Graves J, Harrison LR, McManus R, Minte-Vera CV, Nelson R, Restrepo V, Schratwieser J, Sun C-L, Amorim A, Brick Peres M, Canales C, Cardenas G, Chang S-K, Chiang W-C, Oliveira Leite N de, Harwell H, Lessa R, Fredou FL, Oxenford HA, Serra R, Shao K-T, Sumaila R, Wang S-P, Watson R, Yáñez E (2011) High Value and Long Life—Double Jeopardy for Tunas and Billfishes. Science 333:291–292

Fonteneau A, Chassot E, Bodin N (2013) Global spatio-temporal patterns in tropical tuna purse seine fisheries on drifting fish aggregating devices (DFADs): Taking a historical perspective to inform current challenges. Aquat Living Resour 26:37–48

Geehan J, Hoyle S (2013) Review of length frequency data of the Taiwanese distant water longline fleet. In: IOTC Proceedings. IOTC, San Sebastian, Spain, 23-28 October 2013, p 30

Hampton J, Sibert JR, Kleiber P, Maunder MN, Harley SJ (2005) Decline of Pacific tuna populations exaggerated? Nature 434:E1–E2

Mullon C, Fréon P, Cury P (2005) The dynamics of collapse in world fisheries. Fish Fish 6:111–120

Polacheck T (2006) Tuna longline catch rates in the Indian Ocean: Did industrial fishing result in a 90% rapid decline in the absence of predatory species? Mar Policy 30:470–482

Poloczanska ES, Hobday AJ, Richardson AJ (2008) Global database is needed to support adaptation science. Nature 453:720–720

Pons M, Branch TA, Melnychuk MC, Jensen OP, Brodziak J, Fromentin JM, Harley SJ, Haynie AC, Kell LT, Maunder MN, Parma AM, Restrepo VR, Sharma R, Ahrens R, Hilborn R (2016) Effects of biological, economic and management factors on tuna and billfish stock status. Fish Fish:n/a-n/a

Reygondeau G, Maury O, Beaugrand G, Fromentin JM, Fonteneau A, Cury P (2012) Biogeography of tuna and billfish communities. J Biogeogr 39:114–129

Ricard D, Minto C, Jensen OP, Baum JK (2012) Examining the knowledge base and status of commercially exploited marine species with the RAM Legacy Stock Assessment Database. Fish Fish 13:380–398

Rouyer T, Fromentin JM, Ménard F, Cazelles B, Briand K, Pianet R, Planque B, Stenseth NC (2008) Complex interplays among population dynamics, environmental forcing, and exploitation in fisheries. Proc Natl Acad Sci U S A 105:5420–5425

Satoh K, Minte-Vera CV, Vogel NW, Aires-da-Silva AM, Lennert-Cody CE, Maunder MN, Okamoto H, Matsumoto T, Semba Y, Ito T (2016) An exploration into Japanese size data of tropical tuna species because of a prominent size-frequency residual pattern in the stock assessment model. 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, U.S.A., p 28

Stobberup K, Geehan J (2015) Final Project Report: Collection of Data from Tuna Fisheries in West Sumatra and North Sumatra Provinces, Indonesia.

Worm B, Hilborn R, Baum JK, Branch TA, Collie JS, Costello C, Fogarty MJ, Fulton EA, Hutchings JA, Jennings S, Jensen OP, Lotze HK, Mace PM, McClanahan TR, Minto C, Palumbi SR, Parma AM, Ricard D, Rosenberg AA, Watson R, Zeller D (2009) Rebuilding global fisheries. Science 325:578–585

Worm B, Sandow M, Oschlies A, Lotze HK, Myers RA (2005) Global patterns of predator diversity in the open oceans. Science 309:1365–1369

Worm B, Tittensor DP (2011) Range contraction in large pelagic predators. Proc Natl Acad Sci 108:11942–11947

Figures

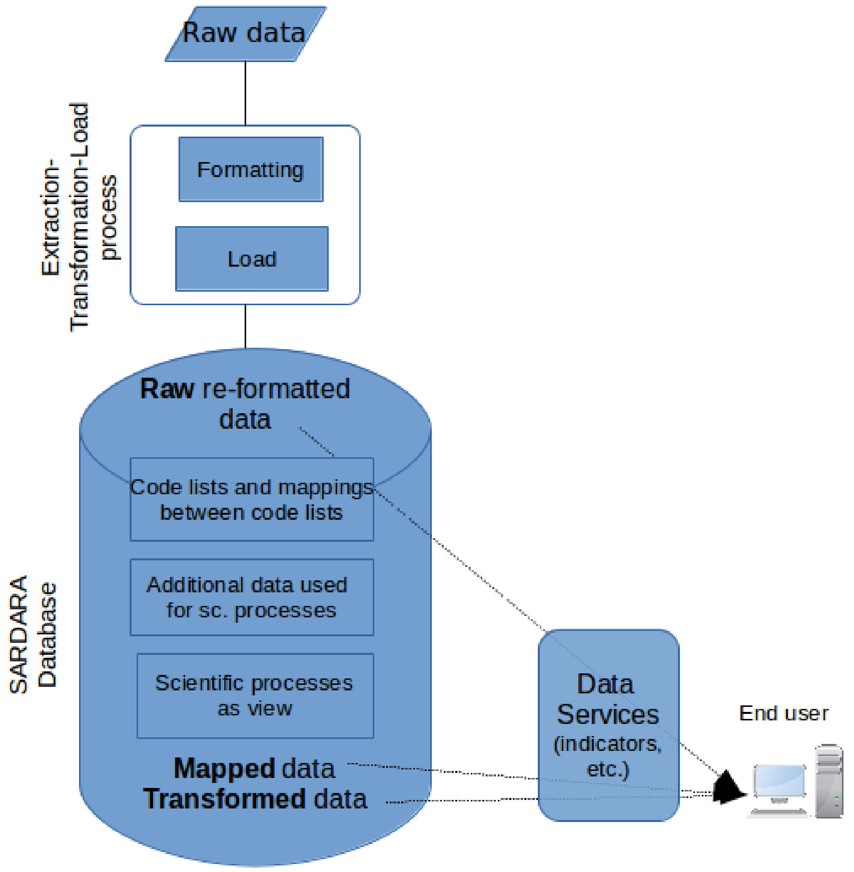

Figure 1. Flow diagram describing the technical choices for coping with the transparency and reproducibility constraints.

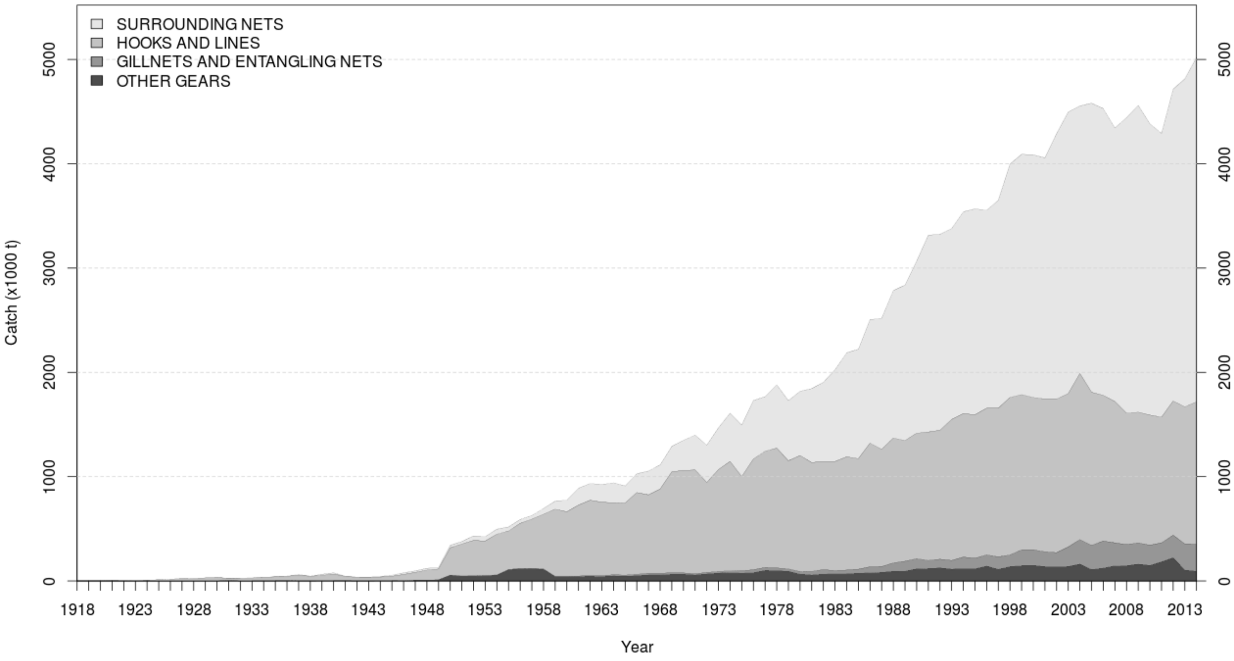

Figure 2. Time series of cumulated global tuna catches by gear over the period 1919-2014.
The species selected are the ones of the "Temperate tunas" and "Tropical tunas" IRD-defined group of species (mapping between TRFMOs species and the IRD-defined species groups are available on this link: http://mdst-macroes.ird.fr/BlueBridge/Tuna_Atlas/sardara_code_lists_and_mappings.zip).

Figure 3. Evolution of the annual catches for each oceanic basin (i.e. tRFMO) according to the data source, i.e. total catches (blue line) and cumulated geo-referenced catches. The difference between the curves combines the fisheries for which no catch-and-effort data with spatial information is available and the fisheries for which geo-referenced catches were not raised to the total catches.

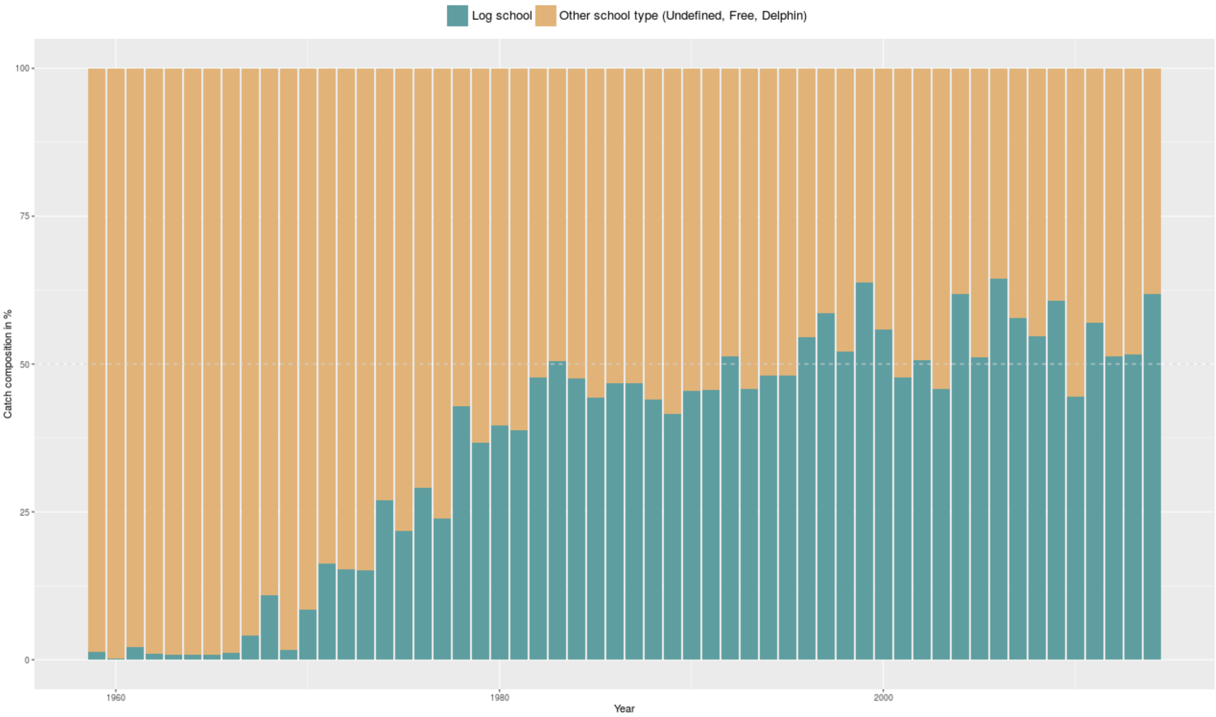

Figure 4. Annual changes in the composition (in percentage) of global purse seine fisheries catch by school association type. Green(ish) represents catches of tunas on log schools which include any type of floating object (artificial fish aggregating devices and natural objects) and yellow(ish) represents free-swimming and dolphin-associated schools.
The gears selected for the figure are the ones that have been mapped to the “Surrounding nets” gear group of the ISSCFG code list (first version of mapping between tRFMOs gears and ISSCFG gear groups are available on this link: http://mdst-macroes.ird.fr/BlueBridge/Tuna_Atlas/sardara_code_lists_and_mappings.zip).