

GLOBAL DATASETS FOR TUNA FISHERIES

Paul Taconet¹, Emmanuel Chassot², Emmanuel Blondel³, Julien Barde⁴

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 Coordinating Working Party on Fishery Statistics (CWP) and the Food and Agriculture Organization (FAO).

In a paper presented in 2016 at the IOTC WPDCS12 ([WPDCS12-27](#)), we had introduced the global harmonized database for tuna fisheries that we have developed by collating the public domain datasets (total catch, monthly-spatially aggregated catch and effort) from the International Commission for the Conservation of Atlantic Tunas (ICCAT), the Indian Ocean Tuna Commission (IOTC), the Inter-American Tropical Tuna Commission (IATTC), the Western-Central Pacific Fisheries Commission (WCPFC) and the Commission for the Conservation of Southern Bluefin Tuna (CCSBT). We had as well introduced the set of open source codes (the « toolbox ») to handle the data, i.e. transform the data formats, load the standardized data into the database, and compute a suite of indicators. However, the project - both the data and the toolbox - was still at an early stage of development. Details on the benefits of this global database for fisheries science and management, as well as technical details on the database, were also exposed in previous papers ([ICCAT SCRS/2016/202](#)).

In this paper, we expose the progresses that we have made in 2017. We first present the diversity of data that are available in the open database, and we then focus on the tools that we have developed to access and exploit efficiently the data.

KEYWORDS

Catch/effort, Fishery statistics, Data Collation, Data Harmonization, Data services, Online processing

--

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

² IRD, Seychelles Fishing Authority, BP570, Victoria, SEYCHELLES.
Emmanuel.Chassot@ird.fr

³ Food and Agriculture Organization of the United Nations, Viale delle Terme di Caracalla, 00153 Rome, ITALY. Emmanuel.Blondel@fao.org

⁴ IRD, Commission de l'Océan Indien, rue de l'Institut, Ebene, MAURICE. Julien.Barde@ird.fr

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 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 Coordinating Working Party on Fishery Statistics (CWP) and the Food and Agriculture Organization (FAO).

In a paper presented in 2016 at the IOTC WPDCS12 ([WPDCS12-27](#)), we had introduced the global harmonized database for tuna fisheries that we have developed by collating the public domain datasets (total catch, monthly-spatially aggregated catch and effort) from the International Commission for the Conservation of Atlantic Tunas (ICCAT), the Indian Ocean Tuna Commission (IOTC), the Inter-American Tropical Tuna Commission (IATTC), the Western-Central Pacific Fisheries Commission (WCPFC) and the Commission for the Conservation of Southern Bluefin Tuna (CCSBT). We had as well introduced the set of open source codes (the « toolbox ») to handle the data, i.e. transform the data formats, load the standardized data into the database, and compute a suite of indicators. However, the project - both the data and the toolbox - was still at an early stage of development. Details on the benefits of this global database for fisheries science and management, as well as technical details on the database, were also exposed in previous papers ([ICCAT SCRS/2016/202](#)).

In this paper, we expose the progresses that we have made in 2017. We first present the diversity of data that are available in the open database (section 2), and we then focus on the tools that we have developed to access and exploit efficiently the data (section 3).

2. The data available: regional and global datasets on tuna fisheries

In a first step, we have collated many public domain datasets from the five tRFMOs - and other organizations such as the FAO. These are the so-called primary (or « source ») datasets. We have integrated them within a database developed on purpose after having harmonized the structure of the datasets so that it fits the database's one. Despite the harmonization of the structures, the content of the data - coding systems, was not modified. Section 2.1 details the data that have been collated and integrated within the database.

In a second step, we have built global datasets on tuna fisheries by merging the primary datasets and applying some scientific corrections (e.g. conversion from number of fishes harvested to weight) to get a more pertinent overview of tuna fisheries at global scale. These dataset were also integrated within the database. Section 2.2 presents the processings applied to the primary datasets to compute the global datasets.

2.1 Data available on the database

A large amount of public domain datasets related to tuna fisheries have been collated, harmonized and integrated within the database. These datasets can be geo-referenced (such as catch-and-efforts) or non-geo-referenced (such as code lists). The following datasets were collated or created:

- [Collated] **Public domain datasets from the 5 tuna RFMOs** (IOTC, ICCAT, IATTC, WCPFC, CCSBT) as they deliver them (i.e. without processings):
 - Nominal catch (spatio-temporal resolution: RFMO area of competence / 1 year)
 - Georeferenced catch (5° or 1° / 1 month)
 - Georeferenced effort (5° or 1° / 1 month)
 - Catch-at-size (5° or 1° / 1 month) to come next year

- [Created] **Global datasets on tuna fisheries** (section 1.2 provides details on how these datasets were computed):
 - Global nominal catch
 - Global georeferenced catch
 - Global georeferenced effort (to come by the end of the year)
 - Global georeferenced catch-at-size (to come next year)
- [Collated] **Code lists used by the 5 tuna RFMOs** (for fishing gears, species, fishing countries, etc.) + global code lists recommended by the Coordinating Working Party on Fishery Statistics of the FAO (e.g. ASFIS List of Species for Fishery Statistics Purposes, International Standard Statistical Classification of Fishing Gear)
- [Created] **Mappings between tuna RFMOs code lists and global code lists**, which are necessary to combine the datasets expressed with sparse code lists. These mappings have been done in collaboration with the RFMOs secretariats.

Section 2.2 describes how the global datasets on tuna fisheries were generated from the primary tRFMOs datasets.

2.2 Processes applied to primary data to generate global datasets on tuna fisheries

Global datasets on tuna fisheries were generated using as input the primary tRFMOs datasets collated - catch-and-efforts, nominal catch, code lists and mappings between code lists.

2.2.1 Global nominal catch

Total catches - also called nominal catches - datasets formats are very similar among the different tRFMOs. A simple merging of the datasets has been sufficient to get the total catches at a global scale, with two additional processes:

- The primary code lists were mapped to the global code lists;
- Southern Bluefin Tuna (SBT) catches are available in both CCSBT and the other tRFMOs nominal catch datasets: SBT catches from the other tRFMOs were removed.

2.2.2 Global georeferenced catch

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. 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.

In addition, the processes to generate the global dataset on tuna fisheries are subjective: they depend on one's "perception" of tuna fisheries. For instance, one might consider that the catches expressed in number of fishes harvested should be converted into weight. In that case, which conversion factors should be used? There can be multiple valid conversion factors dataset.

We have therefore computed three different global geo-referenced catch datasets at global scale (so-called 'tuna atlas' dataset), corresponding to three different visions of the tuna fisheries. The levels gradually modify the primary datasets: level 0 datasets are the closest to the primary data delivered by the tRFMOs, while level 2 are more transformed data that aim, through the corrections applied to the primary datasets, at providing a more "scientific" overview of tuna fisheries at global scale.

The details of the processes applied to the primary datasets are available in the annex 1. Summarizing, the main processes for each level of correction are the followings:

- **Level 0** is the merging of the primary datasets. Duplicated or splitted strata among the datasets are dealt specifically. In addition, original code lists are mapped with standard CWP / FAO code lists (for gears, species, flags) so as to be able to compare the datasets. Other minor corrections regarding overlapping zones between tRFMOs are made (IATTC and WCPFC, CCSBT and the other tRFMOs).
- 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. For **level 1**, level 0 was computed, and values expressed in numbers were converted into weight using simple conversion matrices (A. Fonteneau, pers. Com).
- Geo-referenced catch data and associated effort can represent only part of the total catches. For **level 2**, level 1 was computed, and geo-referenced catches were raised to the total catches.

Global geo-referenced efforts and catch-at-size datasets will be available by next year's first semester.

All the datasets available in the database are exhaustively described through standard metadata. Metadata are very important since they provide information on the dataset: for instance, description of the processes that have been applied to generate the datasets, contacts linked to the datasets, or legal constraints associated to the datasets.

Both data and metadata are primarily stored in the open SQL database. So as to make the data and metadata accessible to the greatest number (not only SQL users), a set of services plugged to the database has been set-up. These services enable to discover the data available, access it in various formats, and process it. The methodological description of the setup of these services, as well as the benefits of describing datasets through standard metadata, are presented in another paper of the IOTC WPDCS13 (Julien Barde et al.). Section 3 of this paper provides details on the services themselves: how to discover the available data, access, process and visualize them.

3. The toolbox: how to access the data, how to create your own regional or global tuna atlas

As mentioned in section 1, we have developed an SQL database to store the data in a consistent and efficient way. Through collaboration between FAO, IRD and technology partners in the context of the BlueBRIDGE project (Grant agreement No 675680) as part of the European Union's Horizon 2020 research and innovation programme, it has been ported and is hosted on the i-Marine platform.

Once this database filled with data and metadata, the key issues are to enable anyone to discover and access the data. We have set-up an online catalogue (section 3.1) and an R library (section 3.2) on this purpose. Another key aspect we have worked on is to enable users of the datasets to understand which processes have been applied to generate a dataset (data provenance, transparency) and eventually reproduce them (reproducibility). Finally, as introduced in the section 1, we propose a set of open R codes for the users to process the data, and especially to create their own tuna atlas dataset (section 3.3).

3.1 Discover and access the data through the online catalogue

A large amount of datasets are available in the database. We have developed an online metadata catalogue to allow users to discover and access data efficiently. For each dataset, an ISO/OGC compliant and comprehensive metadata sheet has been generated and published. An OGC CSW catalogue powered by Geonetwork has been set-up to gather all these sheets. The catalogue is available [here](#). These metadata sheets combined with Geonetwork catalogue offer facilities to locate the right dataset(s) - code lists, mappings, datasets - through filters, keywords, spatial and temporal extents, etc.

For each dataset, the metadata sheet contains many information related to the dataset. Among others:

- a title and a summary;
- a citation including a set of contacts and their roles (originator, publisher, point of contact, etc.);
- a textual description of the processes that have been applied to generate the dataset (lineage) and the datasets that have been used as input (genealogy);
- the spatial and temporal extents.

A set of useful links associated to the dataset are also provided in each sheet, among others:

- links to access the dataset in several formats: CSV, OGC formats, netCDF, ESRI Shapefile;
- link to the R code that has been used to produce the dataset.

At the current stage of the project, many information contained on the metadata sheets still have to be reviewed and validated by the responsible parties (in particular the tRFMOs). The methodology to set-up the catalogue from the data contained in the database is presented in another paper of the WPDCS 13 (Julien Barde et al.). Figures 1 and 2 are snapshots of the catalogue.

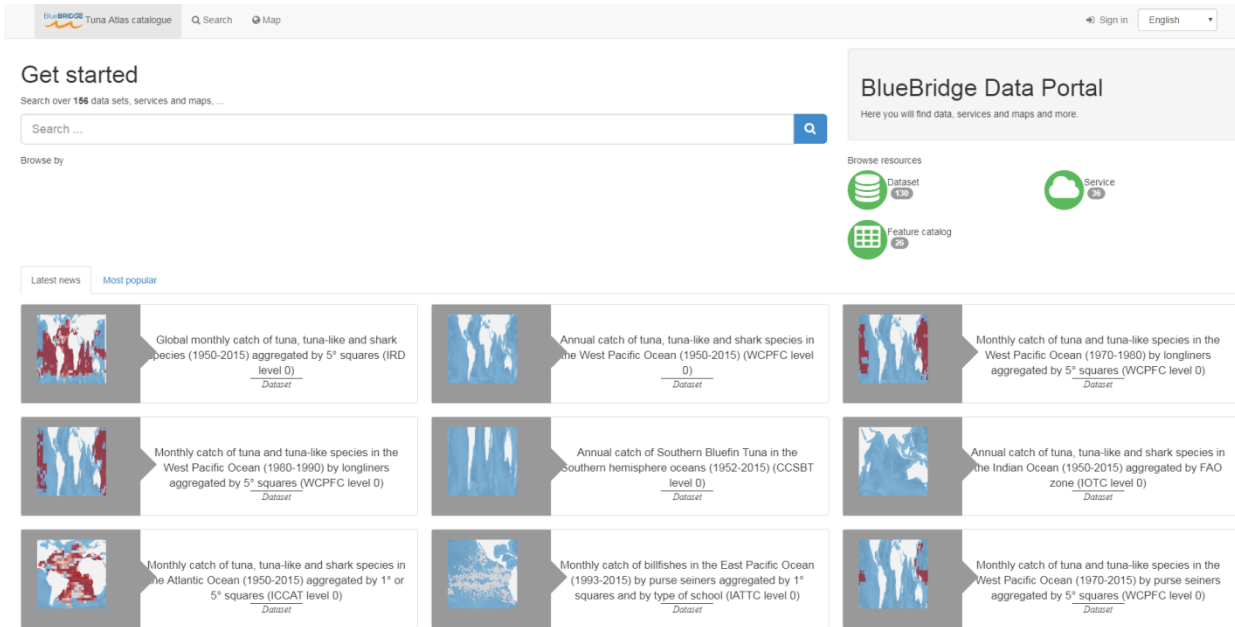


Figure 1: Home page of the online catalogue to browse and locate relevant data

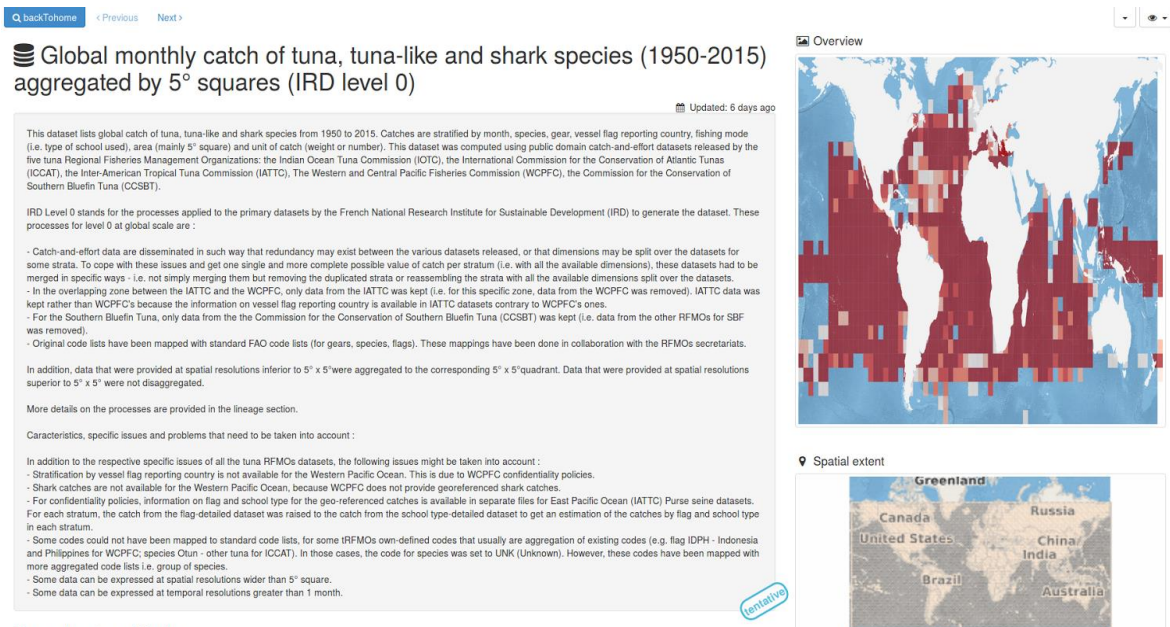


Figure 2: Snapshot of a metadata sheet accessible through the online catalogue

3.2 Discover, access and process the data through the 'rtunaatlas' library

We have developed an R library called 'rtunaatlas' to handle the datasets. This library enables to discover and access data available in the database, and as well to process any dataset properly structured (i.e. with the so-called "harmonized" structured used as input of the database). The library is available on Github [here](#). Five main categories of functions are available in the library:

- Harmonize data structure: functions to convert the structure of the primary tRFMOs datasets to the harmonized structure;
- Load data in database: functions to import a dataset with the harmonized structure (either dataset, code list or mapping between code lists) within the database;
- Discover data: functions to discover which data are available in the database;
- Access data: functions to extract a dataset from the database as R data.frame;
- Process data: functions to process data. Processings are scientific corrections that transform the data, for instance: unit conversions, raisings of geo-referenced datasets, code list mappings, spatial reallocations of data mislocated (e.g. located on land areas), spatial intersection with other spatial layers such as Exclusive Economic Zones (EEZ) , etc;
- Visualize data: functions to create plots or maps from the data.

Annex 2 provides the main functions available in the library (as of 2017-11-26) along with their title.

While the Discover and Access functions are strictly linked to the database, the Process and Visualize functions are independent from the data stored in the database. They can take as input any dataset that is provided with the harmonized structure. In other words, the use of these functions is not stuck to the datasets available in the database: they can be used with the latter, but also with any dataset that is properly structured. The figure 3 summarizes the functions available in the 'rtunaatlas' library and the source of data that they can take.

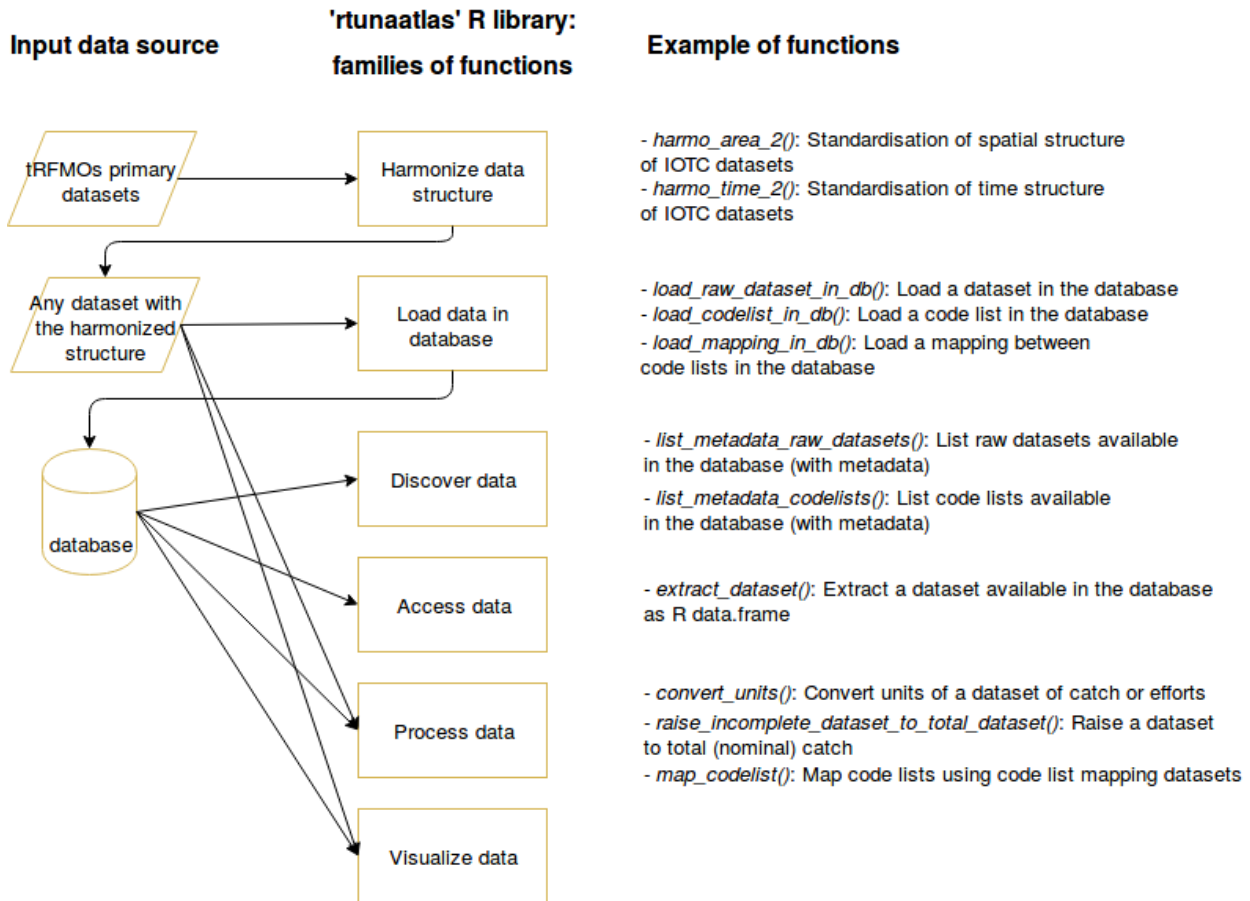


Figure 3: The 'rtunaatlas' library: source of data, families and examples of functions

A tutorial code to start using the library is available in annex 4.

In the section 2.3, we present a more complex code that we have developed and that exploit the functions of the 'rtunaatlas' library.

3.3 Create your own tuna atlas datasets

As introduced in the section 1, there can be multiple ways to apprehend tuna fisheries, whether at regional or global scale. In our work we propose three visions of tuna fisheries at global scale for geo-referenced catch, and one for nominal catch (see section 1.2). When creating a dataset - and especially when combining datasets from multiple sources - a set of subjective choices are made. Examples of these choices in our case - global tuna fisheries - are:

- Unit conversion: should we convert the units of catches (sometimes expressed in weight, sometimes in number of fishes harvested)? If so, which factors of conversion should we use?
- Overlapping zones (e.g. IATTC / WCPFC or CCSBT / other tRFMOs): These data being likely redundant, which source should we keep?
- Code list mappings: Which code list mappings should we use? For instance, one might categorize the fishing flag REU (Réunion Island) as Réunion Island, someone else could consider it as France.
- Data located on land areas: Should we consider that these data are totally wrong (i.e. they have been reported but the actual catch did not take place) or that they are only mislocated (and in this case, using spatial reallocation procedures)?

Based on this, it is important to enable anyone to create its own version of the tuna atlas, with its own perception of the tuna fisheries. We have developed two R codes - one for geo-referenced catch and one for nominal catch - that enable users to create their own tuna fisheries dataset(s), with their own assumptions. Taking as source datasets the tRFMOs primary datasets stored in the database, the codes create the tuna atlas dataset with the parameterization set by the user. Among others, parameters are:

- Spatial extent of the tuna atlas: Choice of the spatial extent of the Tuna atlas: one tRFMO area of competence (e.g. IOTC), multiple (e.g. IOTC and ICCAT), or global (IOTC, ICCAT, IATTC, WCPFC, CCSBT)
- Unit conversion: Choice of the conversion factor dataset to use to convert units. We propose some conversion factors, but the user can provide their own ones, as long as the conversion factors dataset is properly structured
- Code list mapping: As for unit conversion, choice of the code list mapping datasets to use (the user can provide its own mappings).
- Aggregation / disaggregation on 5°/1° grid resolution: Choice on the output spatial grid resolution of the tuna atlas: original (i.e. as the tRFMOs provide the data, mostly mix of 1° and 5° resolution), or all data aggregated in a 5° resolution grid, or all data aggregated in a 1° grid resolution.

The full parameterization, along with the description and possible values for each parameter, are available in the codes. The codes are available [here](#). Annex 3 provides the input parameters used in the code to create own tuna atlas of geospatial catch. The codes to create own tuna atlas of efforts and catch-at-size will be available by next year's first semester.

Use cases – what can be done now with the data?

Global datasets on tuna fisheries have already been used in various studies or works in the more or less recent past. As examples of these projects, we can mention:

- The FAO Tuna Atlas work. Due to the complexity of collating and merging the tRFMOs primary geospatial catch datasets, the project is actually not maintained since several years. Through the “Create your own tuna atlas” code, it will be possible to reproduce the FAO Tuna datasets.

- The FAO data comparison and analysis works. In particular, the FAO/ISSF comparative study of annual tuna catches from two different sources: FAO Global Capture Database vs. Tuna RFMOs Statistical Databases. This work consisted in comparing the tRFMOs nominal catch data with the FAO data and providing some possible reasons explaining the observed discrepancies. The study could be reproduced periodically using the datasets available in the database.

- The IRD Tuna Atlas. This project was initiated in the early 1990's by a researcher from IRD. The objective was to create a global Atlas on tuna fisheries, with indicators (maps, plots) and analyses (expertise) of the data.

Additional works/studies using tRFMOs public domain datasets are provided in the references.

All these studies share the same data collation and harmonization processes – a time-consuming and quite unfriendly task. Generally speaking, the data collation and harmonization processes are likely to prevent works or studies on tuna fisheries at intra-regional or global scale to bring out. One major objective of our project is to enable the reproduction of the global datasets each year. Through this, we aim at making projects more consistent over time, and at bringing out new studies to enhance the comprehension of tuna fisheries at multi-regional or global scale.

It is also noteworthy the recent works aiming at deriving indicators based on public domain datasets to provide a complementary perspective to what is provided from assessment models (Marsac et al, IOTC-2017-WPTT19-09; WPTT18). Such indicators could be calculated from the data available in the database. They could be implemented as R functions and integrated in the ‘rtunaatlas’ library. Automated reports – i.e. sets of graphical indicators - that use these functions could also be implemented. Finally, the functions and automated reports could be ported on web platforms such as Virtual Research Environments that enable their execution through user-friendly interfaces, so as to enable anyone to extract indicators using their own parameterization (filters on species, gears, fishing mode, time and space, etc.).

Acknowledgements

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 are also grateful to our FAO colleagues, particularly Fabio Carocci, Aymen Charef 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.

Annexes:

Annex 1: Details of the processes applied to the primary datasets by IRD to compute the global catch datasets

Level 0:

The following processes were applied to the primary datasets:

- Catch-and-effort data are disseminated in such way that some redundancy occurs between the various datasets released, or that dimensions may be split over the datasets for some strata. To cope with these issues and get one single and more complete possible value of catch per stratum (i.e. with all the available dimensions), some primary datasets were merged in specific ways - i.e. not simply merging them but removing the duplicated strata or reassembling the strata with all the available dimensions split over the datasets.
- In the overlapping zone between the IATTC and the WCPFC, only data from the IATTC were kept, i.e. for this specific zone, data from the WCPFC were removed. IATTC data were kept rather than WCPFC's because the information on vessel flag reporting country is available in the IATTC datasets and not in the WCPFC's ones.
- For the Southern Bluefin Tuna, only data from the Commission for the Conservation of Southern Bluefin Tuna (CCSBT) were kept, i.e. data from the other RFMOs for SBF was removed.
- Original code lists have been mapped with standard FAO code lists (for gears, species, flags). These mappings have been done in collaboration with the RFMOs secretariats.

Level 1:

Level 0 was computed and the following corrections were added:

- 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 kept and the unit of catch was set to Number of fishes harvested.
- Data located at land or without any spatial information were equally redistributed on data at sea on areas with same characteristics (same year, month, gear, flag, species, type of school).

Level 2:

Level 1 was computed and the following correction was added:

- Most catch-and-effort data have catches inferior to the catch available in the nominal catch dataset for a given stratum. Geo-referenced catches were hence raised to the total catches. 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 geo-referenced catches cannot be raised to the total catches because the combination {Flag, Species, Year, Gear} (resp. {Species, Year, Gear}) does exist in the geo-referenced catches but not in the total catches. In this case, non-raised catch data were kept.

Annex 2 : Input parameters of the R code “Create your own tuna atlas” for geo-referenced catch

| <i>Parameter name</i> | <i>Description</i> |
|------------------------------|---|
| include_IOTC | Include IOTC data (Indian Ocean) in the atlas (TRUE or FALSE) |
| include_ICCAT | Include ICCAT data (Atlantic Ocean) in the tuna atlas? |
| include_IATTC | Include IATTC data (Eastern Pacific Ocean) in the tuna atlas? |
| include_WCPFC | Include WCPFC data (Western Pacific Ocean) in the tuna atlas? |
| include_CCSBT | Include CCSBT data (Southern hemisphere Oceans - only Southern Bluefin Tuna) in the atlas? |
| datasets_year_release | Year of release of the datasets by the tRFMOs. First available year is 2017. Usually, datasets released in the year Y contain the time series from the beginning of the fisheries (e.g. 1950) to year Y-2 (included). For instance 2017 will extract the datasets released in 2017 and that cover the time series from 1950 to 2015 (included) |
| iccat_include_type_of_school | Concerns ICCAT Purse Seine data. Use only if parameter include_ICCAT is set to TRUE. ICCAT disseminates two catch-and-efforts datasets: one that provides the detail of the type of school (Fad Free school) for purse seine fisheries only and that starts in 1994 (called Task II catch effort by operation mode Fad Free school) and one that does not provide the information of the type of school and that covers all the time period (from 1950) (called Task II catch effort). These data are redundant (i.e. the data from the dataset Task II catch effort by operation mode are also available in the dataset Task II catch effort but in the latter the information on the type of school is not available). Combine both datasets to produce a dataset with fishing mode information (Fad Free school)? TRUE : both datasets will be combined to produce a dataset with fishing mode information (Fad free |

| | |
|--|---|
| | school). FALSE : Only the dataset without the type of school will be used. In that case the output dataset will not have the information on the fishing mode for ICCAT. |
| iattc_raise_flags_to_schooltype | Concerns IATTC Purse Seine data. Use only if parameter include_IATTC is set to TRUE. For confidentiality policies, information on fishing country (flag) and type of school for the geo-referenced catches is available in separate files for the Eastern Pacific Ocean purse seine datasets. IATTC hence provides two public domain dataset: one with the information on the type of school and one with the information on the flag. Both datasets can be combined - using raising methods - to estimate the catches by both flag and type of school for purse seine fisheries. Combine both datasets? TRUE : both datasets (by fishing mode and by flag) will be combined - using raising methods i.e. dataset with the information on the flag will be raised to the dataset with the information on the type of school - to have the detail on both fishing mode and flag for each stratum. FALSE : Only one dataset will be used (either Flag or Type of school). The parameter dimension_to_use_if_no_raising_flags_to_schooltype allows to decide which dataset to use. |
| iattc_dimension_to_use_if_no_raising_flags_to_schooltype | Concerns IATTC Purse Seine data. Use only if parameter iattc_raise_flags_to_schooltype is set to FALSE. In the case IATTC purse seine datasets are not combined (see description of parameter iattc_raise_flags_to_schooltype) which dataset to use? flag : use dataset with the information on flag. Information on type of school will therefore not be available. schooltype : use dataset with the information on type of school. Information on flag will therefore not be available. |
| mapping_map_code_lists | Map code lists (gears, species, flags, schooltypes, catchtype)? When using multiple sources of data (i.e. multiple RFMOS), code lists used by the various tRFMOs might be different. They should therefore be mapped to single code lists in order to be able to compare the data. TRUE : map code lists. The url to the datasets providing the code list mappings to use must be set in the parameter mapping_source_mappings. See parameter mapping_source_mappings for more details. |

| | |
|--|---|
| | FALSE : do not map code lists. Output data will use input codes. |
| mapping_csv_mapping_datasets_url | Use only if parameter mapping_map_code_lists is set to TRUE. Path to the CSV file containing the dimensions that must be mapped and the name of the mapping dataset for each dimension mapped. The mapping datasets must be available in Sardara database. An example of this CSV can be found here: https://goo.gl/2hA1sq . |
| mapping_keep_src_code | Use only if parameter mapping_map_code_lists is set to TRUE. In case of code list mapping (mapping_map_code_lists==TRUE) keep source coding system column? TRUE : conserve in the output dataset both source and target coding systems columns. FALSE : conserve only target coding system. |
| gear_filter | Filter data by gear. Gear codes in this parameter must be the same as the ones used in the catch dataset (i.e. raw tRFMOs gear codes if no mapping or in case of mapping (mapping_map_code_lists=TRUE) codes used in the mapping code list). NULL : do not filter. If you want to filter, you must write the codes to filter by, separated by a comma in case of multiple codes. |
| unit_conversion_convert | Convert units of measure? if TRUE you must fill-in the parameter unit_conversion_df_conversion_factor. |
| unit_conversion_csv_conversion_factor_url | Use only if parameter unit_conversion_convert is set to TRUE. If units are converted, path to the csv containing the conversion factors dataset. The conversion factor dataset must be properly structured. The coding systems used in the dimensions of the conversion factors must be the same as the ones used in the catch dataset (i.e. raw tRFMOs codes or in case of mapping (mapping_map_code_lists=TRUE) codes used in the mapping code lists) except for spatial code list. Additional information on the structure are provided here: https://ptaconet.github.io/rtunaatlas/reference/convert_units.html |
| unit_conversion_codelist_geoidentifiers_conversion_factors | Use only if parameter unit_conversion_convert is set to TRUE. If units are converted, name of the coding system of the spatial dimension used in the conversion factor dataset (i.e. table name |

| | |
|--|--|
| | in Sardara database). |
| raising_georef_to_nominal | Geo-referenced catch data and associated effort can represent only part of the total catches. Raise georeferenced catches to nominal catches? Depending on the availability of the flag dimension (currently not available for the geo-referenced catch-and-effort dataset from the WCPFC and CCSBT) 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 are kept. TRUE : Raise georeferenced catches to total catches. FALSE : Do not raise. |
| aggregate_on_5deg_data_with_resolution_inferior_to_5deg | Aggregate data that are defined on quadrants or areas inferior to 5° quadrant resolution to corresponding 5° quadrant? TRUE : Aggregate. Data that are provided at spatial resolutions superior to 5° x 5° will be aggregated to the corresponding 5° quadrant. FALSE : Do not aggregate. Data that are provided at spatial resolutions superior to 5° x 5° will be kept as so. |
| disaggregate_on_5deg_data_with_resolution_superior_to_5deg | Disaggregate data that are defined on quadrants or areas superior to 5° quadrant resolution to 5° quadrant? TRUE : Disaggregate. Data that are provided at spatial resolutions superior to 5° x 5° will be disaggregated to the corresponding 5° x 5° quadrants by dividing the catch equally on the overlappings 5° x 5° quadrants. FALSE : Do not disaggregate. Data that are provided at spatial resolutions inferior to 5° x 5° will be kept as so. |
| disaggregate_on_1deg_data_with_resolution_superior_to_1deg | Same as parameter disaggregate_on_5deg_data_with_resolution_superior_to_5deg but for 1° resolutions |
| spatial_curation_data_mislocated | Some data might be mislocated: either located on land areas or without any area information. This parameter allows to control what to do with these data. reallocate : Reallocate the mislocated data (equally distributed on areas with same dimensions (month gear flag species schooltype). |

| | |
|---|---|
| | no_reallocation : do not reallocate mislocated data. The output dataset will keep these data with their original location (eg on land or with no area information). remove : remove the mislocated data. |
| overlapping_zone_iattc_wcpfc_data_to_keep | Concerns IATTC and WCPFC data. IATTC and WCPFC have an overlapping area in their respective area of competence. Which data should be kept for this zone? IATTC : keep data from IATTC. WCPFC : keep data from WCPFC. NULL : Keep data from both tRFMOs. Caution: with the option NULL, data in the overlapping zone are likely to be redundant. |
| SBF_data_rfmo_to_keep | Concerns Southern Bluefin Tuna (SBF) data. Use only if parameter include_CCSBT is set to TRUE. SBF tuna data do exist in both CCSBT data and the other tuna RFMOs data. Wich data should be kept? CCSBT : CCSBT data are kept for SBF. other_trfmos : data from the other TRFMOs are kept for SBF. NULL : Keep data from all the tRFMOs. Caution: with the option NULL, data in the overlapping zones are likely to be redundant. |

Annex 3: Main functions of the R library 'rtunaatlas' as of 2017-11-26

| <i>Function</i> | <i>Category</i> | <i>Title</i> |
|---|-----------------|--|
| list_metadata_datasets | Discover | List metadata of the datasets available in the database |
| extract_dataset | Access | Extract a dataset available in the database |
| get_codelist_of_dimension | Access | Get the metadata of a code list used in a dataset |
| ccsbt_catch_level0 | Process | Extract source georeferenced catch datasets of CCSBT from the database |
| iotc_catch_level0 | Process | Extract source georeferenced catch datasets of IOTC from the database |
| iccat_catch_level0 | Process | Extract source georeferenced catch datasets of ICCAT from the database |
| iattc_catch_level0 | Process | Extract source georeferenced catch datasets of IATTC from the database |

| | | |
|---|-----------|--|
| wcpfc_catch_level0 | Process | Extract source georeferenced catch datasets of WCPFC from the database |
| convert_units | Process | Convert units of measure in a dataset |
| map_codelist | Process | Maps one dimension of a dataset using a dataset of mappings between code lists |
| raise_get_rf | Process | Get raising factors |
| raise_incomplete_dataset_to_total_dataset | Process | Raise incomplete dataset to total dataset |
| spatial_curation_downgrade_resolution | Process | Disaggregate gridded data |
| spatial_curation_upgrade_resolution | Process | Aggregate gridded data |
| spatial_curation_intersect_areas | Process | Perform spatial intersection with a geospatial layer |
| spatial_curation_reallocate_data | Process | Reallocate data using a spatial criterium |
| pie_map | Visualize | Create a pie map |
| time_series_plot | Visualize | Create a times series plot |

Annex 4: Sample R code to use some of the features of the ‘rtunaatlas’ library

Tutorial R script to start handling the 'rtunaatlas' package. The 'rtunaatlas' package enables to handle regional and global tuna fisheries datasets

Paul Taconet, IRD, 2017-11-25

Install packages

```
if(!require(devtools)){
  install.packages("devtools")
}
require(devtools)
install_github("ptaconet/rtunaatlas")
```

```
if(!require(dplyr)){
  install.packages("dplyr")
}
if(!require(data.table)){
  install.packages("data.table")
}
```

```
require(rtunaatlas)
require(dplyr)
require(data.table)
```

1) Connect to global tuna fisheries database
con<-rtunaatlas::db_connection_sardara_world()

```
## 2) List the datasets available in the database (output is a data.frame
providing the metadata):
# 2.1 List all the datasets
metadata_datasets<-rtunaatlas::list_metadata_datasets(con)
View(metadata_datasets)

# 2.2 List only the datasets whose source is IOTC
metadata_datasets_iotc<-
rtunaatlas::list_metadata_datasets(con,source_authority=c("IOTC"))
View(metadata_datasets_iotc)

## 3) Extract a dataset from the database (output is a data.frame providing
the data):
# To extract a dataset, you may put on the argument "dataset_name" the name
of a dataset (column 'dataset_name' of the data.frame output of the function
list_metadata_datasets)
# Extract the Global catch of tuna, tuna like, etc.. .
metadata_global_catch<-
rtunaatlas::list_metadata_datasets(con,dataset_name="global_catch_1950_01_01_
2016_01_01_tunaatlasIRD_level2")
global_catch_tunaatlasIRD_level2<-
rtunaatlas::extract_dataset(con,metadata_global_catch) # (takes time, approx.
2 min in my computer. I suggest to save the file in the computer once
extracted, so as to re-use it without having to query Sardara each time this
code is ran)
head(global_catch_tunaatlasIRD_level2)
# To get more information on this dataset (metadata) execute the following
line:
rtunaatlas::list_metadata_datasets(con,dataset_name="global_catch_1950_01_01_
2016_01_01_tunaatlasIRD_level2")

## 4) Gear codes used in the dataset might not be explicit. Extract the gear
coding system associated to this dataset
gear_coding_system_metadata<-rtunaatlas::get_codelist_of_dimension(con,
metadata_global_catch,
dimension_name="gear")
gear_coding_system<-
rtunaatlas::extract_dataset(con,gear_coding_system_metadata)

View(gear_coding_system)

## 5) Curate a bit the dataset. The dataset contains a mix of 1° and 5°
square resolutions. We re-project all the data on a 5° square grid resolution
# 5.1 Aggregate data that are defined on quadrants or areas inferior to 5°
global_catch_5deg_tunaatlasIRD_level2<-
rtunaatlas::spatial_curation_upgrade_resolution(con,global_catch_tunaatlasIRD_
_level2,5)$df

# 5.2 Disaggregate the data that are defined on quadrant or areas superior
to 5° quadrant resolution by dividing the catch equally on the overlappings
5° x 5° quadrants
global_catch_5deg_tunaatlasIRD_level2<-
rtunaatlas::spatial_curation_downgrade_resolution(con,global_catch_5deg_tunaa
tlasIRD_level2,5)$df

## 6) Visualize the data.

# 6.1 Pie map of catches by 5° square and by gear between 2010 and 2015
```

```
rtunaatlas::pie_map(con,
  df_input=global_catch_5deg_tunaatlasIRD_level2 %>%
filter (time_start>="2010-01-01" & time_end<="2016-01-01" & unit=="MT" &
species_group %in% c("TUNTROP", "TUNTEMP", "BILLFIS")),
  dimension_group_by="gear_group",
  df_spatial_code_list_name="areas_tuna_rfmos_task2",
  number_of_classes=4
)

# 6.2 Pie map of catches by 5° square and by species between 2010 and 2015
rtunaatlas::pie_map(con,
  df_input=global_catch_5deg_tunaatlasIRD_level2 %>%
filter (time_start>="2010-01-01" & time_end<="2016-01-01" & unit=="MT" &
species_group %in% c("TUNTROP", "TUNTEMP", "BILLFIS")),
  dimension_group_by="species",
  df_spatial_code_list_name="areas_tuna_rfmos_task2",
  number_of_classes=4
)

# 6.3 Yearly evolution of the catches of spanish purse seiners by fishing
mode
rtunaatlas::time_series_plot(
  df_input=global_catch_tunaatlasIRD_level2 %>% filter (
gear_group=="PS" & unit=="MT" & species_group %in%
c("TUNTROP", "TUNTEMP", "BILLFIS")),
  time_resolution="year",
  dimension_group_by="schooltype",
  number_of_classes=5
)

## Other available functions in the package:
# - convert units of dataset (e.g. number of fishes harvested to weight)
using your own conversion factors: help(convert_units),
# - intersect a spatial dataset with spatial layers e.g. EEZs:
help(spatial_curation_intersect_areas),
# - raise catch-and-efforts to nominal catch:
help(raise_incomplete_dataset_to_total_dataset),
# - etc...
```

References

Taconet P, Chassot E, Guitton J, Palma C, Fiorellato F, Anello E, Barde J (2016) *Global database and common toolbox for tuna fisheries*. [ICCAT SCRS/2016/202](#)

Taconet P, Chassot E, Guitton J, Fiorellato F, Anello E, Barde J (2016) *Data toolbox for fisheries: the case of tuna fisheries*. [WPDCS12-27](#)

IRD – FAO – Agrocampus Ouest – CNR (2014) *Achievable global research and fisheries management objectives*

Carocci F (2005) *Tuna Catch Data in FAO's Fisheries Global Information System*

FAO Atlas of Tuna and Billfish catches

FAO Global Tuna catches by stocks

Carocci F, Majkowski J (1996) *Pacific tunas and billfishes : atlas of commercial catches*, Food and Agriculture Organization of the United Nations

Lawson T, Garibaldi L (2000) *Comparison of Tuna catch statistics for the western and central pacific ocean held by FAO and SPC*

Garibaldi et al, 2016. *ISSF 2016-15: A Comparative Study of Annual Tuna Catches from Two Different Sources: FAO Global Capture Database vs. Tuna RFMOs Statistical Databases (2000-2014)*

Garibaldi et al (2005) *Discrepancies between the FAO and ICCAT databases for Catches of major tuna species in the atlantic ocean*. [ICCAT SCRS/2005/089](#)