# The FIRMS Tuna Atlas: a scalable open data portal for global tuna fisheries

Blondel Emmanuel[1]; Barde Julien[2], Nieblas Anne-Elise[1], Chassot Emmanuel[3], Fiorellato Fabio[3], Ellenbroek Anton[1], Gentile Aureliano[1], Taconet Marc[1]

(1) Food and Agriculture Organization (FAO) of the United Nations, Fisheries Division - http://www.fao.org/fishery
(2) French National Institute of Research for Sustainable Development(IRD), MARBEC Research Unit - http://www.umr-marbec.fr
(3) Indian Ocean Tuna Commission (IOTC) - https://www.iotc.org

# Introduction

## Rationale of the Tuna Atlas

The five tuna Regional Fishery Management Organizations (tRFMOs) are responsible for collating, analysing and disseminating fisheries data from their member countries required for stock assessment, management, and enforcement of management measures. These datasets generally include catch, effort, and size data covering large spatial scales over several decades. These rich datasets are also very complex, heterogenous, and include some large uncertainties, e.g. in reporting small-scale fisheries. In addition, data structures differ between the tRFMOs, and similar information across tRFMOs may be expressed by using different formats, labels, units, and granularity. These factors make interbasin comparisons difficult, and they can lead to misunderstandings of tuna fisheries.

Recent projects have aimed to harmonise the semantics of tRFMO data in terms of concepts and terminology (e.g. code lists), improve communication, and promote greater transparency and accessibility to their datasets (i.e., GEF funded Common Oceans ABNJ Tuna project, 1,2). These projects were driven by the need to improve data quality and availability in order to address key questions on monitoring and management of tuna fisheries (e.g. monitoring of the global fishing capacity) and ecology such as habitat preferences and the impacts of climate change (Lewison et al. 2004, Reygondeau et al. 2012, Arrizabalaga et al. 2015, Dueri et al. 2016, Monllor-Hurtado et al. 2017). These datasets can also be a valuable source of information for data-poor assessment approaches (e.g. catch time series collapse, changes in mean trophic level of catch).

There is a global call for increased data quality and harmonisation between national, regional and global statistics such that the collation of fisheries statistics is systematic and transparent across all levels. Increasingly, the streamlining of data flows among agencies is being encouraged (ABNJ, CWP), and data sharing arrangements are being promoted among FAO and RFBs. The Coordinating Working Party (CWP) on Fishery Statistics Task Group was established to develop a global standard for Reference Harmonization.

# Objectives

The general policy objectives of the 'Global Atlas of Tuna and Tuna-like species' are to improve the science-to-policy interface and communication on hot issues concerning tuna fisheries to the general public; to improve transparency through open, reproducible and collaborative science; and to foster global research on fisheries related matters by facilitating access to tuna fisheries observations (catches, efforts, FADs...), including in data limited situations.
Specific objectives are to provide a governance umbrella for data sharing agreements among FAO, t-RFMOs and other FIRMS partners, and development of Open Science strategies for global data services. On top of these data services, the Tuna Atlas also aims to provide a visually attractive and technically performant platform, with underlying data available for download, analysis, and integration with data from other domains.
The ambition is that the Tuna Atlas will facilitate better understanding, use and citation of the tRFMO's rich datasets, towards continued improvement of assessment and management of tuna fisheries and ecological research.

# Background to the development of the Tuna Atlas

Several separate initiatives took place during the past two decades to build global tuna atlases, mainly led by FAO and IRD, with work streams made of different objectives and scope but partly overlapping due to the use of common tRFMOs data sources. Among products, we can mention:

- The FAO Atlas of Tuna and Billfish catches, with dissemination of maps of global catches of major tunas and billfishes by fishing gears at 5° by 5° degrees resolution (Carocci et *al.*, 2005).
- The FAO Global Tuna Catches by Stocks, focusing on nominal catches by fishing gear, species, stock, fishing country and year.
- The suite of IRD Tuna Atlases: initiated by Alain Fonteneau (1997), followed from 2001 by the Sardara database (Lechauve et *al.,* 2007) and related websites, with continuous improvements including through the REMIGE project (Trolet, 2007), the MACROES (2010-2013) project, and Sardara World project (Chassot et *al.,* 2015).

For more than a decade, both institutions developed similar approaches to ensure yearly updates of their respective atlases, which involves collation of statistical sources published by the five tRFMOs, harmonization and production of a global dataset. However, although globally available, tuna statistics are reported and disseminated in a heterogeneous way across tRFMOs, on various websites and at different periods of the year but also through different data structures, formats and based on different reference datasets. For both FAO and IRD, this lack of global harmonization led to challenges in proceeding with yearly updates of their respective atlases, due to difficulties in accessing all tuna RFMOs datasets in a timely manner, and dealing with changes to reference data/codes and data structures used by tRFMOs.

With the advent of modern cloud-based e-infrastructures and the opportunity of the EU FP7 and H2020 projects, IRD and FAO decided to join forces to share their knowledge and methods dedicated to global Tuna Atlas data management. For FAO, the goal was to come with an operational workflow to renew the Atlas of Tuna and Billfish catches (not updated after 2012), and the Global Tuna Catches by Stocks (not updated after 2010). IRD, together with tRFMOs and FAO, wanted to enable yearly updates by discussing a schedule for data release and a standardization of data formats, codifications and access protocols. This collaboration

in particular occurred under the EU iMarine (2012-2014) and BlueBRIDGE (2015-2018) projects, together with the IRD FEAMP project (2015-2018). Through these projects, a more modern Tuna Atlas was set-up, taking as backbone the use of international standards for geospatial data (ISO/OGC) and successfully applying FAIR (Findable, Accessible, Interoperable, Reusable) data principles to tuna fisheries data. The approach is generic and this pilot is suitable to provide Atlases for any fisheries data.

## Data harmonization and standardization needs

The challenge behind setting up a global atlas for tuna fisheries data consists in compiling and harmonizing datasets in a sustainable way. Such harmonization does require the promotion and implementation of global standards for fisheries data, a task that is being carried out by the Coordinating Working Party of fishery statistics (CWP). Through its ad-hoc task group on "Reference harmonization for capture fisheries and aquaculture statistics", the CWP parties (that includes the tRFMOs) have been working together to establish standards for fisheries data exchange. With that purpose, the five tRFMOs met in a workshop organized by FAO to discuss tuna data harmonization issues, review proposals of CWP standard for reference harmonization, and review the work on updating the FAO atlas for tuna and billfishes, as result of IRD-FAO joint efforts (FAO, 2018).

In parallel, the CWP recommended the adoption of international standards for dissemination of fishery statistics, in particular from the geospatial domain viewpoint. Hence, general recommendations on geospatial data and metadata handling, and international geospatial standards have been introduced in the CWP handbook, including ISO/OGC standard formats and protocols for handling geographic data and metadata (FAO, 2019b).

## Governance model

In order to foster the sharing, maintenance and update of a single Tuna Atlas workflow, and provide on-line Tuna Atlas data services for the coming years, the FAO, as secretariat of the Fisheries and Resources Monitoring System (FIRMS) (FAO, 2020a), proposed to include the in-kind collaboration between tRFMOS, FAO and IRD under the FIRMS Partnership where most of the tRFMOs are already members. Hence, a proposal of FIRMS Tuna Atlas product (including its workflow, roles, and maintenance scenarios), governed by FIRMS Partners, was introduced and endorsed at the eleventh session of the FIRMS Steering Committee - FSC11, with the participation of IRD as a collaborative institution (FAO, 2019a).

The adoption of the Tuna Atlas under FIRMS intends to formalize data sharing and adopt harmonized data formats and processing, which is expected to enable efficient yearly updates. It also aims to help to collaboratively strengthen data science by improving the collaboration between statisticians, data managers, and data services developers, improve citation of research products (data, code, articles) by fostering management, and disseminate a wide range of operational data services including data management, analysis, visualization and dissemination. A successful Tuna Atlas is expected to constitute an operational proof of concept regarding modern statistical data workflows and exchanges among fisheries agencies.

## Ongoing activities

Within this governance framework, FAO, in collaboration with IRD, established a work plan to release the FIRMS Tuna Atlas. Two preliminary work sessions were held:

> 1)Inventory of the harmonization tasks currently performed for the production of harmonized global datasets foreseen in FIRMS Tuna Atlas (i.e. nominal catches and georeferenced catch datasets gridded either by 1° or 5° squares), including the review of existing material;
>
> 2) In-depth review of existing R codes and refactoring of the R workflow to make it fully reproducible. This second technical session covered the following: i) optimization and deployment of the database model, ii) load of code lists and code list mappings, iii) pre-harmonization and load of tRFMOs datasets, with automated check of code list and mappings, iv) generation and load of global datasets, v) data services enabling for the functioning of the FIRMS tuna atlas, and vi) inception and test of a pilot FIRMS tuna atlas map viewer.

The results of these two preliminary work sessions were shared with tRFMOs, and a release plan was set-up taking as its objective a first release of the FIRMS Tuna Atlas by the end of 2020. In this line, tRFMOs actively contributed to the FIRMS Tuna Atlas, facilitating the access to datasets with data up to 2019.

The work under integration and release in the FIRMS Tuna Atlas includes datasets on catch. The FIRMS Tuna Atlas ad-hoc working group, including the five tRFMOS, FAO and IRD are currently working to further data harmonization, data conversions/raising at global level as well as the integration of other variables (eg. effort, catch at size).

From a methodological and technological point of view, the FIRMS Tuna Atlas activities of FAO and IRD are supported by the EC H2020 Blue-Cloud project (2019-2022). This project also aims to expand the Tuna Atlas to other fisheries, and to re-use several ISO/OGC data management approaches for Aquaculture Atlases, through its two demonstrators: Fish, a matter of scales, and Aquaculture monitor. The level of FAIRness achieved in the Tuna Atlas is exemplary to other initiatives, and has INSPIREd service development in the Blue Cloud / D4Science infrastructure by promoting the use of ISO/OGC metadata approaches related also to the EU INSPIRE Directive.

# Material & Methods

## A metadata driven approach

The Tuna Atlas has been refined in 2020 to continue towards a more robust and reproducible workflow. The method used is driven by rich metadata elements which are used as inputs to initiate the various steps of the workflow (see details in section *Data management workflow*). These metadata elements describe mainly:

- Business metadata: general aspects of datasets: title, abstract, etc.

- Structural metadata: ensuring the good understanding of the data structure and thus the reuse of data once identified (by using discovery metadata). This is basically about writing a data dictionary;
- Provenance metadata to explain how accessible data have been processed.

These metadata elements have to be compliant with internationally recognized and widely used metadata standards to ensure tuna fisheries data interoperability out of Tuna Atlas context. This approach is in line with FAIR data management principles.

## Standards to foster interoperability

Interoperability is a key aspect of FAIR principles which requires compliance with standards. According to domains, various standards can be implemented for both (meta)data formats and related access protocols:

- **Domain agnostic standards**: The use of Dublin Core standard allows to describe easily both tRFMO and global datasets with a simple set of metadata elements required for data discovery.
- **Fisheries domain data standards**: The use of standards from the Coordinating Working Party for fishery statistics (CWP) allows to foster semantic interoperability with data harmonization and dissemination at global scale, in particular by using reference datasets to name key data dimensions and related values (eg *flag state*, *species*, *fishing gear*). The scope of CWP standards is being extended to promote standard data structure definitions in support of fisheries data exchange.
- **Geospatial domain standards**: The use of international standards from the International Organization for Standardization (ISO) and the Open Geospatial Consortium (OGC) allows to describe and expose datasets over the web in a standard way whatever their data structures. These constitute a key pillar for enabling a stable and sustainable way to find, access, query and re-use geo-referenced datasets. More details on the standards used are available in the section *Data Management Workflow - 4.Enabling of (meta)data services based on ISO/OGC geospatial standards.*

## Architecture

The Tuna Atlas architecture consists essentially in a Spatial Data Infrastructure made of:

- a **data warehouse** including a SQL spatial database (powered by Postgres/Postgis) and an online data repository (for the time being based on Google Drive) for data exchange;
- an OGC-compliant **metadata catalogue** (powered by GeoNetwork) for enabling dataset discovery based on dataset metadata;
- an OGC-compliant **geographic data server** (powered by GeoServer) for enabling data access/query and GIS mapping of tuna fisheries datasets;
- an OGC-compliant **geographic data client** (powered by Openlayers and OpenFairViewer) to create a map viewer.

# Datasets

The current version of the Tuna Atlas focuses on aggregated data (multidimensional data cubes). However, the upgrade of the workflow now makes it possible to add other kinds of fisheries data.

## Gridded / Aggregated data

The content of the Tuna Atlas is currently limited to catches but is meant to store and display other available variables with a similar data structure. Multidimensional data cubes of the Tuna Atlas are made of measures of variables expressed along the following dimensions: *species, fishing gear, fishing fleet, period,* and *area*.

We identified different variables to be stored in the Tuna Atlas:

- **Catches** (*species, gear, fleet, period, area, type of school*): current scope of the Tuna Atlas, possibly including discards at sea when available
- **Effort** (*gear, fleet, period, area, type of school*): to be done, with a need of harmonization and likely restricted to the industrial fisheries due to limited and sparse information for most coastal fisheries;
- **Size** (species,*gear, fleet, period, area*): to be done, but complicated because of the variety and heterogeneity of data inputs and related structures
- **Length-weight relationships** or **average weight** (species, period, area): a key variable for the raising used in upper levels of processing.

## Other types of variables

With the recent upgrade of the tuna atlas workflow, other types of data can now be described and made available. Indeed the current workflow can now extract data from various sources and these data don't have to be centralized in the same data warehouse. For example, these variables can now be managed in the Tuna Atlas:

- Trajectories of fishing vessels (Automatic Identification System, Vessel Monitoring System) and drifting floating objects used as Fish Aggregating Devices and equipped with satellite-tracked buoys
- Biological samples, including hard or soft tissue samples as well as individual morphometric data (e.g. fork length, round weight, stomach or gonad weight, etc.)
- Mark-recapture events derived from tagging operations

## Integration of catches

Catch-and-effort and nominal catch datasets were collated from the five tuna Regional Fisheries Management Organizations (tRFMOs), namely: the Commission for the Conservation of Southern Bluefin Tuna (CCSBT), the Inter-American Tropical Tuna Commission (IATTC), the International Commission for the Conservation of Atlantic Tunas (ICCAT), the Indian Ocean Tuna Commission (IOTC) and the Western and Central Pacific Fisheries Commission (WCPFC). Annual time series of nominal catches date back to the late 1910s in the Eastern Pacific Ocean while geo-referenced catch data sets generally start in the

early to mid-1950s with the advent of the Japanese high-seas longline fisheries which quickly spread over the world's tropical oceans.

Datasets provided by tRFMOs are given in different digital formats (Microsoft Excel, CSV) and a variety of data structures. Because of this heterogeneity, setting a global Tuna Atlas requires data harmonization along a common data structure.

Once harmonized, different levels of data processings can be envisaged. According to the level of processing, the global datasets are more or less easy to be reused but imply stronger assumptions. We distinguish 3 levels:

- **Datasets Level 0**: The FIRMS Tuna Atlas data and dissemination current scope only includes the following global data products collated and harmonized from tRFMO data sources:
-
    - *Nominal catches* (Note: tRFMOs level 0 datasets for nominal catches are georeferenced by FAO areas or RFB competence areas or management units, and are already raised);
    - *Geo-referenced monthly catches* (generally available by 5° squares for most gears and 1° squares for surface fisheries, e.g. purse seiners and pole-and-liners);
    - *Regional and global code lists and their mappings*. At level 0, data structure is harmonized on common data dimensions (i.e. aggregated when needed) from regional reference data to CWP reference dataset based on codelist mappings. For geo-referenced gridded datasets, this level 0 does not include any harmonization on the unit of measure (gear-dependant mix of metric tons - *t*, and number of fishes - *no*) nor on spatial resolution (mix of shapes: squares 1°, 5°, 10°, 20°, etc).

And concerning only geo-referenced gridded datasets:

- **Dataset level 1**: At level 1, level 0 data is further harmonized on the unit of measure (aligned on metric tons, based on conversion factors) and spatial resolution. The result is a fully harmonized dataset. Such level is foreseen for dissemination in the FIRMS Tuna Atlas once conversion factors (i.e. average weights for fisheries with catches in numbers) will be reviewed and validated with tRFMOs.
- **Dataset level 2**: At level 2, raising factors have been applied to elevate the level 1 catch values to the total population level. Further estimation processes are also applied to reconstruct geospatial catches whenever these were not originally available. It is worth noting that the level 2 might only be considered for a subset of species (principal market tunas and swordfish) and fisheries (industrial) for which the coverage would be deemed sufficient for extrapolation.

## Data management workflow

The tuna atlas workflow can be summarized as a sequence of four data management steps including: 1) Tuna Atlas database deployment, and upload of reference data (codelists and code mappings), 2) Pre-harmonization and upload of regional datasets from each tRFMO, 3)

Generation of global datasets (including the harmonization vs. CWP standards for fisheries data) and upload, and 4) Enabling of (meta)data services based on ISO/OGC geospatial standards.

All data management steps were performed and automated using the R *geoflow* package (Blondel et *al.,* 2020). The *geoflow* methodology triggers workflows based on dataset descriptions (metadata-driven approach) by means of a domain agnostic metadata model (Dublin Core metadata elements) materialized through a spreadsheet where each column represents a Dublin Core metadata element and each row describes a dataset which is stored or can be extracted from the database. Hence, metadata are rich enough to initiate a workflow which will sequentially (for each described dataset) run different *actions* (ie tasks) as configured within the workflow. For each dataset, several actions can be performed including *local actions* (specific to one dataset, eg. pre-harmonization), and *global actions* (common to all datasets targeted by the workflow, eg. data upload, data services enabling).

Data harmonization tasks (covered in steps 2 and 3) were performed using R scripts produced by IRD as a data toolbox for tuna fisheries (Taconet et *al.,* 2016), and adapted by FAO for the release of the FIRMS Tuna Atlas. A common set of harmonization functions used in these scripts is available in the R *rtunaatlas* package (Taconet et *al.*, 2018). Data upload tasks (covered in steps 1 to 3) include: i) data and metadata upload to a SQL database and ii) data upload to a shared online google drive repository.

1. *Tuna Atlas database deployment & reference datasets upload*

This step is only required to load gridded / *aggregate* data into a common database which is, properly speaking, a data warehouse meant to store multidimensional data cubes. In addition to reference datasets, this data warehouse also stores metadata in a dedicated table which is the input used in the last step of the workflow to create the data services.

The codelist and codelist mappings metadata currently used for the Tuna Atlas database deployment are listed at the following link:
https://docs.google.com/spreadsheets/d/1AGqXnPJvkqUzyhAxLLQ2xH36zfiLLpW6MOPbPlYlyis/edit?usp=sharing

2. *Pre-harmonization and upload of regional datasets*

In order to create the global tuna fisheries atlas, several data pre-harmonization tasks were required to align on a common normalized data structure. These pre-harmonization tasks are materialized with R scripts specific to each tRFMO dataset pre-harmonization. For each tRFMO, a metadata-driven workflow is set to trigger the required pre-harmonization tasks and the upload to the Tuna Atlas data warehouse. The tRFMOs dataset metadata used can be seen at the following links:

- CCSBT:
  https://docs.google.com/spreadsheets/d/1591BqDuehrmTHRN6C8s5ss8Z5jiTwA7jNdfnn_kvVRs/edit?usp=sharing

- IATTC: https://docs.google.com/spreadsheets/d/1G0EhEUC0H89nDFZsZBn7Y6Q9eyX5yR90mdVxrEUXMpY/edit?usp=sharing
- ICCAT: https://docs.google.com/spreadsheets/d/1tHkgvLNAPjAytLaRDqdjtDae_OBO8Yke0XWvgNl7DRk/edit?usp=sharing
- IOTC: https://docs.google.com/spreadsheets/d/1wwovnGbx8VQ4ErC5sGAX_PidSlXP2ATrXMTg42Gdmh0/edit?usp=sharing
- WCPFC: https://docs.google.com/spreadsheets/d/1T-nUq7CLsMUZglfEkYReURHz7UbxalusFdTUR-juUtY/edit?usp=sharing

In summary, the pre-harmonization consisted in applying the following processes to each tRFMO dataset:

- data **normalization** (moving from 'vertical' structure - as columns - to 'horizontal' one - as rows -), eg. converting 'species' value columns to a single 'species' column;
- data **filtering**, including null values, NA values
- data **temporal/spatial dimension harmonization** (for gridded datasets)
- data **aggregation** (sum) based on columns flag / geartype / time (start/end) / area / schooltype / species / catchtype / catchunit.

3. *Generation of global datasets and harmonization vs. CWP standards*

The alignment with CWP standards for fisheries data (FAO, 2020b) consists in harmonizing data dimensions for flagstates (ISO3), species (ASFIS), fishing gears (ISSCFG), geo-referencing (CWP areal grid - in the case of gridded datasets), by means of codelist mappings reviewed by tRFMOs.

For geo-referenced gridded catch data, the harmonization on units of measures was not performed at this stage. The global datasets include a mix of unit of measures as reported by tRFMOS, i.e. metric tons and number of fishes. An ongoing work done in collaboration with tRFMOs consists in reviewing and validating available conversion factors to allow for a catch data variable harmonization in a single unit of measure (metric tons).

4. *Enabling of (meta)data services based on ISO/OGC geospatial standards*

This step constitutes the backbone of Tuna Atlas, as it enables the metadata and data web-services required for the functioning of the Tuna Atlas web data catalogue and map viewer. It is composed of generic actions aimed to feed the Spatial Data Infrastructure, including the following:

- **metadata production**: both for reference metadata - in ISO/OGC 19115/19139 format - and structural metadata - in ISO 19110/19139 format -, operated with R *geometa* package (Blondel, 2020).
- **metadata publication** to Geonetwork (for both ISO 19115 and 19110), using R *geonapi* and *ows4R* packages (Blondel, 2020)
- **data publication** to Geoserver, using R *geosapi* package (Blondel, 2020)

Once published, the Tuna Atlas global dataset metadata can be discovered programmatically by using OGC Catalogue Service for the Web - CSW - (OGC, 2007) made available by the

Tuna Atlas metadata catalogue (Geonetwork). Dataset metadata provides OGC data access endpoints for i) data download with the Web Feature Service - WFS - protocol (OGC, 2010) and ii) web-mapping with the Web Map Service - WMS - protocol (OGC, 2006).

## Dissemination tools

The FIRMS Tuna Atlas (beta) web-interface relies on a set of software to emphasize multiple facets of data:

- metadata catalog: tuna atlas metadata can be accessed programmatically (using CSW Web service) and also displayed with compliant clients. We use Geonetwork to display metadata sheets;
- mapping interface: was deployed mainly as a GIS map viewer, powered by OpenFairViewer open-source software (Blondel, 2020), offering a metadata-driven approach, compliant with FAIR data principles, to find access, query, visualize and share the global tuna datasets;
- Interactive dashboards to display statistical indicators (under development in Blue-Cloud project). This dashboards can be also embedded in the mapping interface to process a set of indicators in given spatial areas;
- data repositories: as best practice for data sharing, previous web-applications are complemented by the provision of DOIs over global datasets in the Zenodo e-infrastructure (to be published with the final Tuna Atlas release).

Additional data access can be operated based on database read-only connection or using OGC WFS access protocol (for advanced users).

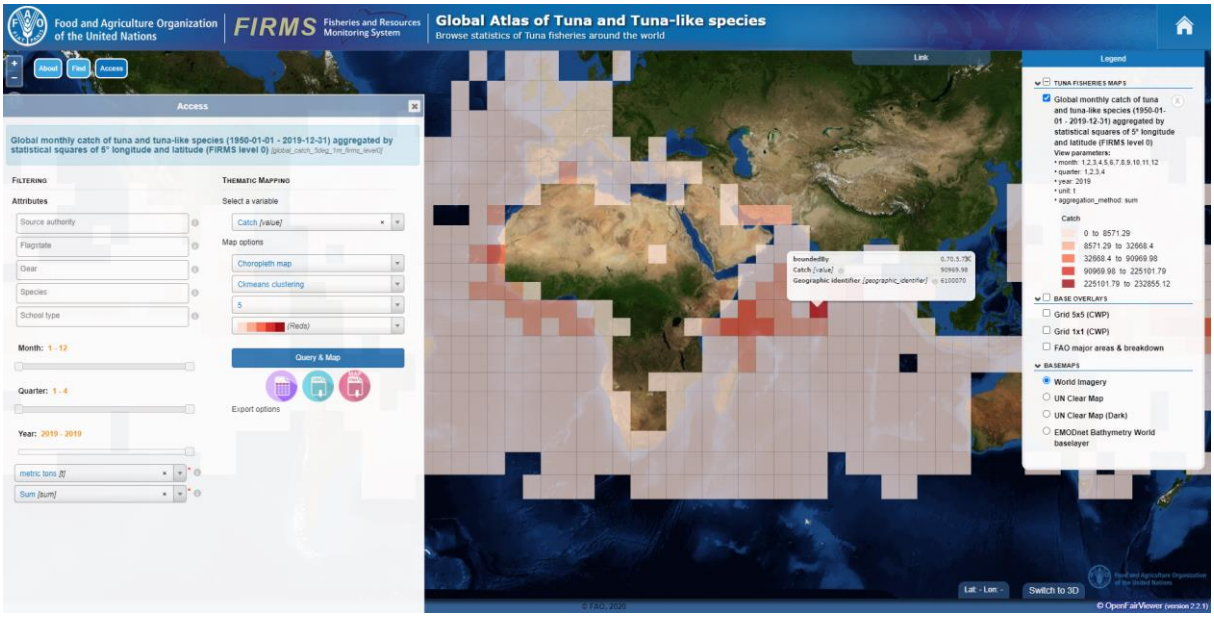
# Results

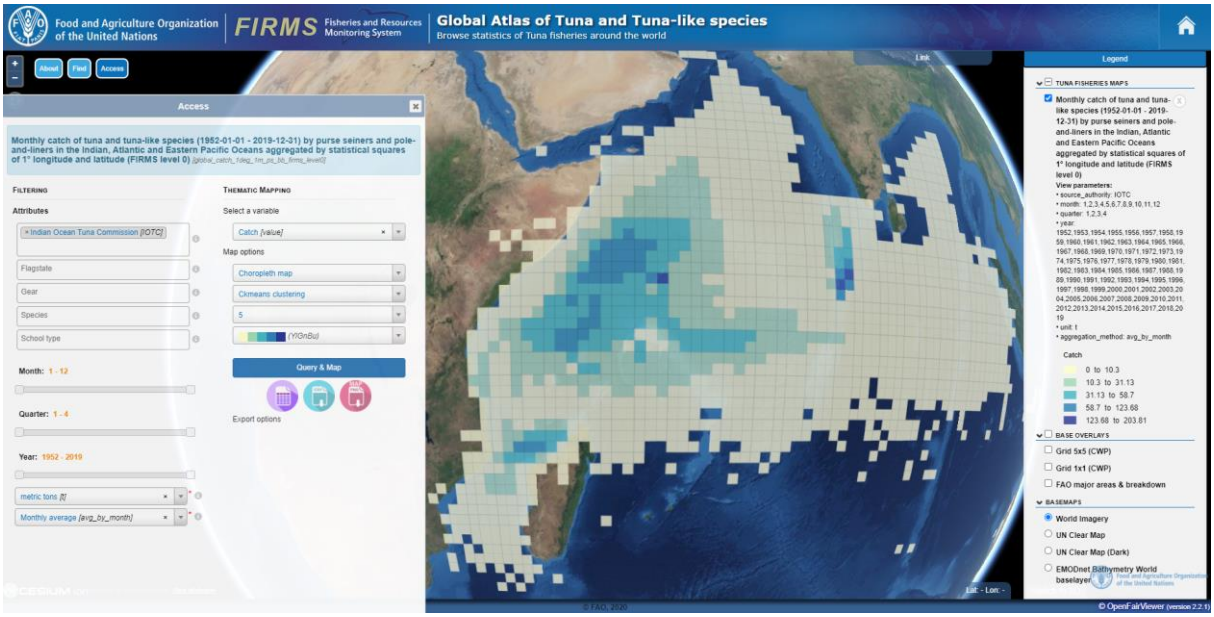
The data management workflow on catch data led to the production and dissemination of four global datasets, all described by means of standard metadata publicly available and published in the Tuna Atlas catalogue at[https://tunaatlas.d4science.org/geonetwork](https://tunaatlas.d4science.org/geonetwork), as follows:

- Global annual catch of tuna, tuna-like and shark species - 1918-2019 (FIRMS level 0) – counting with 186,996 records, and described at https://tunaatlas.d4science.org/geonetwork/srv/fre/catalog.search#/metadata/global_catch_firms_level0
- Global monthly catch of tuna and tuna-like species aggregated by statistical squares of 1° or 5° longitude and latitude - 1950-2019 (FIRMS level 0) – counting with 5,293,199 records and described at https://tunaatlas.d4science.org/geonetwork/srv/fre/catalog.search#/metadata/global_catch_firms_level0
- Global monthly catch of tuna and tuna-like species aggregated by statistical squares of 5° longitude and latitude - 1950-2019 (FIRMS level 0) – counting with 3,766,056 records, and described at https://tunaatlas.d4science.org/geonetwork/srv/fre/catalog.search#/metadata/global_catch_5deg_1m_firms_level0

- Monthly catch of tuna and tuna-like species by purse seiners and pole-and-liners in the Indian, Atlantic and Eastern Pacific Oceans aggregated by statistical squares of 1° longitude and latitude - 1950-2019 (FIRMS level 0) – counting with 1,763,423 records, and described at https://tunaatlas.d4science.org/geonetwork/srv/fre/catalog.search#/metadata/global_catch_1deg_1m_ps_bb_firms_level0

The four datasets can be then discovered, accessed and queried through the beta Tuna Atlas map viewer available at https://tunaatlas.d4science.org/faotunaatlas which offers the possibility to build statistical maps, export, share and reuse the data by means of of standard geospatial data services. The screenshots below shows examples of map queries done on the Tuna Atlas map viewer:



# Discussion

## Data harmonization

In this Tuna Atlas update, we have harmonized the heterogeneous data structures provided by the tRMFOs to a common input and output format, resulting in four global datasets. These datasets represent an important resource to the scientific community, and enable the simplification of global and interregional studies that can shed light on shared patterns between geographically dispersed species. These results are fully reproducible and documented, and have been achieved thanks to an automated workflow designed to standardise heterogeneous datasets.

The beta release of the Tuna Atlas shows that progress has been made (and continues to be made) on the harmonisation aspects of data sets and formats across tRFMOS. The current Tuna Atlas update was built on top of data inputs provided as CSV and Microsoft Excel (.xlsx) files, following the choice of deprecating less performant formats such as Microsoft Access (.mdb / .accdb), as was the case with ICCAT, or DBF files (.dbf) as was the case with WCPFC.

The number of total input data files was also reduced, following WCPFC submission of a single data file for long-line catch and effort data instead of a series of decadal data files. All tRFMOs initiated the delivery of data up to 2019, although the initial commitment was to deliver data up to 2018, and this further confirms the active involvement of tRFMOs and their will to contribute to such global atlas of tuna fisheries.

However, the Tuna Atlas still relies on an heterogeneous and unstable set of data structures that requires several pre-harmonization tasks and *ad-hoc* code adaptations that are detrimental to the concrete data harmonization aspects (eg. unit conversions, data raising factors). Work remains to be done to achieve a true standard data exchange format to consolidate the Tuna Atlas and guarantee its sustainability over the years. This requires that harmonisation and standardisation of data are directly carried on by tRFMOs, in order to simplify the process of yearly updates of the harmonized dataset. On this, the beta FIRMS tuna atlas presented here acts as an ad-hoc pilot to drive the inception, promotion and implementation of fisheries data standards through the Coordinating Working Party on Fishery Statistics (CWP). Hence, a short-term perspective will consist, in tight collaboration with tRFMOs, in drafting an agreed data exchange format, based on data best practices, to be used by tRFMOs to submit up-to-date data to the FIRMS Tuna Atlas. Such a process should be supported by tools that check the submitted data against the foreseen standard format, and according to a timeline agreed with t-RFMOs. A two-step approach may be considered, that starts with the initial alignment to the standard data structure while keeping reporting data with tRFMO reference data; eventually followed by a full alignment to the standard data structure and CWP reference harmonization standards.

Additional key harmonization issues, such as the conversion of units of measure and the incorporation of raising methodologies adopted by tRFMOs need also to be addressed. The conversion of units of measure is crucial for a better exploitation of the datasets, as these are currently not fully harmonized and the user has to select one unit or the other (metric tons vs. number of fish) when querying the map viewer. Without this harmonization, the global

coverage of the datasets will remain incomplete, and the FIRMS Tuna Atlas would only be able to deliver Level 0 datasets.

How to improve the global datasets' coverage is a matter that still requires further investigation, and should consider applying more refined levels of processing, that include raising methodologies that are subject to consensual approval between tRFMOs.

## Data dissemination

The FIRMS Tuna Atlas relies on robust methods and various tools which facilitate data discovery, access and dissemination:

- the metadata driven approach ensures a high level of documentation for all data that can be discovered and accessed in the Tuna Atlas (by using discovery and structure metadata). This is valuable for datasets stored or created by the Tuna Atlas and this is a key issue to properly reuse the data.
- the current workflow can also manage a crucial step of data dissemination by using repositories like Zenodo to assign DOIs. This is a key issue to ensure long term web access, stable URLs, foster data citation and track what usage is done with the datasets delivered by the Tuna Atlas. Beyond technical aspects, DOIs will credit data providers (organizations and related teams) for their work and raise authorship questions. Such a work has been started with IOTC (Nieblas, 2019) and should be pursued in the coming years for all datasets used and produced by IOTC and the Tuna Atlas,
- the data warehouse also relies on a widely used model which is efficient to store and explore the content of multidimensional data cubes (for expert users)
- A set of friendly data dissemination tools with GUIs to easily find and access the data to fit the needs of different kinds of users with interests spanning from data analysis (eg monitoring the status of stocks and fisheries) to communication purposes.

# Conclusion

The Tuna Atlas was set-up using a generic method for fisheries data management. It currently focuses on specific tuna fisheries datasets (mainly catch) from regional to global scale. The underlying approach can also manage additional variables and different spatial resolutions (from national to regional scales). One of the main limitations is the heterogeneity of datasets to be extracted and harmonized from tRFMO's and standardization, as foreseen in CWP and FIRMS aims to reduce (if not remove) this heterogeneity to establish a sustainable fisheries atlas.

# References

Blondel E., Barde J., Heintz W., Bennici A. (2020, November 16). geoflow: R engine to orchestrate and run geospatial (meta)data workflows (Version 0.0.20201116). Zenodo. https://doi.org/10.5281/zenodo.3138920

Blondel, E. (2020, August 4). geometa: Tools for Reading and Writing ISO/OGC Geographic Metadata in R (Version 0.6-3). Zenodo. http://doi.org/10.5281/zenodo.3971960

Blondel, E.. (2020, August 7). geonapi: R Interface to GeoNetwork API (Version 0.4). Zenodo. http://doi.org/10.5281/zenodo.3975454

Blondel, E. (2020, August 7). geosapi: GeoServer REST API R Interface (Version 0.5). Zenodo. http://doi.org/10.5281/zenodo.3975455

Blondel, E.. (2020, May 27). ows4R: R Interface to OGC Web-Services (Version 0.1-5). Zenodo. http://doi.org/10.5281/zenodo.3860330

Blondel, E. (2020, November 17). OpenFairViewer: a FAIR, ISO and OGC (meta)data compliant GIS data viewer for browsing, accessing and sharing geo-referenced data (Version 2.1.1). Zenodo. https://doi.org/10.5281/zenodo.2249305

Barde, J., Chassot, E., Blondel, E., Imzilen, T., Nieblas, A.-E., Taconet, P., 2017. Collaboration between fisheries and computer scientists for improved data description : the case of IOTC data sets. p. 11 p. multigr.

Carocci, F., A. Crispoldi, JIde Leiva Moreno, J. Majkowski, and Rome (Italy) FAO. 2005. Tuna catch data in FAO's fisheries global information system (FIGIS). Available at: http://www.fao.org/tempref/docrep/fao/008/y5984e/y5984e02.pdf

Chassot E., Guitton J., Floch I. Barde J., 2015. Sardara World: A geo-referenced database for global tuna fisheries. Available at http://mdst-macroes.ird.fr/documentation/databases/Sardara/docSardara/doc_database/SARDARA_PROJECT_2014-prepress.pdf

FAO, 2018. Report of the FAO Technical Workshop on Global Harmonization of Tuna Fisheries Statistics, Rome, 19-22 March 2018. FAO Fisheries and Aquaculture Report No. 1239. Available at http://www.fao.org/3/ca3132en/ca3132en.pdf

FAO, 2019a. FIRMS Steering Committee Meeting Eleventh Session Rome, Italy 13–14 and 18 May 2019 REPORT. Available at http://www.fao.org/3/ca5247en/ca5247en.pdf

FAO, 2019b. Report of the Twenty-sixth Session of the Coordinating Working Party on Fishery Statistics, Rome, Italy, 15–18 May 2019. Available at: http://www.fao.org/3/ca6684en/ca6684en.pdf

FAO. 2020a. Fisheries and Resources Monitoring System (FIRMS). In: Fisheries and Resources Monitoring System (FIRMS) [online]. Rome. Updated 25 April 2020. [Cited 30 November 2020]. http://firms.fao.org/

FAO. 2020b. Food and Agriculture Organization of the United Nations - The CWP Handbook of Fishery Statistics. In: FAO Fisheries and Aquaculture Department [online]. Rome. [Cited 18 November 2020]. http://www.fao.org/cwp-on-fishery-statistics/handbook/en/

Fonteneau Alain, Joseph J. (pref.) (1997). Atlas des pêcheries thonières tropicales : captures mondiales et environnement = Atlas of tropical tuna fisheries : world catches and environment. Paris : ORSTOM, 191 p. ISBN 2-7099-1370-4, available at http://horizon.documentation.ird.fr/exl-doc/pleins_textes/divers11-03/010012425.pdf
Lechauve JJ., Pianet R., 2007. SARDARA du modèle conceptuel au modèle physique. Available at: http://mdst-macroes.ird.fr/documentation/databases/Sardara/docSardara/doc_database/doc_sardara.pdf

Nieblas A-E, Fiorellato F, Blondel E, de Bruyn P and Barde J (2019). Assigning DOIs to publically-accessible IOTC documents and their publication on the open-access data repository Zenodo. IOTC-2019-WPDCS15-24. In: IOTC Proceedings. Karachi, Pakistan, 27-30 November 2019, 28p

Open Geospatial Consortium Inc., 2007. OpenGIS® Catalogue Services Specification, Version 2.0.2 - OGC 07-006r1, Corrigendum 2 Release. Available at: http://portal.opengeospatial.org/files/?artifact_id=20555

Open Geospatial Consortium Inc., 2010.OpenGIS Web Feature Service 2.0 Interface Standard - Version: 2.0.0. OGC 09-025r1 and ISO/DIS 19142. Available at: http://portal.opengeospatial.org/files/?artifact_id=39967

Open Geospatial Consortium Inc., 2006. OpenGIS® Web Map Server Implementation Specification - Version: 1.3.0. OGC® 06-042. Available at: http://portal.opengeospatial.org/files/?artifact_id=14416

Taconet, P., Chassot, E., Guitton, J., Fiorellato, F., Anello, E., Barde, J., 2016. Data toolbox for fisheries : the case of tuna fisheries. p. 23 p. multigr.

Taconet P., Chassot E., Blondel E., Barde Julien. Global datasets for tunas fisheries. Victoria Mahé : CTOI, 2017, 10 p. multigr. (Documents de Réunion - CTOI ; IOTC-2017-WPDCS13-32_Rev1). Working Party on Data Collection and Statistics : WPDCS13, 13., 2017/11/26-28, Mahé

Taconet P., Dalleau C.. (2018, April 14). ptaconet/rtunaatlas v0.1.0 (Version v0.1.0). Zenodo. http://doi.org/10.5281/zenodo.1218172

# Acknowledgments

We want to thank the tRFMOs for their collaboration, the FIRMS Partnership, both FAO and IRD Tuna Atlases legacy, the ISTI-CNR (for the maintenance of the Tuna Atlas infrastructure), and the following EU projects for co-funding the work on Tuna Atlas : iMarine (Grant agreement No. 283644), BlueBridge H2020 project (Grant Agreement No. 675680), OpenAIRE-Connect H2020 project (Grant Agreement No. 731011), Blue-Cloud H2020 project (Grant Agreement No.862409), and RDA EU 4.0 H2020 project (Grant Agreement No. 777388).