FIRMS Steering Committee Meeting
Eleventh Session
Rome, Italy 13-14 and 18 May 2019
The Global Atlas of Tuna and Tuna-like species (FIRMS Tuna Atlas)
Authors: FIRMS Secretariat with acknowledgement to IRD for its contributions

Document summary

A global **Tuna Atlas** requires skills to manage both technical and statistical aspects of tuna fisheries data. A collaboration is thus needed between data managers and statisticians. This collaboration between tuna RFMOS, FAO and IRD is proposed to occur under the Governance of the FIRMS Partnership in order to share, maintain and update a single workflow which provides different datasets and data services and make them accessible on-line with the Tuna Atlas in the coming years.

After a review of the rational for an on-line Tuna Atlas and of the background to its development, this document articulates the objectives of the FIRMS Tuna Atlas, and describes the product, the workflow and the roles for its maintenance under various scenarios. The proposed business model for the operations of the Tuna Atlas under the FIRMS Governance is then exposed, as well as the resources envisaged for its maintenance in particular through formal arrangements among public agencies. Citation and acknowledgement of its various components are then reviewed, and the workplan for the official release of the Tuna Atlas is then presented, along with an outlook of potential future developments.

FSC11 is requested to discuss and eventually endorse the proposal for FIRMS to take Governance leadership of the FIRMS Tuna Atlas.

Actions requested from the Committee

1) Provide opinions on the potential use of FIRMS Tuna Atlas, its promotion, and what activities this entails
2) Confirmation of Agencies commitments for sustained operation of the Tuna Atlas under FIRMS governance leadership
   (Paragraph 5)
   a) Expression of commitments by t-RFMOs to provide on yearly basis the required primary datasets i) in their usual format, and ii) in a format compliant with the CWP Reference Harmonization standard; and to update their FIRMS Annex 2 as data sharing agreement for the provision of such Tuna statistics
   b) Expression of commitment by FAO for co-ownership of Level 0 dataset: involvement of a fishery statistician / fishery biologist scientist, either to validate IRD level 0 dataset before its dissemination through FIRMS Tuna Atlas, or to contribute to the co-production of a FIRMS level 0 dataset
   c) Expression of commitment in principle by IRD for becoming a FIRMS Partner under the IRD-FIRMS Collaborative Arrangements
   d) Confirmation by FAO of the status of maintenance agreements: FAO-CNR-ISTI iMarine MoU and SLA on the Tuna Atlas
3) Comment the envisaged resources (Paragraph 6)
4) Endorsement of the IRD-FIRMS Collaboration Arrangement for the FIRMS Tuna Atlas Level 0 (Paragraph 5.2, 5.3)
5) Governance decisions on the operations of the Tuna Atlas:
   a) Short <u>term:</u>
      i) agree on scenario to select between scenario 1 and 2 (Paragraph 3) and responsibilities for the production of datasets and their dissemination through the FIRMS tuna Atlas (Paragraph 4.3.2)
      ii) endorse type of datasets to be disseminated through the FIRMS Tuna Atlas catalogue (Paragraphs 4.1, 4.2)
      iii) endorse proposed citation for the FIRMS maps viewer, and the proposal for datasets to be disseminated building on a DOI (Paragraph 7.1)
      iv) identify arrangements including workplan for the official release of the FIRMS Tuna Atlas map viewer
      v) Agree on a schedule for t-RFMO data delivery
   b) Medium <u>term</u>:
      i) Expression of interest in principle by t-RFMOs to collaborate in the co-production of Level1 and Level2 datasets (scenario 3)
      ii) Expression of interest in principle by t-RFMOs to be involved in the BlueCLOUD project (if approved)

Table of contents

1. The rationale for the on-line Tuna Atlas

The world Tuna fisheries are monitored and managed by five tuna Regional Fishery Management Organizations (RFMOs) on an ocean basin basis, each exerting a massive effort to collate, analyse and disseminate the data required for assessment, management, and enforcement. As a result, the fisheries public domain datasets available from the five t-RFMOs are numerous and include aggregated information on catch, effort, and size which can be expressed in different formats, units, granularity, levels of raising, etc. The data sets are rich and cover large spatial scales over decades and across oceans. But they are also complex, heterogeneous, and uncertain for some fisheries components (e.g. small-scale fisheries), which does not facilitate their manipulation and use. This results in situations where the risk is high for real data to be misunderstood, misrepresented, or even misleading in making claims related to tuna fisheries, and where it is difficult to counter statements.

At a global level the GEF funded Common Oceans ABNJ Tuna project, which objective is to accompany t-RFMOs in the necessary transformational change in the management of tuna fisheries, evoked regarding data matters the needs for harmonization of the concepts and terminology (e.g. the code lists) used in the tRFMO[1], the promotion of communication towards tRFMOs and their member States as well as to the general public[1,2], and the need for tRFMOs for higher transparency and accessibility to their datasets[2].

The necessary transformational in the management of tuna fisheries also implies the need to address several research questions related to the monitoring of fishing capacity and effort (e.g. changes in global longline effort), changes in some fisheries strategies (e.g. expansion of FAD fisheries across the globe) and the provision of catch data for data-poor assessment approaches, e.g. detection of collapse in time series of catches, data-limited assessment models, and changes in mean trophic level of the catch. The detection of common patterns across oceans such as temporal changes in nominal CPUEs (e.g. “jump” in bigeye CPUE in the mid-1970s) can also provide insight into the mechanisms and driving factors of such patterns. The global unified vision of tuna catch and effort data can also be of interest for quantifying the economic value of tuna fisheries and explore global scenarios of management such as the closure of the high seas (White & Costello 2014, Sumaila et al. 2015). More recently, global tuna data sets have been shown to be of major interest to address ecological questions such of the contraction of tuna distribution in relation with the reduction in abundance (Worm & Tittensor, 2011), the habitat preferences of pelagic populations and communities (Lewison et al. 2004, Reygondeau et al. 2012, Arrizabalaga et al. 2015) and the effect of climate changes (Dueri et al. 2016, Monllor-Hurtado et al. 2017).

The Voluntary Guidelines on Small Scale Fisheries also provide a driver for the need to address data limited situations, and attention recently increased in t-RFMOs and other RFBs to properly take into account small scale fisheries. As an example, common issues were found to affect different FAD fisheries in the Caribbean Islands region, and the research required for understanding the impacts of the increasing use of FADs in such data limited situation would benefit from a tool of communication and analytical features between FAO, the t-RFMOs and other parties.

The global debate which challenges the quality of FAO statistics constitutes yet another driver for improving alignment between national, regional and global statistics through a more systematic and transparent collation and assembly of fishery statistics across these levels. Best practices for streamlining data flows among agencies in charge of fishery statistics are being encouraged by the

[1] Report of the Second Project Steering Committee: Sustainable Management of Tuna Fisheries and Biodiversity Conservation in the ABNJ. 2015

[2] Presentation IOTC-2015-WPDCS11- Dissemination of IOTC datasets.pptx, 11th Working Party on Data Collection and Statistics (WPDCS11). 2015

Coordinating Working Party on Fishery Statistics (CWP) including through various data sharing arrangements among FAO and RFBs and the CWP Task Group set-up to establish a global standard for Reference Harmonization.

The above drivers plead for an on-line global Tuna Atlas available to general public and scientists as a transparent and as-objective-as-possible set of data services exposing harmonized tRFMO data with a high degree of reliability, reproducible methodology and accompanied by clear explanations of data coverage and what they actually represent. Such platform should respond to the growing expectation among the general public that data are published in visually attractive and technically performant websites, with underlying data available for download, analysis, and integration with data from other domains. More generally it would also contribute to the improvement of data and information for data-limited species, as well as other services that can be provided to t-RFMOs to improve assessment and management of tuna fisheries.

2. Background to the development of the Tuna Atlas

In response to this general need, FAO and IRD have established since the early 2000s their respective tuna atlases, with slightly different objectives and scope but partly overlapping since relying on same t-RFMOs data sources. During this historical period, both institutions have been working separately.

The FAO Atlas of Tuna and Billfish catches, disseminates maps of global catches of major tunas and billfishes by fishing gears an 5° by 5° degrees resolution. This work was maintained by Fabio Carocci (Carocci et *al.*, 2005)[3]. The FAO Global Tuna Catches by Stocks disseminates nominal catches by fishing gear, species, stock, fishing country and year.

The first IRD Tuna Atlas was initiated in Fortran code by Alain Fonteneau (1997). This was followed from 2001 by the Sardara database and related Web sites, with continuous improvements[4].

For more than a decade, both institutions developed similar approaches to yearly updates of their respective atlases, which involves collation of statistical sources published by the five t-RFMOs, harmonization and production of a global dataset. However since years, both also faced similar limitations making the yearly update a difficult task: ii) the ability to access all tuna RFMOs datasets, ii) the changes in the data structures and codifications provided by tuna RFMOS, iii) the lack of deadline to deliver datasets and lack of standards for data formats. While in principle this task should only consists in executing each year the same code with the newly published t-RFMOS datasets, most of the time IRD has been struggling to get the data and to adapt the code when datasets formats are different. Workflow cannot be maintained yearly if data are not provided on a regular basis and do not comply with standardized data structures.

[3] 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 online at: http://www.fao.org/tempref/docrep/fao/008/y5984e/y5984e02.pdf

[4] - IRD (2001): Jean-Jacques Lechauve (IRD data manager) set up the Sardara database model to manage multiple datasets in a single SQL multi-dimensional data warehouse
- ANR REMIGE project (2007): Julien Trolet uses R (instead of Fortran) to generate the different levels of processing, to load data in Sardara and creates a new Web Site (MDST) to query the datasets stored in Sardara
- Data updates: Emmanuel Chassot, Jérôme Guitton, Norbert Billet,
- ANR MACROES (2010): Norbert Billet and Julien Barde creates a Web service and a new version of the MDST Web site to query Sardara by using rich metadata descriptions (including data structure description).

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 and collaborate with CNR-ISTI (provider of IT technology services) in order to share their knowledge and methods and host their tools, within a Virtual Research Environment (VRE) dedicated to global Tuna Atlas data management. For FAO, the finality was to come with a functional 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 wanted together with tRFMOs and FAO 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[5] (2012-2014) and BlueBRIDGE[6] (2015-2018) projects, together with the IRD FEAMP project[7] (2015-2018).

As a result of these recent projects, the **BlueBRIDGE Tuna Atlas** (i.e. the global atlas for Tuna fisheries delivered jointly by IRD-FAO-CNR under BlueBRIDGE) offers access to one global database for tuna fisheries data - total catch, georeferenced catch and effort, and catch at size - achieved by collating and harmonizing the public domain datasets from Tuna Regional Fisheries Management Organizations (t-RFMOs). This database is accessible online through a dedicated Web portal with various types and levels of access, along with the set of open source codes (a « toolbox ») to handle the data (i.e. transform the data formats, load standardized data into database, compute indicators). Thanks to an underlying Spatial Data Infrastructure (SDI), the database and toolbox also deliver services for metadata management, data sharing in various formats, processing, visualization, and static and interactive reporting. An open policy applies to the components of the toolbox, including data, software, and the services. The toolbox in addition support transparent and reproducible workflows. These workflows are the result of community discussions and require a governance model to adopt harmonized data formats and processing.

In this collaboration, the data collation workflow developed by IRD (FEAMP project[7]) generates datasets with three levels of processing[8] defined hereafter (and referred throughout the document - see also schema in Annex 2, Figure 5b):

- **Level 0 dataset** (harmonized; with no extrapolation) stores catch and effort data as close as possible to primary data collated from countries and made publicly available by t-RFMOs. For level 0, the primary datasets of tuna RFMOs are merged with a set of harmonization rules (see details on the Level 0 dataset, in Annex 3):
    - original code lists of Tuna RFMOs have to be mapped with standard CWP / FAO code lists (for gears, species, flags) so as to be able to compare the datasets.
    - duplicated or split strata among the datasets are dealt specifically.

[5] FP7 i-Marine (2012): Norbert Billet and Julien Barde produce SQL and R code which can be executed online from web-services to replicate the set of indicators displayed in the Tuna Atlas created by Alain Fonteneau (already replicated with SQL and PHP in the online version hosted by Agrocampus). FAO develops and tests a first version of the Tuna Atlas.

[6] BlueBRIDGE: J.Barde (IRD), P. Taconet (IRD) and E. Blondel (FAO) extend the scope of IRD's historical data collation workflow by adding new steps which create data and metadata services on top of Sardara database. By doing so, they set up a Spatial Data Infrastructure (SDI) for the Tuna Atlas, and this infrastructure is a legacy on top of which online Web portals can be built. An example of web portals enabled during BlueBRIDGE is the Global Tuna atlas map viewer where IRD datasets are disseminated.

[7] FEAMP (3 years contract, 2015 to 2018): Paul Taconet, Emmanuel Chassot, Julien Barde. A data collation workflow using R has been set up to generate global datasets with different levels of processing and load these datasets in the Sardara database (new version of the database model).

[8] see Taconet et al, 2017 http://www.documentation.ird.fr/hor/fdi:010071471 ; and Taconet et al, 2016 http://www.documentation.ird.fr/hor/fdi:010071520

    - other minor corrections regarding overlapping zones between t-RFMOs are made (IATTC and WCPFC, CCSBT and the other t-RFMOs).
- **Level 1 dataset**: uses Level 0 as an input of the process. Harmonization of units of measures for catches which can be, according to RFMOS, weight or in number of fishes or both weights and numbers in the same stratum. The values of catches measured in numbers are converted into weight using simple conversion matrices (A. Fonteneau, pers. Com).
- **Level 2 dataset**: uses Level 1 as an input of the process. Geo-referenced catches are raised to the total (nominal) catches since geo-referenced catch and associated effort data provided by RFMOs only represent a part of the total catches. Level 2 is also provided for catch at size data or CPUE.

This Atlas was presented at the FAO Technical workshop on global harmonization of Tuna fisheries statistics (Rome, March 2018) which was organized under the activity of the "task group on reference harmonization for capture fisheries" of the Coordinating Working Parties on fisheries statistics (CWP). The workshop brought together data managers of the five t-RFMOs namely CCSBT, IATTC, ICCAT, IOTC and WCPFC (and SPC) with the aim to harmonize data structures and embedded statistical concepts and codelists used by t-RFMOs. t-RFMOs recognized the usefulness of the presented Atlas to the scientific community and the global user community with further developments of data contents.

The outcome of the workshop includes t-RFMOs pre-agreement to share two public datasets 'nominal catch' and 'catch and effort' together with the respective codelists, and mappings of t-RFMO with CWP codes and any other metadata required. With the goal to set-up a governing mechanism with established data sharing agreement, FAO suggested to obtain this data from t-RFMOs under the FIRMS multilateral partnership and to include provision of data submission. t-RFMOs representatives would come to the next FIRMS Steering Committee with the mandate to provide their final agreement on data provision after discussion and approval with commissions and secretariat. That would proceed with amendment of Annex 2 of the FIRMS partnership arrangement of each t-RFMO. As part of the recommended actions, WCPFC was invited to join the FIRMS Partnership.

The collaborative work on global atlas for tuna fisheries will be effectively operating on the basis of the data structures to be endorsed under the CWP standard for reference harmonization. Indeed the sustainability of the Tuna Atlas depends essentially on the level of standardization of primary datasets collated from t-RFMOs. The adoption of this CWP standard is expected to be a major key for minimizing maintenance burden and cost of the atlas. The data provided by t-RFMOs will be made available according to the harmonized structure together with codelists mappings, with the goal to efficiently allow their interoperability across databases.

The outcome of IRD initiative, BlueBRIDGE project and the CWP Reference Harmonization process provides the foundation for the FIRMS Tuna Atlas, and the expected endorsement by FIRMS SC11 of governance leadership of the global Tuna Atlas.

3. The objectives of the FIRMS Tuna Atlas

The **general policy objectives** of the '**Global Atlas of Tuna and Tuna-like species**' (herewith called, 'FIRMS Tuna Atlas') are, consistently with section 1 'Rational', 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 for placing the Tuna Atlas under FIRMS is 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. The adoption of the Tuna Atlas under FIRMS will help formalize data sharing and adopt harmonized data formats and processing, which is expected to enable efficient yearly update. It will also help to collaboratively strengthen data science by improving the collaboration between statisticians, data managers, data services developers, improve citation of research products (data, code, articles) by fostering FAIR[9] data management and, disseminate a wide range of operational data services across data management, analysis, visualization and dissemination. A successful Tuna Atlas is expected to constitute an operational proof of concept regarding modern statistical data workflow and exchange among fisheries agencies.

In order to achieve these objectives, **several scenarios** can be envisaged:

- **Scenario 1 - FIRMS approved Level 0 dataset, minimum involvement of FIRMS partners**:
  assumes that FAO and t-RFMOs accept the current IRD level 0 dataset (with underlying collation and harmonization criteria) as a valid global level 0 dataset. Implications:
  - Eliminates the need of FAO fishery statistician and t-RFMOs involvement, except to validate the data sources and dataset generation code the first time
  - Dataset ownership/citation: The level 0 dataset remains an IRD data product, cited as such along with a DOI (current situation).
- **Scenario 2 - FIRMS owned Level 0 dataset, optimum involvement of FIRMS partners:**
  FAO and t-RFMOS collaborate with IRD to define consensual harmonization rules to be applied to t-RFMOs data. This will produce a new dataset driven by the needs of FIRMS Tuna Atlas (FIRMS level 0, different from the IRD one); implications:
  - Requires involvement of a FAO fishery statistician / fishery biologist scientist to interact with IRD, and possibly involvement of t-RFMOs data managers (contributions are welcome)
  - Dataset ownership/citation implication: The level 0 dataset becomes a shared FIRMS (i.e. FAO+IRD+t-RFMOs) data product (shared ownership) with proper citation along with a new DOI
- **Scenario 3 - FIRMS owned Level 1 / 2 datasets, thorough involvement of FIRMS partners:**
  in addition to level 0, partners want to generate as well the levels 1 and 2 through collaboration in providing their raising or conversion factors. The data collation workflow has then to be extended with additional steps (using Level 0 as an input).

Summarizing: each scenario will be implemented by using the same sequential workflow with a different parameterization and by customizing or adding steps depending on expected levels of processing. The scenarios are defined with users and how Tuna Atlas data and services are made accessible (i.e. by the spatial data infrastructure, the public data warehouses or the web-viewers).

Whatever scenario is considered by FIRMS partnership, it is understood for this collaboration that each partner is an actor of the FIRMS Tuna Atlas, expected to play one or more roles and contribute with material to the different steps of the overall workflow. The section 4.2.2 details what should be provided by each Partner.

[9] FAIR principles are: data discovery, access, interoperability and reusability

4. Description of the FIRMS Tuna Atlas as a product

This section describes the FIRMS Tuna Atlas as proposed at the time when FIRMS is asked to take it under its Governance. The FIRMS Tuna Atlas inherits from the BlueBRIDGE Tuna Atlas and as such its general description is available in section 2 'Background'. The present section more specifically covers the description of the scope, data products, data services, and role of FIRMS partners.

4.1 The FIRMS Tuna Atlas data products - global harmonized dataset (Level 0)

In terms of content scope, as per the guidance of the CWP Tuna-RFMOs Reference Harmonization workshop, the FIRMS Tuna Atlas primarily targets the global harmonized Level 0 dataset presented in the background section:

- **Level 0 datasets**: The FIRMS Tuna Atlas data and dissemination scope includes the following global data products collated and harmonized from t-RFMO data sources: Nominal catches (note: t-RFMOs level 0 datasets for nominal catches are already raised); Geo-referenced catches [and effort] by 5° squares; and for purse seiners and pole-and-liners by 1° squares; regional and global code lists and their mappings.
- **Metadata**: associated with each dataset, stored in the database and driving the services of the Spatial Data Infrastructure, metadata are the pillar of the Tuna Atlas. At this stage metadata were provided by IRD and have to be reviewed, enriched and validated by FIRMS partners to become a common product.

4.2 The FIRMS Tuna Atlas data services - maps viewer, data catalogue, and data collation workflow

The FIRMS Tuna Atlas will rely on services provided through a set of software components which together constitute a spatial data infrastructure (SDI). The services are available to different user profiles.

The FIRMS **end users** (including Scientists) will have access to public data according to their skills level:

- the ***FIRMS Tuna Atlas map viewer*** (screenshot in Figure 1, click on https://tunaatlas.d4science.org/faotunaatlas/ to access it) presents to end users datasets displayed as maps, and allows them to use the main Tuna Atlas data services: search, query and access datasets visualized on maps (data download facility and a map image export). Each dataset query can be stored (and shared between users) in a persistent and reproducible way by means of web link sharing mechanism. It, allows as well data managers to embed specific data views in external websites. This viewer is powered by an open source software which implements FAIR principles and is entirely built on top of rich metadata and data services. This map viewer can be also re-distributed (e.g. as embedded product) in t-RFMOs websites upon request as well as through the iMarine e-infrastructure.

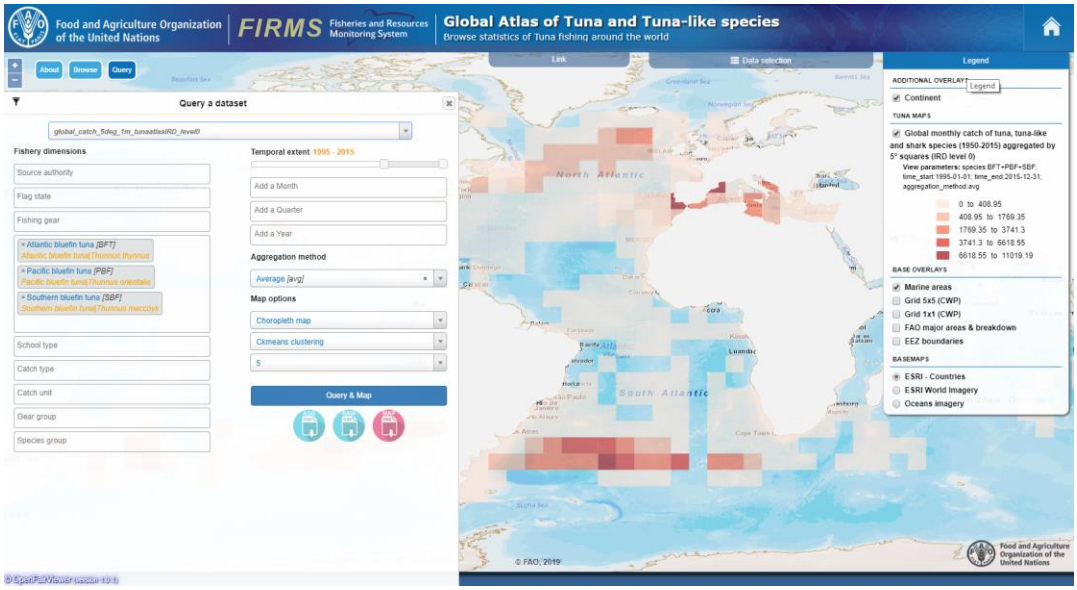

Figure 1: Overview of the FIRMS Tuna Atlas map viewer - with a user query

(Bluefin Tunas, Average catches over period 1995-2015, by 5° squares)

- The ***Tuna Atlas catalogue*** (screenshot in figure 2) which disseminates the (Level 0) datasets listed in Annex 3: the GIS layers, the transformed/harmonized primary sources, the harmonized integrated global datasets, the code lists, the code lists mappings; and their metadata. Currently the number of resources disseminated through the catalogue amount to 400 and some discussion could aim at reducing this number. This catalog provides the underlying services to the Maps viewer.

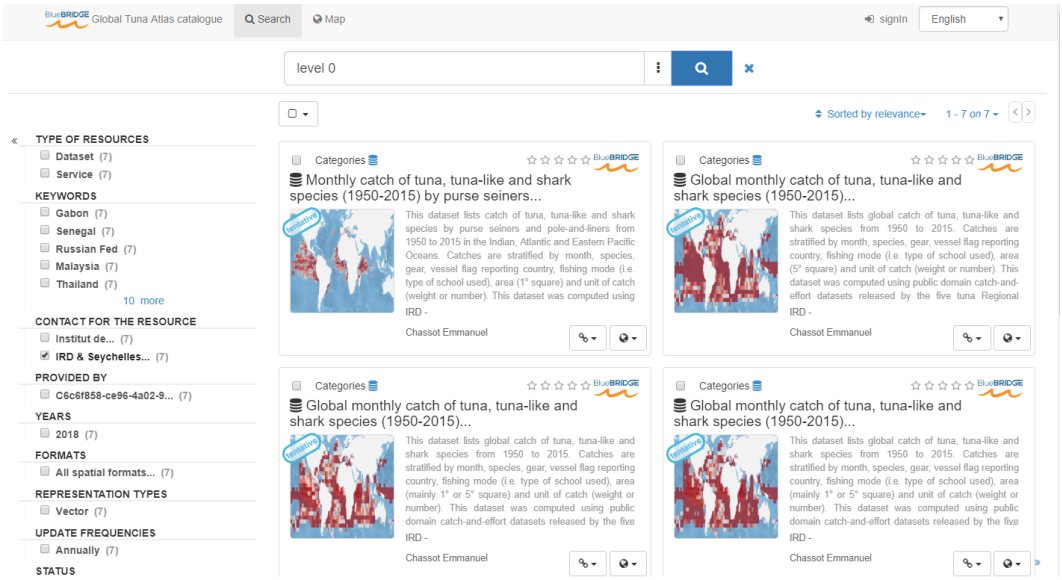

Figure 2: Overview of the Tuna Atlas catalogue

- The ***Tuna Atlas database*** used to store data. Skilled end users have access to this database in read-only mode through various tools (e.g. QGIS) for specific data extractions in SQL.

The **Data managers** in charge of the maintenance of these components (FAO/IRD, in collaboration with CNR-ISTI / D4Science administrators) can access the **Tuna Atlas workflow** to operate the datasets generation and publication (see overall workflow in Annex 2 Figure 5a). This workflow is presently operated with the open-source R software for statistical computing (https://www.r-project.org), data generation being only ensured by IRD at this stage.

Compliance with International standards: the Tuna Atlas set of components is fully compliant with standards for spatial data infrastructures interoperability, as recommended by CWP GIS working group and supported by the participants of the CWP Tuna Workshop (Rome, 2018). Such standards are widely used (INSPIRE directive in Europe, UNESCO Global Ocean Observing System, Integrated Ocean Observing System in USA, Integrated Marine Observing System in Australia).

Hosting: the FIRMS Tuna Atlas relies on data services delivered by software operated on the D4Science infrastructure. It's worth noting that the software might be as well hosted by other infrastructures (IRD, AWS, OVH..). Along with the hosting, the D4Science infrastructure offers specific services, brought within a VRE, that may be used for scientific collaboration (Workspace, social network, RStudio or RShiny hosting..). Notwithstanding, it should be noted that these VRE services are not mandatory for the Tuna Atlas functioning but may be an asset for fostering collaboration between FIRMS partners.

Data publication: the Zenodo (https://zenodo.org) infrastructure is used to assign DOIs to datasets, processing code and scientific publications (which is a requirement to comply with Open Science and FAIR data management principles).

Annex 1 further describes technical aspects, software ownership, and includes a diagram (Figure 4) which illustrates the users and their interactions with the Tuna Atlas software components.

4.3. Arrangement for the collaborative workflow for producing and publishing level 0

This section reviews where do we start from, what are the conditions for efficient update, and how do we propose to work together. The present Tuna Atlas level 0 datasets covers the period 1950 to 2015. A yearly update in joint effort by FAO/IRD is foreseen to include year 2016, depending on the outcomes of the FIRMS Steering Committee.

4.3.1 Current situation

The global level 0 dataset, which is b**ased on regional datasets** provided by tuna RFMOs (aggregation of data provided by each country, see Figure 3 below), aims to keep data as close as possible to these source datasets, and therefore does not apply any extrapolation rule. However, it applies a number of harmonization rules. At this stage, **level 0** is provided by IRD with harmonization rules that need to be reviewed and endorsed by FAO together with relevant t-RFMO FIRMS partners. Indeed, IRD made some choices and there is likelihood that choices made by FAO could be different. This might lead to slightly different versions of level 0. This explains why the title of Level 0 datasets produced by IRD until now highlight the term 'IRD Level 0', which reflects one combination of rules among other possible ones.

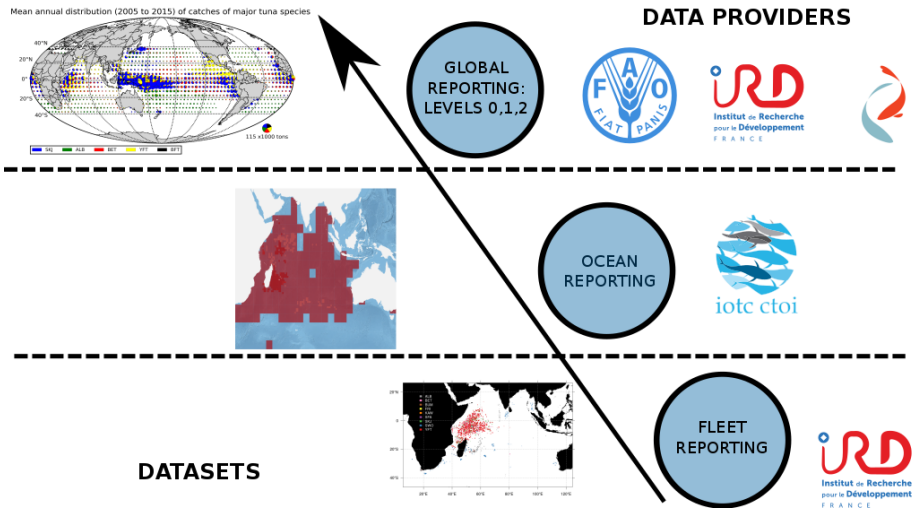

Figure 3: different stakeholders and different levels of processing

4.3.2 Workflow actors, roles and contributions

The success and sustainability of the FIRMS Tuna atlas requires a data management plan (DMP) where actors are clearly identified with roles and expected contributions.

The actors that are required for the Tuna Atlas annual update are:

- t-RFMO data managers
- CWP Secretariat and Reference Harmonization Task Group
- IRD tuna scientists and data manager
- FAO Fishery statistician and/or Fishery biologist scientist; and Statistical data manager or GIS data manager.

The below list indicates the range of expected contributions for each actor:

- The CWP Reference Harmonization Task group will:
  - **provides the specifications** of CWP standards to data formats (data structure definitions, reference Codelists to be mapped with) to be used by t-RFMOs for yearly dataset exchanges (this assumes an endorsement and enforcement of CWP standards by t-RFMOs).
  - **develops general agreements** with t-RFMOs regarding best practices for streamlining statistical workflows from national, regional through to global level, including update calendars and issues related to the implementation of the reference harmonization standards.
  - **contributes** a reference harmonization registry for t-RFMOs datasets registration and Codelist mappings. Tools shall be provided by CWP to check and guarantee the compliance with CWP standards, as prerequisite for the Tuna atlas yearly update
- The FIRMS steering committee (SC):
  - **agrees** on an annual schedule for primary data releases to be achieved by t-RFMOS, and keeps under review the collation, validation and publication of Tuna Atlas updates.
  - **deposits** DOIs as the corporate author of future jointly owned products
- Both FAO/IRD data managers will jointly:
  - collaborate with t-RFMOs for the production of a joint FIRMS level 0 dataset
  - **elaborate** a metadata template and documentation to be shared by FIRMS Secretariat and used by Tuna RFMOS for describing each dataset shared on yearly basis. This metadata template will be aligned with the technical workflow used for generating the Tuna Atlas.

    - **operate** a task force for the corrective, adaptive and evolutive maintenance of the Tuna atlas technical workflow, including the implementation of CWP standard data structures.
- t-RFMOs FIRMS partners will:
    - **provide** access to their public datasets including: Nominal Catch, Gridded catch and effort, Codelists, Codelists mappings with FAO / CWP codelists (with contribution of CWP reference harmonization task group) and metadata associated to each data set. It is understood here that data collation across t-RFMOs will be based on CWP standard for reference harmonization and operated according to the annual data release schedule agreed upon under FIRMS SC. The compliance with CWP standards (based on Quality Assurance controls) should be ensured either by t-RFMOs or by FIRMS Secretariat. Indeed, missing primary data or changes in data structures will block the generation of level 0 and may have adverse impacts on the Tuna Atlas coverage. Mechanisms for gaps management should be sorted out.
    - **fill** metadata associated to each data set, based on the metadata template provided by the FIRMS Secretariat and jointly elaborated by FAO and IRD.
    - **provide** a yearly update of mapping rules developed (maintained jointly by t-RFMOs under the FIRMS partnership) by reviewing IRD mapping rules as a first basis.
    - **collaborate** with IRD and FAO for the production of a joint FIRMS level 0 dataset.
- IRD will:
    - **generate** and provide updated Level 0 datasets ("as usual"). To achieve this task, the availability of primary data to be delivered by RFMOs every year is a prerequisite.
    - **share** and **disseminate** the mapping rules originally established between international and tuna-RFMOs classifications which are required to generate the Level 0 datasets.
- FAO will:
    - **maintain** the FIRMS Tuna Atlas map viewer (GIS data management) on behalf of FIRMS Secretariat. Improvements of this application will be discussed by FIRMS SC and adjusted accordingly, in ways similar to the existing FIRMS web components (website, stocks & fisheries map viewer).

Schedule: The yearly update of the FIRMS Tuna Atlas Level 0 is published according to calendar and rules agreed under the FIRMS Tuna Atlas governance. This is expected to form part of the data sharing agreement as would be reflected in Annex 3 of FIRMS Partners.

5. FIRMS Tuna Atlas business model, proposed governance, status of the agreements

5.1 Business model

FIRMS FSC10 opted for a Public-not-for-profit-Private partnership business model for the FIRMS-GRSF expanded Partnership.

At present the FIRMS Tuna Atlas operates as a Public-only partnership based on the currently involved organizations. Under this model, the involved public Partners contribute their respective in-kind contributions for the operations of the Tuna Atlas.

It is noted that nothing would prevent the FIRMS SC to opt for a model similar to the GRSF case, if e.g. additional NGO partners such as the International Seafood Sustainability Foundation (ISSF) were considered beneficial.

5.2 Proposed governance

The FIRMS Partnership Agreement contains a governance model describing the roles of the intergovernmental Partners, who have voting rights. The FIRMS Collaborative Arrangement is a new instrument which allows membership of 'resource' partners who can support FIRMS objectives and benefit from the products jointly achieved. 'Resource' partners can be members of FIRMS Technical Working Groups and as such have advisory role, but no voting right.

As regards the FIRMS Tuna Atlas, the above FIRMS Governance model will be applied as follows:

- **FIRMS will take 'Governance leadership' of the FIRMS Tuna Atlas,** in accordance with FAO policies[10]. The FIRMS Steering Committee will take decisions regarding strategic orientations and funding, the data standardization and content related activities, the services and who can access them and the responsibilities for daily administration;
- **IRD will become FIRMS member/collaborative institution** (with no voting right) as defined by a specific Collaborative Arrangement (see meeting doc, FSC10/2017/4 "IRD-FIRMS Collaborative Arrangement"). IRD becomes FIRMS collaborative member with the objective to efficiently support the yearly update of the Global Atlas of Tuna and Tuna-like species (herewith referred to as 'FIRMS Tuna Atlas'), and to foster research and development for enhancing collation and dissemination services of tuna fishery statistics. IRD provides human resources (expertise: a data manager to execute the yearly update and a statistician to discuss corrections applied to data) and material (code, data and related metadata).
- The FAO Secretariat establishes the **iMarine MoU** with CNR-ISTI for securing the exploitation of the D4Science Infrastructure in support of the iMarine initiative. Under this MoU, FAO will establish a **Service Level Agreement** (SLA) contract with CNR-ISTI for the operations of the FIRMS Tuna Atlas.
- A FIRMS **Tuna Atlas Technical Working Group** (TA-TWG) will be established with proposed members being t-RFMOs, FAO and IRD. WCPFC (FIRMS Observer) and CNR-ISTI might be invited to attend such TWG as appropriate. This TA-TWG will formulate technical advice for consideration by the FIRMS-SC.

5.3 Status of the agreements with data providers and technical suppliers

FAO/FIRMS Secretariat developed different types of agreements according to the governance model. These documents (MoU, SLA, Collaborative Arrangement) are at a different level of finalization, specifically:

[10] Extract from FAO Terms (http://www.fao.org/contact-us/terms/en/):

Third-party content—FAO does not necessarily own each component of the content contained within the work. FAO therefore does not warrant that the use of any third party-owned individual component or part contained in the work will not infringe on the rights of those third parties. The risk of claims resulting from such infringement rests solely with you. If you wish to re-use a component of the work, it is your responsibility to determine whether permission is needed for that re-use and to obtain permission from the copyright owner. Examples of components can include, but are not limited to, tables, figures, or images.

Non-FAO content: The FAO website may include third party copyright material for which rights and permissions must be obtained from the copyright holder(s) indicated.

- The FIRMS-IRD Collaborative Arrangement is under review by IRD; a final draft is expected to be validated by IRD and presented at FSC11. (Meeting doc FSC10/2017/4)
- The FAO-CNR MoU: a final draft is progressing in the FAO clearance process pipeline; final clearance is expected to be available by the FSC11, with the MoU possibly signed by FAO and CNR-ISTI. The purpose of this MoU is to provide a framework for strategic collaboration between the Parties aimed at (i) securing the current exploitation level of the iMarine platform, powered by the D4Science infrastructure; (ii) integrating and coordinating efforts to increase the exploitation of the iMarine platform; (iii) fostering the development of and coordination with relevant initiatives; and (iv) considering possible options to ensure the continuity of the iMarine initiative. (Meeting doc. FSC11/2019/11)
- The FAO-CNR SLA on FIRMS Tuna Atlas: a first draft is ready and is being circulated in the FAO corporate offices; the SLA is formally expected to enter into force on 1st March 2010, and further reviews are planned in particular related to actual services requested and their costing.

The above agreements will enable to assign efforts to the maintenance and further development of the FIRMS Tuna Atlas under specific arrangements, both in-kind and under future possible funding opportunities from projects or FAO regular program activities.

The proposed CWP standard for Reference harmonization will be presented at CWP-26 and is expected to be endorsed. Draft implementation guidelines will also be discussed. Such guidelines could be matter of a pilot, and the Tuna Atlas could be proposed as such pilot.

t-RFMO FIRMS Partners are expected to revise their Annex 2 as necessary as regards the provision of datasets for the Tuna Atlas. Such revised Annex 2 will constitute the FIRMS Tuna Atlas data sharing agreement for the exchange of tuna fishery statistics between t-RFMOs and FAO.

The FAO Secretariat has presented the FIRMS framework to WCPFC and SPC including through development of case studies, and has invited WCPFC to join FIRMS.

6. Envisaged resources to support the FIRMS Tuna Atlas

As a public partnership business model, the resources to support the FIRMS Tuna Atlas are spread among the FIRMS partners:

- the FAO core budget will support the maintenance of the infrastructure operations through a SLA with CNR-ISTI, for a basic/core configuration; depending on evolution and specific demands for use by Partners, other sources of partner budgets would be sought;
- in-kind contributions are expected to be pledged under the FIRMS partnership and collaborative arrangements, and the iMarine MoU between FAO and CNR-ISTI):
    - FAO: involvement of a fishery statistician / fishery biologist scientist for the validation of rules for production of Level 0 dataset, and of a GIS-R data manager for technical workflow and viewer support;
    - IRD: involvement of a data manager (few days a year), and legacy of previous funding
    - T-RFMOs: staff time for the production of RFMO datasets and associated metadata according to the CWP standard for reference harmonization; participation to TA-TWG;
    - CNR-ISTI: in-kind contributions as part of the iMarine MoU with FAO for the operations of the Tuna Atlas VRE until February 2020, and for the operation of the D4Science infrastructure for the next 5 years.

Specific projects such as EU H2020 BlueCloud (see Annex 4) will allow to fund performance enhancements and new features.

7. Ownership and citation of the FIRMS Tuna Atlas

FAO being custodian organization for the FIRMS Tuna Atlas, the established FAO policies for Partnerships[11] and Terms and Conditions[12] for the use of FAO Web content will apply for ownership and citation of content. Additionally, for third-party content and software assets, the FIRMS Collaborative Arrangement makes provision in Annex 2 for identifying the material, its provenance and the conditions under which it is made available to FIRMS. However, the FIRMS SC will be invited to consider applying best practices for citation such as the publishing in Zenodo of Digital Object Identifiers (DOIs) for each FIRMS product (datasets, software, data): title, authorship, and licensing, branding, ownership.

7.1 Data / content citation

As per recommendation of the CWP t-RMFO workshop on CWP standard for reference harmonization, the bibliographic citation for the FIRMS Tuna Atlas map viewer will be:

Bibliographic citation	**email it**
© FAO 2018. Fisheries and Resources Monitoring System (FIRMS). Global Atlas of Tuna and Tuna-like species. In: Fisheries and Resources Monitoring System (FIRMS) [online]. Version 1.0. Updated dd/mm/yyyy. [Cited dd/mm/yyyyy]. http://firms.fao.org/tuna-atlas	
Source of information: Tuna Regional Fisheries Management Organizations are the provider of data sets and associated reference data. See "Sources citations" for full references.	

Sources citations should indicate that Tuna Regional Fisheries Management Organizations are the provider (source) for primary regional data sets (collated by t-RFMOs from countries) and associated reference data.
Example of Source citation for full reference:

- "© ICCAT 2019. Nominal Catch Information. https://www.iccat.int/en/accesingdb.html [Accessed 21 March 2019]"
- Licence: the derivative Level 0 dataset published under a 'Creative Commons Attribution Non Commercial 4.0 International' license can be downloaded through the FIRMS maps viewer.

Additional citations should be included to the FIRMS Map viewer, such as derived base maps in use particularly when not owned by FAO. The GIS sources in use currently or likely to be used at the time when the Atlas will be officially released are described in Annex 3 and will be properly acknowledged. **A FAO disclaimer** will also be required regarding the addition of country/jurisdictional area base layers: "The boundaries and names shown and the designations used on this map do not imply the expression of any opinion whatsoever on the part of FAO concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers and

[11] http://www.fao.org/partnerships/en/
[12] http://www.fao.org/contact-us/terms/en/

boundaries. Dashed lines on maps represent approximate border lines for which there may not yet be full agreement."

Best practices for citation: for what regards the citation of previous data (and related bibliography), according to FAIR data management principles, citation rules are defined by the metadata attached to the DOI of the resource. This includes the ownership and the licence. Here is an example of good practice (BlueBridge and OpenAire connect) to cite Tuna Atlas datasets and any data extraction (from the viewer): current IRD Level 0 DOI and related citation guidelines and exports. Assigning DOIs to the Tuna Atlas research products will thus foster the citation and facilitate the compliance with partners expectations.

7.2 Software ownership and acknowledgment

The Software components and R & SQL code deployed on the D4Science infrastructure which enable the Tuna Atlas Spatial Data Infrastructure and, thus, the FIRMS Tuna Atlas are bundled together and stem from a mix of authorship including IRD, FAO, CNR and third-party software. These are all licensed as Open Source components (see details in Annex 3). Some of these software have already been assigned DOIs to foster Open Science and citation of these products (eg OpenFairViewer software). DOIs come along with dedicated metadata elements which specify the ownership and acknowledgement. When available, DOIs of underlying software should be displayed on FIRMS website.

Ownership of software components is acknowledged depending on each software licensing types (copyleft or copyright) and specificities. For example, a software licensed with open source MIT license (copyright license) will be acknowledge with copyright.

The D4Science e-infrastructure is acknowledged as hosting the current Tuna Atlas.

8. Research and development perspectives

8.1 Potential research topics and state of the art

The availability of an up-to-date FIRMS Tuna Atlas offers interesting research perspectives regarding tuna fisheries. The resources hosted by the Tuna Atlas VRE potentially allows to support a variety of research topics in support to fisheries management. Examples include:

- assess the size, spatial extent, temporal evolution, and characteristics of tuna fisheries across oceans, e.g. estimation of global fishing capacity of tuna fisheries and understanding of movements of effort among oceans, the increasing use of Fish Aggregating Devices in purse seine fisheries;
- analyse patterns in tuna ecology, e.g. associative behaviour with floating objects and animals (Sequeira et al. 2014[13]) or in the vicinity of seamounts (Morato et al. 2010[14]);
- compare nominal CPUE time series between fleets across oceans as well as with biomass outputs from stock assessment models;

[13] Sequeira AMM, Mellin C, Floch L, Williams PG, Bradshaw CJA (2014) Inter-ocean asynchrony in whale shark occurrence patterns. J Exp Mar Biol Ecol 450:21–29

[14] Morato T, Hoyle SD, Allain V, Nicol SJ (2010) Tuna Longline Fishing around West and Central Pacific Seamounts. PLoS ONE 5:e14453

- assess the global effect of climate change on ecosystems diversity and sustainability through ecosystem models based on global catch and effort data (e.g. SEAPODYM project)
- support research on rights-based management in a multilateral environment, including Fishery Performance Indicators, and the potential impacts of closing the high seas
- address the need for new management approaches concerning By-catch and Discards
- promote collaboration between t-RFMOs by identifying issues in data collection and processing across different fishing areas, e.g. issues in longline size-frequency data (Geehan & Hoyle 2013[15]);

Research topics rather related to IT innovation and OpenScience can also be tackled. These can concern the scope of **data** to be managed with further expansion of content, integration of new variables describing tuna and related fisheries activities such as tagging, biological data, FADs, AIS/VMS..., or environmental variables describing physical and chemical parameters in related tuna habitats. As well:

- the design and implementation of FAIR guiding principles for Fisheries data management, based on agreed licensing and citations;
- give more visibility to the data and more transparency to the data processing steps driving to the datasets used as input in assessment models;
- foster participation and involvement of researchers from coastal countries to the assessments through enhanced understanding of the use of their data.

Considering this potential, the FIRMS Steering Committee could strategically orientate availability of data and services and strategic collaborations.

8.2 Proposed collaboration for FIRMS raised tuna catch and effort statistics (Level1) and for expanding the tuna atlas with additional variables (Level2)

Level 0 is just a step in the workflow for research actors such as IRD to produce more elaborated datasets known as Level 1 and Level 2 datasets. These levels are generated by applying corrections to the level 0 (the processing only consists in SQL queries). Indeed they are known issues in the quality of data delivered by tuna RFMOs datasets (depending on years, areas, flags, species, fishing gears). The corrections currently applied reflect the understanding of Fisheries according to IRD scientists[16]'s view points. Other researchers might disagree and apply other corrections. At this stage, details for the production and distribution of such additional derivative products are out of scope of this FIRMS arrangement. However, at the CWP tuna-RFMO Reference Harmonization workshop, t-RFMOs expressed interest in contributing the parameters used in their science committees for raising datasets to total catches. A collaboration could be envisaged under FIRMS among interested t-RFMOs, FAO, IRD and other invited scientists to develop jointly Level 1 and Level 2 datasets which could eventually be owned by FIRMS. In its FIRMS collaborative arrangement, IRD is opened to such collaboration and for sharing the IRD raising factors and related assumptions and processes considered for each tuna fishery worldwide, and used for generating the Levels 1 and 2 data layers. These will be available for review, modification or uptake by t-RFMOs under the Partnership.

Such collaboration could as well produce research outputs such as **data papers** co-authored by IRD, RFMOs, and FAO to discuss benefits and limitations of levels 1 and 2, and to foster citation of the datasets in scientific articles.

[15] 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, 30p

[16] IRD scientists involved so far were mainly Alain Fonteneau and Emmanuel Chassot

8.3 Tools for scientific collaboration

The ***Tuna Atlas VRE***, inherited from the BlueBRIDGE project, offers access to datasets produced including tuna RFMO's primary datasets and (at current) IRD level 0, 1 and 2 datasets. It is administered by FAO/IRD. This Tuna Atlas VRE includes the map viewer to facilitate data browsing and querying by data managers, and additional services that can be used especially for fostering discussions on Tuna data statistics and knowledge sharing between FIRMS partners e.g. on data raising factors. These services include:

- a workspace: to store data and share it with one or more users;
- a social network: to publish posts of interests for the VRE members;
- an RStudio server: to perform online statistical analyses with the R software for statistical computing;
- an RShiny server to host interactive dashboards;
- a ShareLatex environment: to write papers with LaTeX language in collaborative way.

The prospect of the BlueCloud project (Annex 4) raises a potential framework for achieving some of these objectives, and the FIRMS SC could provide guidance in this respect.

9. Proposed workplan

The workplan presented in this section describes milestones and main activities required in the short term for the release to the public of the FIRMS Tuna Atlas, as well as in the medium term for the enhancement of the workflow. Part of this workplan is likely to occur under the BlueCLOUD project, probably the enhancement of R code, and any steps that would be identified to proceed collaboratively regarding Levels 1 and 2.

Main Milestones:

Short term

1. CWP reference harmonization standard adopted, and pilot phase agreed upon for developing CWP implementation guidelines for Reference Harmonization. The Tuna Atlas is one of the selected pilots.
2. FAO-IRD detailed review of the processing applied to the IRD Level 0 dataset. Findings are shared weith t-RFMOs and decisions are made whether for the public release of the dataset, IRD's Level 0 is chosen or corrections have to be brought as per FAO review.
3. The tuna atlas is publicly released as FIRMS Tuna Atlas

Medium term

4. Enhancements to the workflow – implementation of streamlining rules as per CWP Reference Harmonization standard
    a. t-RFMOs design their datasets according to draft guidelines, and provide feedback; draft guidelines are evolving accordingly
    b. IRD and FAO revise the collation code according to implemented CWP ref harmonization guidelines, in sync with t-RFMOs
5. Steps for uptake of Levels 1 and 2 ?

The work plan establishes the date and the goals to achieve the **yearly update, including contributions required for each actor**. Hence, we may consider this work plan constitutes the drafting of a global tuna fisheries data management plan (DMP) that may defined as sequence of actions. A diagram was drafted and could be a valuable source for understanding implications of all actors and requirements for the yearly update (see Figure 6 in Annex 2).

The first year will require a prior step (ignition phase). Outcomes of the ignition phase will enable yearly updates. The schedule of each yearly update id ruled by:

- **the deadline to get tuna RFMOs primary data complying with CWP standards** (according to the decision made during FIRMS SC meeting),
- **the deadline for FAO and IRD to execute the code and generate global datasets:** at a period to be defined (can only occur few months after primary data delivery by t-RFMOs).

The below table summarizes the sequence of steps required to carry out the yearly FIRMS Global tuna data management (Level 0 dataset).

#	Actor	Action Description	Expected on <DATE>

0	FAO/IRD	Ignition phase (1st year only). Partners:
		- Review, execute and document the existent material (code and data) provided by IRD, with adjustment if required (to ensure that they are modular enough to plug different versions of corrections (so that level 0 can be customized by integrating new corrections),
		- Specify expected limitations and improvements with what regards:
		  ○ extra work if any change in current data structures (cf CWP)
		- first official release of R and SQL code will be disseminated with a DOI (Github & Zenodo)
		  ○ generation of global datasets and related codifications
		  ○ loading datasets in the database
		  ○ plug of data services on top of the database
			according to the decision made during FIRMS SC meeting and CWP meeting

1	CWP Reference Harmonization task group	- Writes and shares CWP standards specifications for (datasets, codelists, mappings) with possible yearly adjustments based on review with t-RFMOs and IRD - Provides i) a service to be used by t-RFMOs to check data structure & content compliance with above standards, ii) a service to submit the data	
2	t-RFMOs	Review code lists and mappings compiled under the FIRMS Tuna Atlas and share RFMOs datasets complying with CWP standards. This data exchange should be ideally done using the submission service made available by CWP (FAO) after data has successfully passed quality control. Upon submission, CWP Reference Harmonization task group is notified, and	

		regional datasets are made available in a standard way (catalogue-based) for use by FAO/IRD data managers in the Tuna Atlas workflow.	
3	FAO/IRD	Adjust R code in case of CWP standards changes, with new release of R and SQL code (new DOIs for the code)	
4	FAO/IRD	Trigger the Tuna Atlas technical workflow, with publication of datasets (regional and global), with DOI (or version DOIs) assigning, and dissemination through the FIRMS Tuna Atlas catalogue and viewer. Each new release of the code resulting from the yearly update will be published on Zenodo as well to ensure reproducibility.	

Annex 1: Tuna Atlas, related applications and users profiles

The FIRMS Tuna Atlas will rely on services which actually refer to a set of services already delivered by pre-existing software components that need to be properly identified and defined to understand the perimeter of the FIRMS Tuna Atlas. This set of components defines a spatial data infrastructure (SDI) compound by:

- a **database**, *used to store data*: This database, named 'tunaatlas', implements the Sardara database model, provided and owned by IRD. It is powered by PostgreSQL open-source software (https://www.postgresql.org), currently administered by IRD, previously hosted by IRD and presently hosted on the D4Science e-infrastructure. In this document, we refer to it as ***Tuna Atlas database***.
- a **catalogue,** referred in the present document as ***Tuna Atlas catalogue*** and compound by:
  - a **geographic data server**, *used to serve data as service (enabling data access protocols)*: it is powered by GeoServer open-source software (http://geoserver.org), administered by FAO/IRD. In this document, we refer to it as ***Tuna Atlas GeoServer***.
  - a **geographic metadata server**, *used to describe and make data discoverable by users and map applications*. It is powered by GeoNetwork open-source software (https://geonetwork-opensource.org, see screenshot in Figure 1), administered by FAO and IRD. In the present document, we refer to it as ***Tuna Atlas GeoNetwork***.

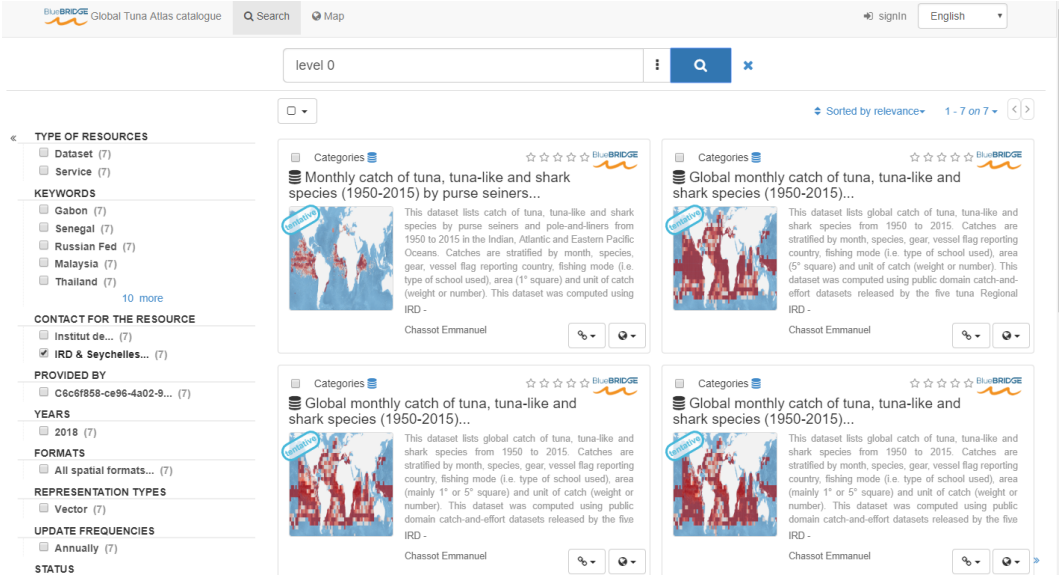

Figure 1: Overview of the Tuna Atlas catalogue (main interface from GeoNetwork and maps overview provided by GeoServer)

- a **technical workflow**, *used by data managers to operate the datasets generation and publication*. This workflow is presently operated with the open-source R software for statistical computing (https://www.r-project.org) and co-developed and maintained by FAO/IRD (data generation being only maintained by IRD at this stage). The SDI web-components are managed in the workflow by means of specialized R packages. In the present document, this is referred as ***Tuna Atlas workflow***.
- a **map viewer:** referred in the present document as the ***FIRMS Tuna Atlas map viewer*** (screenshot in Figure 2) is a gateway for end users to use the main Tuna Atlas data services: search, query and access (data download facility and a map image export) datasets displayed on maps. Each dataset query can be stored (and shared between users) in a persistent and reproducible way by means

of web link sharing mechanism. It, allows as well data managers to embed specific data views in external websites. This viewer is powered by an open source software which implements FAIR principles and is entirely built on top of rich metadata and data services. In this case, both data metadata are delivered by the *Tuna Atlas catalogue*. This map viewer can be also re-distributed (e.g. as embedded product) in t-RFMOs websites upon request as well as through the iMarine e-infrastructure.

-

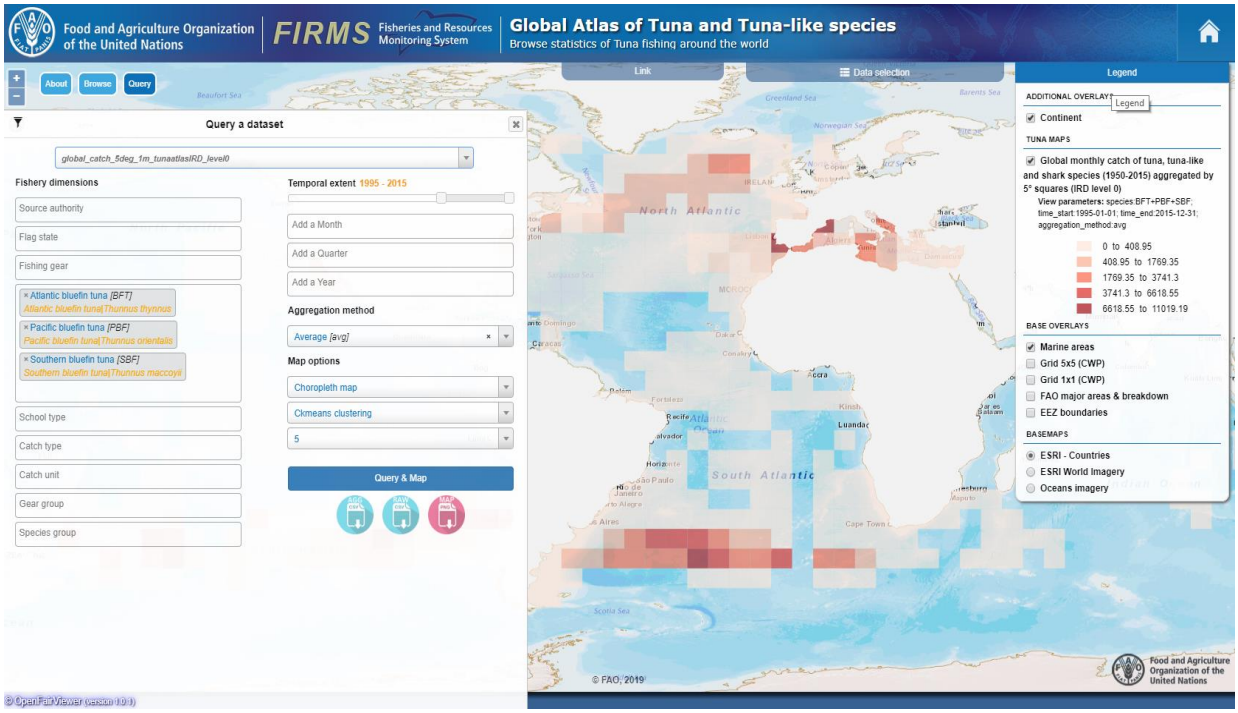

Figure 2: Overview of the FIRMS Tuna Atlas map viewer - with a user query

(Bluefin Tunas, Average catches over period 1995-2015, by 5° squares)

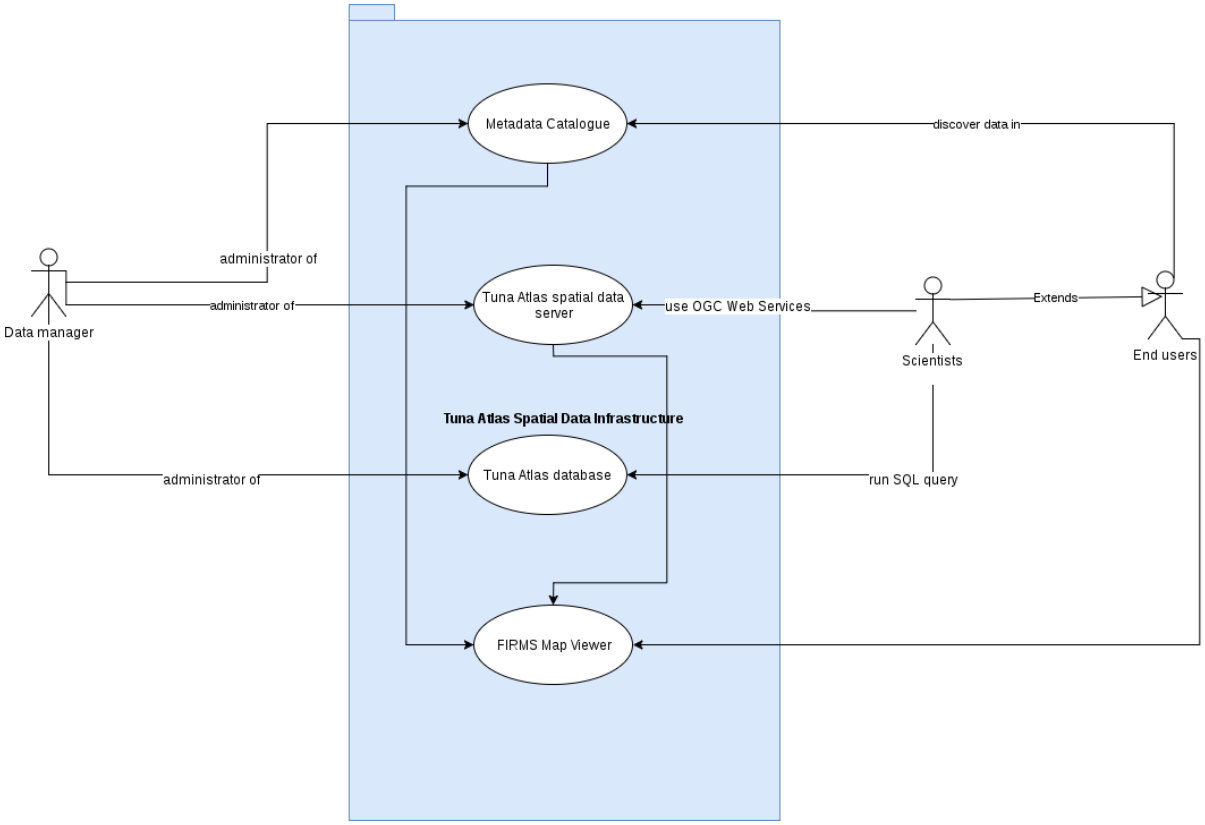

Figure 4: Use cases diagram (application and users profiles)

Annex 2: Tuna Atlas, overview of the workflow

The figure 5a provides a global overview of the workflow to setup and manage the SQL data warehouse, from the extraction of the t RFMOs datasets (heterogeneous, coming from multiple sources) to the management and exploitation of the data warehouse through the web services.

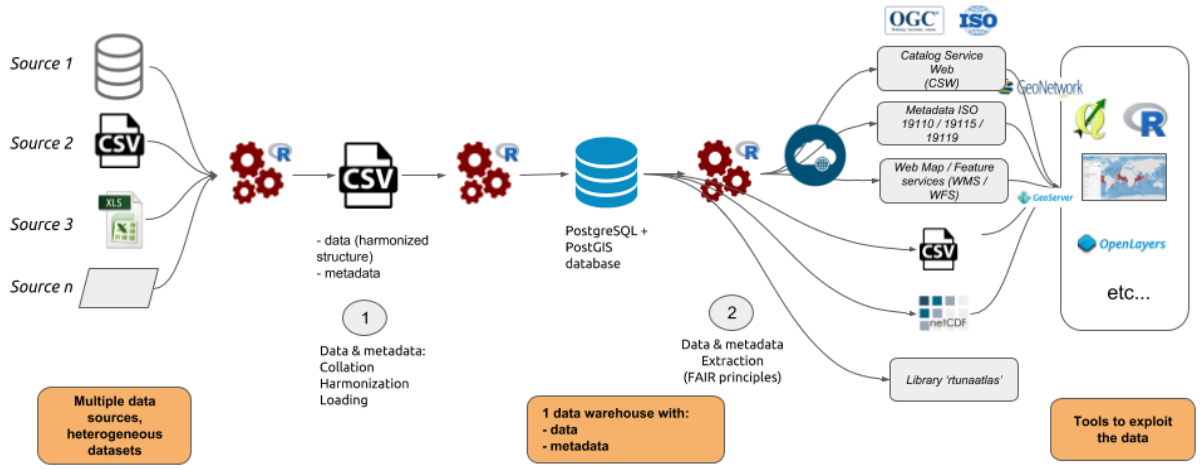

***Figure 5a**: Global overview of the workflow, from heterogeneous datasets to FAIR data services. Full-size figure available here.*

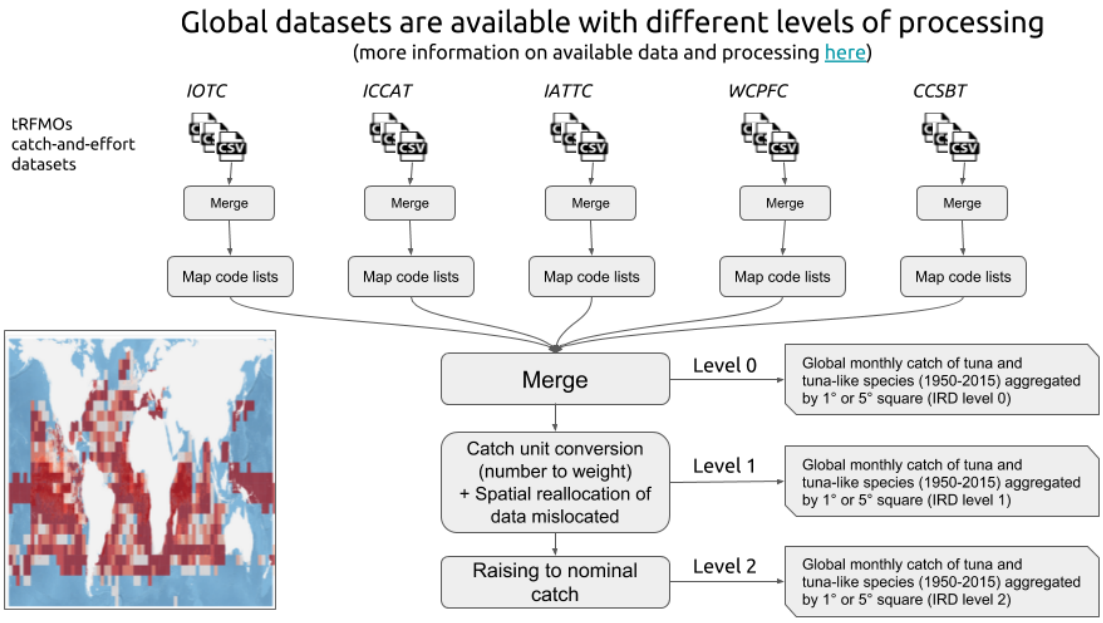

***Figure 5b**: the different levels of processing for Tuna Atlas global datasets (P. Taconet, 2018)*

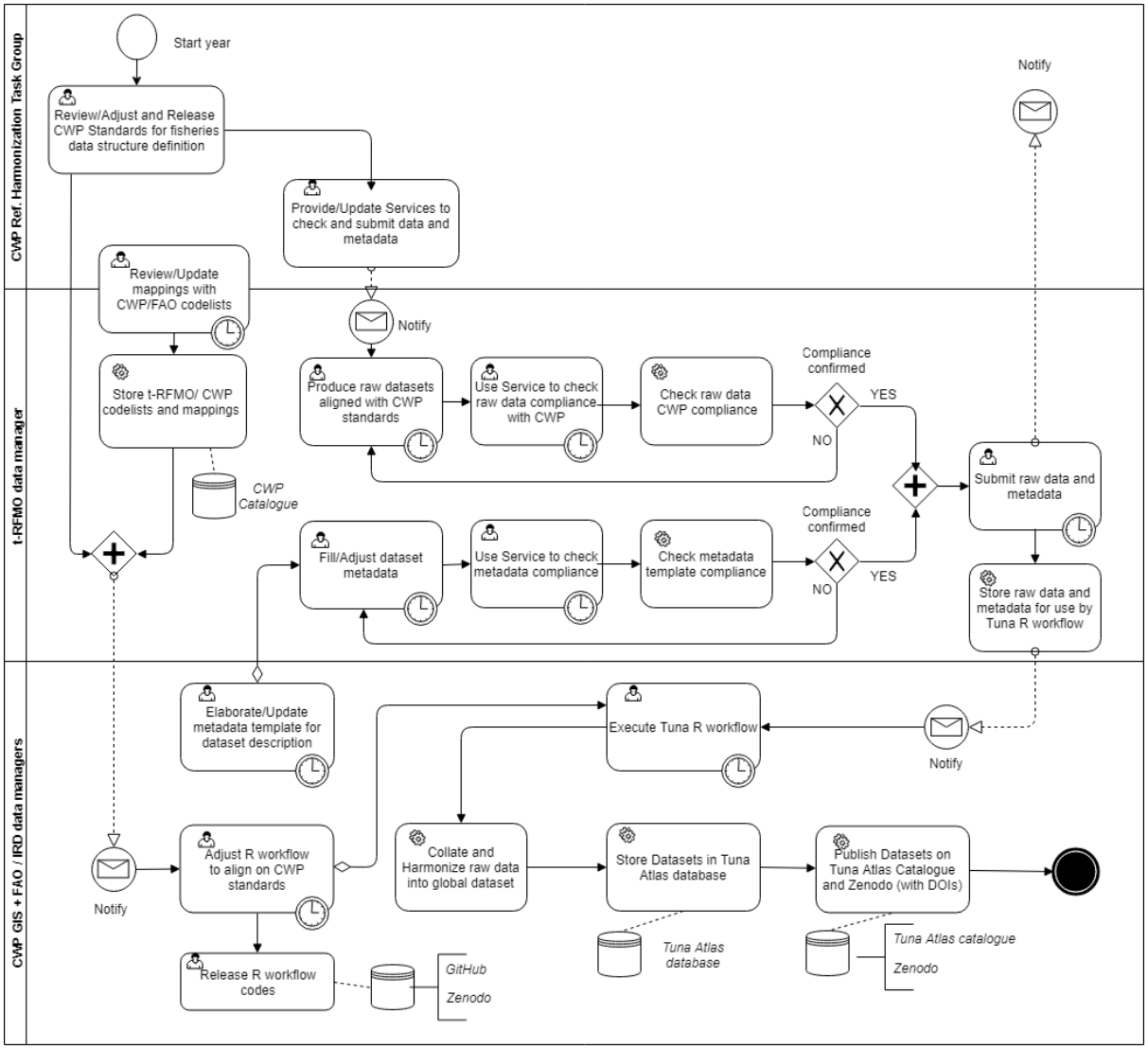

Figure 6: Draft Tuna atlas data management plan diagram

Annex 3: Tuna Atlas, data and software components ownership

The content of the FIRMS Tuna Atlas catalogue, available as of April 2019, will be made of different components and ownerships:

- **Metadata**: currently filled by IRD, these metadata should be reviewed and enriched by other partners (at least for regional datasets),
- **Data**:
  - "primary datasets" refer to datasets provided by tuna RFMOs (primary data providers, even if they are actually aggregations of datasets provided by countries) which are the input of the workflow :
    - nominal catch,
    - catch,
    - efforts,
    - some reference datasets (codelists) and mappings, that in a medium or long term should logically move to a catalogue tailored to CWP reference datasets (the foreseen "CWP Catalogue"),
    - metadata associated to each dataset
  - "global datasets" refer to datasets of the Tuna Atlas:
    - datasets provided by IRD will be made available with DOIs through data repositories (eg Zenodo) with metadata elements indicating the IRD ownership:

      → Reference mappings as handled for tuna datasets, were also produced by IRD (not on Zenodo for now)

      → Different levels of processing (levels 0, 1, 2): some are already available on Zenodo (eg IRD Level 0).

      → datasets provided by FAO: at present, FAO has not provided any dataset, there is however a call by t-RFMOs for FAO to add FAO's conversion factors for commodity to live weight

      → datasets provided by FIRMS Tuna Atlas:

      - if new level 0: the should be stored on Zenodo with a DOI and co-ownership of all partners.
      - if FIRMS reviewed level 1 and 2: the datasets should be stored on Zenodo with a DOI and co-ownership of all partners.
  - Geographic / GIS datasets
    - Ocean base layers:

      → ESRI World Imagery[17]: Copyright:© 2009 ESRI, i-cubed, GeoEye

      → Oceans imagery[18]: Made with Natural Earth. Blended depth colors and relief shading of the ocean bottom derived from CleanTOPO2 data.

[17] https://www.arcgis.com/home/item.html?id=21b4ba14d9e5472d97afcbb819f7368e

[18] http://www.naturalearthdata.com/downloads/50m-ocean-bottom/50m-ocean-bottom

→ Ocean and Sea names[19]: Copyright: © ESRI, DeLorme, HERE, GEBCO, NOAA, National Geographic, Geonames.org, and other contributors.

- Country/Jurisdictional area base layers:

  → ESRI Countries[20]: Copyright:© 2009 ESRI, AND, TANA, ESRI Japan, UNEP-WCMC

  → Marine Regions database[21]: Lonneville, B.; Oset García, P.; Schepers, L.; Vanhoorne, B.; Hernandez, F. & Mees, J. (Flanders Marine Institute) (2019). MarineRegions.org

  Notwithstanding, the use of country/jurisdictional area base layers is very likely to be changed in order to comply with FAO policies on the mandatory use of official UN Maps and boundaries provided by the UN Geographic Information Section; in order to guarantee that UN-FAO neutrality and impartiality principles are not compromised. As of today, sources such as ESRI Countries and Marine Regions database are not endorsed by UN as authoritative sources.

- **Software code and packages**:
  - SQL code: "Sardara" is the legacy name of the SQL database provided and owned by IRD. Sardara is a multi-dimensional data warehouse where regional and global datasets, including spatial data (in Postgres / Postgis). The SQL code of "Sardara" physical model (Tuna Atlas Data Definition and Data loading) are kept in IRD rtunaatlas github repositories.
  - R code rely on CRAN or Github packages with main releases made available on Zenodo. Key packages such as *geometa*, *geonapi*, *geosapi*, *ows4R* and *zen4R* (third-party open-source software), are available under the open-source Copyright MIT license. The code of the Tuna Atlas workflow relies on IRD github repositories like rtunaatlas or R_Metadata. Ownership of code is mentioned in related DOIs.
  - Javascript code: The Tuna Atlas map viewer relies on the open-source web-mapping software OpenFairViewer (third-party open-source software). This software is released under the open-source MIT license, and not owned by FAO (Specific copyright non-intrusive statement for OpenFairViewer). However, the configuration of the Tuna Atlas map viewer instance and its custom website design is owned by FAO, and thus FAO holds a shared copyright on the map viewer application in accordance with FAO Terms. The FIRMS Tuna Atlas map viewer is cited as a FIRMS authored product, which constitutes, along with Zenodo, the primary channels for dissemination of Level 0 data. The map viewer is designed for redistribution (e.g. as an embedded product) in t-RFMOs websites, and is already integrated with the iMarine e-infrastructure.
- **FIRMS tuna atlas** will be built on top of the Tuna Atlas spatial data infrastructure which is made of following softwares:
  - Postgres server to manage IRD Sardara SQL database, hosted by D4science e-infrastructure (previously funded by EC BlueBRIDGE project). IRD hosts a server as well ,
  - Geonetwork and Geoserver instances contain the metadata and datasets produced by IRD, hosted by D4Science in the Tuna Atlas VRE, and exploited by the FIRMS Tuna Atlas map viewer.

[19] http://www.arcgis.com/home/item.html?id=0fd0c5b7a647404d8934516aa997e6d9
[20] https://www.arcgis.com/home/item.html?id=9557906152ac4a51b568c33fb2c5ad04
[21] http://www.marineregions.org/disclaimer.php

- Http server to host the map viewer, currently offered by D4science
- Rstudio server (Optional) to execute R code, currently offered by D4Science

The hosting of the current Tuna Atlas SDI components is currently provided by the D4Science e-infrastructure. FIRMS Tuna Atlas might be as well hosted by D4Science or by other infrastructures (IRD, AWS, OVH..).

Annex 4: Tuna Atlas, brief description of objectives in the BlueCloud project

FAO, IRD, CNR, and FORTH are consortium members in a project proposal to develop the BlueCloud; this is poised to become a part of the European Open Science Cloud (EOSC) that is promoted by the European Commission. The proposal is under review, and if approved will be funded for 3 years with a total budget of €6M. The project will fund 5 demonstrators, including one for fisheries.

The demonstrator related to fisheries builds on the existing Tuna Atlas VRE (discussed in this document) and the GRSF VRE and overall on the iMarine-D4Science platform, and will improve or add features for:

- The Tuna Atlas: manage public fisheries statistical data from ingestion, through harmonization, to publication. The Atlas will offer state-of-the-art monitoring and assessment models, spatial data and mapping tools. Special attention will be given to improve the management of data in a multi-user and multi-purpose environment, to ensure confidentiality and quality of datasets.
- Expanding the Global Record of Stocks and Fisheries (GRSF) with the results of approved status assessments and datasets of fisheries, including those from the Fisheries Atlas.
- Fisheries data services through user dashboards and query panels, developers APIs, and Jupyter notebooks for advanced and registered users. The notebooks are especially useful to capture an entire workflow (of e.g. a working group) from data upload to data analysis, and the notebook can be used to run with a new datasets, e.g. if new data become available.

The expected impact beyond FAO, IRD and the t-RFMOs are improved cloud services for:

- Data services in support of SDG14.4.1 analysis, both using specific well known data and algorithms, and by providing a generic analytic environment;
- On-line training tools for the same SDG for stakeholders to train and learn on the SDG process and data needs;
- The analytical and reporting needs for SDG 13 fisheries related issues under various climate change scenarios.