RDA-WG on Fisheries Data Interoperability

[WG Charter](#_iyy46d9jkvub)

[Value Proposition](#_e2dqc8cikj4)

[Engagement with existing work in the area:](#_pyucqjgcul5w)

[Work Plan](#_pgvcbk9709zx)

[Work plan components](#_2qh9yawqomc0)

[Inventory of existing formats to support solutions (months 1-4)](#_5mbcf518ebkr)

[Defining the reference models (months 3-8)](#_go2f40i34sgs)

[Improve and test the models iteratively (months 7-12)](#_2izn46xlnwi3)

[Promotion of the RDA FDI Model and Reference Adoption](#_9bk0coasa00t)

[(months 8-18)](#_d12thy3ojkmy)

[WGFDI operation](#_ymfyn9irg9jt)

[Form and description of final deliverables](#_x26extpccdsl)

[Milestones](#_5a1j7xpwmzio)

[Communication and outreach](#_76mu3awpo3ki)

[Initial Membership of the FDIWG](#_s42yi7wbx4kk)

WG Fisheries Data Interoperability: Case Statement

# WG Charter

The general **objective** of the Fisheries Data Interoperability Working Group (FDIWG) is to devise a global data exchange and integration framework to support scientific advice on stock status and exploitation that build on fisheries data. Various fisheries data domains utilized in such scientific processes are concerned, including data collected for monitoring control and surveillance, scientific fisheries Data Collection Frameworks, fisheries scientific observers schemes, and statistical or status & trends reporting frameworks. The proposed framework will facilitate the use of de-facto, and preferably open, standards for the identification, description, mapping and publication of fisheries data supporting scientific processes..

More specifically, the fisheries Data WG will **address** the (minimal) metadata requirements to describe fisheries data required for supporting stock assessment and fisheries management. It will also seek to recommend global data standards for topical vocabularies, domain ontologies, and mapping rules and formats (as done for example by [CF Conventions](http://cfconventions.org/) for physical and chemical parameters in oceanography).

Driven by pragmatic considerations, the working group will focus on few selected priority needs expressed by its invited participants, ranging from filling gaps in selected schemes to application of best practices across schemes, through issues of data transformation and harmonization among schemes. In terms of **functionality** and data types, the WG will identify several use cases describing realistic scenarios to **produce and test** fisheries data work-flows. The **result** of the WG recommendations will be captured as a set of best practices.

By including interoperability experts, organisations with standardization initiatives, and standardization bodies, the WG will have key actors to reflect and propose **future governance** of the data framework The focus of this governance is the efficient delivery of interoperability guidelines.

To organize the collaboration and involvement from the community, the WG **co-chair on** Fisheries data structures will oversee the activity of two topical sub-groups, with one **co-chair** responsible for the formulation of a framework for structured fisheries data exchange (data structures), and another **co-chair** responsible for fisheries geospatial explicit data.

To achieve these objectives, the WG will;

1. Promote existing facilities for data sharing on capture, landing, effort, size classes, VMS and production through sharing of structural data definitions. This promotion will be supported by demonstrations of live examples of data sharing;
2. Facilitate access to data by recommending standards such as netCDF, SDMX or UN/CEFACT and assist in adoption of tools and facilities;
3. Recommend existing data tools: Tools for Master Data Management (MDM), database connectors, registries and other assets;
4. Recommend Master Data Management solutions for classifications and multilingual / multi-locale data: The challenge lies in the variety of languages in which the data is stored, and locale specific data types. This requires also mapping between local classifications and regional and global ones.
5. Connect existing data networking initiatives such as
	1. The FAO FiRMS partnership,
	2. The FAO secretariat to the Coordinating Working Party on Fishery Statistics (CWP), which combines 19 global partners such as ICES and IOTC,
	3. The Tuna Atlas initiative (tuna RFMOs, FAO, IRD), to provide examples for storing datasets from various RFMOs within a single gridded data format;
	4. Extend [CF Conventions](http://cfconventions.org/) for biological and fisheries data
	5. EU / DG MARE: DCF, Integrated Fisheries Data Management Programme (FLUX) and INSPIRE directive,
	6. The SDMX community, such as through of Eurostat, FAO, and Worldbank
	7. Unesco’s "International Oceanographic Data and Information Exchange" (IODE) of the "[Intergovernmental Oceanographic Commission" (IOC)](http://ioc-unesco.org/)
	8. Other relevant RDA WG’s and IG’s; such as (Alfabetically) the [Agrisemantics WG](https://www.rd-alliance.org/groups/agrisemantics-wg.html), Data Citation WG, Agricultural Data IG, Geospatial IG Marine Data Harmonization IG, RDA/CODATA Legal Interoperability IG

#

#

# Value Proposition

The WG will provide a negotiation framework on fishery related standards for data storage and exchange structures to improve data analysis. It will benefit organizations in the fisheries sector by providing a reference interoperability framework based on existing initiatives and formats.

In the longer term, implementing a common framework (however small the scale may be) will help to further cultivate a fisheries data ecosystem, based on common tools and services.

* The fisheries data managers and data scientists will have a common and global framework to describe, document, and structure their fisheries data.
* If suitable standards are identified, then the WG can propose generic data storage standards (e.g. for gridded datasets or NetCDF) and services (OGC Web Services for GIS community, facilitating INSPIRE DIRECTIVE compliance..)
* Fishers, traceability organizations, NGO’s, and other data users will have seamless access to a wide range of fisheries data. Data mapping will also ease emergence of new data analyses and knowledge discovery methodologies.
* Other infrastructures data managers and scientists will have the benefit of a reusable data framework. Researchers working on other domains will easily access, reuse and link up fisheries data with their own data.
* Development professionals and policy makers for will be enabled to take informed decisions across multiple data providers.

Expected key impacts of the RDA fisheries Data Interoperability Guidelines

* Reduced costs related to reusing data. The incompleteness of standards (or guidelines) has a cost. Indeed, e.g. data structures can vary a lot for similar data and much time is wasted to transform data from one format into another. Agreement on a set of standards and writing related guidelines is key.
* Increased adoption of existing common standards, vocabularies and best practices related to fisheries data management with new communities, such as regional projects. Increased general awareness about research open data and interoperability standards among the fisheries organizations.
* Enhanced access, discovery (metadata) and reuse of fisheries data, and improved visibility.
* Major fisheries data integration and more effective measure of the of free sharing impact of fisheries data through data provenance attribution.
* Created new opportunities for Data Structure Definition (DSD) and ontology based knowledge management in the fisheries sector.

# Engagement with existing work in the area:

The members of the WG will liaise through their organizations with existing activities in the area of fisheries data exchange and overall activities to foster data interoperability. The engagement will allow the WG to tap into a wide knowledge base of data exchange specialist, and prepare its recommendations that may also be of value to experts beyond the domain of fisheries data exchange, such as legal interoperability and geospatial metadata experts.

* iMarine / BlueBRIDGE: Tuna Atlas use case for RFMOs datasets,
* ICCAT BFT-E Stock Assessment working group to facilitate stock assessment datasets sharing,
* OpenAIRE open data specialists,
* Agroknow network of expertise on open data sharing,
* EGI Engage e.g. for legal interoperability,
* IRD scientific data collection activities,
* FAO and Eurostat SDMX SEIF initiative,
* DG-MARE: FLUX initiative (in particular VMS & elogbooks) and DCF / DCMAP,
* FAO CWP standards for fisheries reference data,
* OGC geospatial standards setting organization.

Through the engagement work, a list of **potential adopters** of the WG products will emerge. Specific statements of interest and priority needs are expected from the invited participants while the WG is established. Examples of interoperable data flows that could benefit from the application of WG reference models and best practices include:

* FAO Fisheries and Aquaculture department:
	+ data ingestion from regional fishery bodies, fisheries organizations or members states; regional databases to support scientific process;
	+ improve the statistical data exchange in line with CWP’s SDMX initiative;
	+ Improve the geospatial data exchange building on CWP’s geospatial standards work group;
* IRD:
	+ improve fisheries observers’ data flows to support regional fisheries bodies;
	+ Improve scientific data flows;
	+ improve the quality of NetCDF metadata;
* EU:
	+ Ease interoperability between FLUX and SDMX;
	+ Ease interoperability between FLUX and FishFrame.

#

# Work Plan

## Work plan components

### **Inventory of existing formats to support solutions (months 1-4)**

The first months after the WG has been established, a consultation of existing formats and activities related to fisheries data will identify:

1. Data formats, existing and proposed,
2. Data exchange needs and examples,
3. Data access and storage existing solutions and development proposals

We will evaluate recommended data exchange approaches for several specific scenarios and select pilot candidates for a demonstration. The selection of these candidates will be in close cooperation with stakeholders and data owners. In this phase, a detailed report of the technical aspects of data sharing approaches will be developed.

Examples of scenarios where the WG can propose data interoperability solutions could be selected for inclusion in the report include:

* Globally established data frameworks interoperability; what are the technical challenges in re-using data collected through e.g. FLUX or SDMX work-flows?
* Improve coverage and re-usability of on-board collected data such as by-catch reports by harmonizing reference data through master data management; can the interoperability of collected data be improved by relying on global reference data for e.g. species names, gear classifications, and area references;
* Legal interoperability requirements; what provisions do exist in current data exchange mechanisms to ensure that the data are properly described from a legal perspective through descriptive metadata on license, copyright, and ownership
* Spatial data interoperability of fisheries geospatial explicit data such as gridded datasets (Tuna Atlas example) through descriptive metadata;
* Identify requirements for additional data formats for activities such as vessel or FAD (Fishing Aggregating Devices) trajectories

The WG will not meet physically, but be consulted on-line with several on-line WG meetings.

This report will be the Deliverable of this phase.

### **Defining the reference models** (months 3-8)

We will develop technical reference models for data exchange based on the inventory above and including possibly Data Structure Definitions for statistical data, and as UML for OGC and ISO standards.

Each model should be open, extensible and, if possible, implementation agnostic. They define how fisheries data can be structured in order to facilitate the sharing of data- and subsets, and how those structured data can be used in interoperable exchanges.

A selected set of DSD’s and UML diagrams or other formalization of fisheries data for exchange, based on the report of the previous phase, will be this activity’s deliverable.

A reference model should address the interoperability issues related to formats, ownership, copyright, data re-use and data quality.

### **Improve and test the models iteratively** (months 7-12)

The models(structure definitions and UML) developed in the previous step will be evaluated against suitable data sets of considerable size from various research organizations. The consortium partners will be asked to provide their real world research sets as a testbed to evaluate each model. This will follow an iterative approach in order to allow improvements.

After the models have been validated, a reference architecture for fisheries data will be implemented. This reference implementation should be based on open source software in order to be usable and improvable by all participating partners. Several implementations of data architectures already exist, and these could be repurposed to also accept the fisheries data models.

* For statistical data;
* For geospatial explicit data;

The implementation has to be generic and flexible enough for being adapted to various purposes. An official release will follow the implementation and iterative improvement phase and demonstrate interoperability between two systems (a producer and a consumer) with a live example of fisheries data.

The evaluation report of the existing reference architectures for suitability to manage fisheries data will be the deliverable of this phase.

### **Promotion of the RDA FDI Model and Reference Adoption**

### (months 8-18)

Promotion activities will include internal and external dissemination about the data structures and architecture. The reference implementation will be accompanied by substantial documentation and use case scenarios in order to increase adoption and encourage contributions.

## WGFDI operation

### Form and description of final deliverables

The deliverables are listed above as activity phase outcomes.

### Milestones

No particular milestones are specified. If needed, the Deliverables of the previous section can be used as milestones.

### Communication and outreach

The entire process will be supported by dissemination activities and community outreach. The dissemination will rely on RDA tools, and include a wiki, documents, and possibly a demonstration site in an EU infrastructure. No developer forum or mailing lists are foreseen.

The outreach will focus on the initiation phase and conclusion phase; the announcement of the activity and pans, the installation of the core team and the resource team, the development of the concrete objectives, and when a result has been obtained, a presentation of progress, and plans for a further development and roll-out phase through the participating members channels.

#

#

# Initial Membership of the FDIWG

The WG organization is specified in the Case Statement. It will be structured with 3 Co-chairs. The first Co-chair will retain overall responsibility on progress and deliverables, and communication with RDA, while the Co-chairs will be responsible for content development and more broadly with technical issues and future collaboration.

Leadership

* Co-chair: Anton Ellenbroek (FAO) - Fisheries data structures
* Co-chair: Julien Barde (IRD) - Fisheries and geospatial data management
* Co-chair: Aymen Charif (FAO) - Statistical data management

Members/Interested (Not formally invited):

* Marc Taconet - FAO Rome - Data Governance and Global fisheries data interoperability
* Donatella Castelli - CNR-ISTI, Pisa - Networking and data interoperability
* Pasquale Pagano - CNR-ISTI, Pisa - Data and infrastructure interoperability
* Yann Laurent - FAO Consultant - Fisheries data exchange and interoperability expert
* Neil Holdsworth - ICES Denmark - fisheries data formats and tools
* Daniel Surany - ESTAT - SDMX Expertise
* Erik van Ingen - FAO CIO Rome - SDMX Expertise, mainstreaming fisheries data in FAO UN statistical data flows
* Fabio Carocci and Emmanuel Blondel; FAO Fisheries - Geospatial data standards expertise;
* FAO ESS - TBC
* Charalampos Thanopoulos - Agroknow Greece - Expert on data interoperability
* Imma Subirats / C.Caracciola - FAO OPCC Rome - Data interoperability experts
* NOAA - TBC
* NAFO - Through FAO FiRMS partnership and CWP (logbook data models)
* David Ramm - CCAMLR Hobart - Fisheries data management expert
* Alicia Mostiero / Dawn Borg Costanzi - FAO Rome - Global Record - Vessel data management expert (UN/CEFACT - FLUX)
* DG Mare - FLUX: Thierry Remy / Eric Honoré (UN/CEFACT business layer standardization)
* DG MAre - DCF: Bas Drukker / Venetia Kostopoulou Venetia.Kostopoulou
* JRC - TBD
* VLIZ - WoRMS Marine species master data, marine georeferences
* [Dimitris Gavrilis](http://www.imis.athena-innovation.gr/en/people/member/23) - Athena RC