منظمة الأغذية والزراعة للأمم المتحدة | 联合国粮食及农业组织 | Food and Agriculture Organization of the United Nations | Organisation des Nations Unies pour l'alimentation et l'agriculture | Organización de las Naciones Unidas para la Agricultura y la Alimentación

COORDINATING WORKING PARTY ON FISHERY STATISTICS

Intersessional Aquaculture and Fishery Subject Groups Meetings

CWP ad-hoc Task Group on "Reference harmonization for capture fisheries and aquaculture"

Copenhagen, Denmark 19-22 June 2017

Author: CWP Secretariat

1- Background

The development of information computer technologies changed the collection' modes of fisheries statistics information from paper to electronic forms. The Excel-based questionnaires which have become widely used enforce specific standards' based reporting formats, however increasingly institutions are acquiring their own information systems with the inventive capacity to produce a multiplicity of dissemination formats. In this entropy context, it appears essential to set a direction for consistency materialized by global standards, guidelines and best practices. These will be precious instruments for addressing the challenge of making data interoperable between data producers and users at national, regional and global levels. In practice, the actions that require some degrees of interoperability include the transferring of data from one repository to another, the harmonization of different data and metadata sets, the creation of new information services such as virtual research environments for data dissemination [1].

The publication of fishery statistics information structured in harmonized datasets is a critical requirement to ensuring time efficient processing of statistical data in support of the best scientific advice and ultimately to improve the fisheries management of marine living resources.

CWP parties collect catch and other fisheries statistical datasets however data do not flow among these entities in a cohesive or standardized way. The exchange and timely submission of collected statistical data from national offices to the regional organizations or FAO remain a struggle for several reasons. Among them the gap in the management of statistical data and metadata between national statistical agencies and the international organizations. Therefore, it is crucial to move towards a harmonization of reference data used across datasets in a rational and efficient way and supported by commonly used standards, formats and software tools.

CWP addressed during the inter-sessional Fishery Subject Group Meeting of February 2015 the need of a unified, coherent and harmonized fisheries data. The reference data need to be structured and be exchanged together with the statistical datasets to enable their identification and interoperability across different organizations' databases.

The interest and incentive of CWP parties on solving these matters have been raised and discussed during several preceding meetings. In CWP 25th plenary meeting held in FAO headquarters (Rome, 2016) participants approved the establishment of the present ad-hoc Task Group on "Reference harmonization for capture fisheries and aquaculture statistics".

In the same context, the collaboration of FAO with the RDA Research Data Alliance during the RDA 9th plenary meeting was an opportunity to increase the visibility of CWP as a responsible institution for fishery statistics and to present the needs for fisheries statistics data interoperability. During this meeting, an RDA Working Group on Fisheries Data Interoperability was kicked off with a main objective to develop FAO MDM activities and BlueBRIDGE services for hosting and exchanging statistical reference data.

2- Ad-hoc Task Group objectives

The overall aim is to present a set of concepts and a structure of datasets that accommodate the codes used across the CWP parties to improve the data reporting and exchange between national, regional and global organizations. Improving this exchange and minimising the time and costs for data mapping would also provide an important service to any user of the data across fisheries data sources. The dissemination of global Standard for Data Structure Definition (SDSD) and related metadata, based on common set of statistical concepts and linked to a standard terminology would enable multilateral exchange of metadata among countries and international organizations.

In particular, the objectives of the ad-hoc task group are:

- providing a global Data Structure Definition and related reference metadata for the aquaculture production and capture datasets.
- assisting national, regional and global institutes in improving and harmonizing information, data and statistics collection, collation and dissemination
- articulating best practices towards usage of the CWP endorsed Data Structure Definition for data collection and dissemination.

3- Ad-hoc Task Group activities

CWP parties that showed interest to contribute to this ad-hoc TG are namely CCSBT, EUROSTAT, FAO, GFCM, IATTC, ICCAT, ICES, IOTC, NACA, OECD, SEAFO, SPC and WCPFC. The teleconference kick-off meeting took place on March 23rd 2017. One-one calls have been ensured with CWP parties that couldn't participate in the Kick-off teleconference.

Following the terms of reference (Annex 1), an inventory was conducted to collect capture and aquaculture data structure and associated reference metadata (dimensions, terminology,..) used by CWP parties. Exchanges were ensured through emails and individual-basis calls depending on the complexity of the collected information.

For the aquaculture scope, EUROSTAT, GFCM and FAO contributed with data structure and metadata. NACA has been solicited to provide feedback on the aquaculture related issues.

The inventory was required to identify the gaps, interconnect and find similarities across reference data used by the parties. Ultimately, the survey inventory served to elaborate first proposals for both aquaculture production and capture global standard for Data Structure Definition.

The issues that have been encountered and raised by the contributors are summarized in the section 5 which represent the setting frame of the discussion during the current inter-sessional meeting. The discussion will also entail the actions to be carried out (section 8) and the agenda of the forthcoming tasks.

4- Scene setting and vision overview

CWP has defined in its handbook a series of concepts used for the purpose of statistical data collection. Examples are "Country", "Flag State", "Nominal Catches", "Landed weight", "Fishing areas", "Currencies", "Time unit", etc… When structuring a statistical dataset, these foundational concepts provide key references for defining its various dimensions: the 'statistical concept' (eg. Capture production) identifies the statistical dataset objective and is generally a key component of the dataset name; the actual measure (e.g. quantity, or value) constitute one dimension; other dimensions identify the breakdown envisaged for compiling this measure, e.g. "Species", "Country", "Fishing areas", "Time unit". However when it comes to actually define the physical structure of a dataset (e.g column names in a CSV file), multiple variants might be found because a range of factors intervene in the decision making: preferred local terminology (for assigning name to the same concept), choice of a specific classification system and linked coding system, scale and level of aggregation, specific attributes to further qualify the dimension.

To reach our objective of elaborating a **CWP global Standard for Data Structure Definition** (**SDSD** - a minimum level of requirements per data domain and related guidelines), we propose to proceed by dissecting and defining the structural elements (e.g classification system, dimensions,..), and we will aim at generically identifying the main structural patterns that are applied for a given data domain: what is fundamentally common, how can the variants be mapped to the agreed commons. Starting from a single apex, these structures should allow to describe in a systematic manner, and to decipher, any dataset. While doing this exercise, the CWP community, starting from the CWP concepts, will strive to agree on applying internationally approved standards to the extent possible.

Our **CWP global SDSD** will therefore describe the structure achieved and the associated standard concepts and terms selected by the group, the rules that should apply for mapping variants to these standards, the affordable degrees of flexibility keeping in mind the need to stick to standard as much as possible (e.g ISSCFG, Areal Grid System,..). It will materialize minimum global requirements, and be accompanied with guidelines describing how, from the minimum requirements, extensions should be designed for catering with other data domains and with local needs.

5- Working terminology for use in this document

Data Structure Definition (DSD) describes how information in a specific dataset is structured in terms of their dimensionality and coding schemes [5].

The structure is composed of a selection of measures, associated dimensions that gather lists of codes. In our context, the **global Standard for DSD** (SDSD) is a global framework/structure enabling to describe

DSDs of CWP parties in a standardized way: the concepts and codelists are given standard names and also comprises the CWP standard classification and relative levels/hierarchies.

Dataset comprises series of observations and it must reference to a DSD [6]. A dataset must conform to a DSD and can only be interpreted by using the DSD and related Concepts and Codelists to decode the dataset information.

Reference data are sets of values or classification schemas that are widely re-used and widerly referenced by systems, applications, data stores, processes, and reports [4]. In our context, reference data of the global SDSD represents the authoritative information to be adopted whenever possible. For example, in the case of the proposed SDSD and under data domain catch for management purpose (annex 2), the reference data is the CWP standard classifications, the sub-classification (aggregations and hierarchies) and the associated codelists listed. When the SDSD reference data cannot meet the requirements of CWP parties, they use **internal reference data** to characterize or standardize their own information [4]. Reference data sets are also defined by external groups, such as government or regulatory bodies, to be used by multiple organizations. For these two later categories, the Task group will have to determine for which purpose (e.g. mapping) and under which status such Reference data could be part of the global SDSD.

Metadata is the data that define and describe other data and processes. Data become metadata when they are used to describe other data in a formalized way and make it easier to retrieve, interpret, or use information [7].

Reference metadata is the metadata of the reference data. It represents the full definitions and terminology used and published by an organization [8]. It provides the detailed definitions (semantics) with the codes (representations) of the reference data items. Reference metadata must be associated with the data to ensure that the data is understood and interpreted by any user.

For instance, in the context of global SDSD, the reference metadata comprises the description of the classification system ISSCFG, the associated gear codelist (named GEAR_CATEGORY) and their other descriptive attributes.

6- Conceptualization of harmonized reference data

In recent years many international organizations have attempted to enhance reference data systems and organizing metadata needed and produced. Efforts to standardise metadata across organizations made progress in the implementation of metadata-driven statistical data management systems[4]. This involves agreement on the metadata components that make up the corporate metadata system, definition of how they are to be generated and presented. Obviously, there are needs for a direct connection between the statistical data themselves and the metadata that describe them, as well as links between the disparate kinds of metadata.

In the CWP context, the data structure and reference metadata required by national statistical organizations and reported to the international/regional organizations are diversified, this primarily because of the diverse objectives (e.g economic, biological, management, environment, control and surveillance) pursued for producing those statistics. For the capture and aquaculture datasets a great diversity of terms and codes are used by the CWP parties. Data sets (with metadata) emanate from a wide variety of contexts and are produced by and used by different experts.

6.1 Rationale

Based on the output of the inventory, it was necessary to develop a hierarchical/systematic structure for reference metadata that defines the ontology of coding, relationships and relative levels/hierarchies of the reference data. The modularity of this structuring is essential so that guidelines can address needs of local extension of DSDs including integration of diverse codelists for different purposes. It is also essential to enable consistent tailoring while being implemented in various standard packaging formats such as Excel, CSV, XML. The conceptual schemes are detailed and illustrated with examples, to describe the proposal of global SDSD for capture and aquaculture (Annex 2).

6.2 Structuring elements of the global SDSD

The terminology presented here emanate from discussions that started since the CWP 24th Inter-sessional meeting and 25th Plenary and continued until recent exchanges with CWP parties. It constitutes a first attempt to define the elements/building blocks of the Global SDSD standard.

Data domain identifies the domains of data collection objectives for which the SDSD minimum requirements can be formulated at global level; within these data domains, produced datasets are expected to share the same dimensions, relations, and semantics that determine the reason of the covered information. Data domain is based on combinations of the triplet:

- target **Indicator**[1] of data collection is the combination of data collected and processed for a clearly defined analytical or policy purpose. e.g. Capture production, Aquaculture production, Catch
- **Indicator type**, i.e. 'statistical' in the context of the current ad-hoc TG work, but could be extended to 'geospatial', 'observational',..
- **Purpose** of the collected data**,** e.g. biological, management, economic, control and surveillance

In the context of this ad-hoc TG, the data domain is oriented toward the collection and exchange of statistical data that covers datasets and associated reference data for capture and aquaculture production statistics.

Concept refer to the terms and concepts defined for statistical purpose in the CWP handbook. These foundational concepts provide key references when structuring a statistical data set.

Concept_Type categorizes the Concept against different elements/roles in the "physical" structure of a dataset:

- **Observation** (also called "primary measure") is the measured/reported value of a particular measure. The observation is fully described thanks to the attributes.
- **Attribute** does not affect the dataset structure itself, but qualifies the observation further. For instance, the attribute named "Unit" provides information about whether the observation value are measured in weight (e.g tonnes) or currency units, and if so which currency.
- **Dimension** identifies the breakdown envisaged for compiling the observation/measure, therefore dimensions together identify each statistical observation. A dimension is implemented through a codelist listing the possible values they can take. For example, the dimension "Country" would

[1] An alternative could be "Descriptor"

explain which country a specific observation refers to. Geographic and Time dimensions are deemed necessary to distinguish separately.

Classification system defines coding and hierarchies used in structuring the reference data. A classification system is generally accompanied of a coding system that gives the rules to assign a unique code for each element of the classification system. The CWP international standard statistical classification systems are primarily used (e.g. ASFIS, ISSCFG, Areal grid system,..). A Classification system is named, owned and maintained by an institution, and certain logics are factored in for the coding, set of aggregations, hierarchies. These logics respond to data collection scope and mandate of the owning institution, etc.

The inventory of usage of standard statistical classification systems by CWP parties as classification system confirmed the necessity of adopting standards to the extent possible or to develop mappings with RFBs' owned classification systems.

Level of granularity/aggregation describes the level of details included in the DSD respectively within/building on the classification system. The decision of the resolution level resides in the choice of the user who wants to report the data. For instance, within or "under" the classification system "FAO Major Marine Fishing Areas", the global SDSD can include breakdowns: Subarea, Division or Subdivision. In the same order of ideas, ICES subareas would be considered as areas at lower level of granularity within the major Fishing Area 37.

Another example would be the aggregations of 3-Alpha code species from the ASFIS classification where aggregation of species are shaped by "building on top"/based on the classification system. ISSCAAP groups are an obvious example of grouping of ASFIS codes used as part of the ASFIS classification system. Another example is encountered in the case of species groupings used by Tuna RFMOs, where a Tuna RFMO specific classification system is built on top of the ASFIS species codes.

In all these cases of whether higher or lower level of aggregation, mapping against standard classification system codes is crucial to be integrated in the global SDSD.

CodeList comprises a set of identifiers/codes enumerating all possible instances of a dimension and responding to a certain coding logic (e.g numeric or alphanumeric) [5].

CodeList_Id: Codelist associate an identifier with a name and optional description. For instance, in the global SDSD the codelist named "Inter-agency 3-alpha code" has the Codelist_id "3alpha_code".

Description provides descriptive information on the codelist and/or related contents.

5.5 Proposal of the Global SDSD for Capture and Aquaculture data

The concepts and structural elements presented below enabled building drafts of the global SDSD, an overarching and modular structuring of common metadata used by national statistical agencies and international organizations.

A modular structuring of the global SDSD allows its extension when applied to different data domains. At this stage, this is illustrated by the three examples of data domains presented in annex 2:

- the capture production statistics for economic purpose, which covers the Quantity and Value compiled according to dimensions represented by concepts Country, Fishing area, Aquatic Species, and Time unit;

- The catch statistics for management purpose, which covers an overlapping selection of concepts to which are added Fishing gear and/or Fishery vessel concepts (however the value Observation concept is not retained here primarily);
- the Aquaculture production statistics for economic purpose, which covers the Quantity and Value compiled according to dimensions represented by concepts Country, Production area, Environment, Aquatic Species and Time unit, and it could be extended for the same purpose with dimensions Farming Structure, Product type;

This proposal has to be discussed with CWP parties to accommodate their needs with their members' countries to the extent possible. Progressively, further efforts will be dedicated to incorporate peculiarities of coding at lower levels of metadata granularity. Data mapping following certain standardized principles and procedures would be an efficient way to data transformation and integration within databases. Simple semantic mapping could ensure direct (one-one) matching for the majority of codes used by RFMOs with CWP classification systems.

7- Implementation - Dissemination and exchange formats

The data format provides the content and the structure of the document sent over the data network. There are several formats and standards of dissemination and exchange which can be used to implement the global SDSD and related reference metadata. Some criteria need to be considered to choose/score the file format:

• It should be spread throughout the CWP parties to minimise compatibility issues.

• It needs to be readable for human and machine, complexity should therefore be kept at an acceptable level.

• Structure standardisation needs to be possible – The format structure must be standardised and the file format must support an open standard.

As for best practices to be pointed out, the writing convention or format could be recommended when exchanging data with specific coding system (e.g FAO areas breakdown,.). In this case, the easy digitalization of the codes should be considered to facilitate the data interoperability and exchange.

7.1 Comma Separated Values CSV

CSV file format is the most commonly used by the CWP parties for dissemination of datasets and metadata. The readability of CSV files is acceptable and facilitate the interaction of human user. It could be a good candidate to exchange the global SDSD, the reference data and metadata.

In FAO, an immediate focus is on the publication of DSDs directly relevant to the statistics disseminated by the FAO FI. The DSDs of the FAO data sets namely Capture production, Aquaculture production, and Global production have been made available in the FAO website in a packaged format comprising DSDs and codelists in csv files and related metadata in text file.

This is an intermedia release of the FAO DSDs towards a dissemination in SDMX format.

7.2 Statistical Data and Metadata eXchange SDMX/SDMX-ML

SDMX standard offers an information model which describes aggregated statistical data sets and the structural metadata needed to exchange them in a standard fashion. The content of SDMX files have visible structure with explanations what is stored where in the file.

The SDMX principles have been applied to fisheries statistics and in particular the catch SDSD for the collection of data in the context of a joint-project SEIF that stands for SDMX for Eurostat, ICES and FAO. The initiative aimed at the alignment and the exchange of SDMX artefacts between the three organizations.

SDMX is being adopted as the data collection format for fisheries in Eurostat, in-line with policy for all statistical domains covered by the European Statistical System. FAO is making progress in the implementation of SDMX principles and acquisition of related tools.

Furthermore, based on the SDMX information model and the XML format, SDMX-ML could be an alternative for exchange of structural metadata, data sets, and queries.

In the CWP context, it remains essential to evaluate the ability of SDMX data model to incorporate the proposed multilingual reference data and the global SDSD that can be expanded with other codelists and enriched with hierarchical codelists.

7.3 Fisheries Language for Universal eXchange FLUX

FLUX offers a protocol to create a secure and configurable network between different parties IT systems. Its standards includes a protocol for the exchange of a request for information, the exchange of the information itself and the acknowledgement and rejection of information exchanged. Built over SOAP and WSDL, this mechanism provides an envelope that can contain a business message, and software which serves as infrastructure to transport the envelopes. FLUX is strongly tied to XML as a data format, and, more specifically, UN/CEFACT standardized XML Schemas.

Adopting the UN/CEFACT standards and usage of the FLUX are to be explored for the purpose of interoperability of standard data structure format. FLUX includes a MDM module and it would be necessary to scrutinize the capabilities of FLUX whether it could accommodate the proposed global SDSDs, the reference metadata and in particular the various levels of hierarchies and granularity. EU DG-MARE implemented FLUX for the data collected for management and control purposes only (e.g Logbook ERS, VMS,..), and a positive conclusion on FLUX's ability to accommodate the proposed global SDSD would make it a good candidate data exchange format while extending the global SDSD to non-statistical domains.

8- Points of discussion/Outlines

The proposed following points have been highlighted by CWP parties during the preceding and recent exchanges. These outlines need to be discussed towards the elaboration of this ad-hoc TG objectives for reference data harmonization.

8.1 Harmonization of data structure and metadata

The data domain of the proposed global SDSD is to be firstly defined to decide the structure and dimensions to be used. Modularity of the SDSD provides flexibility allowing to take into account additional dimensions required to meet different purposes of the data collection. International classifications are essential mechanisms for the harmonization and coordination of data compilation. On the other hand, this

standardization exercise is not only promoting the use of CWP standards across the datasets, but essentially also covering the definition of terminology used in dimensions (e.g names, units,..) and different items of the global SDSD. Harmonizing disparate information systems requires data translation and mapping, as well as procedures that promotes their use.

The discussion should cover the harmonization of the reference data to the extent possible. If internal reference data used by an organization cannot fit within building blocks of the global SDSD, the extension of the structure would be deemed necessary (and should be made possible). For instance the global SDSD could be elaborated to meet the distinctive requirements and reference data used across Tuna RFMOs and should allow shaping a DSD specific for TunaRFMOS data requirements.

8.2 Mapping Code

The inventory revealed that CWP parties are using CWP standards coding system to a certain extent. Some members (e.g, ICCAT, ICES, IOTC,..) adopted extra codes for their purposes or adopted different classification system (e.g DCF for EUROSTAT and ICES). In these particular cases, it is crucial when a coding system other than CWP standards is used to explicit the mapping between coding systems, ie defining semantic relationships between codes of different databases. The semantic relationships can be one to one or more complex to be mapped when it is many to one.

In general, two main situations are encountered regarding the specific codes: these are either built on the codes of CWP standards by aggregating a group of codes (e.g. group of species built upon the ASFIS codelist), or by extending the CWP standards with more details resulting in lower level of aggregation (e.g gear codes that fall within one class/code of the ISSCFG codelist).

Discussing in details all the peculiarities of codes' mapping amongst CWP parties would be challenging. Thus discussion should focus on best practices of generic semantic mapping to start solving the common issues as highest priority.

8.3 CWP registry

After its endorsement by CWP, the **global SDSD** and its structural components would be published through the CWP handbook, FAO FI site as a first step. Subsequently a unified and collaborative **CWP registry** would host the catalog of the global SDSD's reference data and the specific reference data used by the CWP parties. This catalog would disseminate the various DSDs in use following the agreed SDSD. The contents would be made available to be harvested by CWP parties and national authorities to facilitate data sharing and usage.

The **CWP registry** would be as the index of data and metadata hosted in the **CWP repository** [2;3].Two alternatives of CWP repository can be presented. The first proposes a distributed repository architecture which allows interaction among heterogeneous databases of the CWP parties and access to information from dispersed computers. The second alternative consists of a centralized dissemination repository that would host the information made available by the parties. In the context of CWP's work on Master Data Management (MDM), the second option of repository can ensure coordination, cataloguing and dissemination (e.g the contents of codeLists and the mappings among codelists codes) in compliance with the standards of the global SDSD). The MDM dissemination services that FAO envisages to coordinate intend to facilitate countries data management of producing datasets in multiple reporting formats.

8.4 Governance

The role of governance is to define standards and set of best-practices or principles that aim at ensuring that the CWP parties own the process, their information asset, and disseminate and maintain the global SDSD, the international classifications and the codes' mappings. It encompasses the accuracy, accessibility, consistency, and completeness to ensure data quality. At registry level, the maintenance mainly address registry subscriptions regarding new datasets or changes in the reference metadata of datasets already registered (e.g updates to the CWP standards would have to be reflected in the registry) and the mappings among classification and sub-classification systems between any organization when deemed necessary. At repository level, the role of maintaining the individual classifications in the repository and well as the mapping among codelist codes should reside at the level of the CWP party. In the case of any change in the mapping codelists, copies should be made available to the CWP repository for broader dissemination.

Besides building the global SDSD, the discussion should lead to establishing guidelines and a set of best practices for maintenance of the CWP registry's contents and the dissemination workflow across the organizations through the repository.

9- Actions Requested by the CWP-IS

Based on the discussions on the points outlined in section 5, the following actions would be required to achieve the objectives of the ad-hoc task group:

- To discuss and provide feedback on the structural elements and conceptual schemes from an overarching perspective towards the validation of the global SDSD and related reference metadata.
- To address the modular structuring of the global SDSD that could accommodate other domains of data collection, or different aggregation levels to match organization's policies and purposes.
- To provide guidance towards formulation of guidelines and best practices for the governance and the maintenance of the global SDSD through a CWP registry that would host the catalogue of the reference data including the CWP standards and internal reference data used by CWP parties when it cannot be mapped to standards.
- To provide guidance regarding the two implementations' alternatives of CWP repository of codelists and related mappings, and elaborate a roadmap for the further development including the selection of the exchange format.
- To consider the following points for approval:
    - To give the mandate to FAO for coordination and further elaboration of the global SDSD if the proposals are approved.
    - To promote the use of CWP standards to the extent possible and consider procedures of mappings with the organization's coding system, as well as guidelines for implementation of DSDs tailored to local needs.
    - To provide indications towards a mandate of FAO in developing the CWP registry.
- To agree on the forthcoming tasks and agenda and identify any points to be raised under "Any other business" (AOB).

References

[1] Colpaert, P., Van Compernolle, M., De Vocht, L., Dimou, A., Sande, M. V., Verborgh, R., Mannens, E. (2014). Quantifying the Interoperability of Open Government Datasets. *Computer*, *47*(10), 50–56. http://doi.org/10.1109/MC.2014.296

[2] Statistical Data and Metadata Exchange (SDMX) - BIS, ECB, Eurostat, IBRD, IMF, OECD and UNSD - Metadata Common Vocabulary. https://stats.oecd.org/glossary/detail.asp?ID=7078

[3] UNECE (2009) Common Metadata Framework Part A - Statistical Metadata in a Corporate Context: A guide for managers
http://www.unece.org/fileadmin/DAM/stats/publications/CMF_PartA.pdf

[4] McGilvray D, 2008 - Executing Data Quality Projects: Ten Steps to Quality Data and Trusted Information- Morgan Kaufmann Publishers. Elsevier
http://booksite.elsevier.com/9780123743695/10steps_DataCategories.pdf

[5] European Central Bank (ECB), Bank for International Settlement (BIS), Eurostat, International Monetary Fund (IMF), Organisation for Economic Co-operation and Development (OECD), "GESMES/TS User Guide", Release 3.00, February, 2003; unpublished on paper.

[6] Green, T (2009), "We Need Publishing Standards for Datasets and Data Tables", OECD Publishing White Paper, OECD Publishing. doi: 10.1787/603233448430 http://dx.doi.org/10.1787/603233448430

[7] ISO / ISE (International Organisation for Standardisation / International Electrotechnical Commission), n.d., ISO/IEC 11179- (2004) - Information Technology – Metadata Registries (MDR) Part1: Framework, published on the Joint Technical Committee 1 (JTC1) Information Technology Standards website, Available from https://www.iso.org/obp/ui/#iso:std:iso-iec:11179:-1:ed-2:v1:en.

[8] United Nations Statistics Division - Compiler guide for the Manual on Statistics of International Trade in Services 2010. Access on June 2017 (Part IV Cross-cutting topics). https://unstats.un.org/wiki/display/M2CG

Annex 1

Terms of Reference

CWP ad-hoc Task Group on "Reference harmonization for capture fisheries and aquaculture statistics"

1. Background/Context

The CWP 25 plenary meeting held in FAO headquarters (Rome, 2016), adopted the intersessional work plan for the Fishery Group. The meeting approved the establishment of two distinct ad-hoc Task Groups one of which is the ad-hoc Task Group on "Reference harmonization for capture fisheries and aquaculture statistics" which should be led by the Fishery Group. Nonetheless, due to the overarching issues the Task Group will be working with the capture group. The CWP requests this Group to have representatives from Aquaculture Group as well. The first activity of the Task Group on Reference Harmonization, as set out in the CWP 25 report, will be to prepare TORs and work plan and will be submitted to members for validation. The present Terms of Reference describe the set of activities to be carried out to develop Harmonizing references utilized for capture fisheries and aquaculture statistics, including possible dissemination of the global Data Structure Definitions (DSDs) and for which FAO as lead of the task seeks for the collaboration of other relevant agencies.

2. General description of task(s) and objectives to be achieved

The ad-hoc Task Group on "Reference harmonization for capture fisheries and aquaculture statistics" was established to develop proposals of global reference data for statistical data sets of Capture and Aquaculture production.

In the context of reference Geographic data, the GIS group is tasked to elaborate one or more grid reporting systems, their codification, and format. The output of the GIS group on reference data harmonization and codification will be taken as input for the task group.

The task group led by Aymen Charef will proceed as follows:

- Inventory of Capture and Aquaculture datasets structure and associated code lists within FAO statistical databases (FIGIS, Tuna Atlas,..) and among CWP Members
- Prepare a concrete proposal of DSDs, classifications and corresponding codification[2];
- Prepare a proposal for a global DSD including a mechanism for mappings; and guidelines for DSD extensions;
- Follow-up with CWP members seeking their active contributions, and validation;

The results of these activities will also contribute to the revised CWP handbooks, as follows:

- Update CWP handbook (dynamic fetch of reference data from FAO MDM solution);
- Publish reference data for Catch and Aquaculture from FIGIS database in electronic format (CSV);
- Publish the Data Structure Definitions (DSDs) in SDMX format, as well as other relevant reference data services.

[2] The Tuna Atlas is a prime candidate for harmonization that can be taken as a pilot test case, considering the existing product and collaboration and current enhancements to it

3. Timeframe

- Prepare first draft of FAO DSD and reference data (February)
- Contact CWP members seeking for an inventory of their Data Structure and related reference data (Code list and definitions) (March)
- Contact Tuna-RFMOS seeking for validation of the coding system harmonization (March)
- Present the document based on Tuna Atlas pilot test case in the coming mid-2017 intersessional meeting
- Further developments including intersessional meeting feedback and enlarging scope of contributions to other CWP Members, during the second intersessional period
- Present final proposal of DSDs at the CWP-26 session

4. Role of individual group members

Each individual group member will participate actively to the working group by:

- Attending the working group calls and meetings,
- Sharing knowledge and practical examples to illustrate CWP needs
- Reviewing working documents co-prepared by the working group

5. General

5.1 Membership

The CWP Working Group on Reference Harmonization for capture and aquaculture statistics is comprised of the following CWP members that expressed their interest to actively contribute: CCSBT, Eurostat, FAO, GFCM, IATTC, ICCAT ICES, IOTC, NACA, OECD, SEAFO, SPC and WCPFC.

Five Tuna Regional Fisheries Management Organizations (T-RFMOs) are participating and will contribute to the Tuna Atlas harmonization pilot case.

As far as possible, if a member of the working group is unable to attend a meeting, she/he will nominate a substitute to attend the meeting, and inform the Chair of this substitution prior to the scheduled meeting.

5.2 CHAIR/CONVENOR

Aymen Charef, Consultant – Fishery Statistician at FAO/FIAS will chair the ad-hoc Group. Meetings will be convened by the Chair and supported by the CWP secretariat.

5.3 AGENDA ITEMS

Upon request from the chairperson, proposals for agenda items by participants will be forwarded to the CWP secretariat no less than ten working days prior to the next scheduled meeting.

At the end of each meeting, the next meeting will be scheduled (date/time), if possible together ith draft agenda items to be addressed.

5.4 MINUTES AND MEETING PAPERS

The minutes of each meeting will be prepared by chairperson together with the Secretariat, or any other member on a voluntary basis, and shared with all working group members no later than five working days following each meeting.

5.5 WORKING CONDITIONS

Most of activities will be carried out through e-mail exchange and online meeting unless physical meeting are deemed necessary.

5.6 FREQUENCY OF MEETINGS

To be defined

Annex 2

2.1 Proposal of the global SDSD for Capture production/catch data

DATA DOMAIN												
Indicator		Purpose										
Type	Name											
statistical	Capture production	Economic										
	Catch	Management										
Concept		**COUNTRY**	**FISHING AREA**	**AQUATIC SPECIES**	**TIME UNIT**	**QUANTITY**	**OBS_STATUS**	**UNIT**	**FISHING GEAR**	**FISHERY VESSEL**	**VALUE**	**OBS_STATUS**	**UNIT**
Concept_Type		Dimension	Geographic Dimension	Dimension	Time Dimension	Primary measure/ observation	Attribute	Attribute	Dimension	Dimension	Primary measure/ observation	Attribute	Attribute
Classification system		UN Standard country or area codes for statistical use (M49)*	FAO Major Fishing Areas for statistical purpose; Areal grid coding system	ASFIS List of Species for Fishery Statistics Purposes	Calendar year		FAO statistical standard for Observation status flags	UCUM** Unified Code for Units of Measure	ISSCFG International Standard Classification of Fishing Gears	ISSCFV International Standard Statistical Classification of Fishery Vessels by Categories		FAO statistical standard for Observation status flags	UCUM** Unified Code for Units of Measure
Aggregation/ granularity level (Sub_classification)		Aggregated codes (e.g Aggregated member states of EU)	Breakdowns: Subarea, Division (e.g ICES subareas, GFCM GSAs); Areal grid coding system (e.g IOTC 5 degree grid system)	ISSCAAP; Aggregated species (e.g IOTC Group species list)	e.g yearly; monthly; bi-annual		FAO standard symbols		Detailed list of gears; or Aggregated gears (e.g IOTC Fishing gear group)	Detailed list of vessels; or Aggregated vessels (e.g GFCM Vessel Group, OECD Fleet segments)		FAO standard symbols	
Code List		UN code	FAO Fishing Areas; ICES subareas; GFCM GSAs; Statistical quadrangles & quadrant	Inter-agency 3-alpha code	Calendar Year	Quantity	Observation Status Flag	Units of measure	Gear Category	Fishery Vessel Type	Value	Observation Status Flag	Units of measure
Codelist_id		UN_CODE	FAO_AREAS; GRID_SYSTEM	3ALPHA_CODE	YEAR	QUANTITY	STATUS_FLAG	UNIT	GEAR_CATEGORY	VESSEL_TYPE	VALUE	STATUS_FLAG	UNIT
Description		List of countries or areas (three digits code)	FAO major fishing areas; codes for Statistical quadrangles, and for quadrants	Species reference	Reference year	Quantity of production	FAO Observation status codes (e.g "F"Forecast value, "R"Revised)	Unit of measure (e.g tonnes or number of animals)			Value of production	FAO Observation status codes (e.g "O" Missing value)	Unit of measure (e.g 1000 US$)

2.2 Proposal of the global SDSD for Aquaculture data

DATA DOMAIN													
Indicator		**Purpose**											
Type	**Name**												
statistical	Aquaculture	Economic											
Concept	**COUNTRY**	**PRODUCTION_AREA/STATISTICAL_AREA**	**ENVIRONMENT**	**AQUATIC SPECIES**	**FARMING STRUCTURE /PRODUCTION TYPE**	**PRODUCT TYPE**	**TIME UNIT**	**QUANTITY**	**UNIT**	**OBS_STATUS**	**VALUE**	**UNIT**	**OBS_STATUS**
Concept_Type	Dimension	Geographic Dimension	Dimension	Dimension	Dimension	Dimension	Time Dimension	Primary measure/observation	Attribute	Attribute	Primary measure/observation	Attribute	Attribute
Classification system	UN Standard country or area codes for statistical use (M49)*	FAO Major production areas for statistical purpose	TBD	ASFIS List of Species for Fishery Statistics Purposes	TBD	TBD	Calendar year		UCUM** Unified Code for Units of Measure	FAO statistical standard for Observation status flags		UCUM** Unified Code for Units of Measure	FAO statistical standard for Observation status flags
Aggregation/granularity level	Aggregated codes (e.g Aggregated member states of EU)	BREAKDOWNS: SUBAREA, DIVISION (sub-continents or, sub-oceasn, sub-administrative units)		e.g ISSCAAP, Aggregated species			e.g monthly, annual, bi-annual						
Code List	UN Code	FAO production areas	Environment	Inter-agency-3 alpha code	TBD		Year	Quantity	Unit	Observation Status Flag	Value	Unit	Observation Status Flag
Codelist_id	UN_CODE	PRODUCTION_AREA	ENVIRONMENT	3ALPHA_CODE	TBD	TBD	Reference year	QUANTITY	UNIT	STATUS_FLAG	VALUE	UNIT	STATUS_FLAG
Description	List of countries or areas (three digits code)	FAO major production areas	Environment of the production (water type) (e.g brakish water, salt water, marine water)	Species reference	Production unit type (e.g pond, raceway and tank, off bottom, cages,.)	Type of product (e.g eggs, ornamental, food,..)	Time Dimension	Quantity of production	Unit of measure (e.g tonnes or number of animals)	FAO Observation status codes (e.g "F"Forecast value, "R"Revised)	Value of production	Unit of measure (e.g 1000 US$)	FAO Observation status codes (e.g "O" Missing value)

Notes:

- Colored text represent examples given to the related text.
- A single cell contains sometimes different examples of what could be found in this cell; this doesn' t mean that these multiple values would be found concurrently in the same cell when describing DSDs
- [*] The Standard Country or Area Codes for Statistical Use (Series M, No.49/Rev 4) - commonly known as M49 - presents a list of countries and areas used as a common reference for statistical purposes by the United Nations Statistics Division. https://unstats.un.org/unsd/methodology/m49/

  The M49 is maintained in parallel to the International Standards Office (ISO) 3166 list.

  International Standard ISO 3166-1, Codes for the representation of names of countries and their subdivisions-Part 1: Country codes, ISO 3166-1: 2006 (E/F), International Organization on Standardization (Geneva, 2006)
-
- [**] **Unified Code for Units of Measure UCUM** is used for base unit codes and representation symbols. It is a variety of accepted ISO, IEC and ANSI standards (e.g. ISO/IEC 80000 or ANSI X3.50-1986) have been developed to express SI units and/or to build on top of the SI foundation. http://unitsofmeasure.org/