COORDINATING WORKING PARTY ON FISHERY STATISTICS

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

Technical background document Version 2.0

Executive summary

This technical support document provides background, rationale and objectives to achieve the work of the Task Group. The document describes design, content and implementation of the standard for the global data structure definition for capture fisheries and aquaculture. The manuscript is updated with feedback and contributions from CWP members until the next CWP session

Summary of changes
(marked in yellow throughout the manuscript)

Post CWP Inter-Sessional meeting held in Copenhagen, 19-22 June 2017
- Amend objectives and update the mandate
- Change terminology and revise definitions (catch, DSD, module)
- Revise global DSD and add proposals for data collection purpose
- Add text on data exchange formats

Table of Contents

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. 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 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.

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 Research Data Alliance RDA 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 and structure of statistical concepts that accommodate the coding system used by CWP parties to improve the data reporting and exchange between national, regional and global organizations. Harmonization of data structures and related metadata will minimize time and costs of mapping data elements to standard terminology and will improve multilateral exchange among CWP parties.

To reach the objective of elaborating a **CWP standard for Data Structure Definition (DSD)**, we propose to proceed by dissecting and defining the structural elements (e.g classification system, dimensions,..), and we 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.

The **CWP standard for DSD** will therefore describe the structure achieved, the associated standard concepts and terms selected by the Task Group (TG), the use of classification system keeping in mind the need to stick to CWP standards as much as possible (e.g ISSCFG, Areal Grid System,..). It will materialize minimum global requirements, and provide modular flexibility allowing extensions to be designed from the minimum requirements and for catering with other data domains and with local needs.

As a result, the ultimate objective of this TG is to lay basis for establishing data-sharing agreement as practical work arrangements between agencies involved in a data workflow. This is expected to reduce data reporting burden for data producers and to improve data quality by mainstreaming the cross-checking and reconciliation of information from national sources.

Specific objectives of the ad-hoc Task Group are:

- providing a CWP standard for Data Structure Definition and related reference metadata for the aquaculture production and capture datasets.
- supporting CWP Parties and their members in improving and harmonizing data schemes and related data exchange agreements.
- articulating best practices towards usage of the CWP endorsed Data Structure Definition for statistics collection and dissemination.

3. Ad-hoc Task Group activities

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

In the scope of aquaculture inventory, EUROSTAT, GFCM, FAO and OECD 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 draft proposals for global DSDs for aquaculture production and capture fisheries.

Proposals of the standard for global DSD were presented (document and presentation) at the CWP intersessional meeting that was held in Copenhagen, 19-22 June 2017. The meeting's feedback entailed revising terminology and expanding the scope of the DSD domains to data collection and dissemination and to cover nominal catch, catch and effort, logbook. Remarks and actions to be carried out are summarized in the meeting report.

The present document incorporates suggestions and comments from the TG members, which served to elaborate the subsequent proposals of the CWP standard for DSD. Details of the TG's feedback and actions are recorded and circulated in a distinct document.

4. Conceptualization of harmonized reference data

In recent years many international organizations spent efforts to enhance reference data systems and organizing metadata needed and produced. Efforts to standardize reference data across organizations made progress in the implementation of metadata-driven statistical data management systems [2]. This involves agreement on the metadata components that make up the organization metadata system, definition of how they are 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 data domains (e.g economic, biological, management, control and surveillance) pursued for producing those statistics. For the capture and aquaculture datasets, the diversity of terms and codes emanate from a wide variety of contexts and are produced by experts of the CWP parties.

Reference data harmonization is the process of capturing, analysing and reconciling the meaning and representation format of data concepts and codes used by different CWP parties. The harmonization process involves a set of activities undertaken at data structure level and at the semantic codes level. The process starts with an inventory of the data requirements for each domain and used coding system, analysis of the codes definition and the classification system used as basis, reconciliation of the terminology and alignment with CWP standards by developing codes mappings.

4.1 Proposals of the CWP Standard for a global DSD

The inventory's output enabled the identification of concepts of minimum data requirements used by CWP parties. Concepts and structural elements (Table 1) provided the basis to build proposals of the standard for global DSD, as overarching and modular structuring of common metadata for use by national statistical agencies and international organizations. It was necessary to develop a systematic data structure for reference metadata that defines the ontology of coding, relationships and relative structure hierarchies. Based on the first round of the TG discussions, additional proposals of DSDs are built with focus on both data collection and dissemination schemes (Annex 2, Excel file).

A modular structuring of the CWP Standard for global DSD is essential to allow its extension and address needs of local implementation of DSDs including integration of diverse codelists for different purposes. For this, data modules are inserted in the DSD to cluster concepts having the same subject of information (e.g catch module, effort module,..). The DSD is then composed of concepts and data modules that can be added up and extended depending on user's specific needs.

At this stage, this document puts forward for discussion (validation) four DSDs corresponding to different data domains (Annex 2 Excel file):

- Global capture production: this DSD is designed to cover the capture production in volume and value for an economic purpose. Volume and value of nominal catch are compiled according to dimensions represented by concepts Country, Fishing area, Aquatic Species, and Time unit;
- Catch: it covers continuum of concepts (gross catch, discards, nominal catch, etc…) for management purpose to which are added Fishing gear and/or Fishery vessel concepts. The concept "value of catch" could also be included;
- Catch and effort (Logbook): the DSD serves for management purpose and in particular addresses the collection scheme. It contains vessel information, catch and effort for each operation (i.e haul). Information on start and end of time and location of fishing information are also included.
- Global aquaculture production: it is built to cover the aquaculture production and its value for economic purpose. Core concepts compiled are Country, Production area, Environment, Aquatic Species and Time unit, and it could be extended for the same purpose with the other dimensions Farming Structure and Product type;

4.2 Structure elements of a global DSD

Terminology used in this document emanate from discussions that started since the CWP 24th Inter-sessional meeting and 25th Plenary and continued until recent exchanges with CWP parties. This section represents the glossary of terms for use by this Task Group.

Data domain identifies domains of data for which the DSD 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:

- **Indicator**, a clearly defined analytical or policy purpose. e.g. Capture production, Aquaculture production, Catch
- **Indicator type**, i.e. 'statistical' for time series, 'observational' e.g. for logbooks; it could be extended to include e.g 'geospatial' data for VMS.
- **Purpose** concerns two dimensions: positioning in workflow (data collection vs data dissemination); and policy objective for which the indicator is produced (economic vs management or Monitoring-control-surveillance)

Concept refers 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. Description of concepts used in the global DSD for capture, catch and effort data are documented in the Table 1 (Annex **1)**.

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

- **Observation** (also called "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 explain which country a specific observation refers to. Geographic and Time dimensions are deemed necessary to distinguish separately. Present escribe

Module is a building block composed of concepts that can be assembled to describe information. For instance Vessel information module is composed of four core concepts; Vessel flag (country), Vessel Unique Identifier, Length overall and Gross tonnage.

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 statistical classification systems by CWP parties confirmed the necessity of adopting standards to the extent possible or to develop mappings with RFBs' specific owned classification systems.

Level of granularity defines through the name of the sub-classification used by the CWP party for their specific requirements, 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 DSD 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 DSD.

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) [3].

CodeList_Id: Codelist associate an identifier with a name and optional description. For instance, in the global DSD 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.

Data Structure Definition (DSD) describes how information in a specific dataset is structured in terms of their dimensionality and coding schemes [3].

The structure is composed of a selection of measures, associated dimensions that gather lists of codes. In our context, the **global DSD** is a universal 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 referenced by systems, applications, data stores, processes, and reports [2]. In our context, reference data of the global DSD represents the authoritative information to be adopted whenever possible. When the DSD's reference data cannot meet the requirements of CWP Parties, they use concepts for their specific purpose to characterize or standardize their own information [2]. Reference data sets are also defined by external groups, such as government or regulatory bodies, to be used by multiple organizations.

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 [6].

Reference metadata is the metadata of the reference data. It represents the full definitions and terminology used and published by an organization [7]. 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 DSDs, the reference metadata comprises the description of the classification system ISSCFG, the associated gear codelist (named GEAR_CATEGORY) and their other descriptive attributes.

5. Ongoing topics of discussion

The proposed following points have been highlighted by CWP parties during the preceding and recent exchanges. Discussions with particular focus on these outlines are kept toward consensus and achievement of this TG objectives.

5.1 Harmonization of data structure and metadata

The data domain of the proposed global DSDs is to be firstly defined to decide the structure and dimensions to be used. Modularity of the DSD provides flexibility allowing to take into account additional dimensions required to meet different purposes of data collection schemes. 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 DSD. 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 specific reference data used by an organization cannot fit within modules of the DSD, the extension of the structure is necessary. For instance, the DSD for catch and effort does not meet the distinctive requirements for reference data used by Tuna RFMOs and needs to be extended to cover the school type, fishing mode, etc.

5.2 Mapping Code

Discussion should focus on best practices and practical steps for harmonizing reference data by mapping between coding systems i.e defining semantic relationships between codes of different coding systems/dictionaries. Semantic mapping, can be based on automatic routine if necessary, could ensure direct

(one-one) mapping of the majority of codes used by RFMOs to CWP classification standards. Alignment with standards and mapping codes are a prerequisite in multilateral data interoperability among CWP parties.

However, some CWP Parties are using CWP standard coding systems 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 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 higher level of granularity (e.g gear codes that fall within one class/code of the ISSCFG codelist).

In these situations, the peculiarities of codes' mapping amongst CWP parties would be challenging as they require many-to-one mapping. Solving these particular cases would require expert knowledge which entails background on definition of the codes and could result in suggesting best matching.

5.3 CWP registry and repository

After its endorsement by CWP, the global DSDs and their concepts will be published through the CWP handbook, FAO FI site. Two alternatives of CWP repository were presented at the intersessional meeting and FAO was given guidance to work towards a centralized dissemination repository that would host the information made available by the Parties. A unified and collaborative CWP registry would host the repository which would disseminate the various CWP global DSDs, CWP standards, and the specific reference data used by CWP parties. The CWP registry would be the index of data structure and metadata of reference data and mappings hosted in the CWP repository [3]. Contents will be made available to be harvested by CWP Parties and national authorities to facilitate data sharing and usage.

In the context of FAO's work on Master Data Management (MDM), the repository can ensure coordination, cataloguing and dissemination (e.g. CWP codelists and mappings among Parties' codelists) in compliance with the reference data of the CWP global DSDs.

5.4 Governance

The role of governance is to define set of best-practices that aim at ensuring that the CWP parties own the process, their information asset, and disseminate and maintain the CWP global DSD, the CWP international classifications, standards recommended for use by CWP, codelists used by CWP parties and the codelists' mappings. At registry level, the maintenance mainly address registry subscriptions regarding new datasets or changes in the reference metadata of datasets (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 is 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 catalog for broader dissemination.

5.5 Data exchange formats and mechanisms

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 DSD and related reference metadata. Data exchange options were a topic of discussion during the intersessional meeting. The TG recognized the importance of defining and recommending formats and

standards for data exchange. Options put forward should cover the varying capabilities and requirements of CWP Parties, as mechanisms need to be aligned to enable data sharing agreement.

Alternative should be evaluated on the basis of ease of implementation and operation and the following criteria:

• 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 standardization needs to be possible – Structure format must be standardized 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 data interoperability and exchange.

5.5.1 Comma Separated Values CSV

CSV file format is widely used among 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 DSD, the reference data and metadata.

The main advantage of using CSV files lies in the fact that such files can be accessed through common spreadsheet software, making them easily managed manually and a useful option to accommodate data providers without information systems that can generate the data files for transmission automatically.

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.

5.5.2 Statistical Data and Metadata eXchange SDMX/SDMX-ML

SDMX is an international initiative that aims at standardising and modernising the mechanisms and processes for the exchange of statistical data and metadata among international organisations and their member countries. The organizations involved in the SDMX initiative developed guidelines applicable to all statistical domains. Furthermore, the community made available software tools and a registry to host reusable SDMX artefacts [8].

SDMX is not just a technical standard and offers many guidelines such as a Checklist for Design Projects and Modelling Guidelines which are relevant for establishing an SDMX project for a data domain. For a specific data domain (e.g capture data for dissemination purpose), an SDMX project starts by creating a concept scheme that describes this domain and the data flows (e.g Country sends dataset to an organization). The design and creation of SDMX artefacts and the management of such a project are detailed in this standard project workflow. The structure of this checklist is based, to the largest extent possible, on the UNECE Generic Statistical Business Process Model.

SDMX principles have been applied to fisheries statistics and in particular the catch DSD 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 for the Global Capture Production data domain.

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.

Technically, SDMX standard offers an information model which describes 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 usual format in SDMX information model is XML (SDMX-ML) which make it a good option for exchange of fisheries statistical data sets and accompanied metadata.

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 DSDs that can be expanded with other codelists and enriched with hierarchical codelists.

This evaluation is taking place with members of TG using SDMX. As a result background and technical documentation will be provided to the TG on the implementation of SDMX as a standard that would host CWPs DSD and related metadata.

5.5.3 Fisheries Language for Universal eXchange FLUX

FLUX standard, developed and maintained by the Centre for Trade Facilitation and e-Business (UN/CEFACT), provides an harmonized message standard allowing Fishery Management Organizations (FMOs) to automatically access the electronic data needed for stock management, such as vessel and trip identification, fishing operations (daily catch or haul-by-haul) , fishing data (catch area, species and quantity, date and time, and gear used), landing and sales information, license information and inspection data.

FLUX contains two distinct but related parts:

- The FLUX business layer
- The FLUX transportation layer

The core of the FLUX business layer is the detailed and standardised description of each and any data element needed. For the FLUX business layer, standardisation of the data elements and formats is based upon the UN/CEFACT approach of Business Requirements Specification (BRS).

UN/CEFCAT BRS have been defined and endorsed for the following FLUX domains:

- **Vessel Domain:** aims to standardize the exchange of fishing fleet data, and more specifically the information directly related to fishing vessels and vessels supporting fishing operations.
- **Fishing Activities Domain:** is related to data exchanges in the context of fishing activities performed by vessels during a fishing voyage. Fishing activities include all activities of vessels, related to a fishing trip. The domain contains reports related to the fishing trip: departure, arrival, entry and exit from zones, etc.
- **Vessel positions domain:** provides a standard for the communication of vessel position information (e.g. VMS or AIS) between monitoring centers.
- **Fishing licenses, authorizations and permits:** to standardize the exchange of data between stakeholders in the context of request for fishing license, authorization or permit.
- **Aggregated Catch Data (ACDR):** provides standard to exchange aggregated catch data between stakeholders.
- **Master Data Management (MDM):** encompasses exchanges from a Master Data Register to any requester of Fisheries information registered in it.

The focus of the TG should be directed to both Fisheries Activities and MDM domains that are particularly relevant to the proposals of CWP standard for DSDs covering capture, catch and effort data.

Technically speaking, FLUX is a language and not a system. It is a messenger that offers a protocol to create a secure and configurable network between different parties IT systems. UN/CEFACT provides a standardized schema for business process "XML schemas" and a standardized content called "Core Components". The components are harmonized and regularly published in UN/CEFACT Core Component Library.

FLUX offers several advantages, including free, open and global standard to automate the collection and dissemination of the fishery catch data. It provides a common approach towards electronic logbooks for fishing vessels, interoperability between IT systems, and relatively easy exchange of data between parties. FLUX is strongly tied to XML as a data format.

Notwithstanding these benefits, implementation of FLUX are to be further explored for the purpose of interoperability of the CWP global DSDs. Considering the growing importance of UN/CEFACT/FLUX in handling of fisheries data, it is strategically important that the TG output "CWP standard DSDs and embedded CWP standards and metadata" be communicated to UN/CEFACT fisheries group so that this output be up-taken in mainstream considerations by UN/CEFACT FLUX for fisheries data exchange.

The assessment of UN/CEFACT/FLUX to accommodate the CWP standard is taking place. A positive conclusion on FLUX's ability to accommodate fisheries data and requirements of Master Data Management would make it an exchange standard to be recommended and endorsed by CWP.

5.6 Actions Requested by the ad-hoc Task Group

In order to forge ahead with discussions on the topics outlined, following actions are required:

- To provide feedback on the structural elements and terminology used in the new proposals of the CWP global DSDs and related reference metadata.
- To address the modules of the global DSDs that could accommodate other domains of data collection, or different aggregation levels to match organization's policies and purposes.
- To provide guidance and recommendations regarding alternatives for data exchange format and standards.

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] 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

[3] 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.

[4] 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

[5] 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

[6] 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.

[7] 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

[8] 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

Annex 1

Table 1: List of concepts used in each Data Structure Definition

	DSD Global capture production	**DSD Catch data -**	**DSD Catch and Effort data (Logbook)**			
				Catch module	Effort module	Vessel Information module
Aquatic species	X	X	X	X		
Catch type				X		
Country/Flag state	X	X	X			X
Effort descriptor					X	
Fishery vessel		X				X
Fishing area	X	X				
Fishing gear		X			X	
Obs_Quantity	X	X	X	X	X	
Obs_Status	X	X	X	X		
Obs_Value	X	X				
Position (geographic)			X			
Unit	X	X	X	X		
Vessel Gross Tonnage						X
Vessel Length						X
Vessel Name						X
Vessel Identifier						X

Table 2: Description and type of each concept used in the DSD (in the excel file)

Concept	**Type**	**Description**
Aquatic species	Dimension	ASFIS List of Species for Fishery Statistics Purposes
Catch type	Dimension	Catch types (gross catch, retained catch, landings, nominal catch, discards) ftp://ftp.fao.org/FI/DOCUMENT/cwp/handbook/annex/AnnexB1CatchConcepts.pdf
Country/Flag state	Dimension	List of countries or areas (three digits code)
Effort Descriptor	Attribute	Effort category A, B and C (number of effort unit for each fishing gear category, number of days fishing, number of days on the ground) http://www.fao.org/fishery/cwp/handbook/N/en
Fishery vessel	Dimension	Fishery vessel type according to ISSCFV International Standard Statistical Classification of Fishery Vessels by Categories
Fishing area	Dimension	FAO major fishing areas; codes for Statistical quadrangles, and for quadrants
Fishing gear	Dimension	Fishing gear type according to ISSCFG gear category and its standard abbreviation
Obs_Measure	Measure	Amount or quantity of the observation measure (a positive integer number)
Obs_Status	Attribute	FAO Observation status codes (e.g "E" Estimate value, "R" Revised)
Obs_Value	Measure	Monetary value (value of catch, value of aquaculture production)
Position details (geographic)	Observation	Coordinates (Start and End Latitude and Longitude) expressed in WGS84, decimal degree notation, using a precision of at least 3 and maximum 6 decimal positions.
Unit	Attribute	Unit of measure (e.g tonnes, number of animals, 1000 US$)
Vessel Gross Tonnage	Measure	Gross Tonnage of vessel (Gross Tons) refers to the volume of all ship's enclosed spaces (from keel to funnel) measured to the outside of the hull framing.
Vessel Identifier	Measure	The Unique Vessel Identifier (UVI) is established by the Global Record of Fishing Vessels, Refrigerated Transport Vessels and Supply Vessels.
Vessel Length	Measure	Length overall (meters) refers to the distance measured in meters in a straight line on a line parallel to the design waterline between the foremost point of the bow and the aftermost point of the stern of a vessel outside of the main hull.
Vessel Name	Measure	Registered vessel name

Annex 2

2.1 Proposals of CWP global DSD for Capture/Catch data /Catch and Effort data (See Excel file)

2.2 Proposal of CWP global DSD for Aquaculture data (See Excel file)