TOWARDS A COMMON EXCHANGE FORMAT AND A REGIONAL DATABASE FOR LARGE PELAGIC FISHERIES

C. Rodriguez.[1], J. Barde.[2], E. Chassot.[3], P. Cauquil.[4], I. Terrier.[5], and P.Chavance.[6]

SUMMARY

To ensure an efficient use of financial and human resources, European Commission now focuses on improving data-cataloguing and on increasing data-exchanges between national and international partners. This requires more coordination in data collection and some common predefined and validated exchange formats. In that sense, EC establishes the Data Collection Framework, a Community framework for the collection, management and use of data in the fisheries sector and support for scientific advice. Within this new regulatory framework, ICES, founded and supported by the European Union, proposes the use of the FishFrame format as the data-exchange format for sampling, landings and effort data from commercial fisheries between the various institutes of Member States and for their restitution next to data calls. IRD has been working for several years in order to introduce standard formats within its IS and now proposes some possible solutions i) to export fisheries data according to "standard" formats (e.g. Fishframe, SDMX etc.) and ii) to centralize exported Fish-frame compliant datasets in a unique database. The IRD-Fishframe prototype database could be used as support within the framework of the creation of a Large Pelagic Fishes regional database (DCF).

RÉSUMÉ

Afin de garantir l'utilisation efficace des ressources financières et humaines, la Commission européenne (CE) concentre désormais ses efforts sur l'amélioration du catalogage des données et sur l'accroissement des échanges de données entre les partenaires nationaux et internationaux. Cet exercice nécessite davantage de coordination au niveau de la collecte des données et quelques formats d'échange communs prédéfinis et validés. Dans ce sens, la CE a établi le cadre de collecte des données, cadre communautaire pour la collecte, la gestion et l'utilisation des données dans le secteur de la pêche, ainsi que pour l'appui à l'avis scientifique. Au sein de ce nouveau cadre réglementaire, le CIEM, financé et soutenu par l'Union européenne, propose l'emploi du format FishFrame comme format d'échange de données pour les données d'échantillonnage, de débarquement et d'effort en provenance des pêcheries commerciales entre les divers instituts des Etats membres et pour leur restitution suite aux demandes de données. L'IRD travaille depuis plusieurs années dans le but d'introduire des formats standard dans son IS et il propose maintenant de possibles solutions : (i) exporter les données des pêcheries selon des formats « standard » (p.ex. FishFrame, SDMX, etc.) et (ii) centraliser les jeux de données conformes à FishFrame exportés dans une seule base de données. La base de données prototype IRD-FishFrame pourrait servir d'appui dans le cadre de la création d'une base de données régionale sur les grands poissons pélagiques (DCF).

1 Institut de Recherche pour le Développement, UMR 212 EME (IRD/Ifremer/UM2), Avenue Jean Monnet, BP 171, 34 203 Sète Cedex, France; celine.rodriguez@ird.fr; Phone: +33 499 573 204 Fax: +36 499 573 295.

2 Institut de Recherche pour le D´eveloppement, julien.barde@ird.fr; Phone: +33 499 573 215 Fax: +36 499 573 295.

3 Institut de Recherche pour le D´eveloppement, emmanuel.chassot@ird.fr; Phone: +33 499 573 224 Fax: +36 499 573 295.

4 Institut de Recherche pour le D´eveloppement, pascal.cauquil@ird.fr; Phone: +33 499 573 215 Fax: +36 499 573 295.

5 Institut de Recherche pour le D´eveloppement, isabelle.terrier@ird.fr; Phone: +33 499 573 205 Fax: +36 499 573 295.

6 Institut de Recherche pour le D´eveloppement, pierre.chavance@ird.fr; Phone: +33 499 573 254 Fax: +36 499 573 295.

RESUMEN

Para garantizar un uso eficaz de los recursos humanos y financieros, la Comisión Europea se centra ahora en mejorar la catalogación de los datos y en incrementar los intercambios de datos entre socios nacionales e internacionales. Esto requiere una mayor coordinación en la recopilación de datos y algunos formatos de intercambio comunes predefinidos y validados. En este sentido, la CE estableció el Marco de recopilación de datos, un marco comunitario para la recopilación, gestión y utilización de datos en el sector de pesquerías y para respaldar el asesoramiento científico. Dentro de este nuevo marco regulatorio, ICES, con el apoyo y financiación de la UE, propone utilizar el formato FishFrame para el intercambio de datos sobre muestreo, desembarque y esfuerzo de pesquerías comerciales entre varios institutos de Estados miembros, y para su restitución tras solicitudes de datos. El IRD ha trabajado durante varios años en la introducción de formatos estándar en el marco de su IS, y actualmente propone algunas soluciones posibles : i) exportar datos pesqueros en formatos "estándar" (por ejemplo, Fishframe, SDMX, etc) y ii) centralizar los conjuntos de datos exportados que cumplan las normas FishFrame en una única base de datos. La base de datos prototipo FishFrame del IRD podría utilizarse como base en el marco de la creación de una base de datos regional sobre grandes pelágicos (DCF).

KEYWORDS

Data Collection Framework, data exchange, standard format, regionnal database,commercial fishing, landings statistics, effort statistics, biological samplings

1. Introduction

In Europe, the problems of interoperability and improvement in data-exchanges related to environment have been at the heart of the concerns for several years. In particular, the INSPIRE Directive[7] has established since May 2007 an infrastructure for spatial information to support Community environmental policies, and policies or activities which may have an impact on the environment. Since then, many other programs and projects have focused on the issue of data exchanges within the EU and beyond,among which the ERA-NET Scheme[8], the Seadatanet[9], and the Emod-net [10]projects.

Regarding fisheries, the new Common Fisheries Policy (CFP) calls for more coordination in data collection, between European Union (EU) Member States (MS) and with third countries having sovereignty or jurisdiction over waters in the same region (CFP proposal art.37-38 [EC, 2011a]). In that view, the future European Maritime and Fisheries Fund (EMFF) which aims to support projects directly linked to the provision of scientific advice for the purpose of sound and efficient fisheries management decisions under the CFP, will consider eligible from 2014: "cooperation activities between the Member States in the field of data collection, including the setting-up and running of regionalized databases for storage,management and use of data which will benefit regional cooperation and improve data collection and management activities as well as the scientific expertise in support of fisheries management" (EMFFproposal art. 85 [EC, 2011b]).

In that sense, Commission Regulation (EC) No. 665/2008 of the 14 July 2008 [EC, 2009] establishes the Data Collection Framework (EU) (DCF), a Community framework for the collection, management and use of data in the fisheries sector and support for scientific advice. Under this regulation the European Commission requires MS to collect data on Biological and Economic aspects of many European fisheries and related fisheries sectors. The data collected through this program feed the work of the various end-users (i.e., ICES, STECF, RFMOs) that provide the European Commission with advice on the management of fish stocks (**Figure 2**). Discussions are currently being held about the new DCF 2014-2020 and the focus is very much on a unique data-exchange format. This can be achieved by using regional databases for management and use of data and also for data transmission to end-users. The data format used in the regional databases should be compatible with the IT structure DG MARE has started to put in place.

7 http://inspire.jrc.ec.europa.eu/
8 http://cordis.europa.eu/coordination/ era-net.htm
9 http://www.seadatanet.org/
10 http://www.emodnet-chemistry.eu/portal/portal/

Within this new regulatory framework, International Council for the Exploration of the Sea (ICES), founded and supported by the EU, proposes the use of the FishFrame format as the data-exchange format for sampling, landings and effort data from commercial fisheries between the various institutes of MS and their restitution next to data calls (Jansen, 2009).

The Tuna Observatory of the Institut de Recherche pour le Developpement (IRD) has been handling the collection and management of the fisheries data for the tropical tuna baitboat and purse seine (French fleets) since the early 1960s and 1980s in the Atlantic (AO) and the Indian (IO) oceans, respectively. Since 2011, to facilitate its ability to answer DCF data calls, IRD has started to implement the FishFrame format within its Information System (IS). To this aim, IRD has developed a Java Web application (FishframeExport) to automate the export of fisheries data (landings, effort, biological samplings) from several datasources, according to this format.

In the continuity of this work, a database has been set up in order: (i) to store exported datasets from various sources respecting the FishFrame format specifications; (ii) to directly process these data using embedded COST-R functions; and finally (iii) to link a Web portal on top of it to access and visualize data (e.g., DCF data calls management). Besides its internal use by the Tropical Tuna Observatory and its direct partners, this database could be used as a support for the development of a Regional DataBase (RDB) for Large Pelagic Fisheries (LPF) within the DCF, or within the framework of a national or international fisheries observatory.

2. Materials and Methods

2.1 The Fishframe format

2.1.1 Introduction and specifications on the exchange format

The Fishframe format (version 5.0) is a data format for sampling, landings and effort data from commercial fisheries. It has been proposed by ICES, emerged from a long development period led by DTUAqua and has been informally recognized by large parts of the fishery scientific community. The data aggregation level is as low as possible while still respecting data confidentiality issues. It fulfils all the requirements given in the new Data Collection Regulation (EU) (DCR) (EC, 2008a, 2008b) for these data types. Routinely used since 2007 by the institutes in charge of the fisheries monitoring in the Channel and the North Sea, it has given rise to the implementation of a first regional database (Fishframe), hosted and managed by ICES and accessible via http://dmz-web08.dfu.min.dk/NorthSea/FishFrame/. This approach prefigures the new DCF expectations with a network of regional databases.

The FishFrame format model consists of three main entities (**Figure 1**) :

- Commercial Landings statistics (CL): Official landings data (weight, value) with some modifiers for misreporting, aggregated per flag country, landing harbour, month, division 5◦x5◦ or 1◦x1◦,and vessel length category
- Commercial Effort statistics (CE): Effort statistics from logbooks (number of trips, number of sets, daily fishing power or capacity, etc.), aggregated per flag country, landing harbour, month, division 5◦x5◦ or 1◦x1◦, and vessel length category
- Commercial Samplings statistics (CS): Samplings statistics (raw data), obtained from various sampling protocols (on-board observer, auto-sampling, market or seller samplings, etc.).

This entity is composed of 5 sub-entities:

- Trip (TR): Data concerning a commerical fishing trip that has been sampled.
- Haul Header (HH): Detailed information about a fishing operation (a haul or a set).
- Species List (SL): The sampled strata (concerning a fishing operation) defined by species, catch category (landing/discard), landing category (human consumption/industry), sex, etc.
- Haul Length cell (HL): The length frequency in the strata (or subsample of the strata). One record represents one length class.
- Catch Aged (CA): Sex-maturity-age-weight distribution sampled respectively from the length groups. One record represents one fish.

The commercial samplings data are provided anonymised, fishing operation per fishing operation, to the lowest level of aggregation as possible (individual fish sampling). Regarding the landings and effort statistics, they are aggregated (see details above) and can directly be associated to the samplings data.

In other words, landings and effort data only concern sampled fishing operations (not the whole of the trips and operations).

Each entity (CL, CE, TR, HH, etc.) consists of:

- Several primary keys,
- One or more foreign key(s), mandatory or optional (as the corresponding primary key was informed or not in the entity of higher level),
- Attribute data (i.e., statistical measurements)

Nota bene: In addition of the elements presented above, the FishFrame format includes an optional module intended to support the Vessel Monitoring System (VMS) data for boats with an overall length > 12 m. These data are aggregated per month and 3x4 nautical miles square.

The ICES Cooperative Research Report n◦296 (Jansen, 2009) fully describes the format.

2.1.2 Implementing the format

Two implementations are possible for the exchange format: Comma Separated Values (CSV)[11] or extensible Markup Language (XML)[12]. Both formats consist of the same fields and refer to the same reference lists. The CSV version has historically been introduced. It was intended for being exploited using some statistical functions/processings (for ex. the COST-R functions[13]). From now on, the format moved in the direction of XML. The XML version is far stronger, powerful and easy-to-handle in terms of data-exchanges. The XML specifications (structure and reference lists) are given in XML Schema Definition (XSD) files, which can be downloaded via http://dmz-web08.dfu.min.dk/NorthSeaNorthSea/FishFrame/FishFrameNorthSea/FishFrame/.

2.2 Methods

The general method is composed of 3 steps that are detailed below **(Figure 3):**

- Step 1: Exporting data from different sources by using a specific driver for each database model.
- Step 2: Importing Fishframe XML data into the regional database.
- Step 3: Linking a web portal to facilitate data access/response to data calls.

2.2.1 Automating the datasets exports

Exporting the data according to the standard specifications corresponds to the preliminary step in order to correctly process it (RDB-Fishframe loading, COST-R processings, etc.). Data can (i) be exported manually when necessary, or (ii) automatically, using some stored procedures or appropriate tool. File(s) concatenation, database(s) exploitation or codes/values conversions are not always easy to implement. In order to facilitate these operations and to improve DCF data supply, IRD has developed a Java Web application, “FishFrameExport”, which allows users to easily connect a datasource (e.g., PosgreSQL database) and extract FishFrame compliant data (under conditions, see section 4.). This application is free and Open-Source and its use (even its improvement) by partners is encouraged. IRD wanted the application design to be the most generic as possible in order (i) to be able to be used by various end-users, and (ii) to allow the access to various datasources at the expense of a minimum computing development. Besides, the technical choices have been made with the matter of stability and sustainability of the application.

An “on-line” version of the Tuna Observatory application is available via http://vmot-proto.mpl. ird.fr:8080/jaxb-cost-export/Welcome. It allows user to export Fishframe compliant datasets from the IRD-Tuna Observatory databases. It requires user to be aware of the databases connection parameters.

11 ASCII-file with the values separated by either semicolon or comma.
12 XML-file following the W3C-standard for exchanging data over the internet (http://www.w3.org/XML.
13 http://wwz.ifremer.fr/cost/Cost-Project.

The FishFrameExport application has been developed in Java. On line as on localhost, it is thus independent of the operating system (pending availability of a Java Virtual Machine (JVM)). It was built in order (i) to get the application independent from datasources, and (ii) to facilitate the "connection" to multiple sources according to the user's needs. In order to achieve this, IRD decided to implement the Design Pattern "bridge". The bridge pattern is a design pattern which is meant to "decouple an abstraction from its implementation so that the two can vary independently". When a class varies often, the features of object-oriented programming become very useful because changes to a program's code can be made easily with minimal prior knowledge about the program. FishFrameExport is thus structured around 2 entities:

- A "driver", which depends on the datasource from which the fishing statistics need to be extracted. It allows the connection to the datasource, the data extraction as well as their "formatting" for a correct conversion into the FishFrame format

- A "core": central and generic part of the program which is able to call the driver, to convert the codes (code-lists), to export correctly formatted data8 and to validate the exports.

The XSD files and a java library (JAXB) greatly facilitate the data export and conversion in au tomating the XML-objects mappings and XML document (output) validation.

In order to respect the code-lists imposed by Fishframe, several methods for converting code-lists and/or values (attributes) were adopted:

- Simple conversions, within the SQL queries (driver) (e.g. numeric variables or literal results calculated as of several parameters existing in the database),
- Codes conversions using ontologies,
- When possible, spatial extractions (using the PostGIS functions, for example)
  Example: extraction of a geographic area code from some latitude/longitude information. Obviously, it supposes to have geographic objects loaded into the database which contain Fishframe expected codes.

To sum up, to export Fishframe compliant data from a new datasource using FishFrameExport, it is necessary:

1. To compute a dedicated "driver". It consists of writing about 10 SQL queries (in case of database) and incorporate it in the java code (driver),[14]
2. To introduce or more conversion mechanisms[15],
3. To configure the conversions filling in a XML configuration file.

2.2.2 The IRD-Fishframe database

The database creation. In parallel with the FishFrameExport development, in order to merge its "DCF" data and to facilitate the response to EU data calls, IRD created a "Fishframe" database (PostgreSQL) capable to store the exported datasets (XML format) regardless of their origin (datasource) as long as they respect the format specifications. Furthurmore, this database can be used fo merging datasets produced by different partners ; for example: study case IRD-Ifremer on swordfish (*Xiphias gladius*) in the Indian Ocean.

In order to anticipate possible changes in the format (or even changes of format), IRD chose to automate as far as possible the database creation script using (i) the automatically generated java classes within the FishFrameExport project and (ii) the useful Hibernate java framework[16]. The database physical model is given in **Figure 4.**

The database uploads. In order to facilitate the database upload, IRD created a Java tool[17] from which a user can upload his Fishframe XML file. This application could be further extended in order to include database management functions (datasets metadata and data-calls management etc.). Finally, the aim is to link a web portal to facilitate data access/response to data calls.

14 CSV or XML files, according to the user's choice.
15 Within the framework of IRD exports, the conversion mechanism integrate an ontological model
16 Hibernate is an object-relational mapping (ORM) library for the Java language, providing a framework for mapping an object-oriented domain model (java classes and objects) to a traditional relational database. Hibernate is free software that is distributed under the GNU Lesser General Public License.
17 Prototype of Java Web Application.

Note: the two projects (exporting tool and database creation/management) are completely independent from the user's point of view, but are fully linked from the technical point. In other words, a single technical intervention is necessary in case of changes in the format model, consequences will be passed on both projects (java classes).

Embedding the COST-R functions IRD is currently working on embedding the COST-R functions in the database to allow a non-R user to proceed its data using some PostgreSQL stored functions/procedures, without data transformation. This is done translating the R functions into PostgreSQL ones by using the PL/R language ; these PostgreSQL functions are stored in the database.

3. Results

The tools developed by IRD to introduce the FishFrame format within its information system and the testing conduced using the Tuna Observatory data seem to be conclusive and promising to facilitate fisheries data-exchanges. The more significant problems arise from the history of the Fishframe format. Indeed, the format has been created and deployed for/within the European northern countries fisheries. This actually results in some poor code-lists (e.g. species, commercial size categories, countries) and missing or inappropriate fields (e.g no school type information, number of sets for effort statistics not entirely appropriate for longliners data). A collaboration with the ICES team in charge of the Baltic/North Sea RDB has been developed to facilitate the extension of code-lists while discussions are currently conducted within the Regional Coordination Meetings (RCMs) for the Mediterranean Sea and on Long-Distance Fisheries for future model improvements and adaptations. Overall, the technical solutions that IRD has adopted can be easily transmitted to anyone interested in such an approach, even improved or adapted in case of a change in the format.

All source-codes developed by IRD are available via following the subversion repositories:

- the "FishFrameExport" Maven project:
  https://svn.mpl.ird.fr/umreme-ot/jaxbcostexport/trunk/jaxb-cost-export/
- the dependencies[18]:
  https://svn.mpl.ird.fr/umreme-ot/costlibs/maven-cost-commons/trunk/maven-cost-commons/
- some Tuna Observatory drivers (examples): https://svn.mpl.ird.fr/umreme-ot/costlibs/
- associated reports and documentations:
  https://svn.mpl.ird.fr/umreme-ot/costlibs/fishframe-project-report/trunk/

4. Discussion

To ensure an efficient use of financial and human resources, the European Commission now focuses on improving data-cataloguing and on increasing data-exchanges between national and international partners. This requires more coordination in data collection and some common predefined and validated exchange formats. In terms of data-exchanges, it is important to note that there is not a single format but an array of formats, each one with its own specificities. The aim of the new DCF (2014-2020) should be to ensure that everybody involved in the data collection programmes will work with only one data format. This can be achieved by using regional databases for management and use of data and also for data transmission to end-users (e.g., RFMOs). The data format used in the regional databases should be compatible with the IT structure DG MARE has started to put in place. In parallel, wherever possible, the same aggregation levels should be used for all data so that comparability is ensured. Currently, this is not possible for economic data.

No format is currently explicitly named nor approved by the EU with regards to fisheries statistics. The ICES Fishframe format suffers from a lack of visibility within the end-users community and we can expect that new formats will appear in the near future. For example, works on the Electronic recording and reporting system (ERS)[19] are currently ongoing and the EU has launched discussions about the possibility of combining the ERS and the FishFrame formats.

Nevertheless, IRD has been working for several years to introduce standard formats within its IS and now proposes some possible solutions: (i) to export fisheries data respecting "standard" formats (e.g. Fishframe, SDMX, etc.)

18 The dependencies are common librairies used within both the exporting and the database creation/management tools.
19 http://ec.europa.eu/fisheries/cfp/control/technologies/ers/index_en.htm

and (ii) to centralize exported Fishframe compliant datasets in a database, regardless of their origin. The export application can be extended to as many datasources as necessary at the expense of minimum computing development: 10 SQL queries and eventually one or more conversion mechanisms for each source. The IRD-Fishframe prototype database could be used as support within the framework of the creation of a Large Pelagic Fishes regional database (DCF).

For RFMOs obligations, in the framework of the External evaluation of the DCF, an analysis will be carried out in 2012 on the coherence between the current Multi-Annual Programme (MAP) and the data requirements for RFMOs. In the new EU MAP, the data collected aims to be in line with the need of the RFMOs as far as possible. RFMOs data collection obligations are enshrined in international decisions with different deadlines and reference periods. The adjustment of data formats required by RFMOs will only be possible over the long term taking into account that the underlying provisions need to be agreed by all contracting parties. RFMOs will have to be consulted in a more efficient way, by corresponding with the relevant data managers, and it should be considered a priority to harmonize data collection formats between the different RFMOs and the Commission. It could be a possibility to entrust JRC to act as interface between MS and RFMOs. In this way, a regional database could be an appropriate solution.

Acknowledgements

We sincerely thank Henrik Degel (ICES), Henrik Kjems-Nielsen (ICES) and Francois Bastardie (DTU Aqua) for always being disponible to answer our questions and for helping us to implement the Fishfram format within our information system. We thank Norbert Billet (IRD) for being a constant good technical support. Finally, a great thank to Jo¨el Vigneau (Ifremer) and all the “COST” project team for helping us to use the COST-R libraries (useful to link common statistical procedures to the regional database).

References

EC. Commission Decision of 18 December 2009 adopting a Multiannual Community Programme for the Collection, Management and Use of Data in the Fisheries Sector for the Period 2011-2013, 2009.

EC. Proposal for a regulation of the European Parliament and of the Council on the Common Fisheries Policy, July 2011a. COM(2011) 425 final.

EC. Regulation of the European Parliament and of the Council on the European Maritime and Fisheries Fund Repealing Council Regulation (EC) no 1198/2006 and Council Regulation (EC) nº 861/2006 and Council. Regulation no xxx/2011 on Integrated Maritime Policy], December 2011b. COM(2011) 804 final.

Jansen T. Definition of Standard data-exchange Format for Sampling, Landings, and Effort Data from Commercial fisheries. Technical report, ICES Cooperative Research Report No. 296. 43pp, 2009.

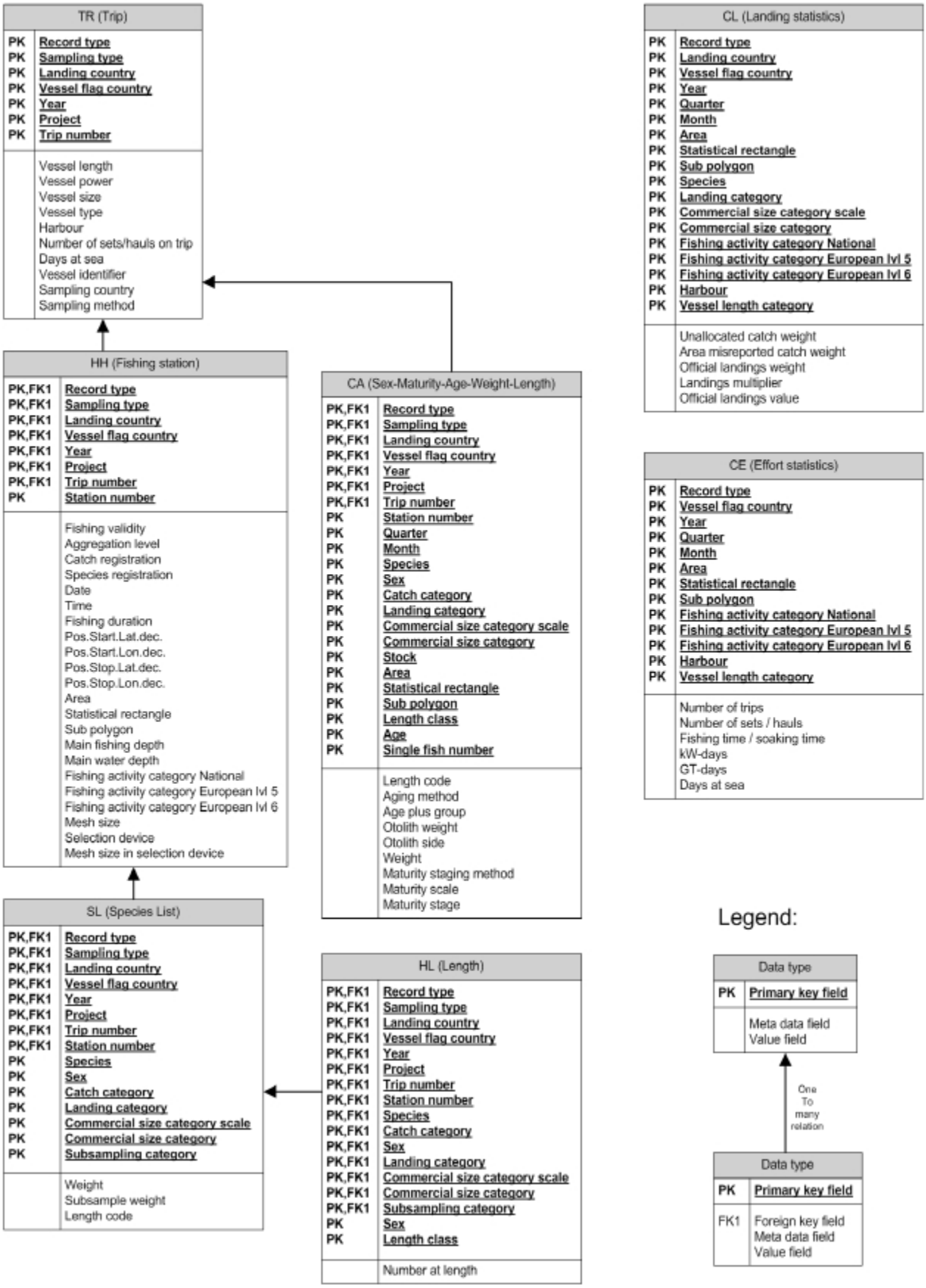

Figure 1: The Fishframe data types.

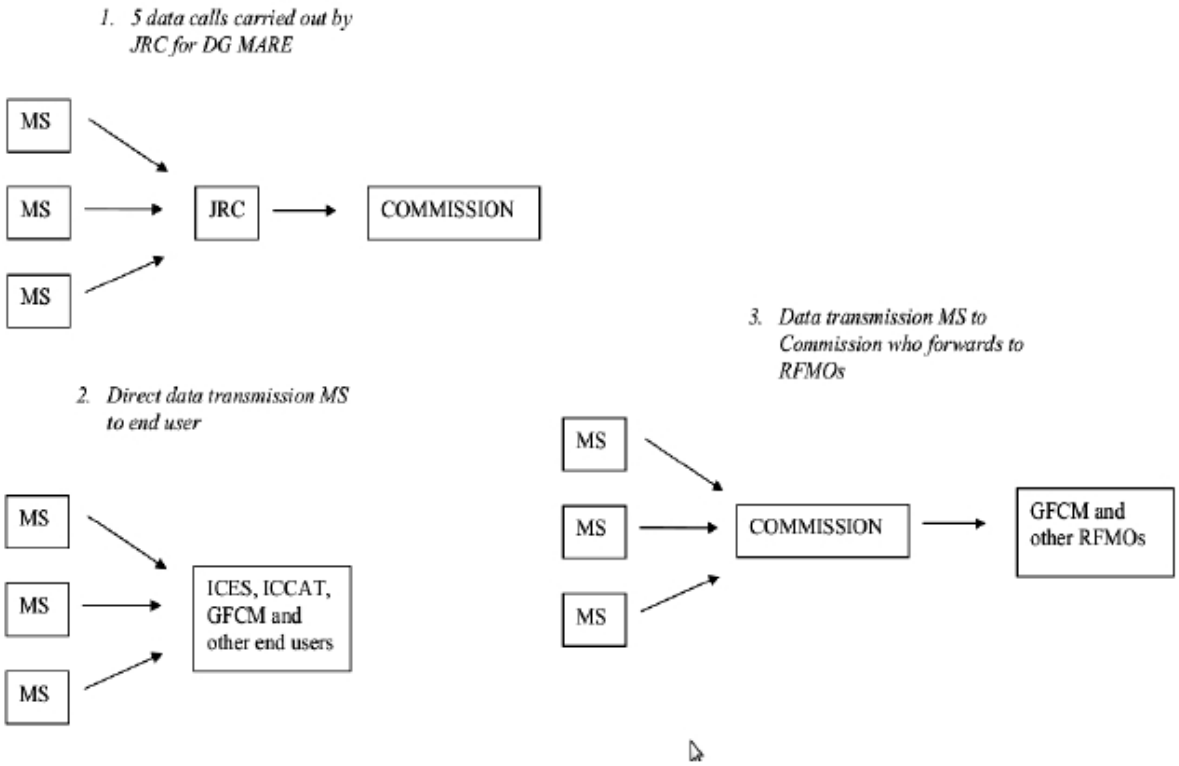

Figure 2. Data transmission flows under the current DCF.

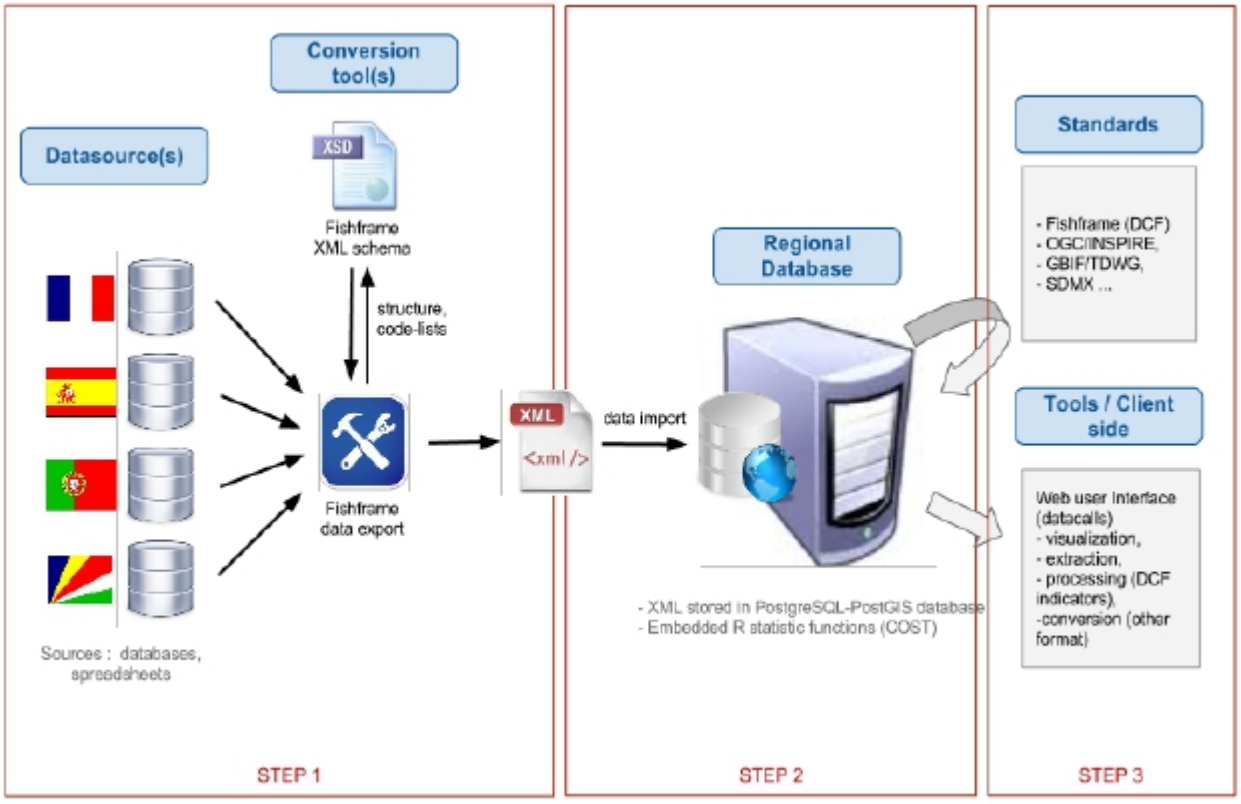

Figure 3. DCF data flows using the Fishframe format and a regional database.

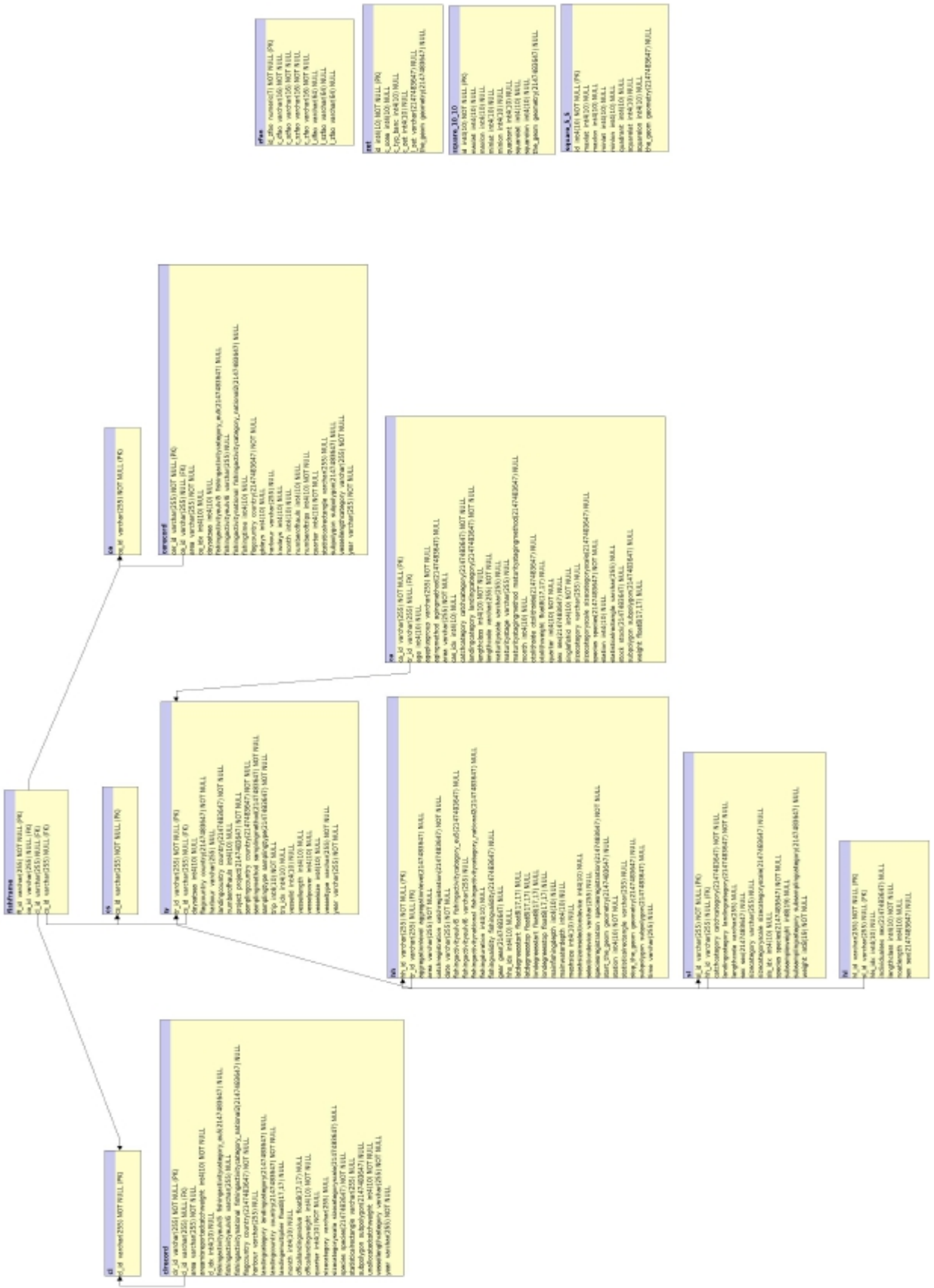

Figure 4. The IRD-Fishframe dababase physical model.