Scientific data storage and transmission under the future Data Collection Framework

Feasibility Study

FINAL REPORT

Part 1

Contract n°:	**MARE/2012/22 – Lot 2 (SI2.656640)**
Date of submission:	**4 September 2014**
Authors:	**Pavel Salz, José Cervera, Christine Alberti, Yann Laurent, Suitbert Schmudderich, Christoph Petereit, Anna Madriles, Ester Azorín, Olivier Roux**
Version:	**V 4**

Table of Contents

List of Acronyms

ACR	Aggregated Catch Report
AMAP	Artic Monitoring Programme
BISE	Biodiversity Information System for Europe
BITS	Baltic International Trawl Survey
BTS	Beam Trawl Survey
BS	Baltic Sea
BSC	Black Sea Convention
BSERP	Black Sea Ecosystem Recovery Project
BSIS	Black Sea Information System
CAMP	Coordinated Atmospheric Programme
CCSBT	Commission for the Conservation of Southern Bluefin Tuna
CDR	Central Data Repository
CEMP	Coordinated Environmental Monitoring Programme
CFP	Common Fisheries Policy
CR	Control Regulation
CSP	Community Statistical Programme
CV	Coefficient of variation
CWP	Coordination Working Party
DAD	Database on Accessions and Documentation (ICES)
DATSU	Data Screening Utility (ICES)
DB	Database(s)
DBMS	Database Management System
DCF	Data Collection Framework
DG MARE	European Commission /Directorate General for Maritime Affairs and Fisheries
DiGIR	Distributed Generic Information Retrieval
DV	Data Validation tool
DW	Distant Waters fisheries
DWH	Data Warehouse
EBB	Editing Building Block (recently renamed to EDIT)
EC	European Commission
eDAMIS	Electronic Data file Administration and Management Information System
EEA	European Environment Agency
EFTA	European Free Trade Association
EIONET	European Environment and Observation Network
EMFF	European Maritime and Fisheries Fund
EMODnet	European Marine Observation Data Network
ESAC	European Statistical Advisory Committee
ESS	European Statistical System
ESSC	European Statistical System Committee
ETC	European Topic Centres
EU	European Union
EWG	Expert Working Group
FAO	Food and Agriculture Organization of the United Nations
FLUX	Fisheries Language for Universal Exchange
GES	Good Environmental Status
GFCM	General Fisheries Commission of the Mediterranean
GSAs	Geographical Sub-Area(s)
HELCOM	Helsinki Commission
IBTS	International Bottom Trawl Survey

IBTSWG	International Bottom Trawl Survey Working Group
ICCAT	International Commission for the Conservation of the Atlantic Tunas
ICES	International Council for the Exploration of the Sea
IFDMP	Integrated Fisheries Data Management Programme
IMP	Integrated Maritime Policy
IOTC	Indian Ocean Tuna Commission
JAMP	Joint Assessment and Monitoring Programme
JRC	Joint Research Centre
LP	Large Pelagic
MAP	Mediterranean Accion Plan
MBS	Mediterranean and Black Sea
MDR	Master Data Register
MDT	Multi Dimensional Table
MEDITS	International Bottom Trawl Survey in the Mediterranean
MEETS	Modernisation of European Enterprise and Trade Statistics programme
MOLAP	Multidimensional, on-line analytical Processing
MoU	Memorandum of Understanding
MPA	Marine Protected Area
MS	Member States
MSFD	Marine Strategy Framework Directive
MSY	Maximum Sustainable Yield
NA	North Atlantic
NAFO	Northwest Atlantic Fisheries Organization
NC	National Correspondent
NS	North Sea and NE Artic
NSI	National Statistical Institute
OLA	Online Analytical Processing engine
OSPAR	Convention for the Protection of the Marine Environment of the North East Atlantic
RCG(s)	Regional Coordination Group(s)
RCM	Regional Coordination Meeting(s)
RDBs	Regional databases
RECO	Reference Code Vocabularies (ICES)
RFMO(s)	Regional Fisheries Management Organization(s)
RSC(s)	Regional Sea Convention(s)
RSP	Regional Sea Programme
SBS	Structural Business Survey
SEIS	Shared Environmental Information System
SEP	Single Entry Point
SMALK	Sex-maturity-age-length keys
SO	Specific Objective
SR	Statistical regulations
SSB	Spawning Stock Biomass
STECF	Scientific, Technical and Economic Committee for Fisheries
SWOT	Strengths, Weakness, Opportunities and Threats
TLW	Tones Live Weight
ToR	Terms of Reference
UML	Unified Modelling Language
UN/CEFACT	United Nations Centre for Trade Facilitation and Electronic Business
VMS	Vessel Monitoring System
WEST	Weight of the stocks

WG	Working Group
WGDIKE	Working Group on Data, Information and Knowledge Exchange
WGIBTS	Working Group on International Bottom Trawl Survey
WGNEACS	Working Group on North-East Atlantic Continental Slope Surveys
WISE	Water Information System for Europe
WKBALT	Benchmark Workshop on Baltic Multispecies Assessments
WKBEAM	ICES Working Group on Beam Trawl Survey
WoRMS	World Register of Marine Species

Part 1

EXECUTIVE SUMMARY

Background

Collection of biological and economic data in support of the Common Fisheries Policy (CFP) has been implemented under a number of consecutive Council Regulations and Commission Decisions since 2002, the most recent being EC Decision 93/2010 and the Commission Implementing Decisions C (2013) 5243 and C (2013) 5568 in relation to the roll-over for the period 2014-2016. During this period a number of important developments took place – the EU expanded from 15 to 28 Member States (MS), various new policies were adopted, including the Marine Strategy Framework Directive (MSFD) and Integrated Maritime Policy (IMP) and the revision of the CFP. In the new CFP, it was recognized that data collection plays a fundamental role for implementation of the CFP process. Financial support for data collection will be assured through the resources available under the European Maritime and Fisheries Fund (EMFF). Last, but not least, parties involved in fisheries data collection, at national and at supra-national level, have gained meaningful experience, ranging from set-up of surveys to software solutions for data processing.

Increasing demands on data quantity and quality coupled with budgetary constraints call for a continued effort to improve effectiveness and efficiency of data collection and processing. Existing duplications should be eliminated. Priorities have to be set focussing on data which is used, especially for fisheries policy decision-making. Supra-national coordination, which is already taking place in the DCF Regional Coordination Groups, allows tasks sharing regarding data collection effort and should be continued and strengthened.

Objectives

The present study pursues three objectives:

1. Review of the current situation at supra-national level in relation to data storage, transmission, quality control and dissemination. The review also considers access to the Fisheries Control Regulation data, which is fundamental for the implementation of the data collection, and interactions with other EU policies (MSFD and IMP) and their data needs. Finally, the study reviews the institutional aspects of data collection, with special attention to assurance of data confidentiality.
2. Design of a number of scenarios for the organization of the fisheries data system at supra-national level.
3. Assessment of the scenarios against a number of policy objectives:
    i. Simplification and costs reduction.
    ii. Adaptations in relation to regionalization and bio-economic integration.

   iii. Strengthening quality assessment.
   iv. Accessibility.
   v. Coherence with the Integrated Fisheries Data Management Programme (IFDMP).
4. Assessment of the scenarios against three constraints:
   i. Legal.
   ii. Administrative.
   iii. Financial.

Methodology

The study was implemented by a team of consultants including experts in fisheries economics, biology, statistics and IT from the consortium composed by DevStat (leader), Framian, COFAD and GOPA. The team acknowledges the fruitful collaboration of experts from the European Commission (DG MARE, Eurostat and JRC), regional and national institutions.

The information for the study was compiled as follows:

- Review of reports from the Scientific, Technical and Economic Committee for Fisheries (STECF), the International Council for the Exploration of the Sea (ICES), the Joint Research Centre (JRC) and the relevant Impact Assessments;
- Detailed questionnaire sent to all DCF national correspondents;
- Detailed questionnaire sent to all supra-national actors: ICES, JRC, Eurostat, DG MARE, DG ENV, Regional Sea Conventions (RSCs: Helcom, Ospar, Bucharest Convention, Barcelona Convention), General Fisheries Commission of the Mediterranean (GFCM) and other Regional Fisheries Management Organizations (RFMOs);
- Individual interviews with ICES, JRC Eurostat and DG MARE;
- A one-day focus group meeting with participants from national and supra-national level of Data Collection Framework (DCF).

The information was processed in a detailed report on the current situation (presented in Part 2 of this report) and summarized (Part 1).

Three scenarios were designed in dialogue with DG MARE:

1. Supra-regional database ("Eurostat model").
2. Regional nodes ("RDB-Fishframe model").
3. Network ("EMODnet model").

The scenarios were developed by the consultants' team and presented to the focus group, which gave a first reaction in writing, based on a set of common questions. The responses were integrated in a document which served as a basis for the one-day meeting. The meeting also discussed the conditions of a fourth scenario, which would combine various characteristics of the

above three scenarios. This scenario is elaborated in the last phase of the project under the name 'Fisheries data hub'.

The summary of the present situation together with the description and evaluation of the four scenarios is presented in Part 1 of this report.

Current situation

Most national correspondents indicated that they have satisfactory access to Control Regulation data, which is a fundamental starting point for the design of surveys under DCF.

DCF data is stored in a number of **databases** held at ICES and JRC:

- ICES – RDB-Fishframe, with regionally flagged primary biological data for the Atlantic areas (including North Sea and Baltic Sea);
- ICES – DATRAS, containing data of surveys at sea;
- ICES – InterCatch, with aggregates for stock assessment;
- JRC – three databases with economic data on fleet, aquaculture and fish processing;
- JRC – several databases with biological data on the Mediterranean and Black Sea, including one on the Medits survey;
- JRC – effort data in relation to various effort regimes.

A number of these databases contain similar, albeit not identical, data related to catches and effort (so called transversal data). The data may be defined differently as it serves different purposes. The main difference lies in the level of aggregation in relation to gear, space and time.

Eurostat compiles and disseminates EU-wide data on catches, landings, fleet, aquaculture and fish processing which is closely related to DCF. There are many similarities and overlaps between the data compiled under the DCF and the Statistical Regulations (SRs). There are also differences in SR definitions (e.g. DCF uses landings by fleet of a flag state independently of place of landing, while SR uses landings into a Member State, independently of flag of the vessels). Streamlining or coordination of data flows would reduce the burden to data providers.

DG MARE holds two separate databases on catches / landings and on effort, exclusively for its own use in relation to the implementation of the CFP.

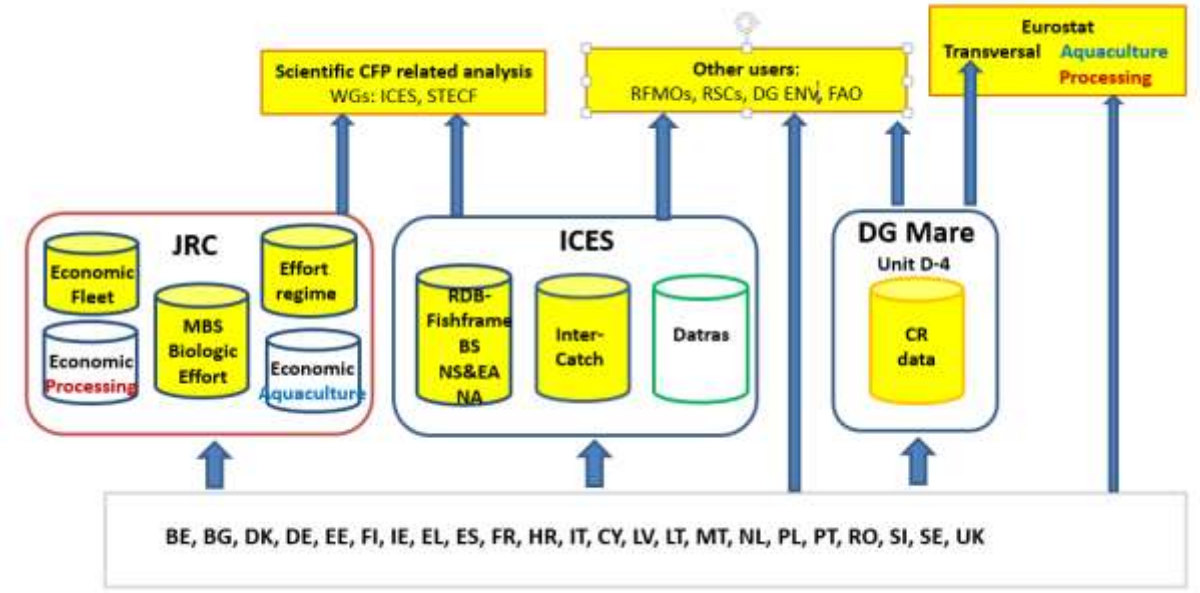
GFCM and other RFMOs hold a variety of data, usually in relatively simple databases. This data originates either from the Control Regulation (transversal) or from DCF. RSCs hold almost no DCF data, as their primary focus is on environment.

Regarding the content of data, MS transmit different data to different databases, although some **overlaps** exist (see also figure 1):

- **Biologic indicators on Atlantic stocks** are transmitted by the MS to RDB-Fishframe and to InterCatch DB operated by ICES. ICES is developing software to transfer data from RDB-FishFrame directly to InterCatch.
- Total MS **landings** (by species and FAO areas) are based almost always on CR data. They are transmitted to Eurostat[1], JRC (fleet economic data), ICES (RDB-Fishframe), FAO, GFCM and the RFMOs.
- **Effort** data submitted to DG MARE under the CR (for compliance purposes) overlaps significantly with the MS submissions of effort data under almost all DCF data calls carried out by JRC and ICES for scientific purposes. However, the purposes of the various DCF calls are very different.
- Eurostat and the JRC both collect data on **aquaculture** on the basis of Statistical regulation 762/2008 and of the DCF Regulation, respectively. The DCF covers also aquaculture production costs, while SR focuses on production volume and value only.
- Eurostat and JRC both collect data on **fish processing**. DCF requires to collect 17 indicators of which 11 are identical or closely related to SBS. The link between DCF and SBS is explicitly stated in Appendix XII of the Regulation 93/2010. Eurostat data calls are very detailed, being part of the broader SBS. Given that the agreed format is used for all other activities under SBS, it seems difficult to envisage that data transmission is changed to include the additional variables collected under DCF. The fish processing industry is a very small part of the SBS.
- Data provided by MS to **GFCM** is based on data collected by the MS under the Statistical or DCF regulation. It is relatively less detailed.

[1] *It is noted that in DCF catches of a MS are sum of landings and discards of the national fleet. However, Eurostat uses a different definition, where landings in a MS are sum of landings in ports of a MS by all vessels, national and foreign. In addition, Eurostat presents landings in product weight, while DCF used live weight.*

Figure 1. Overlaps of data flows



Data calls and data **upload** for the main actors involved in DCF follows similar upload processes. Up-to-date Excel files describe formats and nomenclatures for data reporting. The major difference is in the level of automation of processes. It goes from a manual process of uploading text files in the database, to a semi-automated process requiring some internal manipulation, to automated upload in the database by the provider.

The partial data received by the project from the national correspondents indicates that responding to the data calls requires in total about 2,000 person-days per year (for all MS involved in DCF in 2012-13) Considering that the available information is incomplete, the total effort at EU level-is maybe 2 times as high, i.e. 4,000 person-days per year. Nationally, this ranges from about 10 person days in Sweden to over 1,000 person-days per year in France.

The global level of technology in data upload management is high. Processes and workflows are well identified and managed. Formats are easily accessible. Upload is easily done through secured interfaces or by skilled data managers. In most cases manuals are available and updated regularly.

A large number of **quality controls** are run with different software tools, showing that although a far reaching quality control exists, further standardization should be achieved and the most efficient software tools implementing commonly agreed validation rules should be shared among Member States and supra-national institutions.

Quality reporting and the dissemination of such reports, is limited and fragmented. There are no standards for fisheries data quality reporting shared by supra-national institutions despite the existence of international quality assessment frameworks for a large variety of other data domains.

All DCF data are **accessible** to professional users (i.e. scientific WGs at ICES, STECF, RFMOs, etc.), although at different levels of aggregation. Public access to DCF data is not comprehensive. It is spread over several websites and many separate files. Users who are not intimately acquainted with the system are likely to face problems in accessing the data.

Interactions between DCF and MSFD can be analysed only on a provisional basis as MSFD is still in an early stage. Calculation of MSFD descriptors (mainly descriptor 3) requires a regular "flow" of data from DCF. However, in most cases, it is not possible to derive MSFD indicators and descriptors in an automatic way (e. g. through fixed algorithm) from DCF data, thus, human intervention is required. In view of the role of fisheries in the blue economy, DCF data is likely to become increasingly relevant in the context of IMP, although the exact use still remains to be determined.

JRC, Eurostat and ICES all have a suitable institutional position to play a role in DCF context in the future. In particular they all guarantee the **confidentiality** of the data.

Effectiveness and efficiency of the current system can be increased by:

- Elimination of duplications in data flows, particularly in relation to fisheries (or 'transversal') data, and on a second level, on aquaculture and fish processing industries data.
- Application of common IT tools on all levels – data storage, upload and quality control.
- Sharing primary data, to reduce MS data collection and/or provision burden.

Scenarios

The report presents four scenarios.

1. Supra-regional database ("Eurostat model").
2. Regional nodes ("RDB-FishFrame model").
3. Network ("EMODnet model").
4. Fisheries data hub ("Combination model").

The scenarios are built bearing in mind three dimensions:

- **Geographic** coverage, i.e. should the data system be EU-wide or region-specific.
- **Thematic** coverage, i.e. should biological and economic data be (fully) integrated or not, or to which extent.
- **Functional** coverage, i.e. who should be responsible for what in terms of data collection, storage, transmission and quality control.

The scenarios have a number of common features, which are related to policy objectives and constraints:

Table 1. Common elements in all scenarios

Policy objective / constraint	Common elements to all scenarios
Informing CFP/ IMP / MSFD	The scenarios focus on biological and economic marine data. Ability to inform policy depends on the level of required detail and feasibility to translate the information needs to the collected data.
Regionalization / coordination	RCGs continue playing role in coordination of national data collection programmes and if necessary definition of region-specific data. Coordination is required in most aspects mentioned under 'policy objectives'.
Simplification	Use of existing practices of Eurostat, ICES or JRC.
- Data storage	Experience of Eurostat, ICES and JRC in designing and maintaining large databases. National DBs likely to remain for national purposes. Strict coordination on definitions is required for aggregations, as checks are not possible at higher level. Agreement on aggregation and compilation rules needed.
- Data access / dissemination	Access rights must be defined for different types of users (professional, general public). Standard queries can be developed to generate information required by various users. Users must be informed about data sources and quality.
- Data upload	Automated procedures can be developed for any scenario using FLUX. Availability of EU-wide data set by slowest partner (except if derogations are provided for MS that represent a small contribution to aggregates). Loss of visibility of current data disseminators, i.e. national institutes or JRC and ICES.
Costs reduction	MS will not abandon their DBs until they are convinced that the supra-national system works properly and meets their need. This means that in medium term the costs of the present system will continue and additional costs will be faced to make adaptations for the new system.
- Operation	In the long term costs of operation of an integrated system will be lower than the sum of the costs incurred by the MS and supra-national entities. These costs will probably need to be borne by another institution (e.g. the Commission) and not by the national authorities.
- Investment	National systems will have to be adapted. New EU-wide system needs to be developed.
Quality control	Common quality controls need to be implemented and documented.
Bio-economic integration	Bio-economic integration requires conceptual development and its empirical application, leading to common set of data definitions [i.e. fleet segments (gears and vessel sizes) and spatial and temporal scale of catches and effort]. National fleet segments should be replaced by EU fleet segments, e.g. performance of 12-24m demersal trawlers of individual MS, should be aggregated to one EU segment.
Coherence with IFDMP	Data transmission takes place using UN/CEFACT (United Nations Centre for Trade Facilitation and Electronic Business) certified system FLUX (Fisheries Language for Universal eXchange) presently developed by DG MARE. FLUX is a tool which facilitates transportation of data: it is a piece of software (transportation layer) and a methodology (business layer to standardize the messages sent through the transportation layer). Common nomenclature is based on the Master Data Register.

Policy objective / constraint	Common elements to all scenarios
Legal constraints	DBs in all scenarios contain or have access to primary data. Confidentiality of personal data has to be guaranteed. Providing access to confidential data to institutions other than Eurostat may be rejected by the MS. Close dialogue with the European Data Protection Supervisor, as well as the use of statistical disclosure control techniques, is a must.[2]
Administrative constraints	Administrative arrangements between DG MARE and the institutions responsible for supra-national implementation needed.
Financial constraints	MS which have developed advanced IT systems may be obliged to introduce far reaching / costly adaptations. Financial arrangements between DG Mare and the institutions responsible for supra-national implementation needed.

Other common features are:

- Organization of the work flow is the central theme. IT infrastructure is not mandatorily a copy of the workflow, as illustrated in the following text. Hardware, software and data work flows can be mutually independent. This means that data collection is done by national institutes, but all subsequent data processes may be done by specialized units at supra-national level and the DBs may be located outside the national institutes. For example:
  - Primary data is uploaded to a "cloud DB", composed of a set of national DBs. The "cloud DB" may be located in "A".
  - Quality control (using standard routines) is done by a central unit of the "cloud DB". Quality control unit may be located in "B".
  - Development of routines for data processing is done by specialised WGs (stocks, economics, etc.) located in "C" and "D".
- Physical location of the data (bases) is not relevant. Various stakeholders can be responsible and have exclusive access to parts of the data, located at any place / institution. One system can also consist of several DBs in different locations, linked through Internet.
- Tasks and responsibilities of actors at national and supra-national level must be logically formulated, avoiding duplications.
- Ownership of the data has to be determined and the consequences incorporated in the system.
- Access rights of different users must be clearly defined.
- Transparency is pursued by detailed documentation of all workflows and contents.
- Internationally recognized code lists are used in all elements of the system.

[2] *Individual (confidential) data are already being stored at supra-national level at Eurostat and DG Agri (RICA). This demonstrates that sufficient confidentiality safeguards have been put in place there and can be also applied to fisheries data.*

- Transfer of experience among MS is pursued, including IT tools, statistical methods, quality procedures, etc.
- The scenarios do not cover economic data on fish processing. This area is largely covered by Eurostat's Structural Business Survey (SBS). It is proposed to review in detail whether the indicators required by DCF, but not covered by SBS, are truly essential and justify additional data collection effort and the required resources. Furthermore, SBS does not cover small firms (less than 10 employees). For this group, ad hoc surveys may need to be carried out if data on small firms is considered necessary for DCF end users. Execution of such surveys would have to take place in all scenarios in the same way. The surveys are 'scenario independent' and therefore have not been further discussed.
- Aquaculture data is assumed to be stored in a separate DB, operated by JRC. This DB would also provide input to Eurostat, eliminating submission by individual MS.
- Ad hoc data calls by various scientific working groups cannot be avoided and may require development of specific procedures.

Common definitions and aggregation levels have been already agreed upon in the context of MS responding to ICES and JRC data calls. This means that the tools exist to generate a common data set at national level. However, it is likely that aggregations agreed upon today will be outdated sometimes in the future. Therefore, all scenarios stress the need to create systems on the basis of detailed (anonymised) or even primary (non-anonymised) data. Such systems will always have the flexibility to generate aggregations required at any time.The four scenarios are described in UML (Unified Modelling Language) diagrams in Annex 2.7. The scenarios are briefly presented with their pros and cons below, followed by a graphic presentation and a summary comparison against the policy objectives and constraints.

Summary of the scenarios

Scenario 1 "Supra-regional DB" (Eurostat model)

This scenario proposes to create one DB containing detailed biological, economic and fisheries data. The DB would be operated by an institution, which remains to be designated. The institution adheres to the standard operation of the European Statistical System (ESS), which includes also confidentiality / access rules. MS upload detailed data at an agreed schedule. 'Push' approach (uploading by MS) would be gradually replaced by 'pull' (data extracted from the national DB by the operator of the Supra-regional DB). Web-based tools allow extracting aggregations. Different levels of users (professional, general public) would have different access rights. A main assumption is that the MS would be willing to provide the detailed (anonymized) data. The scenario assumes that data on fish processing would be drawn from Eurostat's SBS and aquaculture data would be stored in a small separate DB at JRC.

This model would offer a number of advantages. Data upload from MS would be simplified as the overlaps between data flows from MS to ICES, JRC, Eurostat and others would be eliminated.

Fisheries data on capacity, effort, catches and landings would be provided by the MS only to the 'Supra-regional DB', from where it would be further distributed to JRC, ICES, Eurostat, DG MARE and other users, according to their specific requirments. Data integration for biological, economic and transversal data would be easier as this data would be stored in one single database. The risk of data incoherence[3] would be minimised, thus increasing the quality of analysis as well as users' trust in the institutions providing the data. There would be no need for developing regional databases, as regional data sets could be generated from the "Supra-regional DB".

The scenario faces also a number of challenges. Development would be needed, for the "Supra-regional DB" and some of the new data transmission procedures. The organisation in charge of maintaining the supra-regional database would have to be designated. New arrangements would be necessary to replace the present system of data calls by regular fixed schedules of automated data uploads by MS to populate this supra-regional database.

Scenario 2 "Regional nodes" (RDB FishFrame model)

This scenario would be based on five regional databases (RDBs) for the Baltic Sea (BS RDB), North Sea and NE Atlantic (NS+NEA RDB), North Atlantic (NA RDB), Mediterranean and Black Sea (MBS RDB) and 'Distant waters' (DW RDB). The DBs would be based on RDB-FishFrame, extended with fleet economic data. They would contain primary biological data and aggregated economic and fisheries data. MBS RDB would be hosted by GFCM. The other four RDBs would be hosted by ICES. In practice at ICES there would be one single DB, in which data is 'flagged' according to the region, which is also the situation at present. Optionally, the Institut de Recherche pour le Développement (IRD, France) could be also a candidate to host the Distant Waters (DW) RDB, if it would expand its Large Pelagics RDB (also based on RDB-FishFrame) to include other EU distant water fisheries. The users would access the different RDBs according to their specific regional needs. The main assumption is that ICES can expand the RDB-FishFrame to include the fleet economic data and new working arrangements can be designed between JRC and ICES (on economics and effort data) and between JRC and GFCM (on MBS data). The scenario assumes that data on fish processing would be drawn from Eurostat's SBS and aquaculture data would be stored in a small separate DB at JRC.

Scenario 2 would offer a number of advantages. RDB structures already exist (RDB-Fishframe) and could be adapted to specific requirements of regions which do not yet work with it (MBS and 'Distant Waters'). Hosting by ICES would ensure the consistency of all RDBs (incl. the MBS[4]). The scenario assumes that the use of a common standard (RDB-FishFrame) would be promoted as much as possible, in order to achieve EU-wide consistency and reduce IT development costs. Initiatives already exist to extend RDB-Fishframe format to Mediterranean countries and to large

[3] *Incoherence may result from different estimation methods, definitions, reference periods, etc. In addition same criteria for data validation can be applied and standard quality reports facilitate users' appreciation of the data.*

[4] *The scenario assumes that ICES would provide a technical service in operating the MBS DB. This does not need to have implications on its regional scope of operation in North Atlantic.*

pelagics (LP) RDB. Developing a Mediterranean & Black Sea (MBS) RDB could provide the EU with an opportunity to take the lead in the MBS region by setting up a sound data system for the EU MS and inviting and supporting other non-EU MBS countries to join it. Such initiative would strongly support GFCM and meet all EU requirements at the same time.

There are also disadvantages to the regional database approach. MS bordering several marine regions, e.g. Spain and France, would have to submit data to several RDBs, under different data calls. Aggregated data at EU level would possibly be more difficult to handle especially if databases are located and separately managed in different places and by different hosts. If the MBS RDB would be operated by a different institution than the other RDBs, then common development will require additional coordination, effort and costs. The coordination burden would further increase if the DW RDB would be operated by yet another institution (e.g. IRD).

At present the economic data is compiled at EU level. The extent to which economic data could or should be separated into regions has to be explored. Sound statistical assessment at regional level may require increasing sample size for some MS (e.g. Denmark, Germany, Spain and France) in order to achieve sufficient regional coverage. Nevertheless, even at the current DCF level of disaggregation DCF economic data allows some regional analysis, as demonstrated by the Annual Economic Reports.

Scenario 3 "Network" (EMODnet model)

In this scenario, the primary biological, economic and fisheries data would be stored only in the national databases, and these databases would be linked through a web-based interface (central platform). The interface would offer users access to the national data aggregated at an agreed level and possibilities of data processing at supra-national level, e.g. comparing characteristics of species in different areas. The system could be operated from the present EMODnet, or in a similar manner by a different host. The main assumption of this scenario is that the MS would be able and willing to provide access to their national DBs and incorporate appropriate safeguards in their systems. The scenario assumes that data on fish processing would be drawn from Eurostat's SBS and aquaculture data would be stored in a small separate DB at JRC.

Scenario 3 offers several advantages. Data in the network, accessible through an Internet platform, would be continuously (and automatically) updated as it becomes available and MS update their national DBs. High flexibility of aggregations at supra-national level would become possible[5], because the user could use primary data to create data products/aggregates (but without seeing or accessing the detailed data itself). Various types of data (biological, economic, fisheries, aquaculture) could be provided from decentralized national DBs.

[5] *Other scenarios also stress the need of compiling detailed or primary data, precisely for the reasons of aggregation flexibility.*

The major disadvantage is that most developments (IT and content) would have to be coordinated among all MS as an important part of the implementation of this scenario would take place at national level. It is estimated that at least 70 DBs would have to be linked (a minimum of 3 DBs in 23 coastal MS). This applies to setting up the system as well as its maintenance and future development. In the other scenarios much more work can be done centrally, which would be more efficient. The access to confidential, secured environment where databases holding primary data should be hosted would certainly cause problems because of the required confidentiality guarantees imposed on various types of users. All MS would need to maintain quite well advanced IT systems to ensure confidentiality of data and to be able to contribute to the platform whatever their IT capabilities/resources are. If some MS would not comply with these technical requirements, then full EU sets of data would not be available. Currently MS fulfil their obligation for transmitting the requested data but sometimes the process is rather manual because their IT systems have not been sufficiently developed. This means that there is a risk that the data would possibly not be available in the expected format without major further IT development in those MS. In this scenario, judgement on quality would be left to the discretion of the responsible staff in the 28 MS. External institutions could carry out certain specific checks on the data quality and availability (as currently done by JRC and ICES), but the platform could not prevent using the data even if not considered reliable, as the national DBs are directly accessible through the platform. In particular, non-professional users may face difficulties in assessing the data quality. Other scenarios are based on a higher level of data centralisation and as such they allow development of common quality checks to guarantee a minimum data quality before it is disseminated.

Scenario 4, the **"Fisheries data hub"**

This scenario would combine three thematically specialized supra-regional DB (biological data, fleet economic data and fisheries data), containing primary data. The DBs are linked, so that fisheries data is stored only in one place and biological and economic data can be aggregated to same levels. Efficient exchange mechanisms between national DBs and the Hub DBs would be established along the lines of the FLUX system developed by DG MARE for control data. The Hub DBs can be placed on one server, but its physical location is not relevant. The three Hub DBs would be managed by three different institutions, based on their current responsibility and acquired expertise: biological DB by ICES (for all regions/MS), economic DB by JRC and fisheries DB by DG MARE. Users could access the data through a dissemination website, having different levels of access rights. The main assumption of this scenario is that the MS would be willing to share their primary data. The scenario assumes that data on fish processing would be drawn from Eurostat's SBS and aquaculture data would be stored in a small separate DB at JRC. This scenario is closest to the present situation so that its implementation period could be shortest period as it would require least organizational or technical adaptations at national and at EU level.

This scenario aims to combine the strenghs of scenarios 1-3 and avoid some of their weaknesses. The system would be integrated at EU level, to assure common approaches to definitions (e.g. using common criteria for measuring effort of fleets segments), to reduce IT development costs

and to avoid gradually divergent regional development. At the same time the system should allow for regional specificities (e.g. different spatial scale in the Altantic and MBS areas). The Hub would also allow linking biological and economic data at a feasible resolution level.This means that each data record could be defined in relation to the EU-wide as well as regional specific dimensions, e.g. EU vessel size of 12-18m and 18-24m, while regional size may be relevant for 12-15 and 15-24m.

From the perspective of data collection and data products there are obvious advantages to integrating biological and economic primary or detailed data in one data hub. The dialogue within RCGs has already demonstrated that higher efficiency (and consequently lower costs and better quality) can be achieved when data collection is planned at supra-national level. The 'Fisheries Data Hub' approach allows data producers to benefit in particular from thematic cooperation within the three specialized DBs. Similarly to scenarios 1 and 2, development and use of common tools (e.g. common quality checks) play also an important role. Increased efficiency can be achieved through cooperation in sampling in collection of biological and economic data. Data users can rely on increased quality, transparency and data consistency, in particular in comparison with scenarios 2 and 3. Creation of a consistent EU-wide system would allow common development in the future, while starting from already established economic and biological DBs. Consequently, this scenario would be probably cheapest to establish in the short term. Long term operational costs are probably comparable to scenarios 1 and 2.

The main disadvantage of this scenario regards the political sensitivity of sharing primary economic and fisheries data. This contrasts with the other three scenarios that rely on compilation of aggregated economic and transversal data according to common definitions, as practiced at present. ICES would be providing a technical service of DB hosting and development to countries which are not ICES members. The scenario assumes that this would be possible, as this would be also most efficient, but in practice this could prove an institutionally and politically sensitive issue.

The four scenarios are graphically presented in figure 2. The scenarios distinguish CFP and non-CFP users, as the first group is assumed to have access to more detailed data, while the second group has access only to higher aggregations. The scenarios deal with data on marine fisheries. Data on aquaculture (marine and inland) and fish processing are not part of this data flow. Therefore the land-locked EU MS are not mentioned.

Figure 2. Overview of the 4 scenarios.

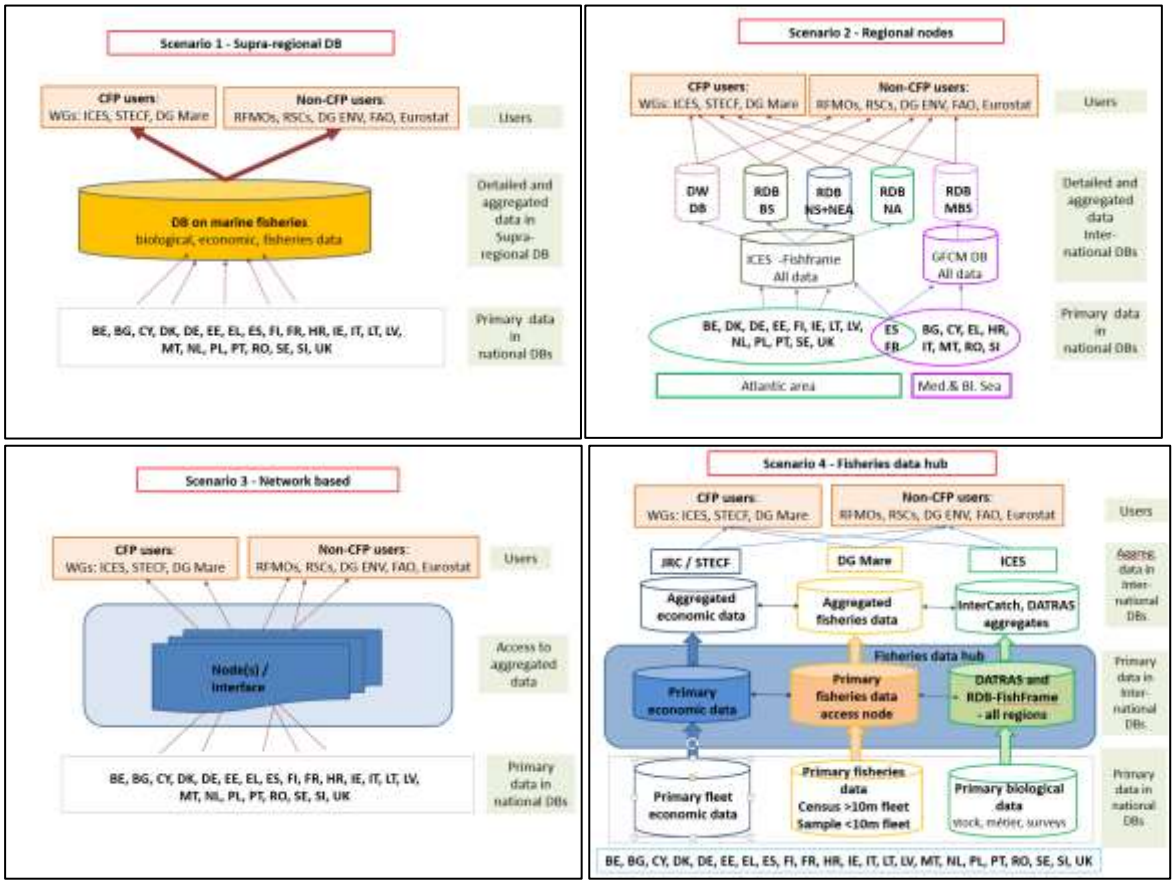


Table 2 presents a summary comparison of the 4 scenarios in terms of their coverage of functional, thematic and regional issues, organization and access.

Table 2. Comparison of the scenarios

	Geographic	Thematic	Functional	Organi-zation	Access
Sc. 1. Supra regional	All regions	All themes	All functions	One	One point
Sc. 2 Regional	By region	All themes in every RDB	By region	Several	By RDB
Sc. 3 Network	MS	National DB by theme	Most at MS level	MS	Node
Sc. 4 Hub	All regions	DBs by theme	By theme	One	One point

Table 3 presents an assessment of the four scenarios against policy objectives and constraints. It must be noted that this assessment is not unambiguous. It depends on the precise perspective taken and the implicit assumptions made. This can be illustrated by several examples:

- What will be the attitude of the MS in relation to provision of primary data? If it is positive, then none of the scenarios will face problems, if not all will face problems, depending on the assumption made regarding the requirement of that data. Scenario 3 is probably least sensitive to this issue, because the primary data would remain at national level and only agreed aggregations would be allowed. Provision of detailed (anonymised) data would be less controversial than full primary data. In the end, this is a political decision.
- Which geographic level should receive priority: EU or regional? If the policy aim is to have EU-wide consistency (and EU-wide data sets), then EU-wide standards should be complied with, so that the system does not end up with a number of different and possibly incompatible regional DBs. However, if the policy priority is to put the regional perspective to the fore, then it could be justified to give precedence to regional specificities at the expense of EU-wide comprehensiveness.
- Costs can barely be specified, because they depend on technical and institutional solutions (e.g. Eurostat vs. ICES, open or licensed software, policy choice of outsourcing or not, direct availability of staff or required new employees and training, etc.).

Because of the 'diffuse borders' between scenarios, it seems relevant to formulate a number of common **principles** (i.e. valid for any scenario) which should be pursued when specifying the future solution for DCF DB set-up:

- Primary or at least detailed **data** should be the basic building block, rather than the present aggregations of economic and transversal data to fleet segments. ICES DBs contain primary biological. This would facilitate linking the three groups of data at a common level of aggregation.
- The system should use common internationally accepted **nomenclatures** (species, gears, areas). This is already the case in most areas, but e.g. not for gears.
- Data and DBs are organized along **thematic modules** (biology, economy, fisheries) and links between the modules would be created. This presents a logical structure for at least two reasons. First, the data has different dimensions (i.e. it presents different types of measurements, e.g. length of fish and revenues of vessels). Second, the data would be managed by different groups of experts.
- **Scientists** who are closely involved in policy related research (ICES, STECF, GFCM) should be also closely involved in development of the system so that it responds to their needs (i.e. to their ability to answer policy questions).
- **Data quality** procedures should encompass the whole process from preparation of the sampling till dissemination. A common IT system can support these procedures, but not replace them. What matters most is the **transparency** of the entire data production process (which is also the basis for ISO certification).
- The importance of data **confidentiality** can hardly be understated. Technical solutions have to be found to meet all legal obligations on one hand and allow full use of the collected data. The technical expertise required to meet these legal obligations will have to be assured in any future scenario that may be chosen. Factors will have to be addressed are: a clear legal basis for handling confidential data, the statutes, reputation and legacy of trust in the organisation(s) hosting the future platform/DBs, mature processes and procedures for storing and handling these data. Distinction will have to be made between personal and commercially sensitive data.
- Proper **documentation** of all processes (including the dissemination of metadata) is the basis for transparency, transfer of experience and further development.
- Clear distinction must be made between **production and dissemination DBs**. This is common practice in all statistical institutes. It allows for processes like correction of errors or estimation of missing data so that a consistent data set is presented to users.
- Data on aquaculture and fish processing should be treated separately because it is not directly related to the biological, economic and transversal data on marine fisheries.

The above 'cross-cutting principles' would address at least some, although not all, **policy objectives and constraints**:

- **Informing CFP / IMP / MSFD** – possibilities would improve from the simple fact of having all the collected data (i.e. primary or detailed) accessible in one system.
- **Regionalization** - regionally flagged data within one system allow regional assessments. Setting-up regional DBs is not essential. Involvement of scientists, who also bring in the regional perspective, ensures that specific regional issues can be addressed.
- The call for **simplification** aims primarily to address the workload resulting from changing needs of the scientific advice (new policy questions) and consequently changing specifications of the data to be provided by the MS. Provision by MS of primary or detailed data, based on common nomenclatures, to the international DBs would resolve this problem, as it would then be possible to derive the needed information from the primary/detailed data.
- **Data upload** – transmission of data should be based on a common standardized format, e.g. FLUX. Uploading primary or detailed data is the only way to avoid repetitive uploading of similar data, based on same primary data.
- **Data storage** – Any future system should be based on 3 modules: biological, economic and fisheries data. The data could be linked by using the common nomenclatures.
- **Data access** – Data confidentiality has to be assured at all times. Data access should be defined by type of users (policy advice, public).
- **Bio-economic integration** – This is primarily an issue of conceptual development of linking biological and economic units (stocks and segments). Availability of primary or detailed data would allow empirical testing of these concepts.
- **Coherence with Integrated Fisheries Data Management Programme (IFDMP)** – can be assured, as long as all involved institutions are willing to cooperate in development and implementation of this common platform.
- **Legal requirements** – Institutions running the system must guarantee data confidentiality. ICES, JRC and Eurostat all have the required legal provisions.
- **Administrative requirements** – High requirements have to be met in any scenario. ICES, JRC or Eurostat have demonstrated that they can meet these requirements.
- **Financial constraints** – Different approaches will have different financial consequences for the MS and the central DBs. Streamlining the data flows is likely to lead to savings at MS level. The financial consequences of the scenarios are elaborated in section 3.6.

Comparing the scenarios

Setting up one EU-wide database (Scenario 1), containing primary or detailed data, which is regionally flagged, would deliver on the needs of ensuring both EU consistency as well as enabling regional specificities to be taken into account. In this scenario one institution would have to be appointed and bear responsibility for the whole system. In Scenario 4, ICES and JRC would continue in their present role, so that from the perspective of institutions and expertise such

approach seems the easiest to implement as it is closest to the current situation. From the perspective of the development needed, scenario 4 allows reusing existing databases while scenario 1 would require developing a new system and hence it may impy higher investment costs.

A major disadvantage of scenario 2 (Regional nodes), as it is envisaged here (with three separate DBs for different regions), is that it leads to unnecessary fragmentation and probably additional IT development costs to establish and further develop three parallel systems. Integration of all regions into a single DB (in which data can be regionally flagged) would address this disadvantage but this is effectively then moving to the set up foreseen under scenarios 1 and 4.

Scenario 3 requires a large number of DBs to be connected to a central node. In most MS, biological, economic and fisheries data is stored in different DBs (even in separate DBs for métier, stock and survey data). Setting up on-line links to some of these DBs would require that file (often in excel) currently stored on standalone PC are at least stored on a server accessible through the internet in a format adapted to the requirement of the network scenario. Finally, Scenario 3 seems to offer the least possibilities for common approaches to the monitoring of data quality. Problems faced by any MS may affect the whole system (the network is as weak as its weakest node).

Decentralised databases may cause various problems of reliability and performance, not only due to complex technical infrastructures, but also because the reliability and performance of the entire system may rely on the weakest element. For these DBs a strict coordination mechanism needs to be put in place and maintained in the long term.

The following table summarises, with a qualitative rating, the capacity of each scenario to address policy objectives and constraints.

Table 3. Evaluation of the scenarios against policy objectives and constraints

(rating: 1 = lowest, 4 = highest rating)

Policy objective / constraint	Sc. 1. One DB	Sc. 2. Regional	Sc. 3. Network	Sc. 4. Hub	Justification of highest and lowest rating
Informing CFP/ IMP / MSFD	3-4	2	1	3-4	Sc. 1+4 Integrate data in a consistent EU-wide system. Sc. 3 Depends on contents of many national DBs.
Regionalization	2	4	1	3	Sc. 2 reflects closest the regionalization objective. Sc. 3 requires intensive coordination in the 5 regions.
Simplification	3-4	2	1	3-4	Sc. 3+4 Mirror national DBs if containing primary or detailed data. Sc. 1 New IT developments or policy questions may required adaptations in 28 MS.
Data upload	2	1	4	3	Sc. 3 Does not require any data upload. Sc. 2 MS upload data to various DBs.
Data storage	3-4	2	1	3-4	Sc. 1+4 Guarantee highest level of consistency. Sc. 3 Large number of national DBs, some not easily accessible on-line.
Data access / dissemination	3	1	2	4	Sc. 4 has strong system embedded access data policy. and can produce various dissemination products.
One-off costs (nivestment) (hard/software)	1	2	3-4	3-4	Sc. 3+4 Reuse existing infrastructures. Sc. 1 Requires development of a centralized DB.
Running costs (operation	3-4	2	1	3-4	Sc. 1+4 Operation at central level. Sc. 3 Requires MS adapt and run the national DBs in consistency with the network.
Quality control	3-4	2	1	3-4	Sc. 1+4 Common procedures implemented centrally. Sc. 3 Requires implementation in each MS system validation rules
Bio-economic integration	4	2	1	3	Sc. 1 Common nomenclature in one DB. Sc. 3 Many DBs have to be linked.
Coherence with IFDMP	4	1	2	3	Sc. 1 One focal point to meet IFDMP requirements. Sc. 2 Relies on regional organizations which will need to justify to their MS legal provisions for coherence.
Legal constraints	3	2	4	1	Sc. 4 Requires countries to provide access to primary data and appropriate legal arrangements. Sc. 3 Relies on MS confidentiality rules already in place.
Administrative constraints	3	2	1	4	Sc. 4 Adminstrations of ICES and JRC already in place. Sc. 3 Continuous cooperation of many administrations is required.

From the study and comparison of scenarios, the following main recommendations can be made:

1. Biological, economic and fisheries data from all MS should be integrated at primary or detailed level. This will reduce costs of data collection, as already demonstrated by regional coordination of the collection of biological data. Analytical value will be increased because greater flexibility to carry out tailer-made aggregations for specific questions.
2. The use of common standards for data definition (Master Data Register) and data exchange (FLUX) will reduce the burden of data transmission.
3. The creation of an EU-wide DB-system (Scenario 1 or 4) requires additional one-off (investment) costs in the short and medium run, but will lead to overall lower running costs in the long run, because economies of scale can be expected (e.g. common IT development).
4. Coordination of definitions and data flows under fisheries regulations (the DCF, the Control regulation and Statistical regulations) should be streamlined, in particular data on fish processing and aquaculture, but also on landings and capacity. This will further reduce data processing and transmission effort.
5. A common set of relevant dimensions (fishery/métier, time and space) should be defined in order to achieve integration of biological, economic and fisheries data.
6. Automated procedures for data upload should be applied in the MS. Such procedures will reduce both errors and costs.
7. Protection of personal data and confidential commercial data must meet all legal requirements.

1 INTRODUCTION AND METHODOLOGY

1.1 Purpose and structure of the document

This draft final report presents the results of the tasks implementated unter the Project: Scientific data storage and transmission under the future Data Collection Framework – Feasibility Study (hereinafter the Project).

This document represents one of the deliverables agreed in the contract between the Client and the Consultant for the project management activity.

The document is structured as follows:

PART 1:

The introductory section and the status of the implementation plan - **Section 1** – presents the background of the project and a brief description of the methodology used for the information analysis and also for the identification of the key elements that a future scenario should have.

Section 2 presents a concise summary of the findings of the study regarding the present situation.

Section 3 presents a description of several possible scenarios for the future for the data storage and transmission set-up as well as the assessment of the effectiveness and feasibility of these possible scenarios.

Annexes Part 1:

Annex 1.1. Comparison of legal requirements under the DCF and Statistical regulations concerning submission of data on aquaculture and fish processing.

Annex 1.2. Information on relevant fisheries related databases.

PART 2:

Section 1 presents the technical details of the current situation (Interim Report).

Annexes Part 2:

Annex 2.1. DCF and Control Regulation Institutions.

Annex 2.2. List of DATRAS data products.

Annex 2.3. Aggregation levels ICES databases.

Annex 2.4. GFCM tables and fields description for data submission.

Annex 2.5. Regional Fisheries Management Organizations – data, data bases and institutional position.

Annex 2.6. Integration of biologic and economic data.

Annex 2.7. Charts in Unified Modelling Lantuage (UMLs).

Annex 2.8. Comparison of Data Calls

Annex 2.9. FLUX presentation.

Annex 2.10.. DiGIR. Distributed Generic Information Retrieval.

1.2 *Background information*

PROJECT OBJECTIVES AND EXPECTED RESULTS

The **overall objective** of the Project is to describe the current situation and elaborate and discuss a number of scenarios for the possibilities of data collection, processing and dissemination that could be effectively put in place to fulfil the requirements of generating policy-related knowledge for the CFP.

Within the global objective, the **specific objectives (SO)** can be formulated into the following actions that have to be carried out:

- **SO1** – Description of the current data storage and transmission set-up (baseline scenario).
- **SO2 –** Development of several (i.e. three to five) possible scenarios for the future for the data storage and transmission set-up which allow achieving a number of policy objectives.
- **SO3** – Assessment of the effectiveness and feasibility of these possible scenarios. The effectiveness will be assessed against the achievement of the policy objectives, and the feasibility will be assessed against legal, administrative and financial constraints to implementation.

As per the Terms of Reference (ToR) the following **deliverables** will be provided:

- Description of the baseline scenario;
- A set of 3 scenarios with schematic description of set-up, processes and infrastructure;
- Assessment of the strengths, weakness, opportunities and threats of the scenarios and a description of the key elements that a future scenario should have.

1.3 Project Phases

The project was carried out in five phases, which are described as follows:

PHASES	DESCRIPTION	SITUATION
INCEPTION PHASE	This phase included the preparatory work (desk research) by the Consultant to prepare a proposal for the structure and methodology for developing the project. It includes identifying the main DCF data, CR data and other fisheries related data providers and end-users at national and supra-national level. The main output of this phase was the Inception Report, considered as the main working instrument and which will be referred to as during the overall implementation of the Contract.	DONE
INFORMATION COLLECTION AND ANALYSIS ON THE CURRENT SITUATION[6]	This phase included the collection, organization and brief analysis of all the relevant information provided by different data providers in order to present a good description of the present situation (baseline scenario). The following outputs were delivered: - Description of the baseline scenario; - Proposals for improvement.	DONE
ELABORATION OF ALTERNATIVE SCENARIOS	This phase consists in proposing several possible scenarios for the future data storage and transmission. Moreover the possible scenarios will be accompanied by a cost evaluation of each one of them. The following outputs were delivered: - A set of three scenarios with schematic description of set-up, processes and infrastructure; - Interim Report presenting the results of the data collection and analysis phase and the preliminary results of this phase.	DONE
EVALUATION OF SCENARIOS	This phase implies the assessment of the effectiveness and feasibility of the possible scenarios developed in the previous phase, considering also their costs. During this phase, the Consultant organized a Focus Groups with people from different institutions dealing with DCF (e.g. DG MARE, JRC, EMODnet, ICES, Eurostat and representatives from national organisations) in order to share conclusions and receive feedback regarding the evaluation of the scenarios proposed. The main output of this phase is an Assessment of the strengths, weakness, opportunities and threats of the scenarios and a description of the key elements that a future scenario should have.	DONE
REPORTING	This phase consists of drafting of the Final Report and the auxiliary material (if needed). The draft version of the report is refined accordingly to the recommendations and comments provided by the Client. A final revision is made before the editing of the final version of the report. The main output of this phase is the Final Report presenting the results obtained in all the phases of this project.	COMPLETED

[6] *Methodology for the information collection is described in section 1.4.*

1.4 Methodology for the information analysis

In order to present a good description of the current situation the Team of Experts prepared a set of questionnaires, based on the model presented in the Inception Report, sent to National Correspondents (NC) and to data intermediaries.

The following methodology has been used for the information collection and analysis of the current situation:

1. ***Questionnaire sent to National Correspondents***

The Team of Experts prepared a questionnaire to National Correspondents which was sent on 23rd September and should be answered by the 7th October 2013.

Although the questionnaires were sent to NC of the 28 MS, in some cases they distribute them to other national Institutions dealing with DCF thus delaying the delivery of the questionnaires.

2. ***Questionnaires sent to data intermediaries***

Adapted questionnaires were sent to the following data intermediaries: GFCM, JRC, ICES, other RFMOs, RSCs, FAO and Eurostat on 23rd September and most of them were delivered on the deadline established (7th October 2013).

3. ***Centralised information.***

Once the Team of Experts received most of questionnaires the information provided was processed and analysed in order to obtain a good overview of the baseline scenario.

4. ***Request for complementary information.***

In some cases a request for complementary information was done.

5. ***Acknowledgements.***

The team of experts acknowledged the fruitful collaboration and effort of the institutions dealing with DCF and Control data for their personal contribution to answering the questionnaire.

1.5 *Focus group - identification of key elements for a future scenario*

The **objective** of the focus group was to identify the key elements that a future scenario should have, by focusing on advantages and/or positive features of the three proposed scenarios and considering how these could be compined.

1. ***Selection of participants of the focus group***

The participants of the focus group were proposed by the experts' team and agreed together with DG MARE. The main characterist of the focus group composition is that all the invited experts come from different institutions dealing with DCF (see table below).

Table 4. Participants of the focus group held on 2nd December 2013.

Institution	Name
DevStat Consortium (Project Team)	**Pavel SALZ (team leader)** **José CERVERA** **Christine ALBERTI-SCHMITT** **Suitbert SCHMÜDDERICH**
DG MARE	**Amelie KNAPP** **Francky CALLEWAERT** **Bas DUKKER** **Iain SHEPHERD**
JRC	**Fabrizio NATALE**
Eurostat	**Vincent TRONET**
ICES	**Henrik KJEMS-NIELSEN**
RDB Steering Committee	**Katja RINDAHL**
National Organisations	**Jørgen DALSKOV (DTU-Aqua)** **Joël VIGNEAU (IFREMER)**

2. ***Relevant documentation:*** In order to prepare the focus group, the following documents were distributed among the participants:

- Agenda for the focus group.
- Preparation focus group document: introduction of the purpose of the focus group and some feedback requested. The feedback requested was received five days before the focus group, so the team of experts compiled and analized them and subsequently forwarded to the participants.
- Summary Interim Report.
- Outline of scenarios: dimensions and description of the three proposed scenarios.

3. ***Focus Group implementation:*** Finally the focus group was carried out on 2nd December 2013 in Brussels from 10.00 to 17.00 in DG MARE.

2. CURRENT SITUATION

2.1. *Overview of the current situation*

2.1.1. Introduction

This section presents a concise summary of the findings of the study regarding the present situation. Section 2 presents the main conclusions. The following sections discuss in more detail the various topics – access to control data at national level, DB structure, data upload, quality controls, dissemination, interactions with other fisheries-related data and institutional considerations. The technical details are presented in the Part 2 of the report.

The following sections summarize the current situation from three perspectives:

- Thematic consideration, dealing with types of data (transversal, economic, biological);
- Functional considerations discuss the present situation by type of activies or functions (transmission, storage, quality control, dissemination);
- Geographic considerations reflect the geographic scope of the various data bases and processes within the DB-system.

2.1.2. Thematic considerations

Transversal data

The three transversal (or fisheries) variables are capacity, effort and landings.

Transversal data is compiled under various data calls at different levels of aggregation, to allow linking to other data:

- ICES/RDB-Fishframe – Biological métier and stock related data.
- JRC - Economic data on fleets.
- JRC - Effort regimes data.
- JRC – Mediterranean and Black Sea data.
- Eurostat – Fleet structure and landings data.

The most important overlaps occur between DBs and data flows in this area. The data provided by the MS originates in most cases from the data collected under the Control Regulation. The differences in aggregation levels are presented in detail in Annex 2.6. of the report. Annex 2.6. shows that for the transversal data it is possible to define one common most detailed aggregation level, from which other (higher) aggregations can be derived.

It is noted that at the moment there not one single DB which would contain the transversal data at this lowest aggregation level, e.g. while ICES has detailed data by métier, but distinguishes

small size vessels with 10 and 12m threshold, JRC has also 6m threshold for the MBS and 8 and 15m for the effort call. This implies that the MS must submit data under different data calls. It would be conceivable that the MS provide catch and effort data only once with length stratification 0-6, 6-8, 8-10, 10-12, 12-15, 15-18, 28-24,24-40 and 40+m. Any required stratification could that be obtained from this single one submission. This would be the 2nd best option, after sharing data at primary or detailed level. In order to reduce the data submission burden to the MS, it is recommended to set up a DB with transversal data at lowest possible level of aggregation. This level will have to be defined by the main users – ICES, JRC and STECF.

Biological data

Biological data is collected in three ICES databases for the Atlantic areas (RDB-Fishframe, InterCatch and Datras). MBS data is compiled by JRC. Biological data on fisheries on large pelagics in non-EU waters is at present compiled ad hoc, but will be centralized in a 'Large pelagics RDB', which is being set-up by IRD (L'Institut de recherche pour le développement, France). At present there is no arrangement for compilation of data on other fisheries in non-EU waters (e.g. small pelagics and trawling off West Africa and South-West Atlantic)[7].

At present there is particularly an overlap between RDB-Fishframe and InterCatch. These two DB contain largely identical information. ICES is preparing an automated routine to upload data from RDB-Fishframe to InterCatch. A manual transfer is already possible, but is not applied, because of existing procedures.

Economic data

Fleet economic data (earnings and costs) is compiled under the JRC data call for all relevant EU MS.

Economic data on performance of aquaculture is collected by JRC, also on EU-wide basis. Eurostat collects also data on aquaculture production (value and volume), by type of production. Eurostat does not collect costs and earnings data. At the level of production the two systems overlap. There is double transmission from MS to JRC and Eurostat, as no data exchange occurs between these two institutions.

Costs and earnings data on fish processing is collected by JRC under DCF and by Eurostat (under the Structural Business Survey - SBS). Some elements of this data collection overlap, while some are unique to each system.

[7] *Oceanic Developpment, Study on the european external fleet, Study for DG Mare Contract FISH/2006/02, Final Report, January 2008, p.3*

2.1.3. Functional consideration

Data storage and access

In relation to functional considerations, distinction is made between production and dissemination DBs. Production database means a database in which data (usually quite disaggregated) is loaded, processed and validated. The dissemination database aims to publish/make available data processed from the production database and usually contains more aggregated data than the production database and non-confidential data (unless restricted access procedures are in place for some users).

Production/dissemination means the same database is used for the production and dissemination. The dissemination can be done to a restricted public or to a wide public through a specific interface limiting the access to the production data.

The main institutions have developped good applications for uploading, checking and disseminating the fisheries data. In general it can be concluded that the system works well, although further improvements are evidently conceivable.

From the available **biological** database system (at ICES and JRC), the RDB-Fishframe is the most advanced in terms of functionalities, level of detailed information storage, implementation of models for the raising of the data, reusability and genericity. Storing biological data in a single environment will make it easier to compare individual data from different sea basins, but this would certainly require having a more "open" access to the data than what is possible now in RDB-Fishframe and InterCatch.

Economic data on fleet, fish processing and aquaculture are stored at JRC in relatively simple databases.

Transversal data on value and volume catches and landings, fishing effort and fleet capacity are stored in different databases of all major institutions involved in fisheries – ICES, JRC, DG MARE, Eurostat and RFMOs. In most cases, the origin of this data is the information compiled by the MS under the control regulation.

Transmission

Data call and data upload for the main actors involved in DCF follow similar upload processes. Up-to-date Excel files describing formats and nomenclatures for data reporting are provided by MS institutions to the responsible supra-national institutions. The major difference is in the level of automation of processes. It goes from a process of uploading text files manually in the database (e.g. to GFCM) to a semi automated process requiring some internal manipulation (e.g. to Eurostat) or to direct upload in the database by the provider.

In general, the technical level of data upload management is high, processes and workflows are well identified and managed. Formats are easily accessible; upload is easily done through secured interfaces or by skilled data managers. The formats of data upload range from text files (TXT, CSV) to XML files, the latter providing better control and checks on the validation of format and metadata.

The IFDMP DG MARE initiative aims to provide a common tool to facilitate data exchange (FLUX transportation layer) and methodologies to standard data format (FLUX business layers). This initiative goes in the same direction of harmonization and standard of data calls to optimize the process and reduce costs for fisheries data exchange.

Quality control

Most important part of quality control takes place at national level in relation to the primary data. All analysed up-loading procedures contain various quality control checks, which are facilitated with specific software tools. Further quality control takes place during (the preparation of) the WGs of ICES or STECF. Eurostat also runs validation procedures for the data it collects, sending them back to or contacting MS for clarifications/corrections.

Dissemination

All involved institutions have developed dissemination DBs, which offer aggregated data to interested users, including general public. An exception to this rule of public access is the RDB-Fishframe. The biological measurements in this DB are not publicly accessible.

Researchers working within ICES working groups have access to the primary biological métier and stock related data.

Transversal and economic data is accessible to researchers as well as to the public mostly at the same level of aggregation.

2.1.4. Geographic considerations

National

All data is collected within the national DCF programmes. All MS have developed DBs of various complexities to store and process the data.

Regional

The biological data is raised to higher geographic level in the ICES DBs, according to the definitions of the stocks in the Atlantic. Aggregation of national data in the MBS is less common, because management takes place within the Geographical Sub-Areas (GSAs), most of which fall under the national jurisdictions. National data on fisheries in non-EU waters is aggregated according to the needs of the RFMOs on an ad hoc basis.

All economic (fleet, aquaculture and fish processing) data is kept in principle at the level of MS. Regional aggregations take place only in specific projects, but not in the existing DBs.

Supra-regional

Aggregation to supra-regional level (e.g. EU total) is mostly relevant for data on volume and value of landings, which is relatively more of interest to general public. Supra-regional aggregations are rarely required for professional scientific use.

2.2. *Access to Control Regulation (CR) data*

2.2.1. Access and aggregation level[8]

In most MS institutes responsible for DCF have real time on-line access to primary logbook data and mainly use Excel/Access or txt/csv formats.

The institutions which do not have on-line access receive data generally frequently (monthly or quarterly) by e-mail. Almost all institutions responsible for DCF have access to primary CR data.

2.2.2. Confidentiality and contacts with data providers

All DCF-related national institutes guarantee confidentiality of the CR data, although the slight majority of them do not have a formal agreement with data providers. In almost all MS contacts between institutions involved in CR data collection and storage are intensive, especially concerning solving of problems and feedback on data and operations.

2.2.3. Storage of CR data

Only five MS store CR data in DCF specific DBs, while most others keep the data in a separate one. In some cases, institutes do not have a CR-related database and/or are not able to link logbooks and sales notes data.

2.3. *Data storage*

2.3.1. Databases and their accessibility

ICES has two own DBs, DATRAS (Trawl surveys data) and InterCatch and hosts RDB-FishFrame.

ICES develops the tools needed from processing of the individual biological measurements to the stock assessment data for the three regions: the Baltic, the North Sea & Eastern Arctic and the North Atlantic.

[8] *The text is based on a survey of the national correspondents by the project. It is noted that a similar overview carried out by ICES concludes that access to CR data is much more limited. (ICES, Report of the Planning Group on Commercial Catches, Discards and Biological Sampling, PGCCDBS 2012, section 7.2).*

The two ICES databases and the RDB-Fish Frame are at the same time production and dissemination databases. They are fully accessible through the intranet but also through the internet with the following restrictions.

- DATRAS database is fully accessible to the public, but data transmitter in the country has specific access right for uploading their data.
- InterCatch and RDB-FishFrame are only accessible on the internet to the special user groups which can upload the data and see the content of the database related to their group.

JRC deals with data on fleet economics, aquaculture, fish processing industry, effort regimes, biological data from the Mediterranean and Black Sea. It provides data to the STECF WGs.

At JRC, there is a production database per data call which is only accessible through the internal network. External users may have access to the production database through a login for uploading the data into the system.The SAIKU dissemination database is freely accessible to the public but contains only data validated by the STECF (not all production data).

DG MARE receives from the MS aggregated information required on the basis of the CR. The set of data received is exhaustive as based on census of all information registered. DG MARE does not maintain any DCF databases but maintains two separate DBs in relation to the CR, on landings and effort. DG MARE does not disseminate any data to the public. DG MARE maintains also the EU Fleet Register. Anonymized data of the register are publicly accessible.

At Eurostat, there is one production database for catches, landings, fleet and aquaculture data (Multi Dimensional Table - MDT). SR use different definitions for catches and landings. Catches are defined as landings by vessels of one MS and registered in live weight, incl. recreational fisheries. Landings regard fish discharged in the ports of one MS, independently of the flag of the vessel. Landings are measured in product weight[9]. This distinction does not exist in DCF and CR. Consequently, there are differences in landings data beween SR on the one hand and DCF and CR on the other hand.

The fish processing industry data are compiled by Eurostat under the SBS Regulation (EBB/Eurocube DB).

[9] *Statistical population for landing as decribed in Eurostat meta data: Under the terms of Council Regulation no 2104/93, the reporting country is required to include data for all products landed by Community and EFTA fishing vessels in ports of that country. Under the provisions of the Regulation the reporting country is not required to report landings by its vessels in ports other than the national ports.*
The data are required to include products discharged within the territory of the reporting country and covered by document T2M referred to in Council Regulation (EEC) no 137/79. Also included are products transshipped to vessels of third countries from Community and EFTA fishing vessels and other components of the Community and EFTA fishing fleet which are discharged within the territory of that Member State.

At Eurostat, the production and dissemination databases are also working independently. The Eurostat production databases are accessible only through the internal network and are updated by internal staff only. The upload of the file transmitted by eDAMIS is made in a semi-automated way by internal staff. eDAMIS is electronic data Data files Administration and Management Information System used by Eurostat which offers standard solutions for collecting data files in the ESS. The dissemination database (Eurobase) is freely accessible.

GFCM maintains databases on fishing vessels over 15m (vessel records database), on economic, transversal and biological data (Task1) as well as on aquaculture (SIPAM). The databases are updated internally by the GFCM data manager with data received by email. The data is partly accessible from the GFCM web site. Helsinki Commission (HELCOM) hosts 10 general marine/maritime databases in Excel, Access and a centralized visualization end point, HELCOM Map and Data Service supported by an MS SQL Server derived from all other databases. These databases contain environmental data but no DCF data.

The Convention for the Protection of the Marine Environment of the North East Atlantic (OSPAR) holds also environmental database but does not require any DCF data.

NAFO receives catches, capacity and effort from area 21 (Northwest Atlantic) directly from Eurostat and processes them in its STATLANT DB. The data are managed in an internal production database. Aggregated data is publicly accessible. It can be downloaded in text format or extracted using the extraction query on NAFO web site.

Biological and transversal data provided to other RFMOs originates from DCF and CR and is compiled according to their data requests. The RFMOs store the data in a variety of DBs. Some data is accessible through queries, while other is published in various reports.

The main actors involved in the data collections and processing of primary or detailed DCF data are JRC and ICES and some RFMOs. Other institutions (Eurostat, GFCM) use DCF related data at higher aggregation level.

2.3.2. DB structure

Trends: Database is developped using up to date SGBD (Postgresql at JRC, Oracle at Eurostat, MSSQL at ICES, GFCM, NAFO, IATTC, SEAFO and SPRFMO) and programming langages. Only JRC opted for a free SGBD solution.

Regarding interoperability: it is a top priority for ICES to connect automatically the RDB-FishFrame and InterCatch, to facilitate raising of biological data to stock level. The ICES-DATRAS datawarehouse is connected to EMODnet Biology and to Geonetwork (a catalogue application to manage spatially referenced resources). At Eurostat, production databases are indirectly connected to dissemination database as functions allow exporting files in a Eurobase format

ready to be imported in Eurobase. For the other institutions, there is no interoperability to be mentioned.

Referential used: it is important to use common nomenclatures and definitions, which should be held in common DBs, to ensure homogeneity. Reference nomenclatures (FAO species coding , ISO country nomenclature) are most of the time defined directly in the database (Eurostat, JRC, GFCM) or managed in dedicated modules (ICES vocabulary: common data coding system for DATRAS and InterCatch). The nomenclatures need then to be manually updated. Rarely, web service connected to the source of the referential are implemented like in DATRAS, where the WoRMS species catalogue from marinespecies.org is used.

It is to be noted that the same nomenclatures are not always used for the gear, fleet segmentation (often similar but not identical in terms of detail level or aggregation). As an exemple, GFCM is using its own classifications for gear type and fleet segmentation (vessel size). Mapping can be made but it would be better to use an international coding, as presented in the EC Master Data Register10, possibly adapted for specific DCF requirements (e.g. mesh sizes).

Maintenance and development: The main institutions have developped good applications for uploading, checking and disseminating the fisheries data and they work well. Some tools developped are generic and can be reused or shared (e.g. MDT, EBB for the validation of data at Eurostat, R developments at JRC, RDB-Fishframe, and COST).

2.3.3. Aggregation level

Biological data can be found at ICES and JRC:

- Under the MBS data call, JRC collects survey data following the MEDITS guidelines as well as aggregated data on métier catch, landing and discards.
- The DBs at ICES (RDB-FishFrame and DATRAS) contain primary data.

Economic data on fleet is only compiled by the JRC at the level of aggregation required by DCF. GFCM stores also yearly information on Mediterranean and Black Sea fleets at fleet segment level in Task1 database (Tsk1_FleetSegments_Economic, see Part 2 of the report, Annex 2.4.) but only a few variables are collected compared to DCF (employment, crew share, variable costs) RDB-FishFrame contains information on value of landings.

Transversal data: the details of the aggregation levels of the transversal data in the various DBs are presented in Annex 2.6.

[10] *http://ec.europa.eu/fisheries/cfp/control/codes/index_en.htm*

The following points can be highlighted in relation to transversal data:

- JRC economic data call is consistent with DCF requirements;
- JRC call on effort regime includes a vessel length of 8m for the Baltic and 15m for the Atlantic areas for the sake of continuation of the exisiting time series (in particular in relation to the Western Waters Regime). These lengths is not foreseen under DCF, which uses thresholds of 10m, 12m and 24m (see Annex 2.6 for detailed comparison);
- JRC MBS data call is consistent with DCF requirements;
- Annual value and weight by species is also compiled by GFCM;
- Information in ICES InterCatch and RDB-Fishframe is more detailed than DCF requirements.
- Eurostat gives information on annual catches in live weight by, FAO main fishing areas. This is similar to information collected by JRC but at a more aggregated level. Value and product weight of landings are available by MS, presentation form (fresh, frozen, etc.) and destination (human consumption, industrial use, etc.). Presentation and destination are not available in DCF data calls. Effort data is collected by month in the FAO area 21 only.

Aquaculture data can be found at JRC, Eurostat and at GFCM:

- JRC aquaculture data call is fully compliant with DCF requirements.
- Eurostat collects information on aquaculture detailled by production method. The production method definitions used by SRs differ from the DCF. SRs require data on fresh water aquaculture, which is not compulsory under DCF. (See Annex 2.6). Eurostat also collects specific information on production of eggs for human consumption and structure of the aquaculture sector (every 3 year). Eurostat collects only value and volume and information on juveniles but not on costs and earnings.
- GFCM stores information on aquaculture production by species. The SIPAM database distinguishes a number of types of production and products. However, despite repeated requests, it was not possible to obtain a full nomenclature presently in use[11].

Fish processing industry can be found at JRC and at Eurostat:

- JRC fish processing data call is fully compliant with DCF requirements.
- Eurostat collects very detailed economic information on fish processing within the SBS survey, including most indicators required under DCF. The following DCF variables are not included in the SBS survey: subsidies, other income, imputed value of unpaid labour, raw material costs, other operational costs, financial costs - net, extraordinary costs, net,

[11] *The GFCM/35/2011/6 on "REPORTING OF AQUACULTURE DATA AND INFORMATION" specifies: Type of culture (cages, ponds, raceways, hatchery, etc.); and - Type of product (ongrowing, eggs, fingerlings, etc.). However, it is not clear what the "etc." exactly means.*

capital value – total value of assets, debt, depreciation of capital. (for detailed comparison see table 2, Annex 2.8).

2.3.4. Data processing and estimations

The data processing is split in three main phases:

Data upload (see section 2.4)**:** transmission of information more or less sophisticated: it can be done by mail (RFMOs), by specific data transmission channel (eDAMIS for Eurostat, Fides and in the future FLUX for DG MARE) and then processed by the institution, or directly uploaded in the database (ICES, JRC). Checks can be implemented at different stages:

- Embedded in the template (Eurostat, JRC).
- At upload, data often go through an error screening procedure and can be loaded into the database only if correct (Eurostat, JRC, ICES).

Data production: data received are further processed and prepared for the dissemination:

- Additionnal checks (see section 2.5) can be performed (ICES, JRC, Eurostat) and feedback on identified errors is requested from the MS to ensure a good quality of the data.
- Preparation of specific aggregates (transversal data to EU level at Eurostat or JRC) or raising to stock level (ICES).

Data dissemination: data to be disseminated are prepared in different manners:

- Directly available after validation when the production database is also a dissemination database (DATRAS at ICES).
- Specific data dissemination files can be built from the production database (Eurostat).
- Files validated by expert groups not directly linked with the original production database can be uploaded for dissemination (SAIKU at JRC).

2.3.5. Conclusions

- The main institutions have developped good applications for uploading, checking and disseminating the fisheries data which work well.

- **From the available biological database system (at ICES and JRC), the RDB-Fishframe is more advanced in terms of functionalities, level of detailed information storage, implementation of models for the raising of the data, reusability and genericity.** Technically, the RDB-FishFrame system is ready for receiving the data for any sea-basin but there is a need for a final validation of the RDB-FishFrame procedures: The national data transmitters should compare the data loaded in RDB-FishFrame raised to the national level with the data they prepared "at home" for InterCatch to check whether the

routines developed in RDB-FishFrame produce consistent results compared with data uploaded into InterCatch.

- **It could be possible to achieve burden reduction by providing an online tool for raising data at national level**. If the RDB-FishFrame aggregation tool is satisfactory, the burden could be considerably reduced (especially for MS which do not have specialised tools developed for processing the biological data). In addition, when the RDB-FishFrame-InterCatch system will be completed, the MS in ICES area will only be requested to fill-in the RDB-FishFrame, not anymore InterCatch which will be fed from the output of RDB-Fisframe.

- **Storing biological data in a single DB could make it easier to compare individual data from different sea basins** (e.g. length or age composition, maturity@age, etc.)**.** Such comparisons may be useful for stock assessement purposes (distinction of sub-population) as well as for the assessment of environmental impact (global warming). This would require having a more "open" access to the data than what is possible now in RDB-Fishframe and InterCatch because these DBs are primarily accessible to the ICES WGs, while scientific work is done in many other institutions as well (e.g. universities).

- **Sharing development effort could be a way to enhance an existing system at a lower cost**. RDB-FishFrame could be used to cover more sea basins (e.g. MBS or distant waters) provided some additionnal modules and ad hoc procedures could be developped. The costs of development could be foreseen and shared among different institutions. A central unit (e.g. ICES) would then review and approve each new development (e.g. addition of data on areas which are not yet covered by RDB-FishFrame) and the compliance with requirements defined by the users (e.g. which data and at which aggregation level, quality checks, standard queries, etc.).

- **Use of common referential (e.g. DG MARE's Master Data Register) would make the data accessible through different networks, comparable and interoperable** (e.g. use of web service against WoRMS species catalogue from marinespecies, org. spatial reference resource through Geonetwork).

- **There are various duplications in the data collected by JRC, ICES and Eurostat.** This is a consequence of additional dimensions or differences in data definitions due to specific logic and purpose of the various data calls. It is then important to stress that calls have sometimes their own specific logic and purpose, which is the origin of the differences of the specifications and need to be analysed depending on the use made before one can really conclude whether there is duplication. This is elaborated in detail in table 1 of Annex 2.8., which is summarized in the following table.

Table 5: Illustration of differences between main data calls

	RDBs – métier related (call 2013)	Economic data – JRC call	Effort regimes – JRC call (2013) Atlantic areas
Catch area	ICES rectangle	FAO Level 4 (Baltic) FAO level 3 (other areas)	FAO and ICES rectangle
Time	Month	Year	Quarter
Vessel size	6 length classes (m)	6 length classes (m)	Atlantic areas : 3 classes Baltic Sea : 7 classes
Gear	58 gears	12 gears (App. III)	9 gears
Target species	DCF level 5 (App. IV)	Not applicable	Description of fishery
Mesh size	Mesh size by mgt regime	Not applicable	Mesh size by to mgt regime
Landings (kg, live weight)	Area misreported catch Official landings	Volume (kg) Value in euro	Landings at age Discards at age
Effort variables	6 variables	13 variables	4 variables

Source: Table 1 of Annex 2.8.

The differences of data calls shown in table 5 have their "own logic", which can be illustrated with two exanples:

- Biological assessments require detailed catch and effort data in terms of time (month), space (ICES rectangle) and 58 gears. This is also feasible, because sampling can be designed accordingly to these detailed requirements. On the other hand catch and effort data under the economic data call is collected on a higher aggregation level (year, FAO areas, 12 gears) because this is a relevant to match with the production costs and feasible in terms of required sample size.
- Effort regimes data call uses its own definition of vessel sizes, gears and time in order to continue with the existing historical time series.

2.4. *Data upload*

2.4.1. Procedures

ICES

Workflow for data upload is very similar in the 3 databases systems managed by ICES (DATRAS, InterCatch and RDB-FishFrame): MS prepare data files based on a detailed format description available on ICES web site and upload them directly via the web site and a specific logging access.

Frequency of data calls varies from one database to another. RDB-FishFrame and InterCatch are updated annually (though sometimes ad-hoc call can be made for InterCatch). Frequency of updates of DATRAS depends on the surveys. In general biological stock-related, biological métier-related, catch/landings and effort data are updated continuously, quarterly or on data call basis.

Manuals for DATRAS data calls are available on ICES websites[12]. A user manual is available for RDB-FishFrame and InterCatch[13].

JRC

JRC has 5 databases related to DCF data. They are updated on an annual basis except for fish processing where data calls are held on a bi-annual basis.

JRC is mandated to collect data on behalf of DG MARE. As stated in the legal framework, the deadline of the data call is one month after the day of the launching of that call.

The upload procedure is identical for all data calls. Only the structure of the Excel templates varies from one data call to the other. The data call web page on JRC web site provides guideline to the data uploading procedure and documentation on the call.

Data are prepared by the MS and submitted to JRC through the same data uploading facility available for all databases.

Eurostat

Eurostat maintains 2 databases of interest in DCF context, MDT (fleet, catches, landing and aquaculture data) and Eurocube (fish processing).

Upload of fleet, catches, landing and aquaculture data in MDT.

Eurostat uses the DG MARE fleet register at the individual vessel level to prepare the data disseminated on the web site. The transmission of landings and aquaculture data from MS to Eurostat is done via the eDAMIS.

Three data calls: catch data, landing data and aquaculture data, are launched every year one or two month before their deadline. For these data calls, Eurostat provides up-to-date Excel templates. Nomenclatures are given in the file as separate worksheets. Files can be converted to SDMX-ML with to meet the format expected by eDAMIS. The upload process using the new SDMX-ML was not easy to implement given MS difficulties to generate the SDMX files. Eurostat implemented an embedded function in the Excel template to convert the file in SDMX-ML to facilitate the MS task in preparing the SDMX-ML files.

Upload of fish processing data in EBB/Eurocube.

[12] *DATRAS Baltic International Trawl Survey (BITS): https://DATRAS.ICES.dk/Documents/Manuals/Addendum_1_WGBIFS_BITS_Manual_2011.pdf*
DATRAS International Bottom Trawl Surveys in the Western and Southern Areas & in the North Sea and Kattegat/Skagerrak area: https://DATRAS.ICES.dk/Documents/Manuals/Addendum_2_Manual_IBTS_Western_and_Southern_Areas_Revision_III.pdf and http://www.ICES.dk/sites/pub/Publication%20Reports/ICES%20Survey%20Protocols%20%28SISP%29/SISP1-IBTSVIII.pdf
Beam Trawl Survey (BTS) "Offshore survey": http://DATRAS.ICES.dk/Documents/Manuals/WGBEAM_Manual.pdf
[13] *http://www.ices.dk/marine-data/Documents/InterCatch%20User%20Manual%20Doc1-10.pdf*

The transmission of all data (and metadata) covered by SBS (thus including fish processing) is given in detail in the technical annexes to Commission Regulation (EC) No 250/2009 of 11 March 2009. Fish processing sector being included in SBS, it follows the usual SBS data collection and upload process.

GFCM

GFCM manages 3 main databases: fishing fleet (>15m), catch and effort and aquaculture.

GFCM current transmission protocol is based on files sent through e-mail. This protocol has significantly evolved within the past 2 years with the implementation of XML files for reporting. The evolution will continue to ward the implementation of a dedicated SSL data submission process through a dedicated SharePoint portal. Once received, data are manually uploaded in the relevant database by the GFCM data manager.

Data file formats are described and available on the GFCM web sites[14] and are now quite stable. As the procedure to upload data in the database is manually done by GFCM staff, there is no technical guideline to the upload process.

The generalization of XML format to upload data from MS will increase the quality of data uploaded in the databases.

DG MARE

Data are provided by the MS to Unit D-4 of DG MARE through the FIDES system in XML format. This system is quite old (implemented in 2002), covers several kinds of data exchange mechanisms and will be replaced by a new transmission system, FLUX (Fisheries Language for Universal eXchange), in the coming years. This activity has already started with CR, no longer using FIDES but a system laying the foudnations of FLUX (not yet FLUX but an evolution of FIDES toward FLUX). No technical guidelines are available for FIDES.

FLUX combines a transportation layer (a communication system to transport safely and securly any message in an envelope and from one end-point to another) and a business layer (definition of the format of the message inside the envelope). FLUX business layer format descriptions will be submitted for adoption as UN/CEFACT standard. It aims to be the universal data transmission system for any fisheries data for DG MARE.

[14] *http://www.gfcm.org/gfcm/topic/16103/en*

2.4.2. Formats and content

ICES

For the 2 main ICES databases and RDB-FishFrame hosted by ICES, the format and content are the following:

DATRAS

The format used for the DATRAS database did not change over the past three year (the list of fields' structure can be found on DATRAS web page[15]).

When delivering the data to the ICES Secretariat one file should only contain data from one survey, year, quarter, ship and gear. The data files structure can handle the different kinds of trawl survey data as described in the corresponding user manual.

There is a relation between the different types of records: HH (record with detailed haul information) refers to several HL (length frequency data) records collected during the same haul. The number and kind of species recorded must correspond to the species recording code as specified in record type HH. Records of type CA (sex-maturity-age–length keys - SMALK's) will also be recorded in the same file.

Nomenclatures used are compliant with DCF.

INTERCATCH

The format used for the InterCatch database did change slightly over the past three years. It is described in a manual.[16]

Two types of data files can be uploaded: "commercial catch" and "survey and logbook" (weight of the stock [WEST] and maturity).

RDB-FishFrame

Since 2013, data are prepared and uploaded by MS using the RDB-FishFrame Exchange Format[17].

Nomenclatures used are compliant with DCF except for species. RDB-FishFrame is using the scientific name, not the ASFIS code. It will to move to WoRMS.

The data formats are given for three data types, each consisting of one or more record types: commercial samplings, commercial landing statistics and commercial effort statistics.

[15] *https://datras.ices.dk/Data_products/ReportingFormat.aspx*
[16] *http://info.ices.dk/datacentre/InterCatch/IC-ExchFormat1-0%20Doc1-8.pdf*
[17] *Jansen , T. (Ed) 2009.Definition of Standard Data-Exchange Format for Sampling, Landings, and Effort Data from Commercial Fisheries. ICES Cooperative Research Report No. 296. 43pp*

JRC

JRC provides Excel templates (Excel 97-2003 or Excel 2007) on the JRC website. The definition of the variables and coding are clearly described on each page from which the template can be downloaded.

Annual Economic Report EU fishing fleet

This call contains fleet economic and transversal data. Nomenclatures are compliant with DCF.

It is composed of 5 Excel templates for fleet data, 3 for transversal data and one for recreational fisheries data.

Each template contains 2 worksheets:

- one normal detailed by supra-region, fishing technique and vessel length, cluster name and sampling information (strategy, achieved sample rate and coefficient of variation);
- one total with main variables aggregated by MS and year and achieved sample rate only.

Aquaculture data call

Aquaculture data collection covers the population of enterprises whose main activity is defined according to the SRs definition under NACE Code 05.02 'Fish farming', but only for marine or brackish waters Fresh water aquaculture is not compulsory.

All DCF data indicated in Appendix X of the DCF regulation are collected. The data upload is organised in two Excel templates.

Nomenclatures are compliant with DCF.

Fish processing industry data call

All DCF data to be collected as indicated in Appendix XII of the multiannual programme are collected for processing industry.

A template is requested to estimate the number of enterprises where fish processing is not the main activity and the turnover attributed to fish processing. The detailed data collection concerns the enterprises where fish processing is the main activity.

Nomenclatures are compliant with DCF.

Mediterranean & Black Sea – "SGMed" data call

Three data calls by JRC for Mediterranean and Black Sea are launched: catch, effort and landing / discard calls.

Nomenclatures are compliant with DCF.

Structure will change in 2014 as stated in "*MEDITS Handbook, International bottom trawl survey in the Mediterranean Instruction manual from March 2013*".

Effort regime data call

Fishing effort regime data call changed due to the introduction of a new species and a new capacity variable for the Baltic Sea fisheries.

Nomenclatures are compliant with DCF except for vessel length.

Eurostat

MDT

The 4 catches data calls are launched yearly: catches in area 27 (NE Atlantic), catches in area (other than those of north Atlantic, i.e. areas 34, 37, 41, 47, 51) and catches and effort in the area 21 (NAFO) – monthly and annual data. Catches are in tonnes of liveweight per year (month when relevant), country, fishing area and species. Each call is an Excel file respectively FISH_ **C27_A**.xls , FISH_34_To_51.xls , FISH_C21A_A .xls and FISH_C21B_A .xls (monthly).

Regarding catches, SR nomenclature of MDT is only partly consistent with DCF: some nomenclatures are compliant with DCF (species, country), some are too aggregated (geographical stratification), some have different classifications (fishing gear category, vessel size[18]) (see Part 2, table 100).

The landing data call, based on regulation 1921/2006, is composed of one template called LANDG_A .xls. The landings data distinguish species, product forms (fresh, frozen, fillets, etc.) and destination (human consumption, industrial use, etc.). Landings are monitored in product weight.

The 6 aquaculture data calls, based on regulation 762/2008, are mostly annual: i) production from aquaculture excluding nurseries and hatcheries, ii) production of eggs for human consumption, iii) input to capture-based aquaculture, iv) production of nurseries and hatcheries, v) structure of the aquaculture sector (once every three years) and vi) annual methodological report of the national systems for aquaculture statistics.

[18] *Classes used by Eurostat are slightly different than the DCF ones. Eurostat length classes [24-30 m], [30-36 m], [30-33 m], [33-42 m], [36-42 m] , [>42] are grouped into length classes [24 < 40 m] and [> 40] under DCF. In addition, Eurostat also uses nine GT-categories.*

Indicators on aquaculture required by the SRs are not the same as those requested by the DCF (see Part 2, table 103).

Fish processing data in EBB/Eurocube

Commission Regulation (EC) No 250/2009 of 11 March 2009 on Strucutral Business Statistics specifies that MS have to send data files with a predefined format to Eurostat. Each series of data has to be sent in a different file.

The breakdown by activity at 4-digit level is required. Therefore the fish processing activity (NACE Rev.2 class C10.2.0 "Processing and preserving of fish, crustaceans and molluscs") has no further breakdown. The contents of the data reflect the requirements of the SBS, which are different with that of the DCF (see Part 2, table 104).

The format and contents are used for all SBS activities. It seems difficult to envisage that data requirements would be adapted to include the additional variables specified by DCF as the fish processing industry is a very small part of the SBS data collection.

GFCM

For 2 GFCM databases, catches (task 1) and vessel record, formats are defined respectively in the GFCM recommendation GFCM/33/2009/319 and GFCM/35/2011/120. SIPAM, the aquaculture database, structure is not described in any recommendation as it has been handed over to GFCM by the SIPAM project when closed.

Call for catches contains data on fleet segments, fishing activities, main resources, biological and socio-economic variables, catch, effort, overall by-catch. Nomenclatures are compliant with DCF or can be mapped with DCF, except for tuna seiner in gear classification.

Vessel record call contains data on all fishing vessels (over 15m) operating in the GFCM competence area. Nomenclatures are compliant with DCF or can be mapped with DCF (see Part 2 tables 82 and 83).

Data on aquaculture are less detailed than DCF, except for production value per species.

DG MARE

The used format is XML with a vast number of different formats depending on the source of the data (Member States – format can vary from one country to another) and the type of data (Fishing effort (art. 26), catch data (art. 33), data on exhaustion of fishing opportunities (art. 34), closure by MS (art. 35), Fleet data from the basic regulation, licenses, vessel notifications and catches from the Fishing Authorisation Regulation, VMS (art 9), eRS (art 15)).

[19] *GFCM/33/2009/3 can be found here: http://151.1.154.86/gfcmwebsite/docs/RecRes/Rec_GFCM_33_2009_3.pdf*
[20] *GFCM/35/2011/1 can be found here: http://151.1.154.86/gfcmwebsite/Docs/RecRes/RES-GFCM_35_2011_1.pdf*

The current activity with the implementation of FLUX and more specifically its business layer is to review these existing formats, harmonize and standardize them through submission to UN/CEFACT group, addressing the main issue of the almost impossibility to summarise all the data calls. This harmonization work has started with Aggregated Catch Report (ACR).

DCF should be part of the process, though no indication on plan to actually do it has been mentioned. Once this harmonization process is completed, it is envisaged by unit D4 to generate DCF data from eRS and VMS, when/if confidentiality issues are resolved.

2.4.3. Resources required for responding to data calls

The survey of the national correspondents, carried out under the study, contained questions regarding the man-power required at the MS level to respond to the data calls. All received information is presented in annex 2.8, table 4 and summarized in the following table.

Table 6 shows that for some data calls there are large differences between the MS. This means probably that the NCs included different actitvities. It is noted that these are mostly 'guesstimates'. Most MS do not monitor precisely time spent on provision of data to data calls, as they have not implemented a system to monitor the time spent by staff on different tasks[21]. While the MS are able to indicate how much time is required to respond to economic data calls on fleet, aquaculture and processing, the information is very incomplete when it comes to all other data calls.

The sum of time required, according to the provided information, amounts to about 2,000 person days, i.e. some 10-12 man-years (with 165-200 effective workding days/year). Considering the incompleteness of the provided information, the total amount of time may well be around 4,000 days, i.e. 20-25 man-years.

[21] *It is noted that this is also the finding of missions carried out to different MS within the project "Assistance for the monitoring of the implementation of the NPs for the collection, management and use of data in the fisheries sector (MARE/2009/2008)*

Table 6. Summary of effort required to respond to data calls (person-days)

Name	BE	BG	DE	DK	ES	EE	FI	FR	EL	IE	IT	LT	LV	MT	NL	PO	PT	RO	SI	SE	UK	TOTAL CALL
Fleet economic data 2008-2013	20	30	20	20	17	20	15	145		30	7	12	60	10		20		60			24	**510**
Aquaculture 2008-2011	15	26	2	10	15	8	10			30	5		30	6		4				13		**174**
Fish processing 2008-2011	22	20	5	10	15	20	5		2	30	5	7	30	2		20				15	15	**223**
Fishing effort	20		15	20	35	100	17	7				15	50		10	20	15			12		**336**
RCM 2013	20		5			100	5	25				13			10		20			27		**225**
ICES various WG	40		9	15			11	210									50			20		**355**
MBS data calls					15			68			15			50				20	5			**173**
RFMOs						20		60				10			5							**95**
Total MS	**137**	**76**	**56**	**75**	**97**	**268**	**63**	**515**	**2**	**90**	**32**	**57**	**170**	**68**	**25**	**64**	**85**	**80**	**5**	**87**	**39**	**2,091**

Source: survey of national correspondents

Comments to the table:

- ✓ Some MS have indicated ranges, e.g. 20-30 days. In those cases, the average number was taken.
- ✓ The information refers largely to data calls either in 2012 or 2013. In some cases it was complemented with data referring to earlier years.

2.4.4. Conclusions

It is estimated that responding to data calls related to DCF requires about 20-25 person-years annually across all involved MS.

Data calls and data upload for the main actors involved in DCF or potentially involved such as GFCM follows similar upload processes. Up-to-date Excel files describe formats and nomenclatures for data reporting. The major difference is in the level of automation of processes. It goes from a manual process of uploading csv files in the database (GFCM), through a semi automated process requiring some internal manipulation (Eurostat) to direct upload in the database by the provider (JRC, ICES).

The global level of technicity in data upload management is high. Processes and workflows are well identified and managed. Formats are easily accessible. Upload is easily done through secured interfaces or by skilled data managers.

The global direction is to go from csv files data upload to upload based on XML files (SDMX-ML being XML), providing better control and checks on the validation of format and metadata.

The FLUX DG MARE initiative, related to CR data, aims to provide a common tool to facilitate data exchange (FLUX transportation layer) and methodologies to standard data format (FLUX business layers). It pursues harmonization and standardization of data calls to improve / optimize processes and costs for fisheries data exchange.

Regarding the content of data, MS transmit different data to different databases, although some **overlaps** exist:

- **Biologic indicators on Atlantic stocks** are transmitted by the MS to RDB-Fishframe and to InterCatch DB operated by ICES. ICES is developing software to transfer data from RDB-FishFrame directly to InterCatch.
- Total MS **landings** (by species and FAO areas) are based almost always on CR data. They are transmitted to Eurostat[22], JRC (fleet economic data), ICES (RDB-Fishframe), FAO, GFCM and other RFMOs.
- **Effort** data submitted to DG MARE under the CR (for compliance purposes) overlaps significantly with the MS submissions of effort data under almost all DCF data calls carried out by JRC and ICES for scientific purposes. However, the purposes of the various DCF calls are very different.
- Eurostat and the JRC both collect data on **aquaculture** on the basis of Statistical regulation 762/2008 and of the DCF Regulation, respectively. The DCF covers also aquaculture production costs, while Eurostat focuses on production volume and value only.
- Eurostat and JRC both collect data on **fish processing**. DCF requires to collect 17 indicators of which 11 are identical or closely related to SBS. The link between DCF and SBS is explicitly stated in Appendix XII of the Regulation 93/2010. Eurostat data calls are very detailed, being part of the broader SBS. Given that the agreed format is used for all other activities under SBS, it seems difficult to envisage that data transmission is changed to include the additional variables collected under DCF. The fish processing industry is a very small part of the SBS.
- Data provided by MS to **GFCM** is based on data collected by the MS under the Statistical or DCF regulation. It is relatively less detailed.

[22] *It is noted that in DCF catches of a MS are sum of landings and discards of the national fleet. However, SRs use a different definition, where landings in a MS are sum of landings in ports of a MS by all vessels, national and foreign. In addition, SRs define landings in product weight, while DCF used live weight.*

2.5. *Quality control*

2.5.1. Quality control procedures

ICES

Data quality controls are driven in the frame of continuous interactions with expert groups in order to support scientific advice process. Controls are done by both manual and automated procedures on all primary and aggregated biological and transeversal data.

All indicators are stored in the offline ICES Quality Control (QC) database. Information on quality indicators is available through WGs annual reports or metadata.

DATRAS database

Quality checks are manually or software based, sometimes both. Expert WGs can rely on a series of possible quality checks concerning comparisons between different datasets (e.g. inter-survey comparisons) or characteristics of a specific dataset (e.g. size range check).

In addition, comparisons of trawl surveys are done and provided to trawl survey WGs.

InterCatch

Missing values, duplicated records, timeliness and coding are checked for all data. Some additional checks are done for catch / landings data (e.g. typing errors, arithmetic checks), for which an "overlap check" has been implemented (and is foreseen to be rewritten) in order to avoid double imputations. InterCatch contains an advanced validation tool.

RDB-FishFrame

Quality check for métier and stock related as well as catch/landings data are manually or software based, depending on their type. For effort data quality tests exist but have not been specified. Several checks are performed during data upload, on relations betweens data (e.g. coherence) and data itself (e.g. enumeration checks, duplicates).

JRC

Concerning the production databases, while all quantitative quality checks are performed by software, qualitative checks are done using both, manual and automated procedures. Qualitative checks of biological data are automated too. A data validation tool (DV) is used for some data calls, in order to upload clean data for which inconsistencies, duplications and problems of format have been previously automatically identified and eliminated.

Concerning the Saiku dissemination database, the STECF WG validates uploaded data. No other tests are done.

JRC has developed coverage reports[23] as part of quality improvement process. Coverage reports analyse issues related to availability, timeliness and completeness of the data detected by JRC and by Expert Working Groups convened under the STECF.

Other reports developed by JRC are those related with the STECF evaluation reports of the MS as well as the Technical Reports (TR) on DCF. These TR contain the required statistical quality indicators (coefficient of variation, samping size, etc.).

DG MARE

DG MARE does not carry out any quality control of DCF data because it relies on procedures estabalished by JRC.

In order to assure quality of the data submitted to DG MARE under the requirements of the CR, DG MARE has started the VALID initiative to address the needs of fisheries data validation at Member State level. The VALID project is assisting MS to implement validation processes on primary data before submission to the Commission. DG MARE will apply these validation procedures after reception of the new ACR (Aggregated Catch Report).

At the same time, a set of tools (based on JBoss business rule engine and Java) which is being developed by DG MARE could be reused by MS as this solution is entirely open source. It could be reused in any DCF scenario to run specific validations processes.

GFCM

Data quality checks are done either manually or automatically in the following three databases: Task 1 (contains data on fleet segment, fishing activities, main resources, biological and socio-economic variables, catch and effort); SIPAM (contains data on aquaculture) and Vessel Records (contains data of all fishing vessels operating in the GFCM competence area).

Regional Sea Conventions

According to HELCOM's principles of Data Information Strategy, HELCOM data will be handled by Data Centres which will be responsible, among others, for ensuring that relevant quality control and validation procedures are in place including quality flagging of data.

HELCOM and OSPAR do not host any DCF-related database, but both share a MoU whereby ICES provides as data centre facilities for the validation of that data and performance of quality checks.

The Barcelona Convention does not collect any data even loosely related to fisheries. The Bucharest Convention does not collect any DCF-related data. Some ad hoc fisheries data is collected by the Bucharest Convention for their Advisory Group on Fisheries, which is transmitted

[23] http://datacollection.jrc.ec.europa.eu/docs/coverage

every year by each country, but serves only for national or regional reporting and is not stored in any database. All fisheries data for the Mediterranean and Black Sea are handled by GFCM. Barcelona and Bucarest conventions do not have any specific quality control policy in relation to fisheries data.

Eurostat

Eurostat has developed a detailed quality assessment and control, based on the European Statistics Code of Practice and ESS Standards for reporting on quality. MS have to submit quality reports on SBS statistics to Eurostat (fish processing). They include, under an agreed report format, methodological information on the stages of data collection, data processing and data output (including the calculated accuracy indicators). Several MS use the same SAS macro (CLAAN) to calculate accuracy indicators. These are examples of good practice that could possibly be replicated for some fisheries data.

MDT (multi dimensional tables)

MDT production database hosts data on catches, landings, aquaculture and fleet.

All data quality checks are software based and share the same kind of test functions on different datasets. However, errors are not reported automatically. When an error is detected, it must be copied manually in Excel file and be reported to MS for correction.

Data validation for fish processing – Editing Building Block (EBB)

Data on the fish processing industry are validated by Eurostat with the EBB internal tool and stored in an Oracle OLAP database.

The EBB tool ensures the validation of the transmitted data based on the application of a series of validation rules. In order to improve the quality at the MS level and to avoid back and forth sending of data, Eurostat has developed a standalone application of EBB to be used by MS in a similar way as the Eurostat EBB tool.

Eurobase

Eurobase is the common dissemination database. The Eurobase displays flags data with comments but does not apply any validation rule.

2.5.2. Conclusions

A large number of quality controls are run with differents software tools; showing that although a far reaching quality control exists, further standardization should be achieved and the most efficient software tools should be compared and shared among MS and supra-national institutions.

Quality reporting and the dissemination of such reports, is limited and fragmented. There are no standards for quality reporting shared by supra-national institutions despite the existence of international quality assessment framework.

2.6. Dissemination

ICES

The ICES DATRAS DB is publicly available online. It allows the free download of primary and aggregated data from trawl surveys. The regional coverage of the data is Baltic, North Sea and the North East Atlantic. DATRAS data download products are in csv format and stored in a zip file containing the data product guide and metadata information.

At ICES, stock and métier sampling data in InterCatch DB and RDB-FishFrame DB are treated as restricted data, with access for the providing countries and the specific assessment WGs. The data can be extracted by authorised users through dynamic tables / filters in Excel/csv/txt formats. RDB-FishFrame allows also extraction in XML.

JRC

At JRC, the dissemination DB is Saiku and is accessible through a standard web browser without requiring a log in. Of all dissemination DBs, Saiku is most directly related to DCF. Disseminated data is validated by STECF.

The online dissemination is divided into two parts. One is for disseminating data on effort, landing, and discards. The data can be extracted as "Data" tab: it presents tables which can be dynamically designed by selecting the column and line dimensions, and even filtering some element in the selected axes. It allows displaying information in graphs, exporting data to Excel or csv. It provides statistics on the selected table (min, max, sum, average, standard deviation).

The second part is about a set of socio-economic indicators concerning EU fishing fleet, aquaculture and processing industry (maximum data coverage is from 2006 – 2010) the "Economic indicators" tab: it presents predefined tables by country and year for dimensions chosen and the sector segment corresponding to the dataset selected. It allows displaying information in chart and exporting data to csv. Information about those indicators is also published, including in-depth analysis, in the STECF reports. Beside the Saiku database/Fishreg web site, the data validated by the STECF are also disseminated on the STECF web site in Excel files and reports.

At JRC, Production data are only accessible internally but an extraction can be prepared for the working groups. Mediterranean and Black Sea biological data and recreational fisheries data are not disseminated on JRC web site.

DG MARE

DG MARE does not disseminate any DCF data.

DG ENVIRONMENT

DG ENV provides an alphabetical list of studies and reports on the environment website, some of those with relevance for MSFD but no specific DCF topics or any DB.

GFCM

GFCM publishes some data through its statistics bulletin on catch, landing and effort (2008 to 2010) from the Task1 DB as pdf files. The GFCM authorized Vessel Record/List is exposed to external users (a secured web service) through an internally developed tool, the AVL browser (Authorized vessel list).

At GFCM there are different groups for data dissemination. One is internal GFCM users with full access to internal databases Vessel Record and Task1 through the LAN (intranet). The 2nd are restricted users that have access to a set of data for professional use (Authorized vessel list for port authorities, Task1 data for scientific groups) and the general public as 3rd group. A 4th group can be added for the SIPAM Aquaculture databases (Sipam – production and Sipam – units), where national coordinators authorized to submit data to the system (actual upload is under GFCM staff responsibility) have also access.

Eurostat

Eurostat uses the Eurobase DB to disseminate data on marine and aquaculture production (from MTD DB) and on fish production (from EBB/Eurocube DB). Eurostat does not do any aggregation of aquaculture data if a single value is confidential. Accordingly, only the data without any confidential issues are disseminated even if further aggregation steps could follow up.

Eurostat has two kinds of data presentations available to the public for fisheries statistics. a) Interactive extractions with customised selection parameters for fisheries (Total fishery production with Catch and Aquaculture, Catches by Area, Aquaculture production by species, Landings of fishery products, Fishing fleet by number, capacity, engine power). b) Main fixed tables which contain the fishery production in all fishing regions (tonnes live weight), catches total and by FAO regions (tonnes live weight), aquaculture (tonnes and euro), fishing fleet tables (total engine power, total tonnage, and total number of vessels). The fixed table formats directly access and extract the latest data from Eurobase DB therefore data provided to the public are always up to date but the extraction parameters are limited. Tables mainly consist in crossing the geographical zone by year. The output can be customised by applying filter, visualising the result in graph or map. Eurobase allows to export results in xls, csv, html, PC-Axcis, SPSS, tsv, PDF. Data

on fish processing industry follows the same presentation as for other economic sectors falling under SBS.

RSCs

Two of the four RSCs, HELCOM and OSPAR have dissemination DBs with lose connection to DCF or related environmental contents. Those DBs are freely accessible through the internet and data is updated on an ad hoc basis when new data is available. Barcelona and Bucharest Conventions do not hold any dissemination DBs.

RFMOs

The extent and the content of data disseminated with public free access differ among the RFMOs considerably. Usually, several access levels for staff, observer, special user groups and the general public exist. Some RFMOs host specific dissemination DBs containing information on annual catches by species, subareas, country, and year or other DBs provide data on monthly catch and effort information by year, country, gear, tonnage, main species, division and year. These DBs use Access, SQL and Windows server technics with a data extraction tool which provides Users online access and the choice for "basic" or an "advanced query". But also other technical solutions are in use in other RFMOs. One RFMO disseminate aggregated annual catch data based on monthly catch returns listed per country in pdf documents only and other RFMOs do not have any dissemination of data to the general public. Examples for restricted data access relate to contents in e.g. GeoDatabase, VMS or vessel registry DB.

2.6.1. Conclusions

All DCF data are accessible to professional users (i.e. scientific WGs at ICES, STECF, RFMOs, etc.).

The DCF data is most directly distributed to the public by the Saiku DB, hosted by JRC. This regards data on fleet economics, effort, landings, discards and some data on MBS. In addition the STECF website offers STECF reports (pdf) and some of the supporting data in Excel. This regards fleet economics, fish processing, aquaculture and various tables on landings.

ICES DATRAS provides the data collected during the scientific surveys at sea in the Atlantic area and the data are publicly available on the ICES website.

There are several areas for which DCF data is not (easily) accessible to the public. This regards mainly the biological data stored in RDB-FishFrame and the biological data compiled under MBS data call of JRC.

In general, it must be concluded that public access to DCF data is not comprehensive. It is spread over several websites and many separate files. Users who are not intimately acquainted with the system are likely to face problems.

2.7. *Interactions with MSFD and IMP*

2.7.1. The Marine Strategy Framework Directive (MSFD)

The purpose of the MSFD (Directive 2008/56/EC, Commission Decision 2010/477/EU) is to establish a framework within which MS shall take the necessary measures to achieve or maintain good environmental status (GES) in the marine environment by the year 2020 at the latest. MSFD is considered "the environmental pillar of the Integrated Maritime Policy"[24] (IMP), but in contrast to the IMP (responsibility of DG MARE), the MSFD is an environmental policy and consequently falls under the responsibility of DG ENVI. MSFD is a cross-cutting policy which concerns a number of sectoral policies, among others fisheries.

In the context of MSFD and for each marine region or sub-region under their sovereignty, MSs have to undertake an initial assessment of the current environmental status, determine the GES and establish a series of environmental targets and associated indicators (2012), set up a monitoring programme for ongoing assessment (2014) and develop a programme of measures to achieve/maintain the GES (2015). For this purpose, a series 11 qualitative descriptors (with associated criteria and specific indicators) has been defined to assess and monitor the good environmental status. Some of these descriptors are in direct or indirect relation to data collected under DCF.

The most relevant is descriptor 3 defined as "Populations of all commercially exploited fish and shellfish are within safe biological limits, exhibiting a population age and size distribution that is indicative of a healthy stock". The indicators defined for this descriptor (i.e. fishing mortality, spawning stock biomass or mean maximum length) can be determined either directly through data collected under DCF or obtained through analytical stock assessments based on DCF data (performed by ICES, GFCM, STECF, ICCAT, etc.).

An indirect connection to DCF can also be observed for some other descriptors of MSFD: descriptor 1 (Biological diversity), descriptor 2 (Non-indigenous species), descriptor 4 (Marine food webs) and descriptor 6 (Sea-floor integrity). In these cases, data collected under DCF as well as the assessments based on these data contribute to the ascertainment of these descriptors (together with data from others sources). Thus, in relation to DCF, interactions can be found not only with regard to métier- and stock-related biological data (including research surveys at sea), but also to the environmental indicators of DCF (e. g. indicator 4, "Size at maturation of exploited fish species" or indicator 7, "Areas not impacted by mobile bottom gears").

In analogy to other environment-related policies and to water-related directives, reporting obligations in the context of MSFD are to be integrated into the Water Information System for Europe (WISE) platform (into a WISE-Marine module) through the EEA (European Environment Agency) ReportNet management system. This is part of the European environment information

[24] *Progress Report on the EU's Integrated Maritime Policy, Brussels, 15.10.2009, COM(2009)540 final {SEC(2009) 1343*

and observation network (EIONet). Moreover, EIONet serves as a repository of general documents and tools on MSFD and of text reports and data uploaded by MSs.

From single cases it is known that DCF data find their way into the reporting obligations under MSFD either at national level (through an exchange between the fisheries and the environmental institutions/ministries, where sometimes fisheries scientists with full access to DCF data and assessments undertake the fisheries-related analyses) or at supra-national level (e. g. through ICES publically accessible databases on international bottom trawl surveys, on data from stock assessments and on commercial landings)[25]. However, there are still a number of open methodological questions in the context of MSFD and its indicators, so the interactions with DCF cannot be thoroughly explored yet. Also, monitoring programmes under MSFD – which will generate a much higher quantity of data at a higher frequency than the earlier reporting obligations – will start only in 2014, and many details on the transmission and storage of these data are not yet defined.

2.7.2. The Integrated Maritime Policy (IMP)

Following a consultation on the future of the European maritime policy, the EU IMP was established in 2007 (COM (2007) 575 final) under the umbrella of DG MARE. The IMP builds on a cross-sectoral and multidisciplinary approach and aims at coordinating the various marine and maritime related initiatives under the different sectoral policies in a coherent way. These range from fishing and aquaculture, shipbuilding, shipping and ports to offshore energy (oil, gas and renewables), blue biotechnology and coastal maritime tourism and recreation. The policy areas addressed in the IMP do not fall only under the competence of DG MARE, but also under that of other EC Directorates such as DGs Environment, Research & Innovation, Mobility and Transport, Energy, Regional and Urban Policy or Enterprise and Industry.

Among the IMP priorities covered by DG MARE (besides blue growth/marine and maritime economy, maritime spatial planning, integrated maritime surveillance and sea basin strategies), a special focus has been placed on marine data and knowledge. In 2008 first actions took place to develop prototype data platforms for the European Marine Observation and Data Network (EMODnet). Six thematic assembly groups have been defined and set up in order to provide specific marine data on: hydrography, geology, physics, chemistry, biology and physical habitats.

The creation of a thematic group on human activities has also been envisaged for a later stage. The thematic platforms are managed by six consortia formed by a network of 53 organisations in total.

[25] *In the context of MSFD, ICES plays a key role in providing scientific advice with regard to the definition of GES indicators and standards. Together with JRC, it has contributed in particular to provide guidance for the implementation of descriptor 3. Besides, it maintains long-term data series on marine living resources and the marine environment which is of prime importance for defining targets and thresholds for MSFD indicators.*

The main objective of EMODnet is to provide access to marine data, metadata and data products for entire sea basins through the six internet portals created for each thematic assembly group. As an important prerequisite, data, metadata and data products shall respect European standards and particularly the INSPIRE Directive (2007/2/EC) which defines common rules concerning metadata, data specifications, network services, data and service sharing, monitoring and reporting with the aim to ensure EU-wide harmonisation and compatibility in spatial data infrastructures. Furthermore, metadata and data products are required to be made available free of charge and without restrictions on use for whatever purpose. This shall apply where possible also to data (despite constraints due to license restrictions by data owners).

The six single thematic portals work as internet gateways and give access to archives managed at national level (i.e. by Member States' institutes), or at regional level by international organisations (i.e. consortia of marine data organisations). To this end, each thematic assembly group has put in place its own IT and software solution and standards. However, similarities are found among portals, e. g. in the use of SeaDataNet[26] common vocabularies.

According to the Interim Evaluation of EMODnet (COM (2012) 473 final), keeping the data at national level (i.e. in the original source) and having the possibility to retrieve them automatically on demand is more efficient than asking the MSs to report to a central body.

In fact, the intention to integrate data collected under DCF into EMODnet has been declared on many occasions. ICES already contributes to the portal on biology with some DCF data from their database of trawl surveys (DATRAS). Besides, no other DCF-related data is accessible through the portal. Dissemination of DCF data through EMODnet could be possibly further developed in the future.

Comparing the type of data available through the various thematic portals with the data collected under DCF, most similarities can be found with the assembly group on biology, where the abundance of living species is reported. DCF biological data (stock-related, possibly also métier-related data) as well as environmental indicators 1 to 4 of Appendix XIII of the Commission Decision 2010/93/EU could be made accessible through this portal. The environmental indicator 7 of the Appendix XIII of the same Commission Decision, "Areas not impacted by mobile bottom gears", has a direct link to the hydrography portal, where information on underwater features (seabed mapping) is provided. DCF could play relevant role in the (planned) thematic portal on human activities, once this will be operational, with the data on fleet activity as well as on marine aquaculture (DCF economic and transversal variables, environmental indicators on the fishing activity, possibly also métier-related biological data). As suggested by the European Parliament in October 2013[27], "(…) given the importance of the fisheries sector, (…) fisheries data should preferably be made an additional specific group within the EMODnet platform or, alternatively, integrated into the newly created 'human activities' thematic portal (…)".

[26] *Infrastructure for the management of marine data sets based on in-situ and remote observations.*
[27] *Text adopted at the sitting of Wednesday 23 October 2013 (P7_TA-PROV(2013)10-23), provisional edition*

2.7.3. Conclusions

Interactions between DCF and MSFD and their relative data reporting obligations and infrastructures can be analysed only on a provisional basis due to the fact that MSFD is in an early stage and still has to resolve issues related to the standardization and consolidation of its methodology. For a series of MSFD descriptors (mainly descriptor 3) a regular "flow" of data and information based on DCF is needed. However, it is not the case of a simple transmission of data from one database to another. For most cases, it is not possible to derive MSFD indicators and descriptors in an automatic way (e. g. through fixed algorithm) from DCF data, thus, human intervention is required. For this reason, there is no need to establish fixed interfaces between DCF and MSFD databases and infrastructures. It is important, though, to ensure full accessibility of DCF data and assessments to those scientists or institutes that are responsible for calculating MSFD descriptors (depending on the established workflow in each MS). For descriptor 3, it is very likely that it is calculated directly by the fisheries institution dealing with DCF, delivering a finished "data product" to the institution or authority responsible for MSFD. However, for the calculation of the other descriptors (where fisheries and non-fisheries data are needed), full access to primary data will have to be ensured to instutions dealing with MSFD. ICES will undoubtedly play an important role in providing input to MSFD monitoring for the marine regions in its geographical scope of competence.

The integration of DCF data into DG MARE's EMODnet platform seems to be in an even earlier phase. DCF data could be integrated within the EMODnet portals in different ways. Confidentiality of primary data will have to be assured. An important question is the continuity of EMODnet which will have to shift from a project-based infrastructure to an on-going sustainable and cost-effective system in order to serve the scientific community, public bodies and the private industry.

2.8. Institutional considerations

2.8.1. Purpose of the institutions legal basis and confidentiality issues

ICES

ICES "is an intergovernmental organization whose main objective is to increase the scientific knowledge of the marine environment and its living resources, and to use this knowledge to provide advice to competent authorities. ICES Science and Advice considers both how human activities affect marine ecosystems and how ecosystems affect human activities."[28]

ICES was established in 1902 as an intergovernmental organisation. It is based on conventions signed by its member states, the latest being the Copenhagen Declaration of 2002. ICES presently has 20 member countries, including all EU members adjacent to the area of ICES competence, i.e. North Atlantic, North Sea and Baltic Sea.

ICES and the European Union have a MoU which forms the basis for the provision of advisory services by ICES to the EU. Deliverables under this MoU include advice and associated support relating to the Data Collection Framework.[29]

ICES statutes contain provisions in relation to confidentiality of the data in their databases. ICES has an 'open' data policy, where users can obtain publicly available data as soon as this is feasible. This excludes, however, the commercial catch databases (InterCatch and RDB-FishFrame), which are subject to the provisions of the DCF regulation regarding confidentiality[30].

JRC

The JRC describes itself as "the scientific and technical arm of the European Commission. It is providing the scientific advice and technical know-how to support a wide range of EU policies. Its status as a Commission service, which guarantees independence from private or national interests, is crucial for pursuing its mission". "As the Commission's in-house science service, the JRC's mission is to provide EU policies with independent, evidence-based scientific and technical support throughout the whole policy cycle."[31]

As a service of the EU Commission, JRC can be entrusted by General Directorates of the EC with specific tasks. Such a delegation of tasks is governed by a specific administrative arrangement. In the case of fisheries data collection, this is e.g. the Administrative Arrangement N °SI2.648254

[28] *http://www.ices.dk/explore-us/what-we-do/Pages/default.aspx*
[29] *http://www.ices.dk/community/advisory-process/Documents/2013_EC_ICES_MoU_WEB.pdf*
[30] *http://ICES.dk/marine-data/guidelines-and-policy/Pages/ICES-data-policy.aspx and http://ices.dk/marine-data/Documents/ICES_Data_Policy_2012.pdf*
[31] *http://ec.europa.eu/dgs/jrc/index.cfm?id=1370*

between the DG MARE and the JRC, (JRC CONTRACT No. 33236-2013 NFP), which defines exact tasks and activities.

JRC has in its statutes provisions regarding data protection and confidentiality. These provisions do not, however, refer explicitly to DCF databases. DCF data/databases have a proper governing framework laid down in Reg. (EC) 199/2008, which is binding for JRC as a Commission service.

DG MARE

The responsibility to ensure the execution/implementation of DCF lies with DG MARE. The technical implementation of data calls, data storage and transmission at supra-national level has been delegated to JRC (for economic, effort and MBS data).

While DG MARE does not maintain any DCF databases, it hosts three Control Regulation (CR)-related databases on fishing effort, catch and the Fleet Register.

DG MARE is the driving force behind EMODnet.

GFCM

The purpose of GFCM is to promote the development, conservation, rational management and best utilization of living marine resources, as well as the sustainable development of aquaculture in the Mediterranean, Black Sea and connecting waters. It has 23 member states, including most riparian countries of the Mediterranean and Black Sea (EU and non-EU countries) as well as Japan. The EU is also member.

GFCM has been established within the framework of Article XIV of the FAO Constitution. It has no direct relation to EU legislation. No formal memorandum of understanding with EC with respect to a formal involvement in DCF exists. EU submits aggregated data on marine fisheries and aquaculture, according to agreements within GFCM. Confidentiality of data is assured by GFCM, but GFCM does not request any primary (possibly confidential) data.

Other RFMOs

The other RFMOs have a specific convention as their principal legal basis, with the exception of IOTC (Indian Ocean Tuna Commission), which is based on the FAO Constitution plus a specific agreement. EU is a contracting party of these RFMOs, or, in case of CCSBT (Commission for the Conservation of Sourthern Bluefin Tuna), has signed up as a cooperating Non-Member. No legal basis for a role in data collection (except for the specific data they require) exists. The area of responsibility of RFMO's is defined in relation to certain seas and/or species.

In general, RFMO's convention legal texts include provisions on data confidentiality.

RSCs

HELCOM is the governing body of the Helsinki Convention[32]. There are 10 contracting parties, incl. the EU. HELCOM has a MoU with ICES[33], regarding, among other things, hosting of some data and research projects related to fish stocks and fisheries. Important relations to the EU legislations exist, in particular to MSFD and IMP legislations.

The HELCOM Data and Information Strategy[34] state that monitoring data should be openly accessible, but data confidentiality is assured in the HELCOM Convention[35] (in particular Art. 18), where data are received under specific conditions.

"The legal basis of OSPAR is the Convention[36] (1992) for the protection of the marine environment of the North-East Atlantic. OSPAR has fifteen members, includding the EU.

Annex V 4(1) of the OSPAR convention[37] states explicitly that "no programme or measure concerning a question relating to the management of fisheries shall be adopted."

OSPAR is bound to the MSFD legislation, but not to the DCF regulation and the data confidentiality is not guaranteed. OSPAR is a (potential) user of DCF data in relation to environmental protection.

The Barcelona and Bucharest Conventions do not deal with any DCF data. The Bucharest Convention provides only some ad hoc fisheries data to the Advisory Group on Fisheries. Data confidentiality is not an issue.

Eurostat

Eurostat is the statistical office of the EU. Its task is to provide the EU with statistics at European level. European statistics are governed by a set of EU legislative acts. Among the data collected by Eurostat are those related to fisheries (catches, landings and fleet), fish processing industry (within SBS), and aquaculture.

Eurostat can have access to primary data for statistical purposes, and guarantees their confidentiality.[38]

---

[32] *http://helcom.fi/Documents/About%20us/Convention%20and%20commitments/Helsinki%20Convention/1992_Convention_1108.pdf*
[33] *http://www.ices.dk/community/advisory-process/Documents/MoU%20ICES%20and%20HELCOM%201999.pdf*
[34] *http://helcom.fi/Documents/Action%20areas/Monitoring%20and%20assessment/Monitoring%20and%20assessment%20strategy/Monitoring%20and%20assessment%20strategy%20Att2.pdf*
[35] *http://helcom.fi/Documents/Abou t%20us/Convention%20and%20commitments/Helsinki%20Convention/1992_Convention_1108.pdf*
[36] *http://www.ospar.org/html_documents/ospar/html/ospar_convention_e_updated_text_2007.pdf*
[37] *http://www.ospar.org/html_documents/ospar/html/ospar_convention_e_updated_text_2007.pdf ; page 28*
[38] *Regulation (EC) No 223/2009 OF of 11 March 2009 on European statistics*

Strengthened collaboration between Eurostat and DG MARE regarding collection, storage and exchange of fisheres data is on the agenda of both institutions.

2.8.2. Budget and Staff of the main organizations

1. **JRC** is a Commission service. It's work on the DCF is funded by DG MARE under an administrative arrangement. JRC is using internal and external staff for database management. It has a dedicated budget for database development (130,000€ in 2010-2012). In 2012, JRC spent about 30,000€ for hosting infrastructure and 100,000€ for IT maintenance.
2. **Eurostat** is the statistical authority of the Union referred to in Article 6(1) of Regulation (EC) No 223/2009. It is a service of the Commission. Its activities have been most recently defined in the Commission Decision 20-12/504. Eurostat has a small unit dealing with fisheries data. EUROSTAT has its own budget (the total budget of the European Statistical Programme 2013-2018 is EUR 299.4 Million) and can in addition manage delegated funds from other DGs. It has limited staff for the management of IT tools (1 internal and 2 external staff related to fisheries). Support for the Eurobase is provided by in house staff, while development is outsourced.
3. **ICES** is financed through the contributions of its contracting parties. The EU contributes to the budget of ICES for recurring advisory deliverables with an annual budget of 1.4 million €. For non-recurring advisory deliverables, the EU reimburses the expenses which are directly connected with the execution of specific tasks. ICES has own staff for database development, but it also relies on the whole ICES community. The total costs for IT maintenance amounted to about 450,000€ in 2012. ICES Data Centre employs 11 people working with information systems/databases.
4. **GFCM** is one of the RFMOs, set up by FAO. It is funded by contributions of the participating countries, incl. the EU. Limited IT resources are available within GFCM, with one person dedicated to IT matters. No additional resources are available for the development of new databases and applications.
5. **Other RFMOs** have relatively little resources available in relation to IT development. Their primary task is to facilitate dialogue among their members.
6. **RSCs** do not dedicate any resources to IT development in relation to fisheries.

2.8.3. Conclusions

1. **DG MARE** has an important task in relation to the compilation of data under the CR.
2. **JRC** is a Commission Service and it has a proper legal basis to deal with DCF data. DG MARE can delegate tasks to JRC. Hence, in case the Commission should decide to give JRC a stronger role in data collection, this would be within the organisation's raison d'être. Funds and staff could be allocated to it. The geographical responsibility could cover the whole EU. Problems about data confidentiality do not exist. The data are provided to STECF WGs for further analysis.
3. **Eurostat** has a proper legal base for a strong role in fisheries data collection and the required provisions concerning confidentiality. Competences in database management are available, and no geographical restrictions exist. However, the main purpose of the organisation relates to statistics – which would largely cover catch and landing data, fleet economic data, fish processing and aquaculture data. To compile and store scientific biological data is not within the main purpose, scope and competences of Eurostat.
4. **ICES** has a conducive institutional setup for a bigger role in the storage and dissemination of DCF data. The MoU with EU could be extended, which would provide a proper legal base, and an adequate financial remuneration could be allocated. ICES guarantees data confidentiality.
5. **GFCM** does not seem to offer a proper basis for an extended role in DCF for a number of reasons:
    a. GFCM has been set up (as other RFMOs) under the framework of FAO constitution (Art. XIV). The decisions are taken by the participating countries. Unlike with ICES, EC could not sign a MoU with GFCM about provision of services related to DCF.
    b. GFCM does not have sufficient provisions regarding data confidentiality.
    c. Competences, staff and budget for database management are limited.
    d. The GFCM's geographic coverage may complicate further strengthening of implementation of the DCF in EU MBS MS, as non-EU GFCM members can be expected to face difficulties to meet DCF standards.
    e. The potential relevance of DCF for non-EU MBS countries seems limited as, even if they could collect the data, they often lack the scientific infrastructure to make use of them.
6. **Other RFMOs** and **RSCs** must be seen as users of data products derived from DCF (and CR) rather than potential hosts for greater amounts of DCF data. Proper arrangements must be made to provide the RFMOs the fisheries data required for their work. RSCs can be expected to make use of information compiled under MSFD, including the fisheries indicators.

3. OUTLINE OF SCENARIOS

3.1. *Descriptive dimensions and common features of the scenarios*

Objective b) of the present feasibility study consists in the development of alternative scenarios for the data storage and transmission set-up. The description of scenarios is a pre-requisite for their assessment, which is Objective c) of the study.

According to the proposal made in the approved Inception Report, the scenarios are described according to three main dimensions: geographic coverage, thematic specialization and functional scope. Each scenario is *in fine* defined by the roles of different institutions and the relations between them.

3.1.1. Geographic coverage

The geographic dimension of scenarios refers to the sea basins/sea areas/Member States on which data are to be stored and /or transmitted for each one of the institutions participating in each alternative set-up. In a way similar to the current set-up, some institutions play a role in data storage at the national, regional or supra-regional level. The level of detail for the geographic dimension may be different for different variables[39].

It has to be recalled that in all cases detailed data are collected by the MS and transmitted to regional and supra-regional institutions. The role of national institutions is not in the scope of this study.

An exclusively EU-wide, supra-regional set-up would require the databases to have a complete geographical coverage, as required by the DCF including all MS and sea regions/basins relevant for EU fisheries, including the non-EU waters where EU fisheries take place. The regional nodes scenario would preclude supra-regional databases in favour of storage and transmission set-ups organized at the regional level for all types of data (transversal, economic, biologic and environmental). A full network-based approach would give a strong role to the MS, with coordination at EU level. These options are elaborated below.

3.1.2. Thematic specialization

To simplify the description, variables and datasets are classified according to broad information blocks: transversal, economic (including aquaculture and processing industries data), biological (métier- and stock-related) and environmental (including physical parameters).

[39] *As a currently existing example of this situation, Eurostat collects from MS and disseminates EU-wide data on catches, but carries out a separate, more detailed data collection for the NAFO area as well.*

The nature and mandate of the intervening institutions determines the main thematic scope of their data collection and storage[40]. However, the importance of integrating economic, biological and environmental data for scientific analyses justifies that procedures are established to combine several thematic datasets, stored either by the same of by different institutions but with the capacity of data integration.

With respect to the thematic specialization, the current situation is somehow split between institutions that mostly cover economic (and transversal) variables, such as JRC and Eurostat, and those that deal with biological (and transversal) variables (ICES, RDB-FishFrame, GFCM). For the different environmental variables (including here the marine physical information), the EMODnet networks, as, well as the European Environment Agency, have a clear specialized thematic coverage.

The scenarios considered therefore include domain-integrated (or "generalist") data storage where all types of variables are collected/stored by the same institution (either at the regional or supra-regional levels), as well as no-overlap, highly specialized ones where different institutions would deal with different sets of variables, without integration of thematic data sets. Further clarification is given in the description of scenarios below.

3.1.3. Functional scope

Functional scope refers to the phases of the statistical processes that are covered, from collection of primary[41] data, to data validation and processing, data storage, data transmission and data dissemination.

The current situation (baseline scenario) is based on the **collection and storage** of detailed data at the national (MS) level, the **validation** of those data in general at the level of the same or a different national entity, the **storage** of data plus some validation at the regional or supra-regional level, and the dissemination at the national, regional and supra-regional levels depending on the type of data.

The transmission or **data upload** occurs directly from the MS to "intermediary bodies" (JRC, ICES) and to end-users (ICES or STECF WGs), as well as from "intermediary bodies" to end-users. In some cases, regional or supra-regional institutions could have direct **access to detailed data** at different levels of security for confidential data.

In the current and alternative scenarios, supra-regional and regional institutions participate, with different intensity, in the **quality control** of aggregated data uploaded from MS.

[40] *By its nature as the statistical office of the EU, Eurostat collects data on a large variety of domains. In a similar way, the JRC also collects data on many different subjects for scientific advice. With respect to fishery statistics, JRC collects mainly economic and transversal data, but also biological for the Mediterranean /Black Sea area.*

[41] *The term" primary data" refers to individual, non-anonymized data, usually only accessible to the institution legally entitled to collect it for administrative or statistical purpose, while "detailed data" refers to individual, anonymized data. These can be shared under specific confidentiality provisions, depending on the legal environment of the countries.*

Finally, at the **dissemination** stage, the various institutions make aggregate data within their geographic and thematic scope available to the public. As this may include "direct dissemination" or "indirect dissemination" via interfaces (such as a web page with links to appropriate URLs) that point to those institutions with the mandate for thematic and geographic dissemination, the scenarios will only describe the "direct dissemination". The thematic coverage of the disseminated data will depend of the thematic specialization of each institution, under the different scenarios.

3.1.4. Common features of the scenarios

3.1.4.1. Data domains and their treatment in the 4 scenarios

The DCF regulation distinguishes six data domains:

1. Biological stock-related data.
2. Biological métier-related data.
3. Fleet economic data.
4. Economic data on aquaculture.
5. Economic data on fish processing.
6. Environmental data.

A 7th domain, namely CR data, plays an important role for the determination of approaches to statistical sampling in domains 2 and 3.

Domains 1-3 deal with marine fisheries and integration of this data is desirable in order to facilitate bio-economic modelling and further improvements in scientific advice. Evidently, these three domains have each value of their own, in the context of stock assessment and economic analysis of fleet performance. The focus of the scenarios lies with these three domains and their relation to CR data.

The domains 4-6 (aquaculture, fish processing and environmental indicators) can be considered in relative separation because of their nature and their use:

- Three reports have been prepared by STECF on **aquaculture**. The aquaculture call generates annually a relatively small amount of data, which can be stored in a small (Access) database. The only link to marine fisheries data is through the 'market', but this link is very weak due to largely differing species. Consequently, it is proposed that, in the different scenarios, aquaculture data be treated and stored separately from domains 1-3. Continuation of the present procedures (JRC) seems more efficient than incorporation of this data in the total system. Data can be disseminated through STECF reports and Excel files on STECF / JRC website. The DCF data collection overlaps partly with the SRs data requirements regarding aquaculture production. It is recommended that the MS integrate the two processes at national level, assuring their mutual consistency and avoiding

duplications of data collection. The detailed data provided under DCF to JRC could then be aggregated (if necessary) and provided to Eurostat. This procedure would eliminate double reporting and could ensure overall consistency. In order to meet also the SRs requirements, the DCF regulation has to encompass all aquaculture, incl. inland fresh water. This will be indeed an obligation under the new DCF programme.

In relation to aquaculture the following 'phenomenon' should be highlighted. The EU intends to stimulate aquaculture production in the future. However, the surveys based on DCF legislation will generate only data on established aquaculture production because farming of new species is done by a low number of enterprises, so that the data remains confidential (on national as well as EU level). It should be evaluated whether the economic performance of aquaculture fluctuates so much that annual monitoring is a necessity for implementation of EU policies. On the other hand is seems relevant to monitor closely creation of new activities, which can be done on the basis of publicly available information (e.g. press releases and articles about new aquaculture farms), although details of economic performance may not be available and are certainly confidential.

- In many MS, economic data on **fish processing** relies either fully or at least partially on data compiled by the NSIs under the Structural Business Survey (and transmitted to Eurostat)[42]. At the EU level there is a clear overlap between the SBS and DCF with respect to the fish processing statistics requirements. Although DCF generates several indicators which are not collected under SBS (as presented in Appendix XII of 93/2010 Decision this concerns unpaid labour, subsidies, net financial costs, net extraordinary costs and debt), it can be questioned whether the value added of these indicators justifies the extra effort. It is recommended to concentrate the data collection of fish processing under Eurostat, for reasons of efficiency. Eurostat could generate annually a special 'file' with fish processing data which could be disseminated through STECF/JRC website. This would provide even better regularity of data provision, as DCF foresees only a bi-annual survey.

- There has been no call by JRC for the DCF **environmental** indicators as such, but a call by ICES/Helcom for VMS data[43] was launched in November 2013. The call facilitates calculation of the environmental indicators 5-7. These indicators are derived from data on distribution of fishing effort compiled with VMS. All environmental indicators are derived from other data collected under DCF or CR / VMS. Environmental indicators are being developed under MSFD in much greater detail, making the added value of the DCF indicators unclear. Maintenance of a specific database with these indicators seems therefore superfluous. They are rather considered as an analytical product from primary

[42] *See findings of the project "Assistance for the monitoring of the implementation of national programmes for the collection, management and use of data in the fisheries sector" (MARE/2009/08).*
[43] *ICES / Helcom, Data call for VMS data for fishing activities in the Baltic in support of ICES/HELCOM advice on the spatial distribution and impact of fisheries. 23 November 2013.*

(or aggregated) DCF data. ICES is certainly the most appropriate body to develop and determine environmental indicators for the Atlantic areas. It already collaborates with DG ENV and various RSCs like Ospar and Helcom on this issue. It is recommended that calculation of environmental indicators for the Mediterranean and Black Sea could be the responsibility of JRC or one of the STECF WGs, as an additional data product of the MBS data call.

3.1.4.2. Primary data collection by Member States

It is hard to imagine a scenario where supra-regional or regional institutions directly collect the data from economic agents and other primary sources in the MS.

In all considered scenarios, the collection of primary data remains the responsibility of the MS, in application of the subsidiarity principle for European statistics. At the national level, a network of institutions is responsible for the collection of data from scientific measurements of fish catches (surveys at sea, measurements in laboratories, on-shore surveys of landed fish) and data from the fleet, the fish processing industry and the aquaculture farms. Data collected to comply with the CR ("Control data") as well as with other statistical regulations (such as the Structural Business Regulation[44] or the different Aquaculture regulations[45]) are collected by MS as well.

Specific bilateral and multilateral arrangements between MS to cooperate on data collection exist. Surveys at sea are coordinated by RCMs. Bilateral agreements between MS have been set up to increase efficiency (and reduce costs) of collecting métier or stock related data.

A major issue for data quality when data are collected by the MS is the harmonization of statistical results, which is ensured in general by either using "ex-ante" common methodologies (comparable target populations, similar questionnaires, common definitions of variables, classifications and nomenclatures, etc.) or "ex-post" adjustments[46] and quality assessment of data by the regional and supra-regional institutions.

The coordination of data collection **within MS** is outside the scope of this study. However, in terms of data transmission to supra-national bodies, it is necessary at the national level to map the full set-up to avoid double reporting by different actors within MS, e.g. NSI may report catches to Eurostat, fisheries department may report to DG MARE and a scientific institutes to ICES.

[44] *Regulation (EC) No 295/2008 of the European Parliament and of the Council of 11 March 2008 concerning structural business statistics (recast).*

[45] *Regulations 788/96 and 762/2008 on the submission by Member States of statistics on aquaculture production*

[46] *Such as the production of EU aggregates, or estimates produced by the regional and supra-regional institutions.*

3.1.4.3. *Integration of biologic and economic data*

Integration of biological and economic data requires a matrix relating stocks and fleets in terms of catches and fishing effort. Both biological and economic primary data must allow aggregation to a common set of dimensions (e.g. at very detailed level time=month; space=ICES rectangle). However, the aggregated data submitted under various data calls are aggregated in a way which does not allow creation of direct link. The following issues have to be considered (details are elaborated in Annex 2.6.).

It has to be decided by experts (and end users) at which aggregation level the integration should take place. The available data has to be (dis)aggregated to that level. In principle this is a conceptual / analytical issue, not an IT / database issue.

At present the transversal data (landings and effort) collected under the fleet economic call entails a number of specific aggregations:

- Time dimension is one year. This means that intra-annual (monthly or seasonal) changes are not reflected. This can be a relevant item in relation to modelling closed seasons, but also in relation to seasonality's and characteristics of fish caught (e.g. spawning).
- Catches are expressed in terms of species by FAO region 3 and 4. Translation to stocks is required. However, stocks which are defined in relatively small areas and the fleets that exploit various areas during one year (e.g. nephrops) may pose special problems.
- Allocation of effort and landings to segments, defined on the basis of 'dominant gear' may lead to significant errors (see Annex 2.6.).

The above 'problems' cannot be resolved satisfactorily with the data collected under DCF. However, the national data compiled under the CR contains all necessary details. Therefore in relation to integration of economic and biological data, all scenarios assume that the required CR data will be made available for DCF purposes and will be submitted along with other data.

3.1.4.4. *Data confidentiality*

In relation to data confidentiality it is necessary to distinguish three distinct areas, which may be relevant to any future scenario:

- Protection of personal data
- Protection of data on a statistical unit ("statistical disclosure control")
- Protection of Intellectual Property Rights (IPRs)

These three issues are discussed below.

Personal data

Directive 95/46/EC is the central piece of legislation on the protection of personal data in Europe. The Directive stipulates general rules on the lawfulness of personal data processing and rights of the people whose data are processed ('data subjects'). The Directive also provides that at least one independent supervisory authority in each Member State shall be responsible for monitoring its implementation.

Two years later, a Directive on privacy and electronic communications was adopted. Updated in 2002 as Directive 2002/58/EC, it regulates areas which were not sufficiently covered by Directive 95/46/EC, such as confidentiality, billing and traffic data, rules on spam, etc.

These two directives created a general and technology neutral system of data protection in all EU Member States. However, protection on the level of the European institutions and bodies was not guaranteed. To remedy this, Article 286 of the EC Treaty was adopted.

Article 286 of the EC Treaty stipulates that the European institutions and bodies shall protect personal data and provides for the establishment of an independent supervisory authority. It was implemented in Regulation (EC) No 45/2001.

Combining the relevant features of Directives 95/46/EC and 2002/58/EC, Regulation (EC) No 45/2001 regroups the rights of the data subjects and the obligations of those responsible for the processing into one legal instrument. It also establishes the EDPS (European Data Protection Supervisor) as an independent supervisory authority with the responsibility to monitor the processing of personal data by the Community institutions and bodies (see also Decision 1247/2002).

According to Article 2 (a) of Regulation (EC) No 45/2001: " 'personal data' shall mean any a information relating to an identified or identifiable natural person, referred to as "data subject" - an identifiable person is someone who can be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his or her physical, physiological, mental, economic, cultural or social identity".

Personal data which is processed in relation to the work of the data subject remain personal/individual in the sense that they continue to be protected by the relevant data protection legislation, which strives to protect the privacy and integrity of natural persons. As a consequence, data protection legislation does not address the situation of legal persons (apart from the exceptional cases where information on a legal person also relates to a physical person).

In its Communication on "A comprehensive approach on personal data protection in the European Union" (COM(2010)609 final), the Commission concluded that the EU needs a more comprehensive and coherent policy on the fundamental right to personal data protection, in particular due to fast development of IT and big data processing. This has led to the Commission

proposal "On the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation, COM(2012) 11 final). The new legislation should come into force in 2015.

Statistical confidentiality

Regulation 223/2009 sets rules on statistical confidentiality which is defined in art. 2e as:

the protection of confidential data related to single statistical units which are obtained directly for statistical purposes or indirectly from administrative or other sources and implying the prohibition of use for non-statistical purposes of the data obtained and of their unlawful disclosure.

The preambles 25-27 are also of particular importance in relation to the future DCF and use of confidential data:

- "25. The availability of confidential data for the needs of the ESS is of particular importance in order to maximise the benefits of the data with the aim of increasing the quality of European statistics and to ensure a flexible response to the newly emerging Community statistical needs.
- 26. The research community should enjoy wider access to confidential data used for the development, production and dissemination of European statistics, for analysis in the interest of scientific progress in Europe. Access to confidential data by researchers for scientific purposes should therefore be improved without compromising the high level of protection that confidential statistical data require.
- 27. The use of confidential data for purposes that are not exclusively statistical, such as administrative, legal or tax purposes, or for the verification against the statistical units should be strictly prohibited."

The Implementing regulation 557/2013 states a number of conditions for access to confidential data (art .3):

The Commission (Eurostat) may grant access to confidential data for scientific purposes held by it for the development, production or dissemination of European statistics as referred to in Article 1 of Regulation (EC) No 223/2009, provided that the following conditions are satisfied:

a) access is requested by a recognised research entity;
b) an appropriate research proposal has been submitted;
c) the requested type of confidential data for scientific purposes has been indicated;
d) access is provided either by the Commission (Eurostat) or by another access facility accredited by the Commission (Eurostat);
e) the relevant national statistical authority which provided the data has given its approval.

In addition Commission Decision 2004/452, last amended by Commission Decisions 2012/200 specifies a list of institutions which may get access to the confidential data. It is noted that none of the research institutions presently dealing with the DCF data is on the list.

In practice, Eurostat and the MS have agreed technical rules to protect the disclosure of statistical aggregates. Examples of protecting rules are:

- No dissemination of aggregates corresponding to less than 3 statistical units (by merging this segment with another one)
- No dissemination of aggregates corresponding to less than 5 statistical units where the highest value is above 70% of the total aggregate (so called "dominance rule").

Similar rules should be discussed for vessel segments, aquaculture and fish processing sub-sectors. It has to be noted that software has been developed for this and is used by several NSIs (e.g. Tau-Argus and Mu-Argus developed by Statistics Netherlands).

Intellectual property rights

Data and databases may be subject to protection under the legislation on Intellectual Property Rights (IPR) as set in the Directive 96/9 on protection of databases. In this respect distinction is made between copy right and database right. The arrangement, selection, and presentation of the data may be protected by copyright, while the database as a whole can be protected by database right.

Conclusions – applications to DCF

All references to confidentiality in this report refer to statistical units, in the sense of the Statistical legislation. However, many small fishing and aquaculture firms are operated by a natural person and not a legal person (company). This creates a grey area, in which distinction between personal data and data on statistical units is not unambiguous (e.g. income of a vessel is equal to personal income).

The EU Statistical legislation (EU regulation 223/2009 and also Commission Regulation 831/2002 concerning access to confidential data for scientific purposes) presents a well-developed system to ensure statistical confidentiality, via legal and organizational provisions (release of anonymized micro-data sets and access to secure premises). It shows in detail which provisions may have to be included also in the DCF legislation in order to protect primary data. The existing practice of developing statistical disclosure control rules and software to apply them may be considered in the DCF-related domains.

Considering that DCF is largely funded by public resources, it is unclear to which extent IPRs are applicable. However, it is certain that the involved research institutes would like to maintain a preferential position. Pursuing broader dissemination and use of DCF data is in the line with the

Eurostat policy of promotion of free access and re-use of data collected under EC programmes and funding.

There are two important obstacles to storage of personal data (economic and transversal) – political will and legislation in relation to data protection / confidentiality. In order to address the second obstacle, close dialogue with the European Data Protection Supervisor, as well as the use of statistical disclosure control techniques, is a must. However, individual (confidential) data are already being stored at supra-national level at Eurostat and DG Agri (RICA), which demonstrates that sufficient confidentiality safeguards have been put in place there and can be also applied to fisheries data.

Accessibility of data is at present restricted due to lack of clarity about the ownership. Who is the owner of the data – the research institutes which collect, the national ministries which provide co-financing or the EC, which provides the EU DCF funding. In addition the DCF legislation does not refer to dissemination. These two issues should be resolved under any future arrangement in order to meet the needs of various end-users.

3.1.4.5. Other common features

- Organization of the work flow is the central theme. IT infrastructure is not mandatorily a copy of the workflow. This means that hardware, software and the organisation of the various tasks (from data collection, through storage and processing to dissemination) can be mutually independent. The hardware does not have to be located within the organization responsible for collection or processing of the data. Different software (SQL, R, Access) can operate on one computer, within ond DB, and be applied for different tasks by different users. Collection and processing of data can by different organizational units. The data collection is done by national institutes, but all subsequent data processes may be done by specialized units at supra-national level and the DBs may be located outside the national institutes. For example:
  - Primary data is uploaded to a "cloud DB", composed of a set of national DBs. The "cloud DB" may be located in "A", e.g. JRC.
  - Quality control (using standard routines) is done by a central unit of the "cloud DB". Quality control unit may be located in "B", e.g. ICES for biological data.
  - Development of routines for data processing is done by specialised WGs (stocks, economics, etc.) located in "C" and "D", e.g. each WG member in his own institute.
- Physical location of the data (bases) is not relevant. Various stakeholders can be responsible and have exclusive access to parts of the data, located at any place / institution. One system can also consist of several DBs in different locations, linked through internet.
- Tasks and responsibilities of actors at national and supra-national level must be logically formulated, avoiding duplications.

- Ownership of the data has to be determined and the consequences incorporated in the system.
- Access rights of different users must be clearly defined.
- Transparency is pursued by detailed documentation of all workflows and contents.
- Internationally recognized code lists are used in all elements of the system.
- Transfer of experience among MS is pursued, including IT tools, statistical methods, quality procedures, etc.
- Ad hoc data calls by various scientific working groups cannot be avoided and may require development of specific procedures.
- In the 2014-2020 financing period, the funding of data collection is brought under EMFF. The MS will receive a contribution of 80% of eligible costs, compared to 50% under the previous programme.

3.1.5. Summary presentation of the scenarios

The combinations of all possible levels for the three dimensions would generate a very large number of scenarios. However, given the current situation and the need for a smooth transition, as well as the experience of existing systems for other statistical domains, it is proposed to restrict the analysis to four main scenarios similar to those proposed in the Terms of Reference and described in the table 7.

For ease of mention, they are referred to as "supra-regional model", "regional nodes model", "network-based model" and "combination model".

- The **supra-regional model** (also called "Eurostat model") refers to a set-up where one (or few) large, integrated database system(s) stores detailed data from the MS and provides aggregate data to end-user on a wide number of thematic issues, allowing for easy combination of data sets for economic-biological-environmental analysis.

- The **regional nodes model** (also called "RDB-FishFrame model") has a strong geographic division of work, where a number of regional institutions collect data on a variety of thematic areas. MS upload their data to all regional institutions to which they belong.

- The **network-based model** (also called "EMODnet model"), which encompasses a larger number of institutional, network-type arrangements, allows end-users to access data from different databases maintained at different geographical levels. There is a thematic specialization of networks.

- The **fisheries data hub** (also called "Combination model") proposes to integrate primary data in three thematically specialized international databases. The databases would be operated by institutions presently active in DCF and CR and mutually linked.

The four descriptions above are very summarized and, of course, there are thematic and functional adaptations to take into account their feasibility. Details are given in next sections for each of the scenarios.

Table 7. Summary description of scenarios[47]

Geographic coverage by topic

Name of scenario	Present situation (Baseline)	Scenario 1 Supra-regional DB (EUROSTAT model)	Scenario 2 Regional nodes (RDB-Fishframe model)	Scenario 3 Network based (EMOD-net model)	Scenario 4 Fisheries data hub (Combination model)
Summary description of the scenario	*The present situation is described in detail in the Interim Report*	One integrated EU-wide DB-system, drawing detailed data from the MS and disseminating to end-users.	Regional databases sourced from MS and disseminating independently to end-users.	All data maintained at MS level. End-users have access to data through an EU-interface.	Several thematically specialized DBs with primary biological, economic, fisheries data. DBs are linked.
- Biological data / related fisheries data	3 regional DB under ICES + LP DB is foreseen + MBS RDB foreseen under GFCM (currently collected by JRC)	All DCF data regarding marine fisheries (biological, economic and transversal) is in one database operated by an **institution X.**	5 RDBs: (A) Baltic Sea, (B) N. Sea and NE Atlantic, (C) North Atlantic, (D) Distant Waters and (E) MBS, ooperated by three **institutions.** Each RDB contains all relevant data. Some MS report to several RDBs. Each stock and fleet segment is allocated to one RDB. The RDBs are based on common concepts and definitions.	All data is maintained in national DBs designed to allow for access to their disseminated data through an "EMODnet-like" portal.	All primary biological data in one DB, with regional flags. Operated by ICES.
- Fleet economic data / related fisheries data.	JRC maintains EU-wide DBs				All primary fleet economic data in one DB, operated by JRC.
- Fisheries data	Transversal data match biological and economic variables in the DBs.				All primary fisheries data in one DB. Operated by DG Mare.

[47] *Common features on aquaculture and fish processing are not repeated in the table.*

Thematic specialization

Name of scenario	Present situation (baseline)	Scenario 1 Supra-regional	Scenario 2 Regional nodes	Scenario 3 Network based	Scenario 4 Fisheries hub
- Biological data / transversal data	ICES (biological data for Atlantic areas) JRC (EU effort regimes and MBS biological data)	Biological, economic, transversal are all in one DB.	Biological, economic and aggregated transversal data in specialized RDBs.	Economic and biological data sets in national DBs.	Biological data at ICES.
- Economic data / transversal data	JRC, fleet aggregates at Eurostat				Fleet economic DB at JRC.
- Fisheries data (CR)	DG MARE	**X** has access to detailed data or low aggregation level. X can guarantee confidentiality	No access to detailed CR data, or access granted to all RDBs under technological and legal arrangements for confidentiality.	No access to detailed CR data or access granted to all institutions (probably requiring high technological investments to ensure confidentiality). Confidentiality cannot be otherwise guaranteed	CR DB at DG Mare

Functional scope

Name of scenario	Present situation (baseline)	Scenario 1 Supra-regional	Scenario 2 Regional nodes	Scenario 3 Network based	Scenario 4 Fisheries hub
- Collection	MS	MS	MS	MS	MS
- Storage of primary data	MS	MS	MS	MS	MS and International DBs.
- Access to primary data	Biological: MS and ICES Economic: MS	Biological: MS and Supra regional DB Economic: MS	Biological: MS and RDBs Economic: MS	Biological: MS and users Economic: MS	Biological: MS and Hub Economic: MS and Hub
- Access to detailed data	Biological: MS and ICES Economic: MS	**X** has online access to detailed data.	Limited by aggregation level required in the RDBs.	Limited by aggregation rules of the interface.	Not relevant
- Data upload	According to data calls	**X** draws data from MS DBs at given moments.	Data calls or via command from the RDBs.	No upload needed, real time accessibility as soon as data are made accessible by the MS-	Continuous up-dating or agreed schedule.
- Quality control	Mostly at MS level. STECF validates JRC calls. ICES validates RCM calls.	MS responsible for primary data All other checks in standard procedures of **X**.	Mostly at MS level. Some at RDB level. Coordination through RCMs.	All at MS level according EU wide guidelines.	Common routines, developed at central level.
- Confidentiality	Assured by MS.	MS and **X** assure confidentiality of detailed and primary data	Assured at MS level and by RDBs.	Assured at MS level.	Assured at MS and Hub level.
- Dissemination [48]	ICES and JRC – various levels of aggregations	**X** disseminates at supra-national level.	Through the RDBs.	Through the interface.	Web-based tools.
- Institutional requirements		**X** needs status similar to Eurostat (i.r.t. quality checks and confidentiality).	RDBs make arrangements with the MS involved.	Access rights to national DBs are defined for various types of users. Possibly not the same access level in the different DBs.	Comparable to Eurostat.
- Development and maintenance of DB systems	Each MS and international DB is responsible for its DB maintenance and development	**X** develops the supra-national system. MS develop systems consistent with **X**.	Some definitions and procedures are developed at EU level and others at regional level, according to need. MS coordinate with RDBs	Common definitions, quality check procedures, etc. are developed at EU level. National DBs must remain compatible with the new developments of the interface.	Development and maintenance at supra-national level.

[48] *MS may disseminate data to their national stakeholders.*

3.2. Scenario 1. Supra-regional database ("Eurostat model")

3.2.1. Summary and conclusions

In the supra-regional scenario, one main database (except those at the national level) contains data for all EU MS and all the waters where EU fleets are operating, in a similar way as the current set-up for Eurostat or JRC data. This would be particularly the case for biological and fleet economic data that would be stored in a supra-regional DB covering all EU MS and marine areas. The data should be stored at the greatest possible level of detail to allow for a broad variety of aggregations.

The institution hosting the biological (stock and métier-related), transversal and economic database would then provide RFMOs, international organizations (such as FAO) and other end-users (including scientific WGs) with the necessary data for their analyses, including selections at particular regional levels.

On the functional side, the institution would be able to run validation tests on data sent by the MS, especially macro-editing validation procedures (i.e. not applied to individual records).

Dissemination from the supra-national DB can be done in various formats, such as predefined sets of tables, tailor-made tables, re-usable datasets (in open, non-proprietary formats). Aggregated data transmission between the centralizing institution and RFMOs and other institutional end-users would be established by data exchange protocols, substituting the direct transmission form MS or the EC to the latter.

This scenario is based on 3 databases:

- A supra-regional DB on marine containing biological, transversal and fleet economic information, allowing for integration of these data domains. It would be hosted by an institution with expertise, resources and endowed with a legal framework that ensures its access to and protection of confidential data.
- A small DB on aquaculture as presented in Section "Common features".
- A DB on the fish processing industry, based on data generated by Eurostat's SBS.

The supra-regional DB would not be created ex nihilo, but be based mainly on the existing RDB-FishFrame. Data validation tools now available at STECF and ICES would guarantee data quality.

The major advantages of this model are the following:

- Data upload from MS is simplified as the overlaps of submitting transversal data by the MS to ICES, JRC, Eurostat and others would be eliminated. The decrease in response burden implies reduction of staff costs in the MS. All data needs of users (ICES and STECF WGs) would be covered by the "supra-regional DB".

- Data integration for biological, economic and fleet data is made easier, and users can find them in a single entry point.
- The risk of data incoherence is minimised by standardising validation rules and quality reporting, thus increasing the quality of analysis as well as users' trust in the institutions providing data.
- There is no need of developing regional databases, as regional data can be generated from the "supra-regional DB".

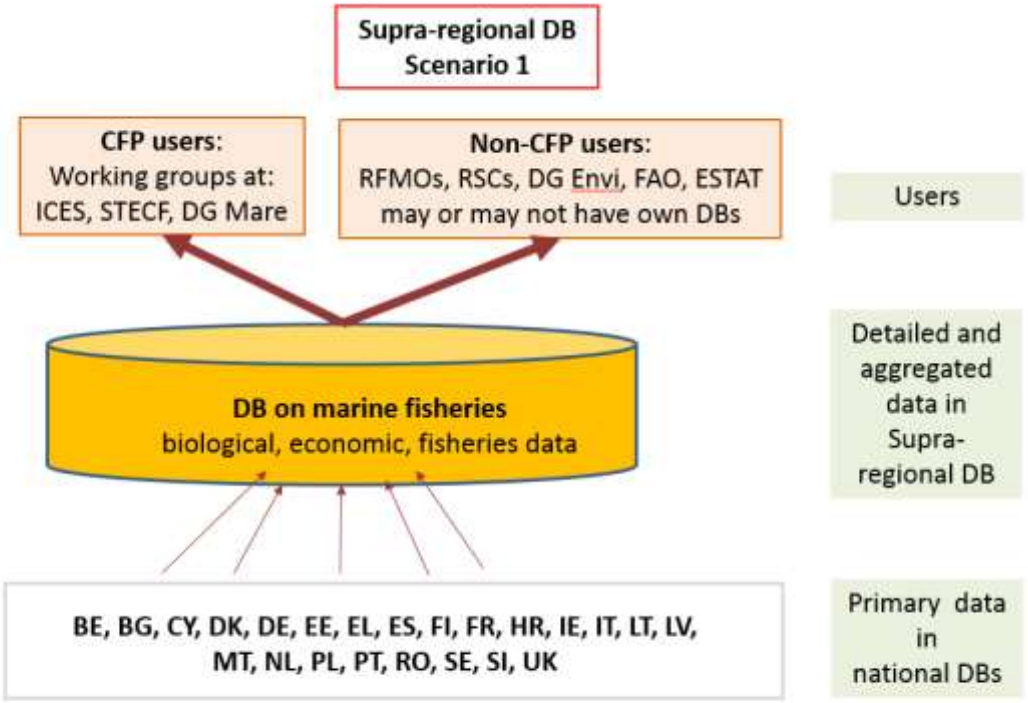
The disadvantages are:

- Development is needed, for the "supra-regional DB", and some of the new data transmission procedures.
- Data calls by JRC and RDB-FishFrame must make arrangements to streamline requests of transversal data (JRC Effort and Economic data calls and RDB-FishFrame métier data call).
- As for the fish processing industry, it is proposed to abolish the data collection under DCF and rely on the data collected under Eurostat's SBS. This implies that some indicators, which should be collected under DCF and but are not part of SBS, would not be available anymore.
- Regional specificities may not be tackled by the "supra-regional DB". In the MBS region, non-EU states may need specific tools for data upload and access which would require development and appropriate assistance. Special issues/problems may also arise in cooperation with other non-EU countries in the Atlantic area (e.g. Norway, Faeroe Islands) as well as under the various RFMOs.

More detailed SWOT analysis is presented in table 8.

The scenario is schematically presented in figure 3.

Figure 3. Scenario 1: "Supra-regional database".



Note: The scenario deals with data on marine fisheries. Data on aquaculture (marine and inland) and fish processing are not part of this data flow (see section 3.1.4.1). Therefore the land-locked EU MS are not mentioned in the diagramme.

Table 8. SWOT Analysis

	Strengths	Weaknesses	Opportunities	Threats
Data storage – production DB	Data are stored in one database. The duplication in stored data is eliminated		Experience of Eurostat, ICES and JRC in designing and maintaining large databases	"Business as usual" in JRC, Eurostat and ICES may imply resistance to merging databases
Data access – distribution DB	Users have to access a smaller number of databases. Integration of data is facilitated by the source	Specificities of RFMOs' information needs may not be tackled.	Needs of specific users (e.g. RFMOs) can be serviced by "supra-regional DB", avoiding data collection from MS	Regional bodies may have the need of developing further RDBs for specific needs.
Data up-load	MS reduce their burden in submitting data to different institutions' calls	MS have routine procedures to respond to current data calls that may have to be changed	MS should support the simplification of data upload procedures	
Quality control	Existing validation procedures in ICES, JRC and Eurostat are merged. Procedures at MS level are standardized for quality control and reporting.		MS should support the coherence of validation rules by supra-national institutions	
Dissemination	Existing practices in ICES, JRC and Eurostat may be re-used	Discontinuation of dissemination products as a result of merger of databases and "loss" by some institutions		
Bio-economic integration	Biological and economic variables are stored in the same database	Breakdowns of variables may have to be revised for integration	Specialized users (STECF, ICES WGs) will have access to integrated data	Regional analysis may suffer with the replacement of RDBs. Lack of data outside EU in the MBS region
Coherence with IFDMP	Consistency with IFDMP can be achieved.	MS with strong IT systems may have to make necessary adaptations.	Common IFDMP rules for data transmission and use of Master Data Register will allow common future development.	
Legal constraints	Eurostat's legal basis for protection of confidential data on aquaculture and fish processing industry	The institution hosting the "supra-regional DB" should be subject to statistical confidentiality	The European Statistical Programme comprises statistical domains partly managed by other DGs[49] (e.g. agriculture statistics)	Transmission of confidential data to institutions other than Eurostat may be rejected by MS unless confidentiality is legally assured.

[49] *The Regulation (EC) No 223/2009 of the European Parliament and of the Council of 11 March 2009 on European statistics constitutes the legal basis for the preparation of the European statistical programme, providing the framework for the development, production and dissemination of European statistics, the main fields and the objectives of the actions envisaged for a period not exceeding five years.*

	Strengths	Weaknesses	Opportunities	Threats
Administrative constraints		Need for administrative arrangements concerning DG MARE, Eurostat, JRC and ICES		
Financial constraints	Reduction of cost as response burden by MS	Development of an integrated database required		General reduction of budgets in international institutions

Table 9. Evaluation against policy objectives

Objective	Evaluation
Informing CFP / IMP/ MSFD	Higher level of data integration may respond easier to information needs of the EU policies.
Regionalization / coordination	Supra-regional databases can disseminate regional selections as it is currently done for ICES RDBs. However, regionalization of economic variables may not be straightforward (namely for fleet operating in different regions). Regions may have different information needs, which are not specifically tackled at the supra-regional level. MBS region has non-EU states which may require specific tools for reporting and data access.
Simplification	Fewer databases imply less management and update cost.
Costs reduction	Costs of IT development and maintenance decreased expect for development of new modules of RDB-FishFrame. Duplication of data upload in response of calls by different organizations is reduced. However, management costs at the institution hosting the supra-regional database may increase (more responsibilities, subject-matter skills required)
Increased quality	One "single source of truth" at the supra-regional reduces the possibilities of incoherence due to different estimation methods, definitions, reference periods, etc. Standardization of code lists facilitates data coherence and integration. Same criteria for data validation. Standard quality reports facilitate users' evaluation of data.
Integration	Biological and economic data are integrated in one database, allowing for better bio-economic analysis. There is a need for detailed methodological work to find common level of aggregation and adaptation of legal requirements (either in "Statistical" or "DCF" legislation (but they are both EU Regulations).
Accessibility	Users need access to a lower number of databases. Data are stored "far from the source" but users can access data from different MS/regions at the same point.
Coherence with IFDMP	The supra-national database should be fed by using standard communication protocols (FLUX). Fish processing data are fully embedded into SBS data collection, for which MS already transmit data via eDamis (SDMX).

3.2.2. Data collection at national level

In the supra-regional scenario (as it will be the case for the other ones), data collection is the responsibility of MS, in application of the subsidiarity principle also applied in the European Statistical System (ESS). In the "Eurostat model", harmonization is mostly achieved by the requirements of statistical regulations, which leave to the MS the decisions on the data collection, but oblige them to submit to Eurostat aggregate results based on the same statistical units, concepts, definitions and classifications, as well as to ensure that the estimates have comparable levels of precision (which have to be reported in addition to the estimates). The "Eurostat model"

has been running for a number of years and is recognised one of the most performing international statistical systems in terms of availability and comparability of the data.

However, the inefficiencies that have been detected for this system, which are described in Joint ESS strategy that supports the so-called Eurostat' Vision for the next decade"[50], are also valid for the fisheries statistics domain: the data collection for different domains is kept in separate "stovepipes" without the easy integration that could be achieved by record-linking, and MS carry out similar statistical processes but don't share IT tools as much as it could be done. All this implies high costs and rigidities to react to rapidly evolving users' needs.

The supra-regional model would benefit of increased standardization of procedures at the MS level, especially with respect to data comparability across MS. Close follow-up of modernization programmes in the ESS (such as the MEETS programme for modernizing enterprise and trade statistics, of the different ESSnet projects[51]) would benefit EU fisheries statistics.

3.2.3. Data storage and access at supra-national level

Overview

As mentioned above, this scenario is based on 3 databases (see table 10)[52]:

- "Supra-regional DB": an EU-wide database including biological, transversal and fleet economic information, covering all MS and relevant marine areas. This implies including the information currently stored in RDB-FishFrame, (with the extension to Mediterranean/Black Sea and fisheries in non-EU waters) plus that contained on the databases *dc-economic*, *dc-effort* and *dc-med* of JRC. In addition, the database should be complemented with link to fleet register and control data. Environmental data can be stored in a specific module.
- Aquaculture database: an EU-wide database on aquaculture including fresh water farms to fulfil requirements currently addressed by the Eurostat compilation.
- Fish processing database: an EU-wide database on fish processing industry fed by the SBS data collection.

[50] *COMMUNICATION FROM THE COMMISSION TO THE EUROPEAN PARLIAMENT AND THE COUNCIL on the production method of EU statistics: a vision for the next decade, COM(2009) 404 final, Brussels, 10.8.2009, and the strategy validated by the European Statistical System Committee at its meeting of 20th May 2010 (document ESSC 2010/05/6/EN).*

[5151] *References to the different projects can be found in the CROS Portal www.cros-portal.eu on modernization of European Statistics.*

[52] *It is noted that with the present technology the physical location of a database is not relevant. It is even conceivable that various parts of one DB are located in different places.*

Table 10. Databases and their domains

Name of the database	Domain covered by the database	Production / dissemination DB[53]
Supra-regional DB	- Biological (stock- and métier-related) - Transversal data - Fleet (fleet register, economic data) - Environmental data	Production and dissemination
Aquaculture database	- Aquaculture	One Production DB. Dissemination through an Excel file provided at JRC-DCF website, similar practice as with current dissemination of fleet economic data. The Excel file contains national data at an agreed level of aggregation.
Fish processing database	- Fish processing industry	Extract from SBS can be provided in Excel file at JRC-DCF website. Approach is identical to aquaculture data.

Connection between national DBs and the supra-regional DB

Data connection procedures currently used with RDB-FishFrame will be applied for the "supra-regional DB".

Physical connection of national databases is not necessary, as the supra-regional DB is fed by uploads from MS following standardized formats.

Database structure

With respect to the integration of biological, transversal and economic data, the RDB-FishFrame would need to be expanded to accommodate transversal (capacity and effort) and fleet economic data.

This would imply a detailed analysis of the structure of the current Saiku database of JRC and of "Control Regulation" DB at DG-MARE which would be integrated/replaced by the supra-regional database.

It is important to enlarge all databases to include metadata. Standardized formats for metadata reporting have to be developed for all variables, using the models currently existing (e.g. in Eurostat).

[53] *In most DB systems a distinction is made between production DB, where data is received and processed (incl. quality checks) and dissemination DB which offers final data at different levels of aggregation to different users.*

The current database storing SBS data at Eurostat (for fish processing) would be kept. It is fed by MS by using SDMX standards via the eDamis system.

Aggregation level (variables and dimensions)

"Supra-regional DB":

The "Supra-regional DB" should allow the integration of biological and economic data. This requires that data are stored at a level of detail which is higher than the current one. Ideally, data are stored at the primary or detailed level, with the necessary legal and technical arrangements for the protection of confidential data. To satisfy the information requirements from RFMOs, data have to be provided according to the spatial definitions of the various RFMOs. Biological data are already geographically defined. As for economic data, the fleets operating in several regions will have to be defined and their costs and revenues may have to be allocated to the specific regions, e.g. on the basis of catches, prices and costs per unit of effort. This methodological work could be developed by the STECF in collaboration with Eurostat which has the practice of geographically distributing economic variables in different data collections (e.g. at the infra-national level or international levels).

Catch data should be required at the level of detail specified in the DCF (level 5) instead of that required by SRs (level 4).

Further harmonization of size classes and gear categories will be required. The DCF regulation specifies size classes, but the JRC effort call uses different classes (with a 15m limit) in order to maintain certain time series. It will have to be considered whether and how this data call could be brought in line with the DCF requirements (or else the future regulation should make an appropriate provision in this respect). SRs use several GT-groups, while DCF uses lengths groups.

Various data calls use also their own specific aggregations of gears, which cannot be always easily related (e.g. it is not clear how the economic data on fleets using 'polyvalent' gears could be linked to the biological métier-related data. It seems desirable to use the FAO gear classification (ISSFSG), as presented on the web page of the Master Data Register, set-up for the purposes of the Control Regulation.

The details regarding harmonization of vessel sizes and gears in various DBs are elaborated in Annex 2.6.

For the aquaculture and fish processing databases, the aggregation level can be the same as the current one. The details on production modes for aquaculture used by Eurostat and JRC have to be reviewed in order to check if bringing them to a common standard would still satisfy users' needs.

Data processing and estimation

Data received from the MS would be treated differently depending on the level of detail:

- Primary biological data would be considered validated by the MS. It would be raised to national and stock level using procedures developed by ICES and integrated in the supra-regional DB.
- Aggregated data (transversal and fleet economic) could be considered validated by the MS but are further validated by developing rules that consider the integration of different datasets. Validation rules applied at present by various institutions are kept and further developed. These rules regard various aspects of quality, e.g. statistical characteristics, arithmetic checks, consistency, completeness, etc.

For fleet data, aquaculture and fish processing data, the current validation rules applied by Eurostat would be kept and combined with those applied by JRC.

All validation and estimation procedures should be documented and disseminated in standardized quality reports. Dissemination databases should allow flagging statistical results with references to quality metadata (as it currently happens in the Eurobase system of Eurostat, which indicates aspects such as provisional figures, etc.).

Interactions with other fisheries-related data

Control Regulation data stored by DG MARE should be accessible by the institution hosting the supra-regional database in a form that respects confidentiality while providing aggregates for statistical purposes (a revision of the Control Regulation should provide for the use of these administrative records for statistical purposes, as in modern statistical systems).

Environmental indicators for MSFD and IMP: until procedures for compiling environmental indicators are standardized and "industrialized", they can be stored in a specific database. Later on, a specific module in the supra-regional database could be developed to store them.

3.2.4. Data upload

Procedures

Ideally, data would be uploaded by MS into the supra-regional, aquaculture and fisheries processing databases (the latter, hosted by Eurostat as part of the SBS) by using (semi-)automatic transmission tools with agreed format. This means that data is submitted by the MS to the "supra-regional DB" through a secured channel (e.g. FLUX, in XML).

A first set of quality checks would take place (i.e. data validation) and if no errors are found, the data could be uploaded in the production part of the "supra-regional DB". In case of errors, the

MS would be requested to submit a corrected data set. Further data processing would take place in the production part of the supra-regional DB, using a 'clean' data set.

Access to economic and control data from all MS could be dealt with by establishing the necessary legal and technical framework to ensure confidentiality and use for scientific and statistical purposes, explicitly prohibiting the dissemination of individual records subject to data protection. MS would maintain national databases compatible with the "supra-regional DB" so that automated transmission ("push" from MS) would progressively replaced by automated retrieval ("pull" from the hosting institution).

Data from the DG MARE Fleet Register should be directly accessible to the "supra-regional DB" hosting institution, i.e. not only through the present web interface.

Following the present practice, the data would be updated once per year. Potentially, the monthly data could be updated monthly, should the need arise. Automatic updating procedures have to be developed.

Formats

As mentioned above, Excel formats and other proprietary formats[54] would be replaced by text files with a pre-defined structure to transmit data and metadata. The experience of various institutions (Eurostat, JRC, ICES) should be taken into account.

3.2.5. Quality control

Data validation

ICES, Eurostat and JRC all apply validation rules (mostly with software procedures) to the data in their relevant domains. In the supra-regional scenario, the validation rules for the different data sets and domains should be concentrated at one institution / database as presented in the following overview:

- Biologic variables by "supra-regional DB"
- Catch, landings data by "supra-regional DB"
- Transversal variables by "supra-regional DB"
- Economic data of fleet by "supra-regional DB"
- Aquaculture data by JRC
- Fish processing industry by Eurostat.

Validation of primary data takes place on MS level, on the basis of EU-wide common software/rules. To increase the comparability of validation rules and a similar level of quality

[54] *Proprietary formats are formats related to specific software and often protected by software licences. In contrast open formats are freely available.*

control at the national stage, standards control should be developed and agreed at the EU-wide level (as it is currently done when developing statistical regulations by Eurostat and the NSIs). They would be established by the appropriate legislative acts.

Standards for quality indicators (such as accuracy measured by CVs, rates of response, imputation rates for missing values, etc.) should be developed in line with the ESS standards, for the different thematic domains.

Table 11. Division of quality control tasks for the economic, biological and transversal data

	National node Primary data	Supra-regional DB*
Availability of data	X	X
Availability of metadata	X	X
Accessibility	X	X
Missing values	X	X
Duplicated records	X	
Timeliness	X	X
Coding	X	X
Std. deviation	X	
Coefficient of variation	X	
Sample size	X	
Sampling rate	X	
Response rate	X	
Coverage rate	X	
Typing errors	X	
Arithmetic checks	X	
Logical checks	X	
Range/ outliers		
- cross section	X	X
- time series	X	X
Other sources		X

Source: Study survey
**For agreed level of aggregation*

Data validation rules currently applied by Eurostat and JRC on the fleet, aquaculture and fish processing data should be combined (checking their coherence) to satisfy all quality requirements.

Estimates produced by the supra-national organizations for any domain, either at national level (when not available) or at higher geographical levels should be clearly identified and their methodology indicated[55].

[55] *The practice of different international institutions of estimating unavailable national data to obtain aggregates at higher levels (e.g. world regions) is often criticized by National Statistical Offices, especially when the estimation procedures are*

Storage of quality indicators

The current practice of ICES in storing quality indicators and the Eurostat practice of requiring quality reports to MS (with the calculation of quality indicators to be described in the reports) should be combined and extended to all variables.

Quality indicators and reports should be made available to users in the metadata part of the "supra-regional DB".

The current "Technical Reports" submitted by MS on DCF activities could be filled in automatically from data stored in the "supra-regional database". This would be an incentive for MS to support the development of such database.

3.2.6. Dissemination

Confidentiality of data

Confidential data on fish processing industry would be "flagged" by sending MS. Eurostat is entrusted to protect statistical confidentiality by law, and therefore it should not be an issue for this domain.

Aquaculture data would be only submitted by MS on an aggregate form. Thresholds for the dissemination of aggregates (such as those applied in some EU business surveys, which establish a minimum number of units – usually 5 – to be combined in an aggregate to be published) should be agreed to allow for granularity while protecting statistical disclosure.

To ensure the protection of confidentiality, the institution hosting the "supra-regional DB" should be subject to statistical confidentiality restrictions.

Access

Restricted access to WGs and internal staff

Internal staff of the institution hosting the "supra-regional DB" should be subject to statistical confidentiality restrictions as well. Members of WGs dealing with confidential data should be treated as researchers, being granted access to data under confidentiality agreements and/or able to work on secure premises.

not clearly made public and agreed with the national authorities. At the European level, estimates produced by Eurostat (for example on crop production) based on statistical models (ARIMA models) are checked with the MS before release. MS have then the opportunity of providing their own estimates.

Public access

All non-confidential data would be disseminated through the web, in a similar way as it is done through *Eurobase* by Eurostat, SAIKU by JRC or DATRAS by ICES.

3.2.7. Institutional considerations

Legal

The status and mandate of the institution hosting the "supra-regional DB" should be adapted to comply with statistical confidentiality protection and to allow for access to confidential data from MS. ICES and JRC are natural candidates, but Eurostat could be also an option to be envisaged.

Access to DG MARE fleet register should be granted to all institutions managing centralized databases, or at least to that hosting the "supra-regional DB".

The European Statistical Programme 2013-2017 and its annual work programmes should be updated to reflect the agreements on fisheries statistics concerning in particular Eurostat's role and the interaction with DG MARE and other institutions.

Administrative

- A MoU between Eurostat and JRC should be agreed to simplify/merge the data collection on aquaculture and fish processing.
- A MoU between the institution hosting the "supra-regional DB" and the different RFMOs (especially with GFCM) should be also adopted.
- A decision should be taken by DG MARE to entrust either JRC or other non-EC institution (such as ICES) to undertake the compilation of the data from the MS to feed the "supra-regional DB". In any case, administrative arrangements between JRC and ICES seem of outmost importance.

Financial

The costs of the current data compilation by JRC should be re-assigned to the institution hosting the "supra-regional DB".

There are additional costs to develop RDB-FishFrame allowing integration of *economic, effort and MBS data calls*.

Costs in compiling aquaculture both in Eurostat and JRC would decrease by greater cooperation between both institutions, in particular by centralizing it in one of them.

Economies can be made by re-using MDT and *Eurocube* for the storage of fish processing data (as part of SBS data). The current dependence of Eurostat on external contractors for maintaining the required software is a weak point.

There should be a reduction of costs in MS especially in staff time to answer redundant data calls.

3.2.8. Evaluation

The "supra-regional DB" scenario streamlines the data collection from supra-national institutions, simplifying the number of databases to be fed by MS, thus reducing the cost at the national level (in terms of staff time). While the institution hosting the "supra-regional DB" would require important resources (to cover IT development and maintenance, subject-matter expertise and coordination tasks), at the global level the resources currently devoted by the EC (through DG MARE, Eurostat and JRC) would be reduced in the long-term, and the governance of the system improved.

Additionally, users would have easier access as biological, economic and transversal data would be stored in the same place.

In the short-term, methodological work is needed to harmonize several legal requirements (SRs and DCF requirements on variables to be collected – especially in aquaculture and fish processing industries), to establish common reference lists for relevant breakdown variables (classification of gears, vessel sizes) and to agree temporal and spatial references.

Further costs would be entailed for developing IT modules for the RDB-FishFrame. Coordination efforts would be needed to streamline data upload (avoiding overlaps).

Specific information needs at the regional level would not be given priority, though. The scenario may be threatened by the will/need of RSCs or RFMOs to develop specific databases, which would invalidate the model, e.g. by using a different classification of gears and vessels sizes, than the DCF standard or by requiring completely different indicators like volume of fish hold or use of fish aggregating devices in the purse seine tuna fishery.

This scenario is also called 'Eurostat model', because of the high level of centralization. However, Eurostat itself does not seem to be the appropriate institution to implement the 'Supra regional DB' for four reasons:

1. The 'DCF data' would have to be included in current statistical legislation which seems difficult and would certainly take long time.
2. Eurostat deals with data which can be broadly used by many users. It is not involved in compilation of specialized scientific data.
3. 'DCF data' would be only a small part of Eurostat activities and consequently it may not get the required attention.

4. The data is not collected by NSIs, which are Eurostat's national partners. Eurostat would have to deal with a large number of speciliazed national scientific institutes and ministries, which may be difficult to integrate in its on-going operation.

3.3. Scenario 2. Regional nodes (RDB-Fishframe model)

3.3.1. Summary and conclusions

This scenario is based on five regional databases, based on the present model of regional DBs:

A. Baltic Sea (BS)
B. North Sea and NE Atlantic (NS)
C. North Atlantic (NA)
D. Large pelagics (LP) – to be gradually extended to important other fisheries of EU fleets in non EU-waters, including fisheries under FPAs, and the relevant stocks (details should be developed in coordination with relevant RFMOs, avoiding duplication with their work)
E. Mediterranean and Black Sea (MBS)

The DBs A to C are operational at the moment. These DBs are based on the RDB-Fishframe DB and are hosted and maintained by ICES. In practice this in one single DB, in which data is 'flagged' according to the region. The LP DB is being developed by IRD (FR). The scenario assumes that the LP DB would be hosted by ICES (however, IRD would be an alternative candidate for hosting or could be involved in the operation and management of the DB). Setting up of a MBS regional DB is presently under consideration. For the MBS DB, this scenario assumes that it would be operated by GFCM[56].

Furthermore the scenario assumes that:

- Fleet economic data will be incorporated in RDB-Fishframe[57] such that regionalized economic data could be generated. As far as necessary and possible, the data structure should be made compatible to the FLUX business layer (endorsed by UN/CEFACT) at the same time. Where the definitions do not exist in FLUX yet, they could be introduced.
- Data on processing would be delegated to Eurostat, with SBS.
- Data on aquaculture would be maintained in a separate DB, as there is no need to integrate them with the data on marine fishing. This may be a centralised DB, as inland aquaculture can hardly be allocated to marine regions. Aquaculture data would be exclusively collected under DCF, not under SRs any longer. The aquaculture DB could be

[56] *This assumption is made as otherwise the MBS RDB would be operated by ICES and the scenario would be very similar to scenario 1. This is also the recommendation of the MBS RCM meegings in 2013 and 2013.*
[57] *Incorporating fleet economic data in RDB-Fishframe is technically possible, but there is a conceptual difference. RDB-Fishframe contains individual measurements, while economic data is aggregated.*

hosted by by JRC for the whole EU. This seems the most logical choice, considering that JRC operates the aquaculture DB at present.

The major advantages of the regional nodes, based on on-going developments are:

- DB structure exists for biological data (RDB-Fishframe and InterCatch) and can be adapted to specific requirements of regions which do not yet work with it (MBS and LP/other relevant fisheries in non-EU waters);
- RDB-Fishframe could cover a significant part of the needs of the EU Mediterranean countries. Initiatives already exist to extend RDB-Fishframe format to Mediterranean countries and to large pelagics (LP RCG).

The disadvantages are:

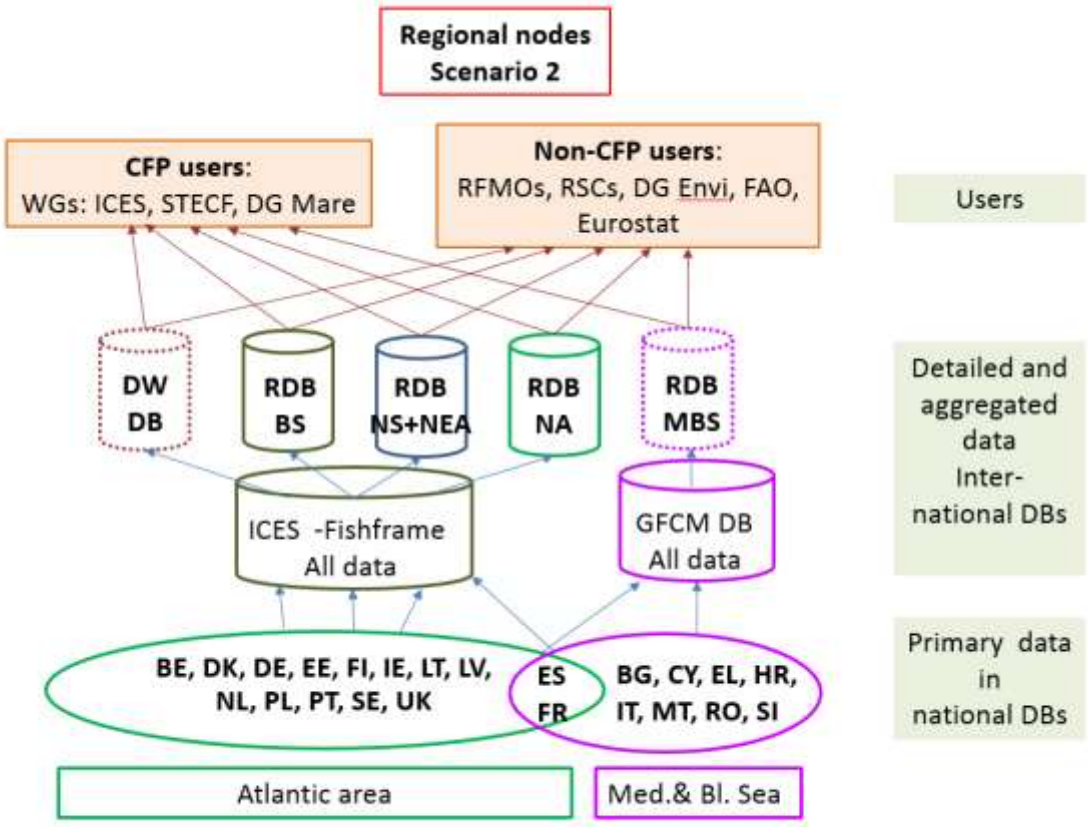
- In theory, MS bordering several marine regions have to submit data to several RDBs (e.g. DK and DE to North Sea and Baltic DB, ES and FR to North Sea and MBS). However, in practice, this could be overcome through joint data calls (the three operational RDBs launched one common data call in 2013).
- If the MBS DB would be operated by a different institution than the other DBs, common development would require additional coordination, effort and costs. The coordination effort would further increase if the LP DB would be operated by still another institution (e.g. IRD).
- RDBs do not contain economic data at present and hence would have to be extended for this purpose. The extent to which economic data can be separated into regions has to be explored. Problems can be expected in particular for Denmark, but also several other countries[58].
- EU needs to assure sound management in its own waters and it may take the lead in the MBS region by setting up a sound data system. RDBs could be opened-up to non-EU countries, especially in the MBS area in order to harmonize data collection procedures and infrastructure between EU MS and non-MS around the same sea basins as far as possible. However, the non-EU MBS countries might have difficulty to follow the RDB-Fishframe standard and to produce the full amount of data required under DCF.
- Still another DB would have to be created and located somewhere to cover other EU distant water fisheries (bottom and small pelagic fisheries off Africa, carried out under the FPAs).
- Economic and transversal data would have to be incorporated in the regional setting.

More detailed SWOT analysis is presented in table 12.

The system of regional nodes is presented in figure 4.

[58] *Some DK fleets operate in both Baltic and North Sea. Determination of their economic performance by region may not be feasible. Fleets in ES and FR which operate in the Mediterranean and Atlantic areas are clearly distinct, so that economic assessment by region should be possible.*

Figure 4. Scenario 2: Regional nodes



Note: *The scenario deals with data on marine fisheries. Data on aquaculture (marine and inland) and fish processing are not part of this data flow (see section 3.1.4.1). Therefore the land-locked EU MS are not mentioned in the diagramme.*

Table 12. SWOT Analysis

	Strengths	Weaknesses	Opportunities	Threats
Data storage – production DB	All data stored in a common format		Common IT development	Different decision making procedures may make it difficult to maintain common standards and processes
Data access – distribution DB		EU data not accessible from one DB (unless specific interfaces or connections are created for this purpose)		Differences in approaches to dissemination and confidentiality.
Data up-load		Some MS have to upload to several DBs		
Quality control		The development and maintenance of common quality rules requires additional coordination and creates additional costs. In case of economic data, aggregation and quality control was undertaken up to now by MS and STECF in close contact. An additional regional level may make this direct link more difficult.	Greater transparency of common quality control rules.	
Dissemination	Biological data located 'close to users'; economic data become available on regional level.	Regional economic DBs located further away from end user STECF. It will be difficult to derive national totals (where data of one MS are stored in different DBs) and EU-totals from the system, unless such totals are automatically calculated from the different DBs.		
Bio-economic integration		For some purposes, primary data may be needed for bio-economic integration, but it may be difficult to make them available.	Common solution to the problem	Lack of data outside EU in the MBS region
Coherence with IFDMP				ICES and GFCM are not EU institutions, so that adherence of IFDMP, developed by DG MARE, is not self-evident.
Legal constraints		Legal provisions for data confidentiality do not exist at present for all nodes involved (operating on different legal bases) and have to be designed.		Agreement on confidentiality may take time.
Administrative constraints				Availability of required staff depends on funding which is organized differently in different host organizations. Consequently, sufficiency of funding is not assured.
Financial constraints		Additional EU funding required.		See above

Table 13. Evaluation against policy objectives

Objective	Evaluation
Informing CFP / IMP/ MSFD	Higher level of data integration, in particular at regional level, may respond more easily to information needs of the EU policies.
Regionalization / coordination	The DB system allows regional evaluation of biological, fleet economic and transversal data. DB structure would correspond to an important role of RCGs in data collection.
Simplification	Simplification and higher efficiency can be achieved through delegation of tasks from national to regional level (e.g. quality reporting). Production of regional economic data may cause some additional effort. Allocation of processing data to Eurostat and aquaculture data to JRC will lead to simplification, as will the use of common coding, through Master Data Register.
Costs reduction	Opportunities of common development of IT, quality check routines, raising routines, etc. will reduce costs in the medium and long run. In the initial stages there will be additional start-up and adjustment costs.
Increased quality	Common routines will check data quality, which is expected to increase quality and transparency. Concerted attention can be given to documentation and meta data. National data providers are close to / involved in quality control as well as raising and estimates; their background knowledge ensures high quality (in particular for biological data).
Integration	Options can be explored to define common denominator for biological and economic data. The availability of (as far as possible detailed) economic data on regional level facilitates integration on this level. Integration of these two areas for bio-economic modelling requires proper theoretical concepts, which will determine the optimum level of aggregation at which the integration is feasible and relevant.
Accessibility	A small number of regional DBs (or of central DBs using regional flags) will be easier accessible than a large number of national and supra-national databases. It may become more difficult to access national or EU totals when data are spread over various RDBs. This may be even a problem for national fisheries administrations when their data are stored in different RDBs. To avoid this problem, specific mechanisms to access totals across regions have to be implemented.
Coherence with IFDMP	The scenario could be combined with the FLUX concept (business and transport layer) and then would be coherent with IFDMP.

3.3.2. Data collection at national level

Data collection at national level would largely remain unchanged. In some cases, production of regional data might have some repercussions on details of data collection (sample sizes for biological data may be reduced; the generation of economic data at regional level may require to review sample sizes or clustering procedures).

Collection of processing data under DCF would become unnecessary, but responsibility for the collection of some aquaculture data would have to be taken over from national statistical offices.

3.3.3. Data storage and access

Overview

This scenario is based on RDB-Fishframe DB, which would have to be expanded to incorporate:

- Specific requirement of the MBD MS.
- Information on LP fisheries and other EU fisheries in non-EU waters.
- Economic data to allow region-specific extraction; e.g. required for bio-economic modelling of specific regionally defined fisheries[59]. At the same time, it should still be possible to retrieve national or EU totals, e.g. by storing data in one central DB with regional flags or by allowing joint queries across all RDBs).
- Present effort data call of DG MARE / JRC.
- Present MBS data call of DG MARE / JRC.

Table 14. Databases and their domains in the 'Regional nodes' scenario

Name of the database	Domain covered by the database	Production / dissemination DB
ICES RDB-Fishframe, incorporating RDB BS, RDB NS+NEA, RDB NA (one DB using flags for each region) and RDB LP[60]	All domains related to marine fishing (biology, economy, transversal) Incorporates JRC data calls for: - Effort - Economic data (all EU MS, including MBS)	Production and dissemination
RDB MBS GFCM	Biologic and transversal data (economic data stored in ICES RDB-Fishframe, in order to ensure overall consistency of economic data). Incorporates present JRC data calls for MBS.	Production and dissemination

Connection

In this scenario there would be two servers operating in parallel: one operated by ICES and the 2nd one operated by another institution, potentially GFCM as recommended 2012 and 2013 report of the MBS RCM. Both institutions would formulate their conditions of accessibility.

RDBs under ICES would be based on presently used technologies: data format in txt and transmission in the secured environment of https:/ASP.NET. Migration of data format to FLUX business layer and transmission to FLUX transport layer may be envisaged in future (independently of the scenarios).

[59] *see also Chapter 7/p. 232 of the AER 2013 concerning the need for regional analyses of economic data;* http://stecf.jrc.ec.europa.eu/documents/43805/581354/2013-09_STECF+13-15+-+AER+EU+Fleet+2013_JRC84745.pdf

[60] *The LP RDB could alternatively be hosted by IRD.*

GFCM would have to assure interoperability not only with the EU, but also with non-EU Mediterranean countries.

At present most data is submitted to GFCM by email in Excel/Access of xml files. By the beginning of 2014 a secured (SSL) data submission will be made available to GFCM members through dedicated SharePoint portal.

This scenario proposes to store economic data also from MBS in a central DB at ICES (with regional flags); otherwise, an interface for communication between ICES and GFCM RDBs would be required to allow EU-wide analyses.

Database structure

The two databases (ICES and GFCM) are likely to have partly different structures because they serve different purposes.

RDBs under ICES focus primarily on compilation of scientific measurements to facilitate planning of survey and (in the future) also more directly stock assessment. Economic data will have to be incorporated, at a suitable aggregation level to allow linking it to aggregated biological data.

The MBS RDB would be presumably at least partly related to the biological parameters DB of GFCM. However, the latter DB contains measurements of growth, maturity and length of GFCM priority species[61]. It would have to be expanded to include all species required under DCF (p.51-55 of the Decision 93/2010), and biological variables in conformity with DCF provisions.

It would be essential that all DCF-related RDBs are based on a common format. This also applies to future DB systems and structures: development of DCF DBs has been a continuous task, where experience gained during collection, processing and using of data led to initiatives to improve the DBs. For regional DBs, it would be important that they stay compatible with each other even in the course of further developments and improvements. Otherwise, new developments may not be applicable to all RDBs, and also data may lose compatibility. This is relevant in particular where data are not stored in one physical DB, and where DBs are adapted to the needs of the specific fishery, such as MBS and LP.

Nomenclature to be used should be based on international standards, as presented in the Master Data Register (MDR)[62] – in relation to species, areas, ports and gears.

[61] *See: http://www.gfcm.org/gfcm/topic/166221/en*
[62] *http://ec.europa.eu/fisheries/cfp/control/codes/index_en.htm*

- ***Recommendation: All RDBs related to DCF should be based on one common format and developed along common lines.***
- ***Recommendation: The needs of MBS MS and LP fisheries (RFMOs) must be analysed in detail, in order to determine the adaptations which must be implemented in RDB-Fishframe to make it fully compatible with the needs of these fisheries / areas.***

Aggregation Level (variables and dimensions)

The aggregation levels should meet at a minimum the DCF requirements in both DB systems (ICES and GFCM).

RDB-Fishframe currently contains individual measurements. Other RDBs (MBS and LP) should be also based on this principle.

Economic data submitted by the MS are at present aggregated at MS level. In order to allow regional disaggregation, some MS would have to either adapt their data collection and processing accordingly or design estimation procedures to derive the required regional data. The future DCF regulation would have to be adapted to distinguish 'regional segments', e.g. on the basis of dominant region (parallel to dominant gear).

Following the present practice, the data would be updated at least once per year. Potentially, the data could be updated as soon as they become available. Automatic updating procedures would have to be developed.

Data processing and estimation

At present, supra-national data processing (biological and economic) takes place at ICES and JRC/STECF; thiswould be continued in a similar manner, with involvement of GFCM for the MBS. For economic data, however, some aggregation and estimation processes would be transferred from MS to regional level. Technological improvements would be implemented, e.g. linking RDB-Fishframe to InterCatch at ICES and required adaptations at GFCM.

Apart from an already on-going data processing, an important part of processing would be to bring the biological and economic data to same definition at spatial and temporal level. This is discussed in detail in chapter on 'Common features'.

Interactions with other fisheries related data

Data from CR would still be required for the purpose of sampling design. CR would be also the source for aggregated data on catches for Eurostat and FAO.

Eurostat and FAO compile national totals of landings by MS and marine region. Eurostat also compiles data on fleet capacity. This information is based on census of the CR data. DCF data is

based on samples, derived on the basis of the CR. It is necessary to check on consistency between these two data sets, i.e. extrapolation of the sample data should give a national total in the range of the census results.

Evaluation and conclusions

In this scenario, two identical RDBs would operated by different institutions in parallel. All new developments should be undertaken in common to avoid gradual divergence. The different institutional set-up and decision making procedures are potentially a major weakness. For economic data, the workflow of processing, aggregating, estimations and quality control (see below) – activities that are now undertaken in dialogue between MS and JRC/STECF[63] – would have to be adapted in order to involve a regional level.

3.3.4. Data upload

Currently the various types of data (biological, economic, transversal / effort) become available in the MS at different times. This is a consequence of the way in which the data can be collected and analysed. Economic data on "year 1" can be only collected starting in April of "year 2" (or even later) because the firms must close their books first. Biological data is collected in the course of the on-going year. Effort and catch data from logbooks and sales notes is collected almost in real time, and therefore becomes available shortly after the end of any period, be it month or year.

Consequently, ICES and GFCM might have to launch different data calls for different data domains in different time schedules. Further analysis of the contents of the data calls will have to show whether the timing can be coordinated between the two institutions. This would be particularly useful for France and Spain, which would have to report to both.

ICES RDBs would continue using the present data format in txt and transmission in the secured environment of https:/ASP.NET, but also for upload the FLUX transport layer may become relevant.

By the time that a GFCM RDB would be in place, its SSL data submission system through a SharePoint portal will be operational.

Evaluation and conclusions

The data upload may take place as soon as data becomes available, and at least annually. Some MS (ES and FR) would have to upload separately to two different DBs or more, e.g. if LP is set up elsewhere. It is uncertain whether the protocol and format could be identical.

[63] *According to information received from JRC, about 50% of the fleet economic data are changed between their first submission and final publishing in the AER.*

3.3.5. Quality control

Data validation

ICES would continue with present quality checks in relation to the biological data. The manual checks would be gradually automatized by being translated into software. These quality checks would be also developed for economic, transversal and MBS data, using (if possible) routines developed by JRC (see JRC section).

Table 15. Present quality checks – ICES RDBs

	Biological – métier related data			Biological – stock related data			Catch / landings		
	No	Manual check	Soft-ware check	No	Manual check	Soft-ware check	No	Manual check	Soft-ware check
Availability	X			X				X	
Accessibility	X			X				X	
Missing values			X			X			X
Duplicated records			X			X			X
Timeliness		X			X			X	
Coding			X			X			X
Std. deviation	X			X			X		
Coeffic. of variation	X			X			X		
Sample size		X			X			X	
Sampling rate		X			X			X	
Response rate		X			X			X	
Coverage rate		X			X			X	
Typing errors			X			X	X		
Arithmetic checks	X			X					X
Logical checks	X			X					X
Range/ outliers	X			X					X
- cross section	X			X					X
- time series	X			X			X		
Other sources	X			X			X		

Source: Study Survey

Most quality checks in relation to stock related biological data done at GFCM are manual (see chapter on GFCM). The only software check regards coding. This implies that significant effort would be required by GFCM to develop software checks for its RDB.

In addition to the quality checks stated above, procedures to validate aggregations used for analytical purposes (stock assessment, AER, etc.) need to be formalized.

Storage of quality indicators

ICES maintains all quality checks of RDBs in a database, except those regarding effort. This practice would continue.

GFCM does not store any quality indicators in a DB at the moment. It would be necessary to develop an approach for this purpose.

3.3.6. Dissemination

Confidentiality of data

ICES pursues an 'open' data policy, where users can obtain publicly available data as soon as this is feasible. The data policy states, however, that "All data provided to ICES are considered to be publicly available unless otherwise explicitly specified and agreed"[64]. The latter restriction is applied to the commercial catch databases (InterCatch and RDB-FishFrame), which are subject to the provisions of the DCF regulations regarding confidentiality. This means that DCF data can be made public only if the confidentiality rules are not compromised, i.e. only at an aggregated level. A proper framework to ensure confidentiality accordingly appears to exist on the side of ICES, but it would be necessary to include details concerning the confidentiality of economic data (in particular if non-aggregated economic data should be stored at regional level) in a specific MoU for the storage of such data.

GFCM indicates that its statutes would have to be adapted to guarantee data confidentiality and such process would take about 12 months.

Eurostat's provision on the access to micro data for scientific purposes[65] may serve as an example for respective agreements with ICES and GFCM[66].

Restricted access to WGs and internal staff

At the moment, ICES RDBs are mainly used for coordination of survey activities, but not for scientific stock assessment work. It seems useful to develop further functions of the RDBs.

In both cases (ICES and GFCM) access to most detailed data (production DB) would have to be granted to members of scientific WGs and selected staff members in order to facilitate their scientific work. A system of flexible queries would have to be developed.

[64] *Art. 4 a; see http://ices.dk/marine-data/Documents/ICES_Data_Policy_2012.pdf*
[65] *see e.g. http://epp.eurostat.ec.europa.eu/portal/page/portal/microdata/introduction*
[66] *It should be noted that FADN-RICA DB (operated by DG AGRI) on agricultural farms also contains primary data.*

Public access

The public should be provided access to aggregated data through a dissemination DB. Dissemination DB will have to be developed in both institutions.

Lowest level of disaggregation for public access should be determined.

3.3.7. Institutional considerations

General

Each RDB would be coordinated by one RCG. The RCG would also need to include economic competences.

Legal

ICES has a status of international organization. It is related to the EU through a MoU.

GFCM is an UN institution. It was established within the framework of Article XIV of the FAO Constitution and it was approved at the Fifth Session of the FAO Conference in 1949. The EU is a member of GFCM, and so are the individual EU MS adjacent to the MBS. In order to formally involve GFCM in the DCF process, the EU would have to make a MoU in this respect and provide financial support. It is unclear at this stage how easily such MoU could be agreed upon and implemented.

Administrative

ICES would have to exempt economic data from their open data policy, where the DCF requires confidentiality; this could be defined in a respective MoU, which would be also needed to delegate storage of economic data to ICES.

GFCM would have to introduce specific confidentiality rules before handling primary or detailed data originating from DCF.

In particular for GFCM, it would be necessary to make available sufficient staff for the tasks related to DCF. In ICES, additional staff with economic competences would be needed, provided the organisation takes over more than just storing economic data.

Financial

ICES is funded through the contribution of its members. Additional funding from the EU is provided in the context of the MoU in view of the provided services.

GFCM is funded by contributions from its members. Out of the total budget of 1.9 mln USD, 0.6 mln is provided by the EU and another 0.6 mln is provided by the individual EU MS in the MBS region.[67]

Additional funding would have to be provided to both institutions if they would be required to assume additional tasks of development, maintenance and operation of regional DBs and if they would have to support scientific WGs requiring the data stored in the RDBs.

3.3.8. Evaluation

This scenario offers a number of advantages. DB structure exists for biological data (RDB-Fishframe and InterCatch) and can be adapted to specific requirements of regions which do not yet work with it (MBS and LP/other relevant fisheries in non-EU waters). RDB-Fishframe could cover a significant part of the needs of the EU Mediterranean countries. Initiatives already exist to extend RDB-Fishframe format to Mediterranean countries and to large pelagics (LP RCG).

However, there are also some disadvantages. In theory, MS bordering several marine regions have to submit data to several RDBs (e.g. DK and DE to North Sea and Baltic DB, ES and FR to North Sea and MBS). In practice, this could be overcome through joint data calls (the three operational RDBs launched one common data call in 2013). If the MBS DB would be operated by a different institution than the other DBs, common development would require additional coordination, effort and costs. The coordination effort would further increase if the LP DB would be operated by still another institution (e.g. IRD). RDBs do not contain economic data at present and hence would have to be extended for this purpose. The extent to which economic data can be separated into regions has to be explored. Problems can be expected in particular for Denmark, but also several other countries[68]. Still another DB would have to be created and located somewhere to cover other EU distant water fisheries (bottom and small pelagic fisheries off Africa, carried out under the FPAs). Finally, economic and transversal data would have to be incorporated in the regional setting.

[67] *GFCM, Report of the Thirty-seventh session of the Commission Seventh session of the Committee on Compliance Fourth session of the Committee on Administration and Finance Split, Croatia, 13-17 May 2013, p.126*

[68] *Some DK fleets operate in both Baltic and North Sea. Determination of their economic performance by region may not be feasible. Fleets in ES and FR which operate in the Mediterranean and Atlantic areas are clearly distinct, so that economic assessment by region should be possible.*

3.4. Scenario 3. Network-based (EMODnet model)

3.4.1. Summary and conclusions

This scenario is based on a network of databases. A comparable system is being developed today by the EMODnet initiative[69]. In this model, the national DBs or primary data allow extracting an output based on common format and definitions. The DBs are linked through a web-based interface which offers users access to the national data and possibilities of data processing at supra-national level.

EMODnet focusses on providing spatial information through mapping facility. This is an important aspect of fisheries data in terms of migration of stocks, distribution of fishing fleets, effort and catches / landings. However, spatial distribution is given at aggregated level and must be complemented with other relevant data to allow for full stock assessment and economic analysis.

The major advantages of the Network approach are:

- The data is managed by the MS Meta data description is formalised and describe the available data which can be requested at raw data level or disseminated at aggregated level or presented in maps.
- Data is continuously up-dated, as it becomes available in the MS.
- There is no duplication of data and the compilation of European or regional data is automated limiting the burden to MS.
- MS having developed their own system covering the DCF and most of the time other domain can continue using it and could expect less additional work as soon as the transmission procedure are in place.
- More flexibility of aggregations at supra-national level is possible as the national data are available at raw level.
- Various types of data (biological, economic, transversal, control regulation, aquaculture) can be provided from decentralized DBs.
- Standard queries can be developed for the central interface to generate information required by various users.

The disadvantages are:

- Most developments (IT and content) have to be coordinated among all MS as implementation takes place at national level. This implies duplication of processes, need of national IT resources to maintain the system and consequently unnecessary inefficiencies and more labour input.
- The EMODnet model is very suitable for compiling basic data in real time in maps and simple data streams etc. where checks are simple and expected changes in the data are

[69] *http://EMODnet.eu/*

very few (e.g. control regulation data). But for some specific work dealing with complex raising methods of biological fish data from individual fish to a complete stock, the model becomes a bit unrealistic when new data are made available to calculate the stock on the fly (i.e. as soon as new data becomes available) as some corrections /raising needs to be coordinated (e.g. ICES work performed in RDB-FishFrame or InterCatch).

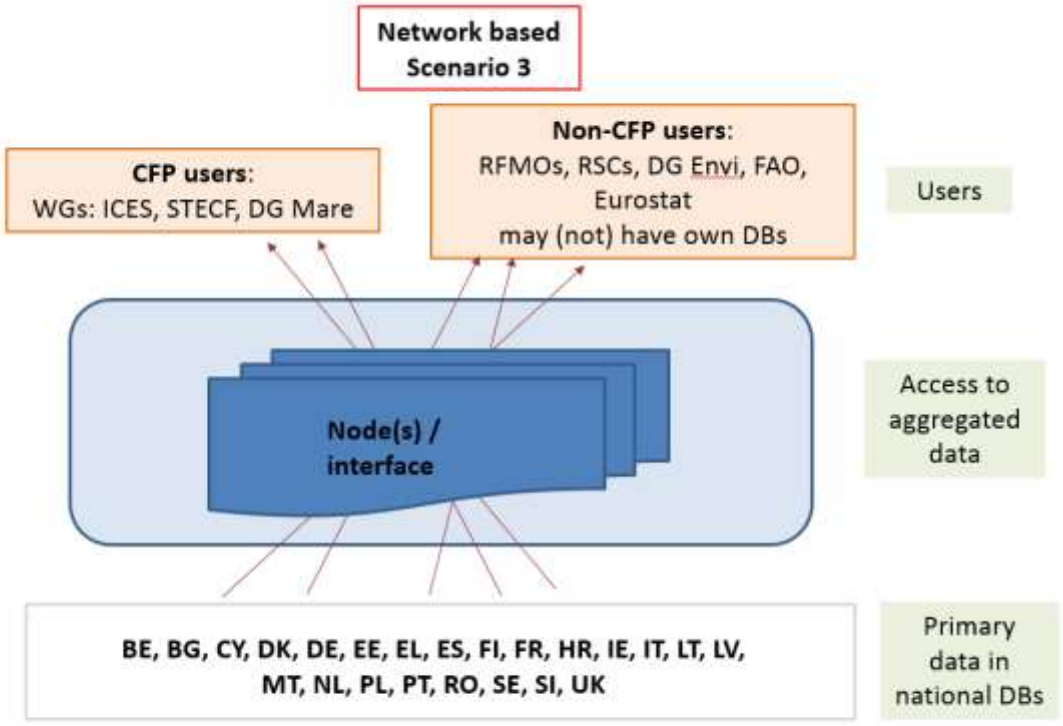
- The different database systems connected to EMODnet are fully independent and only final data are visible. The system relies on the quality of data and meta data provided as a limited number of complementary tests are performed on the data made available to associate a quality label (for EMODnet biology, it concerns the check on taxonomy compared to WoRMS, plotting of sampling locations on a map to check for odd locations. If there is any doubt, or if errors are suspected, this is communicated to the data provider so corrections can be made). The quality of the data and correction/estimates depends on the tools used in each database, and there is no guarantee of having the same checks applied in all places.
- The lesson and difficulties learned from EMODnet biology project[70] is that meta data are usually documented but the sharing of the national data is often prevented by legal acts or absence of official permission. The institutes those were positive about data sharing mostly opted to share the data through web-services. In most cases, these web services are however still under development at the different institutes.
- The system requires the final data to be made available in due time but also all the national systems to be never offline, else compiled data will be incomplete.

More detailed SWOT analysis is presented in table 16.

The Network scenario is presented in figure 5.

[70] *http://bio.EMODnet.eu/documents/Final-Report/Executive-summary/*

Figure 5. Scenario 3: Network



Note: *The scenario deals with data on marine fisheries. Data on aquaculture (marine and inland) and fish processing are not part of this data flow (see section 3.1.4.1). Therefore the land-locked EU MS are not mentioned in the diagramme.*

Table 16. SWOT analysis

	Strengths	Weaknesses	Opportunities	Threats
Data storage – production DB	No replication of the data Data close to the source. Compilation of detailed data provided by different MS will lead to better aggregation at regional level as some missing information in one country could be completed/correlated with available information in another country, making it possible to reach an acceptable level of completeness.	Reliance on many national DBs, which are also driven by national requirements and priorities. The EMODnet scenario requires to support countries facing difficulties with their existing system, and possibly replace their current system with a generic system compatible with EMODnet approach. The level of explanation on the meta data needs to be made very clear at national level to ensure a correct understanding of the automatically compiled information. For some specific biological work dealing with complex raising methods of biological fish data from individual fish to a complete stock, it becomes unrealistic to calculate the stock on the fly as some corrections /raising needs to be coordinated and decisions saved.	Promotion of the data collected at national level (by giving the information on the source of the data)	Coordination among all involved DBs operators may face human and technical problems. Some countries are not able to support the maintenance/evolution of their database.
Data access – distribution DB	The level of details of the data accessible to the different kind of users needs to be discussed/agreed as well as compilation rules ensuring the confidentiality of the data The source data for the compilation being the primary /detailed data, the compilation provided to user will be based on detailed data and provide more relevant aggregates depending on the dimension of aggregation chosen by the user.	The full exploitation of primary/detailed data will depend on agreement on the level of dissemination of data which may be different from one country to the other	The access to primary/detailed data and connection with different domain open the possibility to run ad hoc studies	Failure of national DBs – strength of the system is its weakest link.

	Strengths	Weaknesses	Opportunities	Threats
Data up-load	Data upload process in different databases is not needed anymore, as source data are directly accessible Burden at MS level is reduced.	Relies mainly on national procedures for the collection, compilation of data	Data updated as soon as available at national level	Lack of data in time in the national DBs impedes to obtain relevant result at European or regional level.
Quality control	Quality checks continue to be managed by the national institutes who could develop supplementary checking procedures, bring innovation which could be generalised. If the network approach is combined with an opportunity to provide generic tools to small countries, the small country could benefit from improved quality procedures as well.	A limited number of checks are performed at supra regional level Common quality rules standard have to be developed and maintained. Implementation of new rules in each MS system will be time consuming and would need a lot of effort. If errors are identified at supra national level and new checks/corrections need to be implemented in all MS, it will take time before the correction are completed on primary/detailed data to be usable at supra national level Checks on MS data can not be done before the data are exposed to the network.	If the network approach is combined with an opportunity to provide generic tools, the difference in the data quality between the MS could be reduced.	The quality mainly relies on MS. Each national institute implements the 'agreed' checks" individually according to his understanding of the check and the quality level to reach.
Dissemination	Lots of opportunity for using the fisheries data	Primary data located 'far from the users'.	Possibility to design common dissemination policy	
Bio-economic integration	Could be ossible through central queries.	Pre-defined aggregation level is a limitation by definition. Needs to find an acceptable gateway between biological and economical data ensuring a proper link between the two domains.	Opportunities of new studies.	Lack of data outside EU in the MBS region The level of interoperability of the system and database schema may prevent using existing databases
Coherence with IFDMP	Use of Master Data Register is feasible.			
Legal constraints	Confidentiality to be assured at national level			
Administrative constraints		Defintion of the sharing of the data		Varying administrative rules among DB operators.
Financial constraints				Most costs incurred at national level. Danger of budget cuts.

Table 17. Evaluation against policy objectives

Objective	Evaluation
Informing CFP / IMP/ MSFD	Higher level of data integration (linking various types of data) and greater detail (low aggregation levels) may respond more effectively to information needs of the EU policies.
Regionalization / coordination	A very high level of coordination is needed between the different actors to apply the same level of tests and make the data available in due time, as in most MS several institutes are involved in various DCF tasks.
Simplification	No replication of data, no need for data upload in supra national database resulting in a reduction of burden for MS. This is possible if assistance in developing capacity is foreseen for small MS not having a robust system in place.
Costs reduction	The cost reduction is not obvious as there is a need to maintain/adapt existing application, to develop new generic systems for MS not having their own system and finally for all MS, the needs remain to develop modules /web services, registry allowing to connect the national database by respecting the EMODNet requirements.
Increased quality	Dependence on skills, funds and ambitions in MS; the scenario is as strong as the "weakest" national institute. Supra regional aggregation are based on raw data and should provide more accurate information as it could be possible to combine some raw data provided by different MS Standardisation based on international coding improve the quality of the data.
Integration	The network option open the perspective of connection of different databases but this will become possible if the logical schema of the database makes it possible to define a gateway between the domain, else new development and structures needs to be envisaged to make the database interoperational
Accessibility	Continuous update making it possible to compile directly the source data in real time in maps and simple data streams Nevertheless agreement on the use of the primary data and the level of dissemination of compiled data needs to be designed..
Coherence	The data are taken from a single place so the risk of inconsistency in disseminated data is limited. But this means also that data are exposed to the public without prior checks, and correction of errors identified later by supra regional organisation may take time

3.4.2. Data collection at national level

Data collection and processing at national level would play key role in this scenario as supra-national procedures and quality control would be relatively weaker than in the other scenarios. Intensive cooperation among the MS and harmonization of national approaches would have to be assured.

3.4.3. Data storage and access

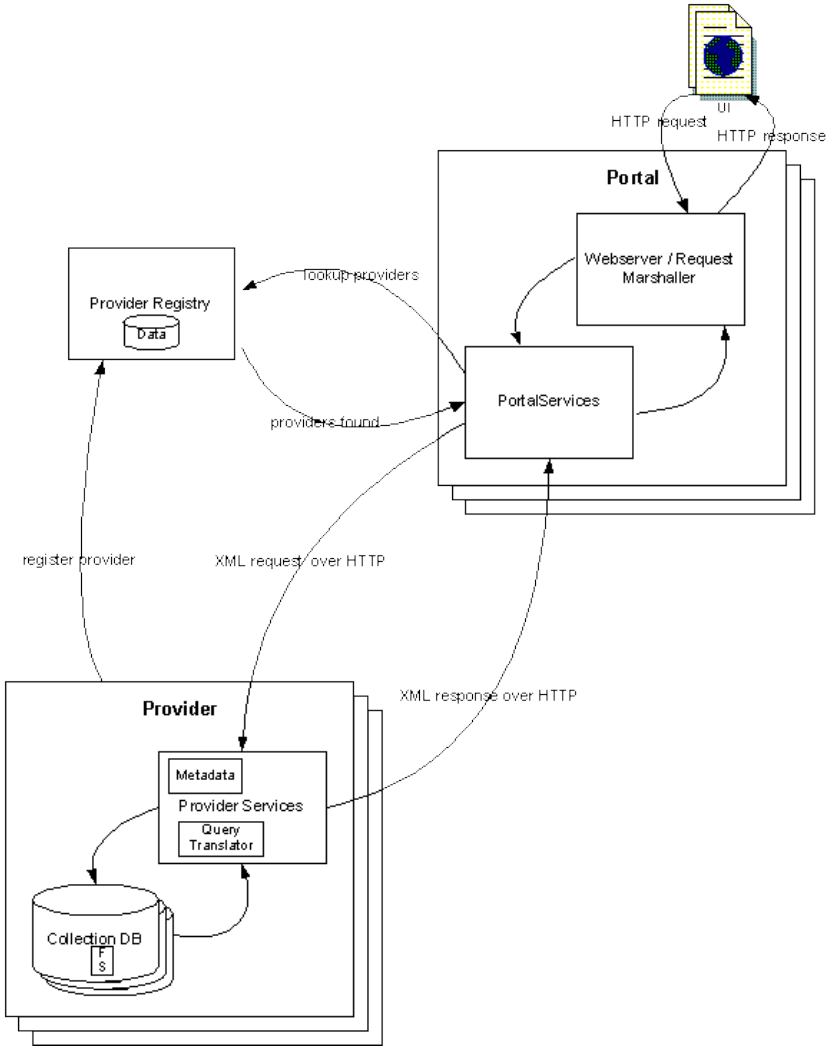
Overview

The network based model would aim at assembling fragmented data into interoperable, contiguous and publicly available datasets for whole basins. The system should be considered as a network of distributed data systems, is OGC [71]compliant, containing standardized data and integrateing data with different levels of accessibility. The portal should include different functionalities such as the online data catalogue, a data querying service, a taxonomic ontology, a data mapping interface, a data downloading service and a help and feedback function. A full description of an experimental portal system can be found in the final report on biology portal[72].

The distributed system would make use of the Distributed Generic Information Retrieval (DiGIR)[73] protocol (an open standard capable of supporting many communities, without regard to discipline or domain. See Annex 2.10.) and would be fully platform independent.

[71] *OGC® Standards support interoperable solutions that "geo-enable" the Web, wireless and location-based services and mainstream IT. The standards empower technology developers to make complex spatial information and services accessible and useful with all kinds of applications. (http://www.opengeospatial.org/ogc)*
[72] *http://bio.EMODnet.eu/documents/Final-Report/Final-report-Biological-Lot/*
[73] *http://digir.sourceforge.net/*

Figure 6. DiGIR high level architecture diagram

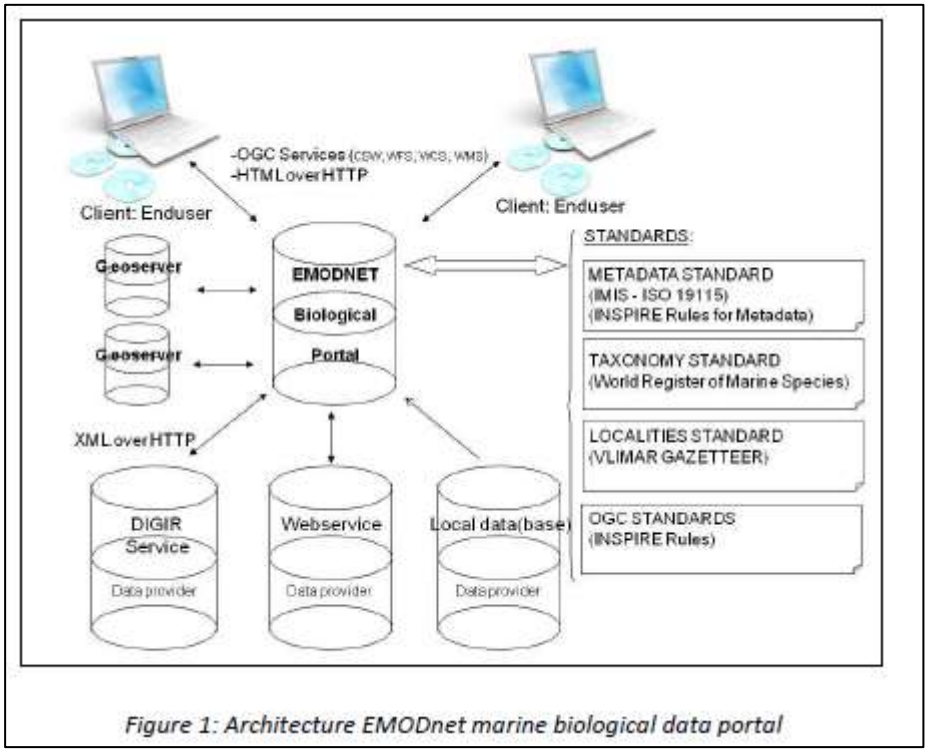


Explanation on the terms used in the above diagram:

- Protocol: the definition of message formats (i.e. requests and responses) by which components must communicate.
- Provider: an application that makes structured data available to compliant portals.
- Provider Metadata: a structured description of a provider's database, including description of the database, URI to federation schema, supported columns, supported operation types, and summary data (such as geographic and/or taxonomic index info).
- Portal: an application that communicates with multiple providers and performs operations to retrieve and integrate data (and metadata). It functions as the point of access for users.
- Registry: a centralized (public) repository of available providers.

The database management remains under the responsibility of the database owner, but there is a need to develop the data provider part for communicating with the portal, to document the provider registry and to comply with the protocol requirements as defined in the Annex 1.2.

Figure 7. Representation of EMODnet marine data portal key elements



Source: Lot N°4 – Biology (SI2.531562) of the Service Contract No MARE/2008/03
Note: In this figure the MS DBs are connected at data provider level, either via DIGIR (see above) or specific web services

Connection

In this scenario, the national DBs would continue on the basis of their present technologies. The distributed system would make use of the Distributed Generic Information Retrieval (DiGIR)[74] protocol and would be fully platform independent The connection between the portal and the national DB would be based on XML over HTTP.

The access to confidential secured environment of databases holding primary data is likely to present problems.

[74] *http://digir.sourceforge.net/*

Database structure

A sustainable solution would need to be developed to ensure the maintenance/evolution of the portal, registry, and provider support in a long term perspective. Strong organisation and coordination aspects would be covered.

This scenario does not foresee a central DB. The national DBs remain in their original architecture. This leaves the maintenance and update of the databases in the hands of their owners, but there would be a need to develop the data provider part for communicating with the portal and to document the provider registry and to comply with the requirements as defined in the Annex 1.2.

Nevertheless, with such a network system would be difficult to automate the calculation of some complex product like the raising part of the national data or stock assessment calculation which requires integration of the relevant data as well as the input of the experts.

Table 18. Databases and their domains

Name of the database	Domain covered by the database	Production / dissemination DB
National DBs	All domains, possibly in different DB	Production, primary and aggregated data
'EMODnet – Fisheries'	All domains	Dissemination

Aggregation level (variables and dimensions)

The EMODnet based scenario implies that several WGs would first have to be set-up to define the IT characteristics of the data at national level. Their task would be to develop further the approach formulated by EMODnet Biology group[75].

The database system would be integrated at the level of primary or detailed data to allow for greater analytical flexibility in scientific applications. The primary data would be updated continuously, as they become available. Data extracted will be compiled and anonymized following the confidentiality rules defined.

Data processing and estimations

At present all required data processing (economic and biological) takes place at national level and would continue in a similar manner.

It would be difficult to implement complex raising methods for stock assessment without centralising the data in regional databases.

[75] *Lot N°4 – Biology (SI2.531562) of the Service Contract No MARE/2008/03 on the Preparatory Actions for European Marine Observation and Data Network*

Certain specific users (e.g. RFMOs) have specific data requirements. In the definition phase of the national module, these requirements must be taken into account so that these users can be serviced through the central node, using predefined queries. This may regard vessel size class or gear definitions, e.g. ICCAT distinguishes 3 sizes of purse seiners according to the volume of their fish hold and effort units like successful fishing days and number of sets[76]. This means that the common nomenclature must be developed accordingly.

Interactions with other fisheries related data

Data from CR remains required for the purpose of sampling design. The national node should be linked to the CR data. Automated procedures will allow cross-checking CR data with extrapolation of the survey results. The maximum difference between the two values should be within a pre-defined range.

3.4.4. Data upload

The fundamental difference with scenario 1 and 2 is on the data flow. The various types of data (biological, economic, transversal / effort) would be available in the MS at different times. This is a consequence of the way in which the data can be collected and analysed. Economic data on "year 1" can be only collected starting in April of "year 2" (or even later) because the firms must close their books first. Biological data is collected in the course of the on-going year. Effort and catch data from logbooks and sales notes is collected almost in real time, and therefore become available shortly after the end of any period, be it month or year.

As in the present situation, a final deadline would have to be agreed among the MS (or operators of the national DBs) when specific data would have to be uploaded to the national node / module. The procedure in which the upload takes place can be decentralized, as long as data quality is assured (see below).

The data would be updated continuously, as they become available. This is particularly interesting for aggregates of the CR data (effort and landings) because it would allow continuous scientific monitoring. e.g. having monthly aggregates of landings and effort would allow up-to-date estimations of economic performance as well as better appreciation of the role of new age-classes of various stocks entering a fishery

[76] *http://www.iccat.int/en/Stat_Codes.htm*

3.4.5. Quality control

Data validation

Most quality controls would have to be done at national level regarding primary and aggregated data. Common procedure should be agreed and followed, (e.g. the COST model[77]). The central node checks the certain aspects of the aggregated data.

Table 19. Division of quality control tasks

	National node Primary data	Central node
Availability	X	X
Accessibility	X	X
Missing values	X	X
Duplicated records	X	
Timeliness	X	X
Coding	X	X
Std. deviation	X	X
Coefficient of variation	X	X
Sample size	X	
Sampling rate	X	
Response rate	X	
Coverage rate	X	
Typing errors	X	
Arithmetic checks	X	
Logical checks	X	
Range/ outliers		
- cross section	X	X
- time series	X	X
Other sources		X

Source: Study survey

Considering that national DBs are based on different software, application of common automated procedures for quality checks of primary data may require substantial resources. Checking aggregated data could be done by one common programme as the architecture of the national modules would be identical.

Storage of quality indicators

All quality indicators should be stored in the national module, along with the metadata, to be accessible for users.

[77] *COST: Common "Open Source" Tool for assessing the accuracy of the biological data and parameters estimates collected for stock assessment purposes within the framework of the Data Collection Regulation. http://wwz.ifremer.fr/cost*

3.4.6. Dissemination

Confidentiality of data

The primary data would remain at all times in the national DBs and would not be directly accessible to the public through the central node.

The dissemination level of the data would be implemented through specific rules agreed with MS. Even if detailed data could be retrieved for specific calculation allowing better estimates at regional level, they would never be presented at the primary level. In addition, the aggregation would always ensure the confidentiality on the data.

Restricted access to specific users

Depending on the specification of the national nodes, it would be conceivable that specific users obtain access to more detailed data than the public.

Public access

The public should be provided access to aggregated data through the central node. Dissemination policy should be formulated, to promote the use of the data.

Lowest level of disaggregation for public access should be determined.

Technical functionalities

There is a broad spectrum of possible technical solutions. EMODnet – Biology or other EMODnet portals could be used as a relevant example / experience.

3.4.7. Institutional considerations

Legal

Legal considerations do not seem to play a major role in this scenario. In comparison to the present situation, the MS only provide their DCF data through a different procedure.

Administrative

Administrative implications of this scenario regard in particular the operation of the central node. This could be probably included under the operation of EMODnet as a whole.

Financial

The scenario implies a number of new activities which will require additional funding to design and implementation of (cf. Annex 1.2):

- The national data provider modules
- The portal central node and query interface
- Common quality checks procedures at national and portal level.

Activities at the national level could be possibly financed through DCF funding under EMFF, where data collection is a priority. IT development at the central level may follow similar procedures as other EMODnet portals.

3.4.8. Evaluation

The data storage foreseen under this scenario would allow access to primary or detailed data in all domains. The main obstacle is probably the uncertainty about political will in the MS to cooperate at this level of detail. Overcoming this obstacle could generate significant efficiency gains in terms of better quality of the data products and / or lower costs.

In this scenario, small MS lacking of resources for IT development may require continuous support to ensure their contribution to the system. In order to provide this technical support it could be foreseen to have the "national data stored in cloud databases administered by the central organism but the data being managed by the national entities". This illustrates that there is only a thin line between network-based and more centralised approach.

Even in this scenario, supra-regional database cannot be avoided, because biological data must be integrated during the stock assessment process. Some checks on the data can be performed on the fly (at InterCatch /RDB–Fishframe level to determine if the quality of the data is sufficient or not) but some rising to stock level cannot be fully automated. In addition, according to ICES, some specific corrections may take a lot of time to be implemented individually in each MS instead of in a central location.

Data upload would be simplified, compared to the present situation. Provision of primary or detailed data would imply that data calls do not need to take into account possible additional requirements of new data products. Scientific WGs, having access to the most detailed data, would be in the position to produce new types of analysis, without requesting the MS for adapted specification of data (e.g. in terms of aggregations).

In this scenario, the quality of the system would largely depend on the implementation at national level. The design of data flows matters a lot as well as the quality assurance and the reactions to quality issues. The implementation of specific corrections (as a solution to identified data problems discovered at supra-national level) may take a lot of time as each MS would make corrections on its primary data. This is different from the other scenarios, where aggregates can be corrected by MS before uploading to supra-national level.

Dissemination of the data would be subject to rules regarding access and aggregation levels. Confidentiality of individual data would be guaranteed at all times.

It should be noted that Eurostat considered a similar approach when setting up international trade statistics. However, for technical reasons it decided against it (e.g. a querry requiring information from all MS has to be routed to all national DBs and integrated at central level).

3.5. Scenario 4 "Fisheries data hub"

3.5.1. Summary and conclusions

Scenario 4, the "Fisheries data hub" pursues to combine the strenghs of the scenarios 1-3 and avoid their weaknesses. The system is integrated at EU level, to assure common approaches to definitions, to reduce IT development costs and to avoid gradually divergent regional development. At the same time the system should allow to incorporate regional specificities (e.g. different spatial scale in the Altantic and MBS areas). The hub should also allow linking biological and economic data at a feasible resolution. The IT system and workflow organization would offer possibilities for large scale data processing, including monitoring of quality. The appropriateness of the IT system and workflows needs to be (continuously) evaluated in the light of its merits, i.e. whether it expands opportunities for scientific work (which it should) or on the contrary whether it imposes restrictions (which it should not). Such evaluation must be done by those most directly concerned, i.e. the scientific community (primarily ICES, STECF and GFCM) and the main users of the scientific analysis (policy makers).

The Hub would constitute the overall structure allowing to build common referentials (nomenclatures, metadata description, GIS, etc.) and it would ensure interoperability between the different types of databases (biological, economical, transversal) which could be maintained by different institutions. The MS work would be facilitated as they would have to use only a few well defined and stable data transmission formats embedded in their databases to feed transparently the different Hub databases. Any output formats would be defined directly from the accessible databases by end users reducing the MS burden of responding to data calls.

The IT system and workflow organization would offer possibilities for large scale data processing, including standardized monitoring of quality at EU level because all tools and quality checks could be developed and implemented centrally in one system. This is also feasible in scenario 1, but not in scenarios 2 and 3. The appropriateness of the IT system (e.g. using push or pull for transferring data from MS to the Hub) and workflows (e.g. timing of data availability or implementation of monitoring of data quality) would need to be (continuously) evaluated in the light of its merits, i.e. whether it expands opportunities for scientific and policy work (which it should) or on the contrary whether it imposes restrictions (which it should not). Such evaluation must be done by those most directly concerned, i.e. the scientific community (primarily ICES and STECF) and the main users of the scientific analysis (policy makers).

This scenario is based on several principles overarching the differences between scenarios 1-3:

- This scenario explores the options of a **'Fisheries data hub'**, in which hardware, software and data workflows are mutually independent. This means that data collection is done by national institutes, but all subsequent data processes may be done by specialized units at supra-national level and the DBs may be located outside the national institutes. For example:
  - Primary data is uploaded to a 'cloud DB', composed of a set of national DBs. The 'cloud DB' may be located in "A".
  - Quality control (using standard routines) is done by a central unit of the 'cloud DB'. Quality control unit may be located in "B".
  - Development of routines for data processing is done by specialised WGs (stocks, economics, etc.) located in "C" and "D".
- Physical **location** of the data (bases) is not relevant. Various stakeholders could be responsible and have exclusive access to parts of one central database, located at any place / institution. The system could also consist of several DBs in different locations, linked through internet as if it were one single DB.
- **Tasks and responsibilities** of actors at national and supra-national level would be clearly defined, avoiding duplications.
- **Ownership** and access to data are well defined to assure confidentiality of personal data.
- **Transparency** is pursued by detailed documentation of all workflows and contents.
- Standardisation of the **coding**: use of internationally recognized code lists in all elements of the system.
- **Transfer of experience** among MS, including IT tools, statistical methods, quality procedures, etc.
- **Gradual** process, rather than rapid / radical change.
- **Intensive international cooperation** in data collection reduces the costs. This has been already demonstrated at the Regional Coordination Groups Meetings.

The "Fisheries data hub" is based on transfer of detailed/primary data to supra-national DBs and in the establishment of more efficient exchange mechanisms along the lines of the FLUX system developed by DG MARE for control data:

- Data would not depend on the availability and performance of the national databases spread over EU. The complexity and performance of the query for retrieving information would be simpler as focused on a limited number of supra national databases.
- The hub will constitute the overall structure allowing to build common referential (nomenclatures, meta data description, GIS ...) and will ensure interoperability between the different types of databases (biological, economical …) .
- Only a few well defined data transmission format will be required to feed the "hub" databases: the database could then be updated with a minimal effort from the MS, potentially at different period in times (different revisions depending on the need for availability of the data for specific users). The system should in the long term reduce

highly the burden at national level (no provision of administrative report, only one format to provide the data which will remain stable in the long term)

- Any output formats will be defined directly from the hub accessible databases by end users reducing considerably the burden for MS and the cost for data calls.

One single, perfect solution for all situations does not exist. Therefore it is proposed to **prioritize** the areas to which the future set-up of DCF IT systems should respond. EU institutions have primary responsibility and decision powers over fish stocks and fisheries within EU jurisdiction, i.e. in Atlantic, Mediterranean and Black Sea waters. This is also the area where by far the largest part of the EU fishing fleets is operating. Therefore the first priority of the future DCF IT systems should respond to the information requirements stemming from the management of these areas. While information needs on fisheries in non-EU waters should be adequately addressed, they should not compromise the first priority. It can be assumed / expected that the data needs of management of non-EU waters will be similar to the internal EU needs.

From the perspective of data collection and data products there are evident advantages to **integrating biological and economic primary or detailed data** in one database. The dialogue within RCMs has already demonstrated that higher efficiency (and consequently lower costs and better quality) can be achieved when data collection is planned at supra-national level. Not all MS need to collect all types of data, as long as sufficient number of biological observations is achieved. Similar argument applies also to fleet economic data.

'Fisheries data hub' approach offers various advantages for data producers and data users:

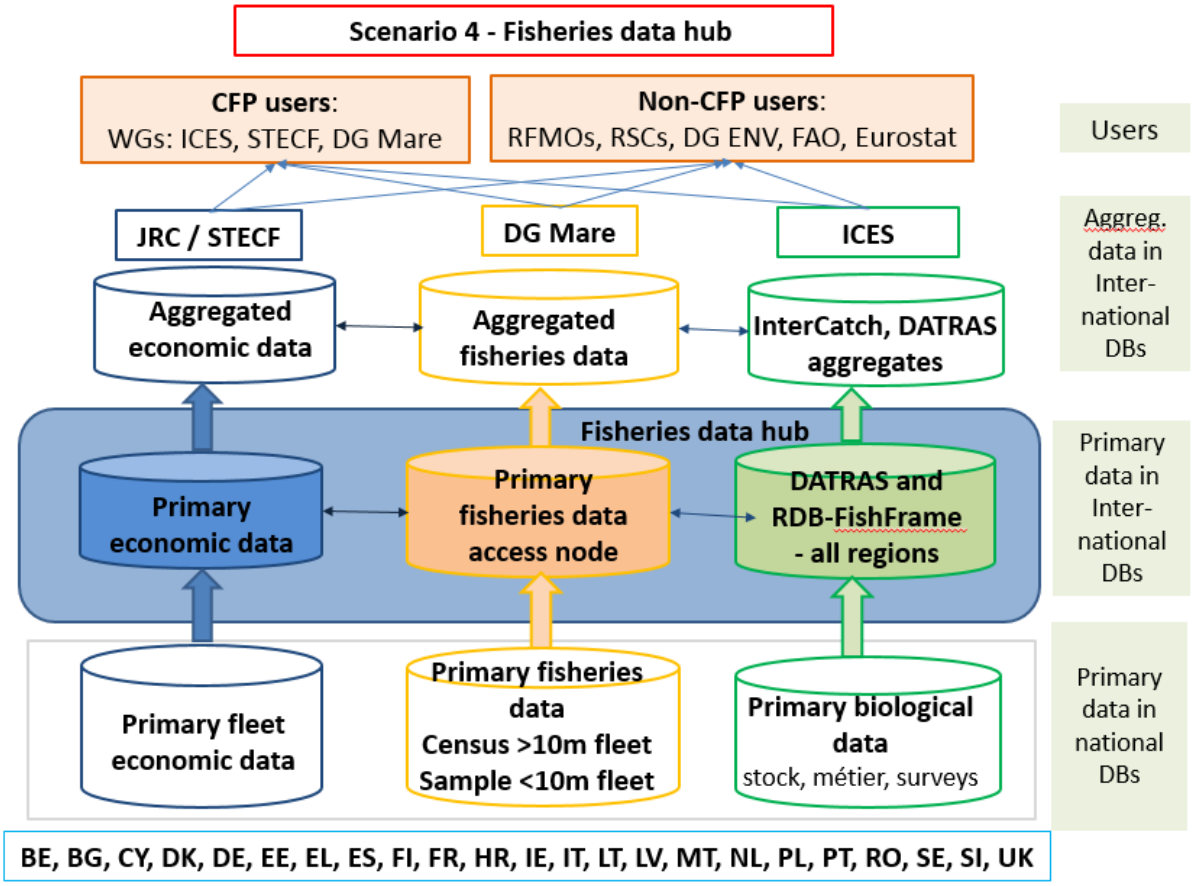
- Data producers benefit from use of common tools, increased efficiency and lower costs
- Data users can rely on increased quality, transparency and data consistency.
- Creation of a consistent EU-wide system allows common development in the future.

The disadvantages are:

- Political sensitivity of sharing primary or detailed economic data.
- The hub system needs to be defined in terms of access/confidentiality of the data, security storage, use of the existing databases and possible revision of their structure to handle the primary/detailed data and compliance with the defined referential.

The scenario 4 is presented in figure 8. Detailed SWOT analysis is presented in table 20.

Figure 8. Scenario 4: "Fisheries data hub"



Note 1.: The scenario deals with data on marine fisheries. Data on aquaculture (marine and inland) and fish processing are not part of this data flow (see section 3.1.4.1). Therefore the land-locked EU MS are not mentioned in the diagramme.

Note 2: Primary data in national databases is identical to the primary data in international databases.

Table 20. SWOT analysis

	Strengths	Weaknesses	Opportunities	Threats
Data storage – production DB – primary data	Comprehensive EU-wide approach. Transparency on all levels. Well defined access rights.	All participants are obliged to adhere to same standards, which may cause resistance. Weaker relation to national reality.	Common IT development.	Threat of a rising supra-national 'bureaucracy' in the management of data storage.
Data access – distribution DB	Well defined access rights for various users. Incorporation of raising procedures in a new system		Common development of data products.	Institutionalization of the 'Hub' may increase distance to main users (science).
Data up-load	Standardized procedures and formats for the transmission all data.		On-going data compilation replaces data calls.	
Quality control	Independent evaluation. Use of common validation rules. Use of common "clean" datasets.	Specific national issues (related to the data collection system) may not get sufficient attention at hub level, but have to be addressed at national level before the transmission to the hub	Common procedures. Strong involvement of data collection experts is required. Share of subject matter expertise to validate data.	Too high reliance on automated quality checks.
Dissemination	Higher consistency in disseminated data.	The system is as timely as the slowest partner.	Design of a common dissemination policy.	Loss of visibility of current data disseminators.
Bio-economic integration	National fleet segments replaced by EU (or regional) fleet segments.	Integration of data does not solve conceptual problems.	Development of a common denominator for economic and biological data.	Danger of misinterpretations.
Coherence with IFDMP	Use of IFDMP / FLUX transmission procedure.	Obligation to advanced MS to shift to a new system.	Achievement of UN/CEFACT standard.	
Legal issues	Institutes involved in DCF are able to guarantee confidentiality.	Sharing primary data faces confidentiality problem.	Increased harmonization of legal requirements for data.	Data ownership may be disputed.
Administrative issues		High requirements on the institution responsible for operating the whole system.		
Financial issues	Lower costs through common infrastructure and procedures in long term.		Lower costs through common development in long term.	Higher costs of adaptation in short term in a situation of scarce public resources.

Table 21. Evaluation against policy objectives

Objective	Evaluation
Informing CFP / IMP/ MSFD	Higher level of data integration may respond more easily to information needs of the EU policies.
Regionalization / coordination	The DB system allows regional evaluation of biological, fleet economic and transversal data. Higher efficiency of data collection can be achieved.
Simplification	Data calls are replaced by continuous data flows / update of the DB. This process eliminates the work required to respond to the data calls. Use of common coding, through Master Data Register.
Costs reduction	Opportunities of common development of IT, quality check routines, raising routines, etc. will reduce costs in the medium and long run. In the initial stages there will be additional start-up and adjustment costs.
Increased quality	Common routines will check data quality. Concerted attention can be given to documentation and meta data.
Integration	Options can be explored to define common denominator for biological and economic data. Integration of these two areas for bio-economic modelling requires proper theoretical concepts, which will determine the optimum level of aggregation at which the integration is feasible and relevant.
Accessibility	Accessibility to one "Fisheries data hub" can be organized centrally, rather than having to define accessibility to a large number (20+) individual national and supra-national databases.
Coherence with IFDMP	Using the IFDMP / FLUX transmission protocol may be expected to simplify the transmission procedures between various stakeholders.

3.5.2. Data collection at national level

Collection of primary data is the responsibility of the MS. Without compromising this principle, costs of data collection (incl. required IT systems) as well as data quality can be improved with more intensive cooperation among the MS and harmonization of national approaches. Problems identified in relation to DCF exist also in other statistical domains. Possibilities for solutions have been analysed by Eurostat in its 'Vision for the next decade'[78].

Sound data quality starts with the statistically robust data collection at national level. Reliability of primary data can be promoted through sharing experiences and tools (approaches to surveys, tools for quality checks and their results). An important aspect is the transparency about (meta) data, data collection and processing, which allows in particular the professional users (ICES, JRC, STECF) to make their own quality assessment. At present it is not always clear which data, uploaded to supra-national DBs, has been collected and which has been estimated (and how).

[78] *COMMUNICATION FROM THE COMMISSION TO THE EUROPEAN PARLIAMENT AND THE COUNCIL on the production method of EU statistics: a vision for the next decade, COM(2009) 404 final, Brussels, 10.8.2009*

One of the fundamental problems with sharing fisheries data is their political sensitivity in relation to the implementation of the CFP. This problem can be resolved by full separation of scientific evaluation from control purposes.

Most important efficiency gains can be realized at national level, where also the highest costs are incurred. This has been already demonstrated by the results of the RCMs in relation to collection of biological data and by setting up the RDB-FishFrame, which will facilitate the processing of the primary data. Cooperation among MS in collecting economic data may also lead to higher efficiency.

3.5.3. Data storage and access at supra-national level

Overview

This scenario foresees in principle one "Fisheries data hub", encompassing primary biological as well as fleet (and aquaculture) economic primary (or detailed) data.

All primary **biological** data would be stored in RDB-FishFrame and DATRAS. Specific RDB-FishFrame modules would be dedicated to MBS data and to data on fisheries in non-EU waters. ICES WGs would be responsible for further processing (raising / aggregations) of data related to Atlantic stocks. STECF WGs deal with the MBS stocks, having access to the regional MBS module of RDB-FishFrame and DATRAS. RDB- FishFrame and DATRAS would be also in charge of providing data to GFCM and other users. Provision of data to other RFMOs (outside Atlantic and MBS) could be delegated either to IRD or to STECF, which would obtain the primary or detailed data from RDB-FishFrame. RFMOs would also get direct well defined access to the data related to their area of responsibility.

All primary (or detailed) fleet and aquaculture **economic** data would be stored at supra national level by JRC and accessible to STECF WGs.

All **transversal** data (volume and value of landings, volume of catches and fishing effort) would generate through the data compiled under the control regulation. DG MARE should have direct access to the respective national DBs. Access through a hub, rather than creation of a large DB, seems preferable. The hub contains access rights for ICES and JRC staff to generate transversal data required to match biological and economic data.

Confidentiality of the primary data would be assured by appropriate allocation of access rights.

Raising routines and aggregations would be developed in a modular way, generating intermediate data products required for further scientific analysis, e.g. stock assessment for CFP and estimation of environmental indicators for MSFD.

Table 22 presents an overview of the DCF data, the database where it would be stored and the main responsible institution.

Table 22. Databases and their domains

Name of the database	Domain covered by the database	Supra-national operator
RDB-FishFrame, DATRAS, InterCatch	Biological data – all areas	ICES
Saiku	Economic data (fleet and aquaculture)	JRC
"Control Regulation"	Transversal data	DG-Mare
"SBS"	Fish processing	Eurostat

Connection

In this scenario, the national DBs could continue on the basis of their present technologies. The MS which do not dispose of a sufficiently robust IT system would be able to use a system which would be centrally designed, an adaptation of a system developed and implemented by some MS. As an example, the national sample data can be uploaded in the RDB FishFrame and directly process/checked there without previous processing at MS level if a MS has not the capabilities to perform these checks in house. Same kind of facilities could be envisaged for other data as far as MS agrees to have his own data accessible but located outside the country.

All DBs would be 'equipped' with a FLUX (IFDMP) module which allows mutual communication, based on standard technology and set internationally certified codes (Master Data Register). FLUX is first implemented to the control regulation data flows and subsequently to other data flows, according to relevance. The current experience in data transmission under SDMX standards between NSIs and Eurostat proves that such system is feasible.

The communication between the DBs would be based on XML over HTTPS.

Database structure

The fundamental idea of this scenario is the creation of "Fisheries data hub". Efficiency is improved by focussing on improvements of the organization of the workflows.

Specialized DBs (or modules of the DB system) deal with specific data domains: biological, economic and transversal. Transversal data required at present within biological, economic or effort data calls will be drawn from the transversal DB, by staff having specific access rights. This approach allows central development of specific queries and reduces the effort required from the individual MS in data provision.

Aggregation level

The "Fisheries data hub" is integrated at the level of primary or detailed data. The primary data would be updated continuously, as they become available. This is particularly interesting for aggregates of the CR data (effort and landings). Confidentiality would be assured by applying principles of Eurostat (thresholds for micro-aggregates dissemination). Economic data on fleet

and aquaculture could be also provided in anonymized form, which would (combined with other confidentiality safeguards) further guarantee that individual data cannot be traced to private persons or firms. A comparable approach could be also applied to transversal data.

Data processing and estimations

Raising and aggregations to national or supra-national level take place according to needs. They can be viewed as intermediate or final data products. This means that the present practice of raising biological data to national and stock level can be continued as well as calculations of the economic performance of national fleet segments.

Certain specific users (e.g. RFMOs) have specific data requirements. In the definition phase these requirements are taken into account so that these users can be serviced, using predefined queries.

Application of MOLAP (Multidimensional, on-line analytical processing) / Cube structure, including cube-slicing (2-dimensional view) and drill-down (from summary to detail) is continued as common approaches in present DBs.

Interactions with other fisheries related data

The hub would provide access to CR data of the MS and would also generate transversal data required for DCF. Specific queries would be developed for this purpose.

Specific arrangements are made to provide data / indicators for MSFD and IMP. These arrangements fall under MoUs between EC and the responsible institutes, in particular ICES and JRC. Once the information requirements of MSFD and IMP would be defintitively defined, it would be possible to prepare queries and/or data models to calculate the various indicators. Additional financial arrangements may be needed.

3.5.4. Data upload

Procedures

In this scenario, national differences in time schedules of data collection do not play a major role as long as deadlines are respected when specific data must be available. Procedures of data upload are gradually adapted to the FLUX system. Biological and economic data is uploaded (push approach). Transversal data is extracted at the required level of detail (pull approach).

Formats and content

IFDMP / FLUX system together with the Master Data Register would be used to define formats and contents during the upload process.

3.5.5. Quality control

Data validation

Most data quality controls could be done at supra-national level. Common procedures could be followed, e.g. the COST model. Automatically generated validation reports would inform the MS about the required corrections and the data users about still pertaining data issues / problems.

Table 23. Quality control at supra-national level

	Economic data	Biological data	Transversal data
Availability of data	X	X	X
Availability of metadata	X	X	X
Accessibility	X	X	X
Missing values	X	X	X
Duplicated records	X	X	X
Timeliness	X	X	X
Coding	X	X	X
Standard deviation	X	X	X
Coefficient of variation	X	X	X
Sample size	X	X	X
Sampling rate	X	X	X
Response rate	X	X	X
Coverage rate	X	X	X
Typing errors	X	X	X
Arithmetic checks	X	X	X
Logical checks	X	X	X
Range/ outliers	X	X	X
- cross section	X	X	X
- time series	X	X	X
Other sources	X	X	X

Source: Study

Storage of quality indicators

All quality indicators along with the metadata would be stored in specific modules of the supra-national DBs, to be accessible for users.

At present the MS must produce annually 'Technical reports' about their DCF activities. In this scenario, most of the information in these reports could be generated by the supra-national DBs. This would also significantly simplify the evaluation of the TRs by STECF.

3.5.6. Dissemination

Confidentiality of data

Data confidentiality would be assured by the responsible organizations and could be further strengthened by implementation of procedures applied by Eurostat. In order to overcome specific national sensitivities, economic data could be provided in detailed (i.e. anonymized) form instead of primary form. It is expected that detailed data could be potentially traced back to individual legal or private persons only in very few exceptional cases for which further special arrangements could be designed (such as the thresholds for the number of units contributing to an aggregate, as mentioned above).

Access

Access to primary and/or detailed data would be allowed only for specific staff members that would have special access rights. These staff members would prepare aggregated data to be used by various end users eg. ICES, WGs. Providers of the data have full access to the data which they have submitted or made available.

Public access would provide to aggregated data, either through pre-defined tables or queries. Lowest level of disaggregation for public access would have to be determined.

3.5.7. Institutional considerations

Legal

The four main institutions currently responsible for data storage and processing (ICES, JRC, DG MARE and Eurostat) have a legal position which is fully compatible with dealing with sensitive data, incl. provisions to guarantee data confidentiality.

Administrative

The present tasks of the four responsible institutions specified in table 24 (DG Mare, ICES, JRC and Eurostat) are such that new activities under DCF would not require any adaptations of their administrative procedures or set-up. They are all already treating at least part of the data for which they would become responsible under this scenario. Only the volume of the data flow would increase.

MoUs between JRC and Eurostat (on aquaculture and fish processing), ICES and RSCs, (on environmental indicators) and RFMOs (on non-EU waters) may be necessary to define the roles in data exchange.

Financial

The "Data fishery hub" scenario is close to scenario 3 in terms of architecture for the control regulation data, as well as a combination of scenario 1 and 2 as the primary data are centralised by themes (biologic, economic) but in different databases which can be potentially connected.

The "Fisheries data hub" combines new developments with the reuse of existing well developed components (RDB-FishFrame, DATRAS, FLUX…) meaning low implementation risk compared to the expected improvements. The solution could be implemented gradually and ensure flexibility in the use of existing data and possibly combination of data.

Table 24 summarizes costs reduction for the main actors in the short and long term and points to some additional costs in the short term. It is not expected that cost would increase in the long term.

Table 24. Evaluation of costs

Actor / Data processes	Costs reduction		Costs increase
	Short term	Long term	Short term
Member States			
Collection	Smaller national samples due to coordination in RCGs, possibly for biological and fleet economic data.		
Processing	Use of standard routines, agreed at EU-wide level, replaces partly the development of statistical processing software at national level (in line with developments in the European Statistical System). Development and maintenance done centrally, not in concerned EU MS.		Standard routines have to be implemented in national systems, replacing current statistical processing systems. Additional routines may have to be designed again to meet specific national needs.
Storage			Some adaptations will be required.
Upload	General routines for continuous (or regular) upload replace data calls. Direct links between national and supra-national DBs are created. Use of Master Data Register simplifies / streamlines data definition issues.		Implementation of the FLUX for the data transmission requires installation of specific components in the MS.
Quality control	Primary data checked in the supra-national DBs (as in Eurostat procedures) and MS only carry out corrections of the identified problems. Validation reports generated from supra-national DBs.		Ad hoc development needed (quality report…) Development of standard quality reports may need additional methodological work.
ICES			
RDB-FishFrame			Expansion to MBS and non-EU waters. Routines for quality reports.
InterCatch			Need for development of routines for raising data to stock level for Atlantic stocks.
DATRAS			Expansion to include MEDITS survey. Routines for quality reports.
JRC			
Fleet economics	Easier central analysis of various issues. No data calls.		DB expansion to deal with primary data, incl. routines for data processing (aggregation) and quality control.
Aquaculture			
Fish processing	Transferred to Eurostat. No further costs.		
Biology – MBS	Arrangements to manage the RDB MED (hosted by ICES) Data calls abandoned.	No further IT development for a MBS database	Routines for data processing.
Effort	Data calls abandoned.		Query to be developed to draw data from DG MARE CR hub.
Eurostat			
Fishing	Data on catches and fleet through DG MARE hub		Query to be developed to draw data from DG MARE CR hub. Update of existing system for feeding the databases.
Aquaculture	Uses DCF aquaculture data: data call abandoned		Update of existing system for feeding the databases.
Fish processing		MEETS programme should reduce cost by shared procedures and better access to administrative data from enterprises.	

3.5.8. Evaluation

The data storage foreseen under this scenario allows access to primary or detailed data in all domains. The main obstacle is probably the uncertainty about political will to cooperate at this level of detail. Overcoming this obstacle could generate significant efficiency gains in terms of better quality of the data products and / or lower costs.

Data upload is simplified, compared to the present situation. Provision of primary or detailed data implies that data calls do not need to take into account possible additional requirements of new data products. Scientific WGs, having access to the most detailed data, are in the position to produce new types of analysis, without requesting the MS for adapted specification of data (e.g. in terms of aggregations).

Most quality checks and reporting can be done at supra-national level, easing the burden on national institutes.

Dissemination of the data subject to rules regarding access and aggregation levels. Confidentiality of individual data is guaranteed at all times.

The implications of the scenario remain largely within the present legal, administrative and financial scope of the involved institutions. The main changes of tasks and responsibilities can be summarized as follows:

ICES:

- Expands RDB-FishFrame to include data on MBS and non-EU waters fisheries. Draws data on transversal variables through DG-Mare hub on CR data.
- ICES continues stock assessment of Atlantic stocks.

It must be noted that MBS is not ICES' area of competence. Whether ICES members and non-members (MBS EU MS) would agree that ICES also stores data on the MBS region is uncertain. In this respect ICES would have to be considered only as a 'service provider' to MBS countries, given its technical and organizational capacity. ICES would not be involved in any further data processing or analysis, but exclusively in data storage.

JRC / STECF:

- Sets-up fleet and aquaculture economic DBs to deal with primary / detailed data.
- Transversal data for fleet economics and effort call is drawn through DG MARE hub.
- MBS data call is transferred to the RDB-FishFrame MBS module.
- Data call on fish processing is terminated. Data is drawn from Eurostat's SBS.
- STECF continues its present tasks on economic performance, effort regimes, MBS stocks and DCF evaluations, although the data is differently sourced.

DG MARE:

- IFDMP / FLUX creates access to primary CR data of the MS (either through a hub or a DB).
- DG MARE is the primary source of transversal data for ICES, JRC and Eurostat. This means that special access rights are granted to the DG MARE's hub / DB to these institutions and specific queries are designed to serve their needs.

Eurostat:

- Data on catches / landings are obtained through DG MARE CR-hub.
- Aquaculture data is obtained from DCF survey, the result being compiled by JRC.
- Provides fish processing data to JRC.

Member States:

- Adopt international standards for classifications and codes.
- Adopt international standards for data transmission (FLUX).
- Grant access to primary or detailed data to specific supra-national institutions.
- Discontinue current IT developments in view of adapting to new standards.

Other:

- Specific arrangements must be developed for evaluation of EU fisheries in non-EU waters. This can be delegated to IRD (France), STECF or another suitable institution. The data is provided by ICES, JRC or DG MARE.
- Assistance is provided to small MS lacking of resources to implement and maintain the new system.

3.6. Comparison of the scenarios

The development of an EU-wide IT system for DCF, as proposed by the scenarios, will be only possible if the MS demonstrate the political will to share data at a greater level of detail than is the case at present. This regards particularly economic and transversal (fisheries) data, which is in its primary and even detailed form, are subject to the legislation regarding protection of personal data. But also primary biological data is only accessible to a select group of researchers (ICES). It must be stressed that this political will is fundamental to all four scenarios, as they all assume that the data will become available at a greater level of detail than at present. This means ideally at detailed or even primary data. Broadening analytical possibilities can be probably achieved only very partially when the present data is organized in a different set of databases, presented in the scenarios. The greatest benefit would be achieved if more detailed data become accessible, which would indeed call for a new IT structure. The economic and fisheries data is there, but cannot be exploited because of the perception of its confidentiality. However, Eurostat (and all national statistical institutes) demonstrates every day that it can deal with detailed individual data, without compromising the confidentiality rules.

The tables below present a summary evaluation of the scenarios against the policy objectives and constraints, specified in the ToR.

Comparing the four scenarios should be ideally done on the basis of set of common indicators against which they would be benchmarked. However, the scope of the DCF and of policies to which it should contribute is very broad, making the choice and definition of such indicators rather arbitrary. Ranking the scenarios in a multi-criteria decision does not seem not possible, because their pros and cons are often on different and incomparable aspects.

The main advantages and disadvantages of each of the four scenarios compared to the other three are summarized in the following table:

Table 25. Advantages and disadvantages of the scenarios

	Advantages	Disadvantages
Scenario 1: Supra-regional DB	Only one DB, allowing concentrated IT development. One single institutional focus point on all DCF matters.	New responsible institution has to be appointed and experience transferred.
Scenario 2: Regional DB	Data reflect regional needs.	Without EU-wide coordination several specific, but different systems may arise.
Scenario 3: Network	Data stored only in one place (national DBs), no replication of data storage.	Development of interface requires intensive coordination with all MS and provision of IT support to MS lacking the IT expertise.
Scenario 4: Fisheries Data Hub	Primary data in specialized linked DBs, eliminating the need for data aggregation prior to upload. Institutions presently involved in DCF get opportunity for further development and exploitation of their experience.	All primary data stored in two DBs (national and hub). Political will to share primary data will be a major issue.

A truly objective selection of the 'best' scenario is difficult if not impossible in the multi-criteria assessment of the four scenarios, in which many different aspects (IT, legal, institutional, etc.) play a role. Depending on precise implementation, the four scenarios could all fulfil similar requirements.

In addition, in some respects the distinction between the results obtained using one or the other scenarios may be in practice relatively small:

- All scenarios can work with primary data and perform the relevant aggregations;
- At present the Regional DB (RDB-FishFrame) is in fact one single DB, containing regionally flagged data for all Atlantic areas. Such DB could be expanded with regional data on MBS and distant waters. In that situation the difference with the Supra-regional DB becomes very small;
- The Supra-regional DB or the Regional DB could be built in thematic modules (biological, economic and fisheries data), which would be linked but possibly operated by different group of experts. Such set-up would also reduce the difference between these two scenarios and the Fisheries Data Hub.

Because of this 'diffuse borders' between scenarios, it seems relevant to formulate a number of common **principles** (i.e. valid for any scenario) which should be pursued when specifying the future solution for DCF DB set-up:

- Primary or at least detailed **data** should be the basic building block, rather than the present aggregations to fleet segments;
- The system should use common internationally accepted **nomenclatures** (species, gears, areas). Common nomenclature has been largely used across biological, economic and transversal data. The only area where certain discrepancies exist is the definition of gears. Extensive nomencaltures have been formulated by FAO and are reflected in DG Mare's Master Data Register;
- Data and DBs are organized along **thematic modules** (biology, economy, fisheries) and links between the modules would be created;
- **Scientists** who are closely involved in policy related research (ICES, STECF, GFCM) should be also closely involved in development of the system so that it responds to their needs (i.e. to their ability to answer policy questions);
- **Data quality** procedures should encompass the whole process from preparation of the sampling till dissemination. Data must produce complete and standardised results comparable across all feasible axes, which will also allow creation of links between the various data sets. A common IT system can support these procedures, but not replace them. What matters most is the **transparency** of the entire process of the data production (which is also the basis for ISO certification) to allow for scrutiny, control and improvement;
- Proper **documentation** of all processes (including the dissemination of metadata) is the basis for transparency, transfer of experience and further development;
- Clear distinction must be made between **production and dissemination DBs**.

The above 'cross-cutting principles' would address at least some, although not all, **policy objectives and constraints**:

- **Informing CFP / IMP / MSFD** – possibilities would improve by having all the collected data (i.e. primary or detailed) accessible in one system.
- **Regionalization** - regionally flagged data within one system allow regional assessments. Setting-up regional DBs is not essential. Involvement of scientists, who also bring in the regional perspective ensures that specific regional issues can be addressed.
- The call for **simplification** originates from changing needs of the scientific advice (new policy questions) and consequently changing specifications of the data to be provided by the MS. Provision of primary or detailed data, based on common nomenclatures, by the MS to the international DBs would resolve this problem.
- **Data upload** – transmission of data should be based on a common standardized format, e.g. FLUX. Uploading primary or detailed data is the only way to avoid repetitive uploading of similar data, based on same primary data.

- **Data storage** – Any future system will be based on 3 modules: biological, economic and fisheries data. The data is related by using the common nomenclatures.
- **Data access** – Data confidentiality has to be assured at all times. Data access has to be defined by type of users (policy advice, public). It should be clear for any user where exactly the data they are requesting is coming from, whether from one supra-regional database, a national database or anything in between, so they can judge its quality, reliability etc. for themselves.
- **Bio-economic integration** – This is primarily an issue of conceptual development of linking biological and economic units (stocks and segments). Availability of primary or detailed data would allow empirical testing of these concepts.
- **Coherence with IFDMP** – can be assured, as long as all involved institutions are willing to cooperate in development and implementation of this common platform.
- **Legal constraints** – Institutions running the system must guarantee data confidentiality. ICES, JRC and Eurostat all have the required legal provisions.
- **Administrative constraints** – High requirements have to be met in any scenario. ICES, JRC or Eurostat have demonstrated that they can meet these requirements.
- **Financial constraints** – Different approaches will have different financial consequences for the MS and the central DBs. Streamlining the data flows is likely to lead to savings at MS level. The financial consequences of the scenarios are elaborated in table 26.

It should be noted that some of the above issues are scenario specific, while others would not be affected by the selected scenario. Scenario specific issues are: simplification, costs reduction, strengthening of quality assessment and accessibility as well as legal, administrative and financial constraints. Other issues – informing policy, bio-economic integration, coherence with IFDMP and regionalization – are relatively less dependent on the scenarios.

Choosing among the scenarios

Setting up one EU-wide database, containing primary or detailed data, which is regionally flagged, seems the best option as it meets the EU consistency as well as regional specificities. Considering that the Supra-regional DB would be probably also built around thematic modules, the scenarios 1 and 4 seem to be the most appropriate options. ICES and JRC would continue in their present role, so that from the perspective of institutions and expertise such approach seems the easiest to implement. From the perspective of the development needed, the scenario 4 allows reusing existing databases while scenario 1 implies developing a new system. The greatest challenge regards the creation of 'fisheries (transversal) data module'.

Scenarios 2 (Regional nodes) and 3 (Network) seem to offer less specific advantages. As stated, the RDB-FishFrame is already one DB, covering three regions and putting data on MBS and distant water fisheries in separate DBs seems to lead to unnecessary fragmentation and probably additional IT development costs. Integration of all regions in one DB dilutes the specificity of Scenario 2 compared to scenarios 1 and 4.

Scenario 3 obliges to connect a large number of DBs to a central node. In most MS, biological, economic and fisheries data is stored in different DBs (even in separate DBs for métier, stock and survey data). Setting up on-line links to some of these DBs would require that file (often in excel) currently stored on standalone PC are at least stored on a server accessible through the internet in a format adapted to the requirement of the network scenario. Finally, Scenario 3 seems to offer the least possibilities for common approaches to the monitoring of data quality. Problems faced by any MS may affect the whole system (the network is as weak as its weakest node).

The following tables summarise and classify the features of the 4 scenarios in relation to the policy objectives and constraints (table 26), and the potential costs to be incurred in (table 27). For ease of reading, positive (+) and negative (-) signs have been added to highlight the favourable or unfavourable aspects.

Table 26. Comparison of scenarios by policy objective and constraint

Policy objective / constraint	Sc. 1. Supra-regional database	Sc. 2.Regional nodes	Sc. 3. Network-based	Sc. 4. Fisheries data hub
Informing CFP/ IMP / MSFD	+ Data integration allows easier response to EU-wide needs.	+ Improved response to regional policy needs.	+ May generate more detailed data.	+ Data integration allows easier response to policy needs.
Regionalization / coordination	+ Regional data can be generated. - Regional focus is not assured.	+ Regional focus is assured. - Regional coordination is required.	-Coordination required in order to meet specific regional needs.	+ EU-wide and regional evaluation feasible.
Simplification	+ Fewer databases imply less management and update cost.	+Possible simplification from regional perspective, but -not for the EU as a whole.	+No uploading, no replication.	+One system, regular data flow, automatic upload
- **Data upload**	+MS burden in submitting data reduced. +Overlaps eliminated. +Simplification of data upload procedures.	-Some MS upload to several DBs.	+No upload needed, burden to MS reduced.	+Standard procedures and formats. +Data calls replaced by automatic upload.

Policy objective / constraint	Sc. 1. Supra-regional database	Sc. 2.Regional nodes	Sc. 3. Network-based	Sc. 4. Fisheries data hub
- **Data storage**	+Few DBs. +Duplications of data storage are eliminated. -Data stored far from the source. -JRC, Eurostat and ICES may resist to merging DBs.	+Regional biological DBs are operational for an important part of EU fisheries. -Maintaining common standards and processes requires intensive dialogue.	+No replication of data storage[79]. +Data close to source. -Reliance on many national DBs. -Some MS require IT support.	+Comprehensive and transparent EU-wide approach. +Common IT development. -Weaker relation to national and regional specificities -Risk of supra-national 'bureaucracy' in data management
- **Data access / dissemination**	+One DB or DB-system. +Needs of specific users (e.g. RFMOs) serviced by "supra-regional DB", avoiding data collection from MS. +New data products can be developed. -Regional information needs may not be fully met. -Data stored far from the source. -The continuity of dissemination of some products is lost.	+Data closer to producers and users. +Regionalized economic data requires adaptation of sampling. -Calculation of certain totals (e.g. landings) for some MS and EU requires consulting several DBs. -Differences in dissemination policy and confidentiality assurance may arise.	+New data accessible as soon as it is available in the MS. +Common dissemination policy of the nodes. +Primary data close to national users but -Far from EU-level users. -Access rules in different MS may be different.	+Centrally defined access rights. +Common development of data new products. +Consistency, common dissemination policy. -Increased distance from national users.

[79] *The main advantages of the Network scenario is that the primary data is stored only in one place (the national DBs), but the operation of Eurostat's SBS demonstrates that it does make sense to integrate primary data from the MS in one supra-national DB.*

Policy objective / constraint	Sc. 1. Supra-regional database	Sc. 2.Regional nodes	Sc. 3. Network-based	Sc. 4. Fisheries data hub
Quality control	+Common quality rules. +Common nomenclature, quality reports and metadata. +One single source.	+Common quality rules. +Common nemenclature, quality reports and metadata. +Greater transparency on regional level. -Coordination of quality rules required at EU level.	-Coordination of national procedures needed. -Implementation of common rules may be time consuming. -Procedure for centralised checks and corrections cumbersome.	+Common quality rules. +Common referential, quality reports and metadata. +Strong involvement of experts. -Specific national issues must be addressed before upload to hub. -High reliance on automated quality checks.
Bio-economic integration	+Biological and economic variables stored in one DB. +Users access integrated data. -Regional bio-economic analysis may suffer due to lack of regional specificity.	-Obtaining primary data may be difficult because several institutions must provide sufficient guarantees. -Feasibility of regionalization of economic data uncertain.	+Could be achieved through central queries if predefined aggregation levels agreed upon and available in the national DBs.	+Development of common denominators.
Coherence with IFDMP	+Consistency is feasible. -MS IT systems need adaptations.	-Coherence is not self-evident because of different positions of regional institutions.	+Consistency feasible.	+Consistency feasible.
Legal constraints	+Eurostat / JRC / ICES guarantee confidentiality.	-Legal provisions for data confidentiality must be put in place for all regional nodes. -Achieving agreements may take time.	+Confidentiality assured at MS level.	+Eurostat / JRC / ICES guarantee confidentiality.
Administrative constraints	-High requirements on the institution responsible for operating the whole system.	-Availability of required staff depends on funding which is organized differently in different host organizations.	-Varying administrative rules among MS DB operators.	-High requirements on the institution responsible for operating the whole system.
Financial constraints	+Costs of MS reduced. -Development of the DB required – costs of investments. -Overall budget cuts.	-Sufficiency of funding is not assured. -Costs for some MS down and for others up.	-Most costs incurred at MS level.	+Lower costs in long term. -Higher costs in the short term.

Table 27. Evaluation of costs related to the scenarios

	Supra regional	Regional	Network	Fisheries hub
Investment / short and medium term costs				
Hardware	The storage needs are important as all primary data are stored out of MS in a secured environment. Other infrastructure costs should be low Low cost	The storage needs are important as all primary data are stored out of MS in a secured environment. Other infrastructure costs should be low Low cost	The storage needs are minimal; data is stored securely in the MS systems. Other infrastructure costs should be low Low cost	Primary data are stored in the existing structure and in the newly secured storage created. Other infrastructure costs should be low Low cost
Member state side	+Use of available infrastructure.	+Use of available infrastructure.	-Use of available infrastructure but very depends on its quality. Budget should be foreseen to upgrade some MS servers if needed.	+Use of available infrastructure.
Commission side - Access	-Web access to the supra regional system to be developed.	-Web access to the regional system to be developed.	-Fisheries EmodNet portal to be developed in a similar way as the existing ones.	-Fisheries hub portal to be developed.
- Storage	-Supra regional database secured storage.	-New regional databases secured storage where needed.	+Not needed.	+Use of available storage system. -And develop secured storage of data in the fisheries hub when needed
- Communication system between the databases (semantic, vocabularies …)	+Not needed.	+Not needed.	-Emodnet experience and system to be expanded to include fisheries.	-To be developed.
Other IT infrastructure (firewall, LAN, WAN , ….)	+Upgrade available infrastructure if needed.	+Upgrade available infrastructure if needed.	+Upgrade available infrastructure if needed. See the option of collaboration with iMarine for holding the system on sustainable environment.	+Upgrade available infrastructure if needed.

	Supra regional	Regional	Network	Fisheries hub
Software developments (databases)	To be developed on the basis of existing experience. High cost *Software licences/SGBD if needed will depend on the software language, DBMS and server operating system selected.*	Development on basis of RDB-FishFrame. Medium cost *Software licences/SGBD if needed will depend on the software language, DBMS and server operating system selected.*	Development on the basis of EMODnet similar projects. Medium-Low Cost *Software licences/SGBD if needed will depend on the software language, DBMS and server operating system selected.*	Development of a system to link the existing DBs. Medium-Low Cost *Software licences/SGBD if needed will depend on the software language, DBMS and server operating system selected.*
Development of data transmission (upload) format and tools	-Needed	-Needed	+ Not needed	-Needed
Transmission system	-Flux: implementation in each MS. Harmonised standard procedure for the set up.	-Flux: implementation in each MS. Harmonised standard procedure for the set up.	-Implementation of web service and related work (interface, etc.) for connecting to EMOD-net.	-Flux: implementation in each MS. Harmonised standard procedure for the set up.
Support / training	-Needed	-Needed	+Not needed.	-Needed
New software development	-Development of a new system able to manage all kind of data and provide any kind of information on fisheries (including quality insurance procedure and tools to import FLUX compliant format in the supra national database).	+Reuse of existing tool like RDB-FishFrame. -Update RDB-FishFrame to handle economic data. -Extend RDB-FishFrame to MBS. -Develop tools to import FLUX compliant format in RDB-FishFrame.	+Not needed.	+Reuse of existing tool like RDB-FishFrame, DATRAS. -Update of existing tools at JRC to handle the primary economic data. -Develop tools to import FLUX compliant format in the different databases.
Connectivity between the databases	+Not needed, everything in one database.	-Regional databases are not directly connected +but should be interoperable as based on the same package.	+Implemented in the data retrieval system.	+Implemented in the hub portal.

	Supra regional	Regional	Network	Fisheries hub
Confidentiality (especially raw data)	-Access to confidential data (raw data) or restricted data (for scientists) is handled at the software level.	-Access to confidential data (raw data) or restricted data (for scientists) is handled at the software level	-Develop the central node to exploit MS data - This central node handles confidentiality issues, tailored data to the needs of the registered user groups.	-The system handles confidentiality and way to retrieve information from the different systems
Dissemination of restricted or public data	-Development of new web- based dissemination components. -Storage of possible final product (specific aggregates already compiled, ready for use)	-Development of new web- based dissemination components. -Storage of possible final product (specific aggregates already compiled, ready for use	-Adaptation of EMOD-net dissemination interface -Storage of possible final product (specific aggregates already compiled, ready for use).	-Storage of possible final product (specific aggregates already compiled, ready for use.
Operational / long term costs				
Costs	Low	low	medium	low
FLUX support	+Ensured by the Commission for other project as well.	+Ensured by the Commission for other project as well.	+Not needed	+Ensured by the Commission for other project as well.
Web site /database maintenance	+Only one database/web access to maintain.	+Only one package/web site structure shared between different regions to maintain.	-Every upgrade needs to be implemented in all MS. -All (inter)national and institutions face hosting and maintenance costs.	+Same package as now to maintain (ICES, JRC, Eurostat).
Support to MS with low IT resources	-MS has to continue dealing alone with the national system.	-MS has to continue dealing alone with the national system.	-MS has to continue dealing alone with the national system.	+MS could use the online tool to process the data, have a secured place for data storage, benefit from the common tools and process its own data through the hub.
Connection to non-fisheries data	+Low effort - only the supra-regional DB to be connected.	+Possible, only the regional nodes to be connected.	+Connection through other EMOD-net portals.	+Low effort - only the hub to be connected.