

# ICES SC-RDB REPORT 2012

ICES ADVISORY COMMITTEE

ICES CM 2012/ACOM:84

## Report of the Steering Committee for the Regional Database (SC-RDB)

6-7 December 2012

ICES HQ, Copenhagen, Denmark



**ICES**

International Council for  
the Exploration of the Sea

**CIEM**

Conseil International pour  
l'Exploration de la Mer

## **International Council for the Exploration of the Sea Conseil International pour l'Exploration de la Mer**

H. C. Andersens Boulevard 44–46  
DK-1553 Copenhagen V  
Denmark  
Telephone (+45) 33 38 67 00  
Telefax (+45) 33 93 42 15  
[www.ices.dk](http://www.ices.dk)  
[info@ices.dk](mailto:info@ices.dk)

Recommended format for purposes of citation:

ICES. 2013. Report of the Steering Committee for the Regional Database (SC-RDB), 6-7 December 2012, ICES HQ, Copenhagen, Denmark. ICES CM 2012/ACOM:84. 24 pp.

For permission to reproduce material from this publication, please apply to the General Secretary.

The document is a report of an Expert Group under the auspices of the International Council for the Exploration of the Sea and does not necessarily represent the views of the Council.

© 2013 International Council for the Exploration of the Sea

## Contents

---

Executive summary .....	1
<b>1 Introduction .....</b>	<b>2</b>
<b>2 Background .....</b>	<b>3</b>
<b>3 The Regional Database in the Regional Coordination Meetings 2012 .....</b>	<b>5</b>
<b>4 Recommendations to the RDB-SC from the Liaison Meeting 2012 .....</b>	<b>6</b>
4.1 Recommendations from the RCMs supported by the Liaison Meeting .....	6
4.2 Recommendations from the Liaison Meeting .....	8
<b>5 Strategy towards a roadmap for development of RDB-FishFrame, taking requirements of a design based approach to sampling and raising into account. ....</b>	<b>9</b>
5.1 Vision .....	10
5.2 Strategy short-term .....	11
5.3 On line pivot tables are not supported anymore .....	11
5.4 Design based <i>estimation</i> and the RDB .....	11
5.5 Envisaged Roadmap for the adaptation of the RDB for design based sampling and estimation .....	12
5.6 Strategy Long-term .....	13
<b>6 Update of Data Policy Document .....</b>	<b>16</b>
<b>7 RDB Workshops 2013 .....</b>	<b>18</b>
<b>8 Terms of References for RDB-SC 2013 .....</b>	<b>19</b>
<b>Annex 1: List of Participants .....</b>	<b>20</b>
<b>Annex 2: RDB-SC, Fourth meeting, Terms of Reference .....</b>	<b>21</b>
<b>Annex 3: RDB-SC, Fourth meeting, Agenda .....</b>	<b>22</b>
<b>Annex 4: Study Proposal on Development Needs .....</b>	<b>24</b>



## Executive summary

---

The steering committee for the regional database (RDB-SC) met 6-7 December 2012 in Copenhagen, Denmark. It was the fourth meeting of the committee. The RDB-SC consists of representatives from the RCM Baltic, RCM North Sea and Eastern Arctic, RCM North Atlantic, ICES and the Commission and meet 1-2 times each year. The RDB-SC is responsible for strategic planning, technical governance, operational issues and estimates of costs in the overall governance of the regional database (RDB). The RDB-SC interacts with the Regional Coordination Meetings (RCMs) and Liaison Meeting (LM) on other tasks such as development needs and content governance.

In 2012 were the RCMs for the first time working on the basis of the RDB-FishFrame. This meant that their meeting time could be utilized in a more effective way. The RCMs did also put forward recommendations via the LM to the RDB-SC, which subsequently met to respond to these recommendations. The main topics which are discussed in the report are long and short-term development strategies and needs. In short term is the strategy to facilitate data uploads while the RDB in the long term needs to be adapted to a statistically sound design based sampling and estimation approach. Possible development is though dependent on funding.

The RDB-SC did further initiate a revision of the data policy document. The revision will be based on comments on the document from the National Correspondents as well as experiences from the past year. In particular need different types of end-users be better defined and their access to data clarified.

## 1 Introduction

---

The steering committee for the regional database (RDB-SC) is responsible for strategic planning, technical governance, operational issues and estimates of costs in the overall governance of the regional database (RDB) (Fig 1). The RDB-SC interacts with the Regional Coordination Meetings (RCMs) and Liaison Meeting (LM) on other tasks such as development needs and content governance. The RDB-SC consists of representatives from the RCM Baltic, RCM North Sea and Eastern Arctic, RCM North Atlantic, ICES and the Commission and meet 1-2 times each year. The meeting was held in Copenhagen 6-7 December and was the forth meeting of the RDB-SC. Terms of reference, agenda and list of participants are found in annex 1, 2 and 3.

## 2 Background

---

Regional coordination of the data collection underpinning assessment of marine resources and fisheries are primarily handled by the five Regional Coordination Meetings (RCMs). These take place every year to review past sampling and to lay down the rules for sampling coordination for the next year in the region. The aim of the meeting is to achieve adequate international sampling coverage, task sharing and cost efficiency.

The work of the RCMs has not been easy, partly because of the complexity of data collection, but also because no central source of data has been available to perform the analysis necessary for optimization of the sampling schemes and quality of the data achieved, at a regional level. In every case it has been necessary to request data from each country in the region in order to carry out basic analyses, which are necessary for coordination. This process is error prone and also time consuming both for the national institutes and the actual meetings of the RCM which also is reflected in several of the recommendations in the reports of their meetings. This situation has led several RCMs to express a strong need for a Regional Database (RDB) as a data source and tool for their work.

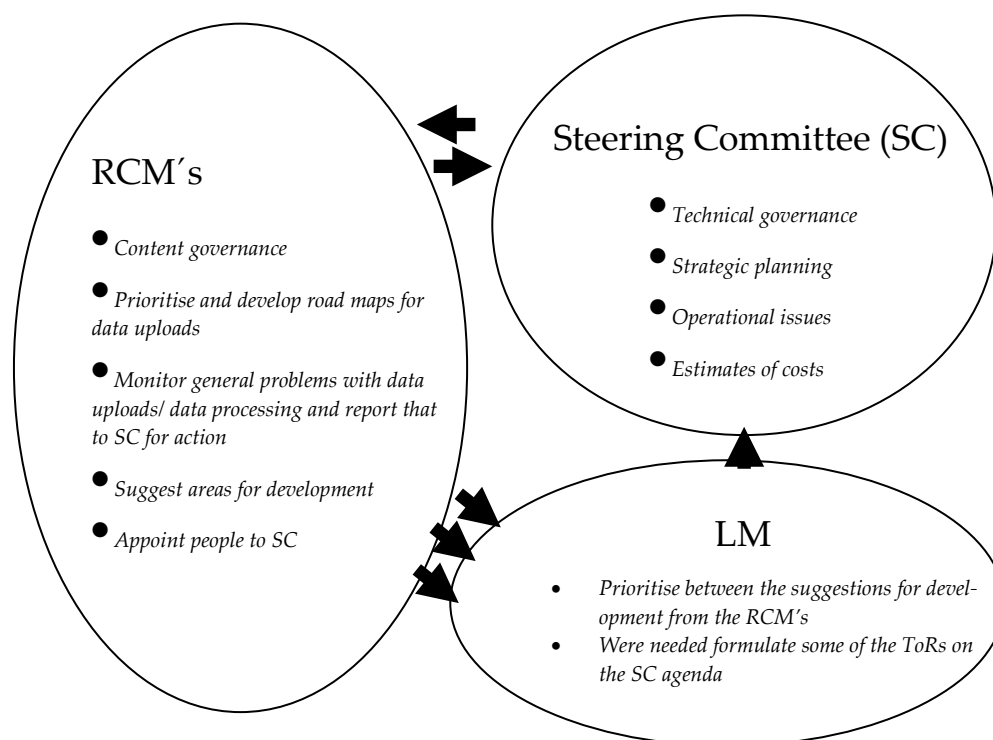
A RDB would also facilitate transmission of data to end-users from an institute perspective where work power can be saved as well as from an end-user perspective where more transparency on the compilation and quality of the data could be achieved. Potential end-users that will benefit from a RDB are thereby all groups which want to make use of tabulations, analyses and graphic presentation of fishery information across countries within a region.

Following a recommendation from the Liaison meeting in 2009 the Commission organized the workshop “Regional scenarios and Roadmap on Regional Database” in 2010 (Anon., 2010). A strong need for a regional database (containing biological and transversal data but also VMS data) was expressed by participants from the Baltic and North Sea regions. For the North Atlantic region the opinions were divided. Participants from some MS saw the possibility to improve the quality of data and data management through a RDB while others considered the present situation with national databases satisfactory and saw a risk with increased workload. The Workshop recommended the development of a roadmap on a regional level to be addressed by the different RCMs giving each region the ability to act on different scenario options. The RCMs (Baltic, North Atlantic and North Sea and Eastern Arctic) responded in their meetings during 2010.

All the three RCMs considered that a database with “disaggregated” (sampling data in detailed form and transversal data in a low aggregated form) data would fulfill most of the needs of the RCM. Such database would facilitate analyses on a regional scale and it will give MS a tool to coordinate their programmes. Also, in order to be able to reply to data requests and transfer data routinely to end-users, it would be more cost efficient to use a RDB and it would provide better quality standards compared to the present situation. RCM NA recognized that not all MS agree to share their data in such a RDB, but expressed that this should not hamper the establishment of a RDB for the North Atlantic region. In the Baltic region MS had already used a regional “disaggregated” database for several years. This database, FishFrame, was developed for this purpose. The experiences with FishFrame were positive and the RCM Baltic decided in 2009 to continue to use FishFrame in the future. In the 2010

meeting of the RCM NS and EA FishFrame was adopted as platform also in this region.

In 2010, the RCM Baltic and the RCM NSandEA recommended an interim steering group to be set up with clear terms of references and mandates in order to start the implementation of a RDB including a Steering Committee (SC). The RCM NA proposed items to be discussed in such a SC. The 7th Liaison meeting endorsed this recommendation. As a consequence an interim steering group consisting of representatives from the three RCMs, ICES and the Commission was put together. This steering group had a meeting in February 2011 in order to elaborate on a governance model for the RDB but also to suggest road maps on how to proceed towards implementation of a RDB from a content point of view as well as from a technical point of view. The outcome of the interim steering group was adopted by the RCMs which also appointed participants to the RDB steering committee (RDB-SC). The first RDB-SC meeting was held in December 2011.



**Fig 1. The RDB; tasks for and interactions between The Regional Coordination Meetings, the Liaison Meeting and the RDB Steering Committee.**



### **3 The Regional Database in the Regional Coordination Meetings 2012**

---

MS participating in the RCM Baltic, RCM NS&EA and RCM NA uploaded data in the RDB-FishFrame as a response of a data call launched by the RCM chairs. The purpose of the data call was twofold:

- To facilitate analyses for regional sampling strategies at the upcoming 2012 RCMs.
- To gain experience in uploading data to RDB-FishFrame and to discuss these experiences at the RCMs

In general, most MS uploaded all requested or at least some data in RDB-FishFrame. This meant that the meeting time in the RCMs could be used in a much more effective way. In previous years a large part of the meeting time have been spent on harmonizing national datasets and compiling them into international overviews. In 2012 meeting time could be spent on analysing data uploaded in RDB-FishFrame and discussing sampling and outcomes of sampling based on those analyses.

In particular the data uploaded by MS to the RDB facilitated analyses on:

- Landings in foreign countries – were do we need to have bilateral agreements in place
- Ranking of métiers to sample
- Regional overviews of sampling intensity
- Data quality at a regional level- simple plots of e.g length-at-age reveal if there are differences in contries or not

One major issue that became apparent is that the RCMs do not know if the datasets are complete or not. This knowledge is of cause essential to the RCM work.

Only two MS did not uploaded any data. Most MS managed to upload at least part of the requested data. There were several reasons for MS failing to deliver all the requested data. These reasons are described in the RCM reports and include:

- Some MS were unable to load specific data types because of coding issues (reference tables in FishFrame)
- Problems to convert data in national databases into the FishFrame format in an efficient way
- Impossibility to report missing values on mandatory field leading to entire sets of valid data not to be uploaded.
- Privacy issues; The fields Vessel\_length, Vessel\_power, Vessel\_size are mandatory in the TR file, MS expressed concerns that these values could matched with fleet registers and individual vessels be identified.

## 4 Recommendations to the RDB-SC from the Liaison Meeting 2012

The liaison meeting (LM) compile all the recommendations from the different regional coordination meetings, comment upon them and proposes points of action. Several recommendations were, during 2012, dealing with different aspects of the regional database and were consequently directed towards the RDB-SC for action. The LM did further make recommendations on its own behalf. Some of these recommendations dealt with the RDB and one of them were directed to the RDB-SC. The recommendations are showed below including action of the SC. It is however important to underline that there presently are no means to fund development of the RDB. The RDB-SC has compiled a proposal for a small-scale study to explore and solve some urgent development needs. The study proposal was put forward to the LM, which gave the proposal a high priority for funding. The LM did further, based on the outcomes of the RCMs, add an extra Terms of Reference dealing with privacy issues, to the proposal. The possibilities for the SC to act on the recommendations are in many cases dependant on that the proposed small-scale project is funded. The study proposal on development needs is found in annex 4.

### 4.1 Recommendations from the RCMs supported by the Liaison Meeting

<b>Regional Database: Access rights to data in the regional database; specific role in the RCM</b>	
RCM NS&EA 2012 Recommendation	Access to data hold in RDB-FishFrame is restricted to persons with a password. Different roles are defined within the system and different users have access to a certain level of data and functionalities. To facilitate future regional coordination work it is recommended that members in the RCMs are given a specific role in the system in accordance with their needs.
Followed actions needed	SC need to identify and specify a role for RCM work and suggest level of access rights. Depending on the work required this may be included in the study proposal for development needs.
Responsible persons for follow-up actions	RDB-SC
Time frame	2013
LM 2012 comments	LM forwards this to RDB-SC and suggest to cover this under the 2013 study proposal if needed regarding technical impact of the implementation of the access rights.
RDB-SC comments	See section 6

<b>Regional Database: Review of the Data Policy Document</b>	
RCM NS&EA 2012 Recommendation	<p>In respect of the development of the RDB and the protection of the data and the ownership of the data, a draft Data Policy Document has been established. The data policy document is based on the current situation but need to reviewed in all its aspects in order to be satisfactory for all MS. The data policy document is a “flexible” document and must be updated as the needs and the development of the RDB are changing. For example, a new data policy document will be prepared if there are changes to the exchange format (update is needed). The document is available in Annex 5 of of the RCM report or through link:</p> <p><a href="https://groupnet.ices.dk/rcm2012/nsea/Report%202012/Forms/AllItems.aspx?RootFolder=%2frcm2012%2fnsea%2fReport%202012%2fToR6%20%2d%20RDB&amp;FolderCTID=%2f61b3b9fd9d2%2d7943%2d41b0%2dae26%2d53e95ed5d50a%7d">https://groupnet.ices.dk/rcm2012/nsea/Report%202012/Forms/AllItems.aspx?RootFolder=%2frcm2012%2fnsea%2fReport%202012%2fToR6%20%2d%20RDB&amp;FolderCTID=%2f61b3b9fd9d2%2d7943%2d41b0%2dae26%2d53e95ed5d50a%7d</a></p>
Follow up actions needed	<p>The National Correspondents (NC) from all MS are requested to read through the document, and sent all remarks and/or suggestions for improvements to the chair of the relevant RCM and to the RDB Steering Group (RDB-SG). Even if the NC has no specific remarks or suggestions, it is recommended to send a notification that the document has been read. Based on the input from the NCs, an updated version will be presented at the next NC Coordination meeting organized by the EC.</p>
Responsible persons for follow-up actions	Chair RCM & RDB-SG, National Correspondents of all MS, EC
Time frame	Before the 15th of November 2012
LM 2012 Comments	LM agrees with this recommendation and recommends the Commission to forward the request to the NC's.
RDB-SC comments	See section 6

<b>Data quality: Standard reports in the Regional database</b>	
RCM Baltic 2012 Recommendation	RCM Baltic recommends that some standard reports should be established in FF that present overview of sampling intensities in maps, tables and figures. The reports would give the regional coordination, assessment working groups and other end-users an overview of the quality of the data in an efficient way.
Follow-up actions needed	A list of useful standard reports should be suggested and discussed in several fora. Input needed from WKPICS, RCM and ICES.
Responsible persons for follow-up actions	RDB-SC
Time frame (Deadline)	As soon as possible. To be considered by the RDB-SC in the further development of new functionalities in RDB-FishFrame.
LM 2012 Comments	LM endorses this recommendation for inclusion in the study proposal by the RDB-SC taking into account the suggestions done by the RCMs, ICES expert groups, RDB WK3 and methodological groups like WKPICS.
RDB-SC comments	See section 5

<b>RCM Baltic 2012 – RDB-FishFrame development: Standard reports in the Regional database</b>	
RCM Baltic 2012 Recommendation	<p>The RCM Baltic recommends that in order to facilitate the data upload process it should be possible to download the look up tables.</p> <p>In addition, for the purpose of the RCM-Baltic report with non-processed data should be developed. As a start very simple reports where it is possible to tabulate the results are needed, see “Overview of used data” for data needed by the RCM-Baltic</p> <p>Pure ‘Data dump’ as raw as the data policy allows could also be a quick way to enable work with the uploaded data.</p> <p>More sophisticated reports with maps and graphs should also be developed, see RCM Baltic 2012 report for inspiration.</p>
Follow-up actions needed	A list of useful standard reports should be suggested and discussed in the RDB Steering Committee.
Responsible persons for follow-up actions	RDB Steering Committee.
Time frame (Deadline)	April 2013 at latest
LM 2012 Comments	LM endorses this recommendation and forwards this to RDB-SC to take into account the suggestions done by the 2012 RCMs.
RDB-SC comments	See section 5

## 4.2 Recommendations from the Liaison Meeting

As regards the future work related to RDB, the LM recommends the RDB-SC to develop a timeline for user objectives and developments in the RDB-FishFrame. There should be two scenarios, one with money and one without.

Action by RDB-SC: see section 5

## **5 Strategy towards a roadmap for development of RDB-FishFrame, taking requirements of a design based approach to sampling and raising into account.**

---

The RCMs summarized their positions on the forthcoming DC-MAP in the “Oostende declaration”. In accordance with this declaration future sampling programmes should be statistically sound and regional coordination strengthened. The RDB is considered a prerequisite for planning and coordination. The RDB need to be developed to meet the requirements of “statistically sound sampling schemes” and consequent estimation processes. The RDB-SC was further informed by the Commission that development aiming to increase support to regional coordination of sampling programmes, strengthen regional data quality checks as well as adaptation to statistically sound sampling programmes were of high priority. The RDB-SC discussed based on these position as well as the recommendation from the LM, short and long-term visions for development.

### **Preamble**

There are three items that must be considered, understood, and agreed on before any meaningful progress can be made on envisioning how a regional fisheries science and advice database for the Northeast Atlantic can be scoped strategically. These items may seem at first glance to be outside any terms of reference that might be considered for a group such as this steering committee. However, they affect the ability of this steering committee to discuss strategic possibilities relevant to the concept of a “regional database”.

The items are:

(i) trust; (ii) allowing legislative and regulatory framework; and (iii) national interest.

(i) Trust: it is necessary that fisheries scientists working in the context of ICES Working Group assessments have access to the complete and any set of data that is required to run and evolve sampling and advice programmes. It is acknowledged that the nature of the engagements and relationships between the players in fisheries advice makes for a rather porous structure. However, there are ways of configuring an IT infrastructure to ensure the data can flow into a virtual advice environment but not out except under privileged accounts (for example, <http://www.lumension.com/Solutions.aspx>). In any case the issue around trust has to be dealt with in a manner whereby it ceases to interfere with the system design process.

(ii) Allowing legislative and regulatory framework: the provisioning of fisheries advice does not have any specific legislation enacted that underpins the development of a system that can give effect to sustainable and repeatable scientific advice. Instead fishery scientists must rely on interpretations and opinions derived from legislation that was designed for other purposes, mainly control and enforcement. This is unsatisfactory and creates an air of uncertainty that detracts from the purpose. Unless fisheries scientists can assert their requirements with those agencies that create the data necessary for generating fisheries advice then it will not be possible to build something that approximates a system. So far, it has been the case that the necessary data are not forthcoming in a continuous, integrated fashion.

(iii) National interest:

Every flag member of ICES has their own distinct implementation of sampling schemes in terms of resource management, logistics, financing, and subsidiary interests. The views and preferences of the flag members must be recognized and if it is not possible for them to adapt, then they must be accommodated. In simple terms, sampling designs that are appropriate to giving advice about the overall position might not be appropriate to subsidiary interests which are equally important from the perspective of the flag member.

The National Interest is often invoked as a stumbling block on the way towards creating a system that will support a Northeast Atlantic fisheries advice system. It is often perceived that there is something slightly atavistic and negative about a flag member maintaining a strong national interest when in fact it is a perfectly valid and understandable position to hold. The existence of a regional database system should not negate the validity of any national interests and such considerations are already protected under the principle of subsidiarity which has a preeminent place in EU legislation.

These items outlined above have been and are particularly difficult for the committee to discuss. They represent a major block towards reaching an agreed long-term strategic vision for a regional database and compromise the short and medium term efforts by obliging flag members to adopt approaches that are by their nature limited in scope and fail to deliver on expectation. It is important that these issues be discussed.

## 5.1 Vision

Statistically sound sampling schemes (SSSS) by which we mean the move to design based regional sampling as envisaged under the DCMAP and outlined in the Oostende Declaration (RCM NS & EA 2012).

It is envisaged that a development of RDB for SSSS runs parallel with the continued and existing use of the RDB to

- 1 ) provide a raising (= estimation) method for countries using FishFrame at present (mainly in the Baltic)
- 2 ) serves the needs of the Regional Coordination Groups (RCGs)

It is envisaged that the process toward SSSS will be a gradual stepwise approach that will take differing periods of time in different nations. That said substantial progress toward implementation is to be anticipated by the review of DC MAP (2017?) as outlined in Oostende declaration. It is probably realistic that at this time design based sampling can be accommodated via, or within, the RDB for at least some nations and some stocks.

We make an important distinction between three distinct aspects of statistically sound sampling schemes; sampling design, data collection, and estimation.

- *Sampling design* is the overall design used to gather data (involving sampling frames, sampling protocols etc)
- *Data collection* is the process of recording data at a sampling event (observers collecting age data, measuring fish and shellfish at sites onshore and at-sea)
- *Estimation* is the process of deriving estimates relevant to the population of interest from the collected sample data.

In what follows we are primarily concerned with estimation, but estimation cannot be considered in isolation from sampling design or data collection.

## 5.2 Strategy short-term

The short-term strategy is to support all countries in importing data into the RDB FF, so the Regional Coordinating Meetings (RCM) can use the RDB FF as the main source for gaining information of the national sampling and fisheries ongoing. This means the problems some countries have regarding mandatory fields for which there is no national data, should be solved. To make information/data in the RDB FF available for the RCMs, it is needed to make some standard reports, which allow the RCM to use the information in the RDB FF. For those stocks which are using the RDB FF for preparing data for the stock assessment, the 'length-overwriting' bug should be fixed. All of the goals below are subject to the very limited budget:

- 1 ) Solve the overwriting length data bug (completed to be determined (before the WGBFAS 10<sup>th</sup> April))
- 2 ) Solve issues with mandatory fields (completed before June 2013 because of a planned work shop)
- 3 ) Update reports for the RCMs (completed before the RCMs (before September 2013))

## 5.3 On line pivot tables are not supported anymore

In the short term the standard output reports from the RDB FF, will be focus on creating standard tables in comma separated values files (CSV-files). The 3-party component in the RDB FF, which make it possible to work with a pivot table inside the RDB FF, is not supported anymore. The 3-party pivot table is also based on a technology, which are using the Microsoft cubes, this technology have also be abandon by Microsoft, so it not supported in the programming language ASP.NET, in which the RDBFF is programmed. On top of this the nightly recalculations of the cubes is taking so long time, that it is start to be a problem, there is not enough hour in the night. So it is obvious that in the short-term future pivot reports will not be created. In the future pivot tables can be supported again if prioritized. Probably a new 3-party component which fulfils the needs will be used. For the standard output reports needed by the RCM, it is suggested that two related standard reports, CSV-files, are created; one with the table, exactly as it is needed for the RCM report, ready to paste directly in the RCM report, and another table with all detailed data to be used for deeper analysis in R or Excel.

## 5.4 Design based *estimation* and the RDB

The fundamental purpose of the collection of data are to provide statistically sound estimates to expert working groups. This depends on having an appropriate design for the particular fishery, collecting data using probability based sampling *and* using the appropriate estimation process for the design under which the data were collected. The estimation stage is an integral part of each specific catch sampling scheme and while there is much in parallel in the estimation process between schemes, the specifics of the estimation process are dependent on the specifics of the catch sampling programme. This is recognized in the envisaged role of national sampling schemes as strata within a regional design (elaborated upon in reports from PGCCDBS 2012, RCM NS& EA 2012, RCM NA 2012). Provided national designs and estimation is done correctly, regional estimation should be a relatively simple process of summation across correctly weighted national estimates.

Design based estimation is not currently possible in RDB FishFrame, Intercatch or COST (WKPICS 2), all of which are based on the (over stratified) métier system, where sampling probabilities are largely ignored and there is extensive reliance upon the use imputation for domains lacking sample data.

## **5.5 Envisaged Roadmap for the adaptation of the RDB for design based sampling and estimation**

### **1. Revision of the csData exchange formats**

Revision of the csData exchange formats for sampling data (csData) is needed to create a common data template for the recording of sampling data collected using probability based sampling. This needs to encompass all the various collection scenarios encountered in the diverse national fisheries operating at national levels (as identified in the 2012 RDB workshops) and enable the calculation or recording of sampling probabilities.

This revised data template should be extensively tested prior to an envisaged switch from the existing data exchange format.

### **2. Revision of the cl and ce data formats**

Design based sampling involves the use of sampling frames and primary sampling units (PSUs). For onshore sampling these frames will be based on landing and auction sites, and for at-sea sampling the frames will usually be based on vessel lists. Including the data relevant to these frames and allowing aggregation of data to stratifications specified in sampling designs has implications for the exchange structures of the landings and effort data structures (cl and ce data) used in the RDB at present. Additionally the requirement to be able to calculate sampling probabilities necessitates additional fields, and a level of disaggregation of landings and effort data, currently not supported by the cl and ce exchange structure RDB has inherited from FishFrame. A revision of landings and effort data structures is thus required for design based sampling.

There are potential confidentiality issues that complicate the use of landings and effort data at the most disaggregated form within the RDB. However these issues were considered primarily to be technical and political in nature and are not addressed here.

### **3. Estimation**

For design based estimation WKPICS2 suggested the use of the R statistical language and the package "R survey" (CRAN, Lumley 2010) be explored as a cost-effective and practical means to allow estimation of design based surveys in the medium term. Whether this is done outwit the RDB or as some module called within the RDB in the future has yet to be determined.

### **4. Basic database housekeeping**

Issues such as version control, data integrity, code lists, etc need to be addressed.

The data flow pathway is star-shaped meaning that data inputs flow from the points of a star to the hub or centre of a star. The points of the star represent the flag members of ICES and the hub represents ICES itself. From the hub, data flows back out to the points of the star. The data flows are asymmetric. This means that the data that flows from the points to the hub is not the same type of data that flows from the hub to the points. Specifically, flag members supply raw data sufficient to run the centralized system and in return benefit from the analyses of the hub and all the time being



entitled to re-retrieve their own raw data if required, along with any aggregated analyses. Critical to the smooth operation of such a system is the ability to verify and synchronize with certainty data differences between the hub of the star and the points of the star. Data feeds are effectively snapshots at points in time and require that the temporal nature of the snapshots be maintained. By implication, this means that the data management systems of the hub, at the very least, have this ability. Ultimately, the member flags will acquire the same capability to manage data by collaborating with each other and with ICES to evolve a software framework (i.e. working code) that will underpin the data management of raw data (i.e. all relevant data collected by member flags) and meta-data between the hub and the points of the star. This can be supported by using decentralized software version control tools like *git* or *Mecurial* which allow and encourage open participation and development while still permitting centralized control of an official version of the code.

## **5. Engagement with end-users**

The migration of national sampling schemes to design based sampling and SSSS will have positive benefits to for assessment working groups. However it will also involve a change of perception by these EG of the data with which they are provided. This is most pertinent to MIXFISH and STECF because these groups are used to receiving raised estimates at fleet or métier level. Currently Intercatch and FishFrame use extensive imputation to estimate for métier based domains, under SSSS this practice would no longer be needed, being replaced by domain estimation based on the realized samples. Reliable estimates for domains require sufficient samples sizes, and which domains currently have reliable estimates needs to be explored in a dialogue between the data scientists and assessment scientists.

## **6. Interactions with other database developments**

The establishment of other regional databases, for example for the Mediterranean and for large pelagic long distance fisheries, are likely to have many parallels with the RDB. RDB SC considered that liaison with other developments to be prudent, but also that the focus of the SC should remain with fisheries within the areas (Baltic, North Sea, and North Atlantic) served at present.

## **7. Timeline and Funding**

It is envisaged that development work on RDB will be undertaken by individuals working in or in collaboration with national instates and coordinated through ICES. This requires funding in some form and the mechanisms of this need to need to be worked out. The timeline for development is linked to the funding available to the RDB.

### **5.6 Strategy Long-term**

The long-term strategy is to make the RDB work according to the principles of the statistical sound sampling schemes. The following approach for the development of the RDB was embraced, with focus on harmonization, standardization, security and being able to document and audit, see the figure below. To be able to document 'who, what, where, when' and to ensure only approved standardized methods and algorithms have been used to raise and aggregate data, a system model, where data goes in one line from national data input to international output is proposed, just like the supply chain model (see Connolly & Caffrey 2011). The main purpose is to ensure only standardized methods and algorithms are used for all countries and stocks. So independent of countries and stocks only the highest level of raising methods are

used, so those countries or stock coordinators, which do not have the expertise and tools to make and use the latest best practice standard methods and algorithms safe and easy can use the methods and tools from the experts implemented in the RDB. In this way raised/estimated data for all stock assessments can fulfil the requirements for being statistical sound/best practice and documented. Another advantage in using the RDB is that an enormous amount of resources are freed, which can be used for further improvement of the assessments, and not spend in individuals just get up to a decent level in data raising. The goal is to let the RDB produce as many of the data requests as possible – both international and national, again for better cost benefit. Of the international there are two obvious ones; standard sampling tables for the RCMs and stock data for the assessment expert groups, but more will come. Preferably all relevant output formats should be supported; CSV, R objects, XML and JSON. To make data available for analyses and to ensure continues development, data can be accessed and downloaded at any node/link in the chain according to the data policy rules. This will allow members of the Regional Coordination Meetings and stock assessment expert groups to analyse data in the chain and suggest new functionalities/develops, which will be included in the RDB after approval, to ensure the validity of the suggested methods. Preferably a repository of software packages like the COST, which could be used for analysis could be set up.

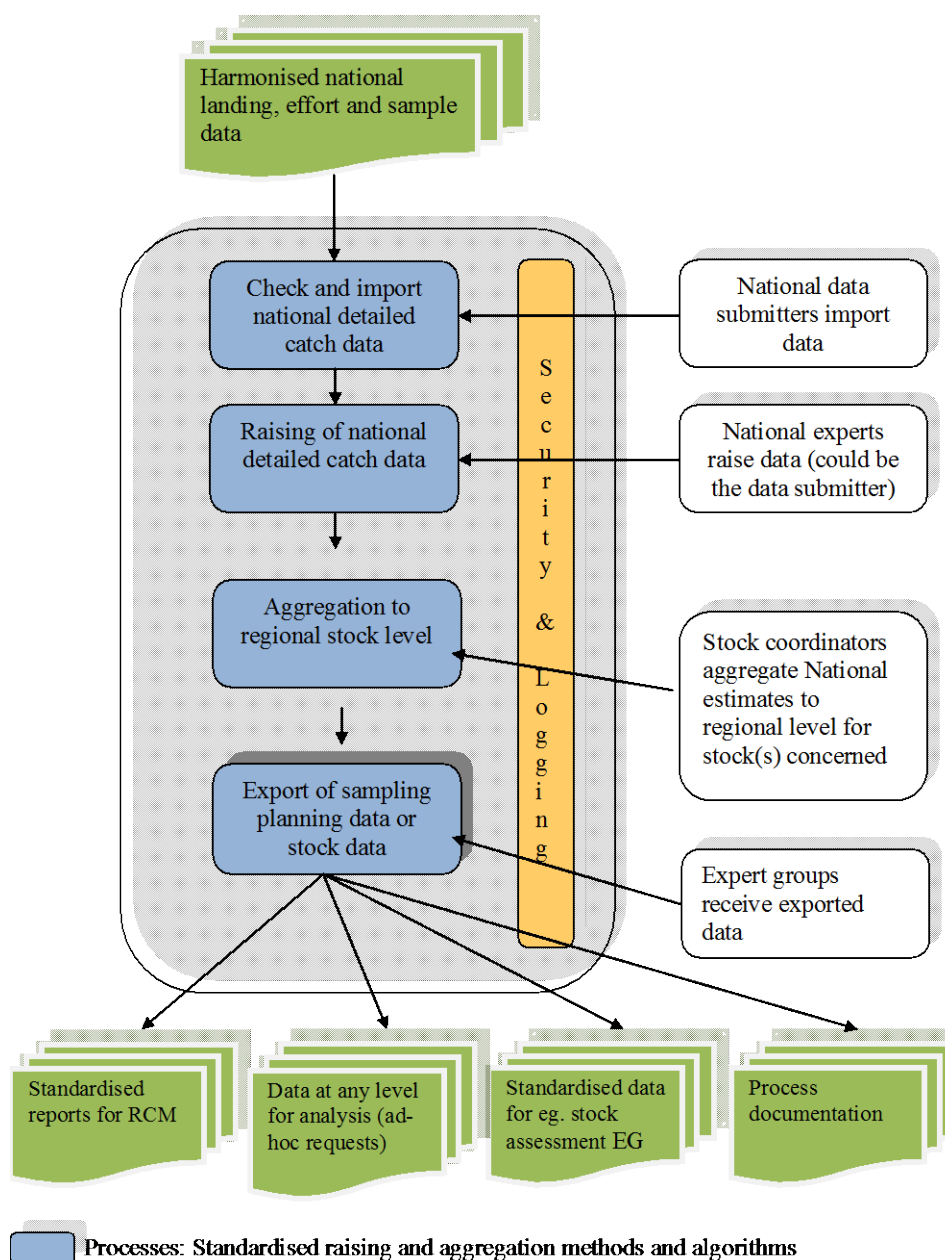


Fig 2 showing the long-term vision of the RDB

## 6 Update of Data Policy Document

---

The RDB Steering Committee developed and adopted earlier in 2012 a Data Policy Document. The goal of this document is to define how the data uploaded into the RDB-FishFrame are stored and used in accordance with agreement made between the data submitters, data users and the RDB host (ICES). The document was provided to all MS to facilitate the data upload for the 2012 RCM data call. The data policy document is based on the current situation but may need to be reviewed in order to be satisfactory for all MS (RCM NS&EA 2012). It has further been recognized during the year that different types of users, including RCMs, and their access rights need to be better defined in the document. This is in order to clarify what expectations different user types currently can have on access to the RDB data as well as to support the host when data requests are coming.

All national correspondents were in response to a recommendation from RCM NS&EA (section 4) approached to comment upon the present data policy document. The RDB-SC received comments from Belgium, Spain and France. A subgroup consisting of Els Torreele, Neil Holdsworth, Jörgen Dalskov, Christian Dintheer, Jose Rodriguez and Katja Ringdahl was put together to intersessionally summarize the comments and suggest changes in the document.

The tasks for the subgroup include:

### **Summarize and respond to comments from the NCs**

- Summarize comments from NCs by category e.g.. Exchange format, lack of knowledge, privacy/legal issues, editorial etc. This summary should be presented at the next relevant NC meeting and should also be made available as an FAQ at the RDB website.
- Propose update of the text in accordance with relevant comments

### **Definition of End-users and access rights**

- Map roles in FishFrame system to the groups of users that we should define, so it is a clear matrix for managers and users to see where they fall into and what access they will have. The matrix need to be focused and practical, so that the policy acts as guidance to users/managers so it is unambiguous when a request comes in
- Simplify access roles, so that it is perhaps 3 types of users mapping to the following roles in RDB:
  - RCMs + DG-MARE + ICES Advisory Working = National Data editor
  - STECF = Data reader
  - Everyone else = guest role
- Reports on what is available/overview in the database should be made public through 'guest' role and therefore a way of asking for access knowing what is available and what is relevant
- Rewrite the section on access rights in the data policy document

### **Other issues to be considered in the revision of the data policy document**

- Definition of terms can be used from legal text plus existing definitions from ie. ICES Data policy, ISO 19115 Metadata standard

- There needs to be a transaction log that records all extractions so that on request, a National Correspondent can have a list of all extractions from the RDB that have occurred involving national data
- Need to add some text about this i.e. section on data security in ICES Data policy for example

## **7 RDB Workshops 2013**

---

During 2012 were three training workshops held in order to allow institutes to get familiar with the present functionalities in RDB-FishFrame. All three workshops were chaired by Henrik Degel and Kirsten Birch Håkansson, DTU-Aqua. The first workshop were dealing with the present processing functionalities, the second on data uploads and the third on data raising and output modules. It was also discussed during the workshops how the present functionalities meet the requirements of statistically sound sampling. The general view expressed at various RCMs meetings in 2012 was that the workshops were beneficial and appreciated. The workshops increased the overall knowledge in the facilities within FishFrame.

It is evident from the 2012 workshops but also from other groups such as ICES WKPIC2 that some of the present data processing and raising modules will need to change when data collection moves from a métier based system to a design based approach. The RDB-SC thereby agreed that there is no need to have more training workshops on data processing and raising until these adaptations are developed (see section 5). For 2013 will only one workshop be held. This is the workshop on data uploads that support MS to upload data into RDB-FF prior to the RCM data call. The training workshop will have the same ToR as the workshop in 2012 and will be chaired by Henrik Degel, DTU-Aqua and Henrik Kjølms Nielsen, ICES.

## **8 Terms of References for RDB-SC 2013**

---

The terms of reference will be prepared by the chair and ICES secretariat prior to next RDB-SC meeting.

## Annex 1: List of Participants

NAME	ORGANIZATION	E-MAIL
Neil Holdsworth	ICES	<a href="mailto:NeilH@ices.dk">NeilH@ices.dk</a>
Jörgen Dalskov	RCM Baltic	<a href="mailto:jd@aqua.dtu.dk">jd@aqua.dtu.dk</a>
Henrik Degel	RCM Baltic	<a href="mailto:hd@aqua.dtu.dk">hd@aqua.dtu.dk</a>
Katja Ringdahl (chair)	RCM Baltic	<a href="mailto:katja.ringdahl@slu.se">katja.ringdahl@slu.se</a>
Sieto Verver	RCM NS & EA	<a href="mailto:sieto.verver@wur.nl">sieto.verver@wur.nl</a>
Els Torreele	RCM NS & EA	<a href="mailto:els.torreele@ilvo.vlaanderen.be">els.torreele@ilvo.vlaanderen.be</a>
Christian Dintheer	RCM NA	<a href="mailto:christian.dintheer@ifremer.fr">christian.dintheer@ifremer.fr</a>
Liam Caffrey	RCM NA	<a href="mailto:lcaffrey@marine.ie">lcaffrey@marine.ie</a>
Alastair Pout	RCM NA	<a href="mailto:a.pout@marlab.ac.uk">a.pout@marlab.ac.uk</a>
Kirsten Birch Håkansson	chair WKRDB	<a href="mailto:kih@aqua.dtu.dk">kih@aqua.dtu.dk</a>
Henrik Kjelms Nielsen	ICES	<a href="mailto:henrikkn@ices.dk">henrikkn@ices.dk</a>
Cristina Mogado	ICES	<a href="mailto:cristina@ices.dk">cristina@ices.dk</a>
Jose Rodriguez	Observer, Spain	<a href="mailto:jose.rodriguez@st.ieo.es">jose.rodriguez@st.ieo.es</a>



## **Annex 2: RDB-SC, Fourth meeting, Terms of Reference**

---

**The steering committee (SC) for the regional database (RDB) will meet the 6-7<sup>th</sup> of December 2012 in Copenhagen in order to**

- a) Respond to recommendations put forward to the SC by the Liaison meeting and summarize how FishFrame has been used in the regional coordination meetings
- b) Update the data policy document dealing with access rights, data confidentiality and data ownership issues.
- c) Develop a strategy including a work plan for a road map on development of FishFrame, taking requirements from a design based approach to sampling and raising into account.
- d) Plan and agree on ToRs for the RDB workshops 2013
- e) Agree on ToRs for the SC 2013

### **Annex 3: RDB-SC, Fourth meeting, Agenda**

---

#### STEERING COMMITTEE FOR THE REGIONAL DATABASE

COPENHAGEN 6-7 DECEMBER 2012

#### *AGENDA*

Achieve

- Data Policy Document - Summarize comments from NCs and update data policy document were needed
- Status of FishFrame – identify for what purposes FishFrame can be used today
- Development of FishFrame - Strategy and a work plan for a road map on development of FishFrame, taking requirements from a design based approach to sampling and raising into account. Timeline with or without funding from study proposal. We need to agree on way forward and compile a subgroup that can work intersessionally.
- Agree on subjects for RDB workshops 2013
- Agree on generic ToRs for SC meetings 2013

6 December

- 13:00 Opening of the meeting and appointment of rapporteurs
- 13:10 Presentation of outcomes RDB-3 (Henrik D or Kirsten)
- 13:30 Presentation of outcomes (relevant to the SC) from WKPICS2 (?)
- 13:40 Feedback from RCM NS&EA (Els)
- 13:50 Feedback from RCM Baltic (Jörgen)
- 14:00 Feedback from RCM NA (Sieto)
- 14:10 Feedback from LM (Katja)
- 14:20 Status of FishFrame - identify for what purposes FishFrame can be used today
- 15:00 Coffee break
- 15:30 Status of FishFrame (continue)
- 16:00 Strategy and workplan for road map on development of FishFrame.....
- 19:00 End for the day

7 December

- 8:00 Feedback on data policy document from National Correspondents (Jörgen?)
- 8:30 Discuss and update data policy document. Agree on way forward
- 11:00 Agree on subjects and ToRs for RDB workshops 2013
- 12:00 Lunch Break
- 13:00 RDB workshops 2013 (continue)
- 14:30 Agree on generic ToRs SC meeting 2013
- 15:00 End of meeting

## **Annex 4: Study Proposal on Development Needs**

---

### **Proposal: Exploration and Development of new facilities in RDB-FishFrame 5.0**

**Duration: 18 months**

**Proposed budget: 450 000 Euros**

#### **Background:**

The demands from the users to a regional Database is under constant change; in the first hand, because the users discover new possibilities in the use of the data as they get more familiar with the use of the database and second because the fish stock management and modelling environment changes and new data types become important. The first one mostly requires design of new output reports to tabulate new combination of the existing variables, while the second one quite often requires adding of new variables and processing functionality. Furthermore, RDB- FishFrame has now been introduced to additional regions. This has given rise to additional requests how data should be centrally processed due to new sampling stratifications practiced in the member states included compared to the existing. It is essential that a database reflects on new demands and not act as a straightjacket preventing new progressive initiatives. A constant development is therefore very important in order to keep the momentum.

The development will be outsourced to the extent that external expertise is necessary in order to follow the time schedule.

#### **Development**

The main fields for development in 2013-14 are identified by the RDB-Steering Committee and presented in no specific order of priority:

- 1 ) Development of additional tools for analysis and data tabulating to support regional coordination. (20% of total budget)

Outputs: Technical report, programming development

Development of output reports which provide:

- Overview of data status by region; data coverage;
- Support the planning of future regional based sampling schemes;
- Overview of potential areas for task sharing between member states.

- 2 ) Explore options and cost implications of implementing of external tools (i.e. COST) in the RDB-FishFrame. (35% of total budget)

Outputs: Technical report, Technical Workshop(s), programming development

Such analysis should include the following elements:

- An inventory to collate and examine the tools present but also tools missing
- What level of documentation/quality controls would be required of a tool to be accepted into the RDB?
- What exports should the RDB provide to other formats/tools?

- What changes need to be made to the COST format/coding to comply with the RDB?
- Is COST sufficiently documented (methods, quality controls etc.)?
- Which level of integrating should the RDB-FishFrame provide to COST (just export to COST or an interface that allows users to manipulate RDB data using COST tools/functions)?
- Proof of concept of programmatic interface to RDB-FishFrame

3 ) Requirements and automatisisation of Data calls procedures. (20% of total Budget)

Outputs: Technical report, programming development

- What is formally required from the regional database to reply to data calls?
- What data calls can we respond to at present/future? (The present functionalities and documentations in the regional database need to be compared with most common data calls)
- Alignment with FLUX developments

4 ) Development of more flexible structure to handle correct processing of design based sampling schemes to address regional differences in approach. (25% of total budget)

Outputs: Technical report, Technical meetings/workshops covering all regions

- What changes need to be made in the Exchange Formats in order to comply with design based sampling schemes?
- Which additional processing functionality need to be developed in order to comply with design based sampling schemes?

5 ) Development of procedures to ensure confidentiality on individual vessel level for CL, CE and on value.