EU request on quality assurance for data collected under the DCF, with the specifications of the RDBES for commercial catch sampling data

Summary

Two workshops were conducted in 2018 to provide new specifications for the RDBES data model that determines which data is transferred, stored, and used for estimations. The data model is the fundament for the database and the web system surrounding it. The Regional DataBase (RDB) stores the commercial fish catch and sample data at detailed level for the Regional Coordination Groups for the North Sea & Eastern Arctic, the Baltic Sea, and the North Atlantic. However, the current RDB does not support storage and estimations based on statistical data. The RCGs has therefore requested that specifications are developed for a new version of the RDB; the new version will have a new name, Regional DataBase and Estimation System (RDBES), to clarify that it also contain an estimation module. The RDBES will support the countries' shift in moving towards a design-based sampling and estimating the catch and sampling data using statistical estimation methods. The aim is also that ICES stock assessment groups and other ICES groups working on improving the assessments will use the detailed data in the RDBES such that data collection and the estimation/raising of data at the detailed level will, for the first time, be transparent and quality assured.

The main goal of the workshops was to further specify the data model and fill in the various sampling design/hierarchies, and to test if the data model could contain commercial fisheries sample data for the typical sampling designs or whether the data model needed to be improved. Testing the data model with real data allowed for missing information and fields to be identified, e.g. the group of fields presenting clustering sampling, as well as weight frequencies and several other fields. The data model was tested in various stages; in the first stage typical sample data for one of the hierarchies was used to fill in just a few lines of data in each of the tables of the hierarchy in a spreadsheet. The last test was to create a small data file with real data in the correct format. A very important part of the work has also been to ask countries to look into the RDBES data model and get feedback from the countries, because the use of the RDBES will have a huge impact on data and assessment work in the countries. The RDBES should be able to handle sampling designs from all countries and data for all species; this turned out to be a much more difficult and complex task than first expected. The feedback from the countries was used to make changes to the data model, and answers to the countries were given. After the first workshop the data model was used in the development of the RDBES web system and database, which means that changes to the specifications in the data model also had to be made in the database and system. That has been a more time consuming process, but that has been the case, because it has been so difficult and taken longer time to specify the RDBES data model than expected. The data model for the commercial fisheries data is at a stage where it is finalised for now. However, most of the second workshop, originally intended for requirement specification of the RDBES system, was spent finalising the data model, which is the crucial part. At the second workshop the first schematic overview of a statistical estimation procedure was outlined and the specification of code lists, unique indices, user roles, and overwriting rules was only discussed shortly. However, in a third workshop funded by the countries participating in the specifications of the RDBES system and by ICES Secretariat, the first specifications of code lists, user roles and overwriting rules were produced, enough to continue the development of the RDBES web system and database. In summary, the specifications of the "Input" parts of the RDBES have been specified to a high degree, while the specifications of the "Processing and Output" part of the RDBES have been specified to a low degree. The code and test the feasibility of the outputs from the workshops in the implementation of the RDBES have been done to a high degree.

Request

Background

Quality assurance of advice starts by quality assuring the input data to stock assessment. There are several stages and tasks underlying this; first steps being to assure that all data are held in databases with transparent and quality assured outputs (e.g. survey indices, raised catch data) and even before this that the data are produced in a quality assured way.

ICES was requested by EU to engage in

3 steps of the quality assurance of data collected under the DCF:

a) development of the SmartDots platform for age-data,
b) specifications of the RDBES for commercial catch sampling data, and
c) including DCF-funded fisheries independent surveys/variables not yet held in DATRAS into the DATRAS database.

All three developments would enable ICES to reduce uncertainty/error, but also more precisely quantify errors in input to assessments.

The present Technical Service refers specifically to step
b) specifications of the RDBES for commercial catch sampling data
Support the development of a full specification of the RDBES, including a regional data model for commercial catch design based sampling and agreed estimations methods

The existing RDB does not support statistical estimations and design based sampling, therefore the development of a new version of the RDB, the RDBES (RDB and Estimation System), has been started. The RDBES will fully support the RCG BS, RCG NS & EA and RCG NA and potentially RCG LDF, RCG LP and RCG MED with modifications. The specifications depend entirely on the availability of the few experts in the ICES community who need to be allocated specifically for this task by their National Institutes and as well on the ICES secretariat beyond regular maintenance.

ICES is requested to carry out the

- specification of the 'Input' part of the RDBES
- specifications of the 'Processing and Output' part of the RDBES
- to code and test the feasibility of the outputs from the workshops in the implementation of the RDBES

Elaboration on the advice

The Workshop on new data model for the Regional Database (WKRDB-MODEL) met in Copenhagen on 15–18 January 2018, chaired by Henrik Kjems-Nielsen, ICES Secretariat. The meeting had six attendees, representing five countries (Denmark, Ireland, the Netherlands, Norway, and Sweden).

The Workshop on new specification for the Regional Database (WKRDB-SPEC) met in Copenhagen on 3–6 April 2018, chaired by Henrik Kjems-Nielsen, ICES Secretariat. The meeting had five attendees, representing four countries (Denmark, Ireland, the Netherlands, and Sweden).

Links to reports from the meetings

ICES. 2018a. Report of the Workshop on new data model for the Regional Database (WKRDB-MODEL), 15–18 January 2018, ICES HQ, Copenhagen, Denmark. ICES CM 2018/ACOM:41. 40 pp.

http://www.ices.dk/acom/2018/WKRDB/wkrdb-model_2018.pdf.

ICES. 2018b. Report of the Workshop on new specification for the Regional Database (WKRDB-SPEC), 3–6 April 2018, ICES HQ, Copenhagen, Denmark. ICES CM 2018/ACOM:62. 55 pp.

http://www.ices.dk/sites/pub/Publication Reports/Expert Group Report/acom/2018/WKRDB/WKRDB-SPEC.pdf.