# ICES WKIACTDB REPORT 2015

STEERING GROUP ON INTEGRATED ECOSYSTEM OBSERVATION AND MONITORING

ICES CM 2015/SSGIEOM:35

REF. SCICOM AND DIG

# Report of the Workshop on the review of the ICES acoustic-trawl survey database design (WKIACTDB)

1–2 October 2015

ICES Headquarters, 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
info@ices.dk

Recommended format for purposes of citation:

ICES. 2016. Report of the Workshop on the review of the ICES acoustic-trawl survey database design (WKIACTDB), 1-2 October 2015, ICES Headquarters, Copenhagen, Denmark. ICES CM 2015/SSGIEOM:35. 17 pp.

# Contents

# Executive summary

## Highlights

The WKIACTDB workshop is part of the Horizon 2020 (H2020) project AtlantOS task 2.4, and the objective was to review and consolidate a model for storing and extracting acoustics data at the ICES Data Centre, to support storage and processing of acoustic survey data from the ICES coordinated acoustic surveys.

Two post-processing software systems (StoX and EchoR/Echobase) were presented and the data requirements for these packages was taken into account when reviewing the requirements to the data storage at the ICES Data Centre.

The ICES Data Centre presented a first version of the data model, both for acoustic data by category and the DAtabase of TRAwl Surveys (DATRAS) extension for pelagic hauls, and the details for each model was discussed in detail.

After the workshop the ICES Data Centre compiled the information and set up a first specification. This specification is the AtlantOS internal milestone 2.4.1. that includes the data model and specifications of the API's, and the document is included as Annex 4 to this report.

During the workshop few recommendations came up to survey groups and other Expert Groups (EGs) that need attention for the system to be operational.

# 1 Terms of reference

A **Workshop on the review of the ICES acoustic-trawl survey database design** (WKIACTDB), chaired by Neil Holdsworth*, Denmark, and Nils Olav Handegard*, Norway, will meet in Copenhagen 1-2 October 2015 to:

a) Review the proposed design for the ICES acoustic trawl survey database. The first step is to evaluate the existing databases that are currently in place, including the IMR acoustic database (IMRdataBase.doc), PGNAPES db (WGIPSdataBases.doc), Ifremer dB, and FishFrame. The report will evaluate to what extent the proposed design will be meeting the requirements from:

i ) the ICES working groups responsible for carrying out acoustic surveys (WGIPS, WGACEGG, WGBIFS and WGIDEEPS)

ii ) international standards including those adopted by the Atlant OS project.

iii ) The ICES acoustic metadata standard developed by WGFAST.

WKIACTDB will report by 1 November 2015 for the attention of SGIEOM and DIG.

## Supporting information

| | |
|---|---|
| Priority | ICES have got funding from the Atlant OS project under task 2.4 to develo an acoustic trawl survey database for ICES. To ensure a wide adoption of the system within ICES, an ICES workshop is proposed to ensure that the needs for those not within the project are taken into account. |
| Scientific justification | To improve the fish survey data availability through the ICES data center there is a need to host interpreted acoustic data (NASC values) and corresponding meta-information, in addition to archiving biotic sampling data. This is important for secure storage and consistency between groups doing acoustic surveys, and for data availability. This is one of the main datasets for fisheries advice on pelagic species, including herring, blue whiting, sardines, sprat etc. This is relevant to expert groups that coordinate surveys and feed the ecosystem assessment process in the Nort Atlantic.<br><br>Software that runs on top of the database will be developed based on best practice techniques as part of the AtlantOS project. |
| Resource requirements | Funding is covered by the H2020 project AtlantOS. |
| Participants | Representaives from WGFAST, WGIPS, WGACEGG, WGBIFS and WGIDEEPS |
| Secretariat facilities | The datacentre is part of the project and will be part of the WK. |
| Financial | No financial implications. |
| Linkages to advisory committees | Supports the expergroups that give advice on pelagic species observed by acosutics. |
| Linkages to other committees or groups | WGFAST, WGIPS, WGACEGG, WGBIFS and WGIDEEPS |
| Linkages to other organizations | It is part of the AtlantOS project |

## 2 Introduction and background

Nils Olav Handegard, SSGIEOM Chair, and Neil Holdsworth, head of ICES Data Centre, welcomed the participants.

The meeting was opened by short round table introductions of the participants, followed by an introduction about the aim of the Workshop and overall activity in relation to the AtlantOS project and specifically to task 2.4. The head of data support introduced the overall dataflow and technical development of the Acoustic DB, and how it fits in with the other system at the ICES data centre (Figure 1).

*An important aspiration is that simple input and output procedures is a key to success.*

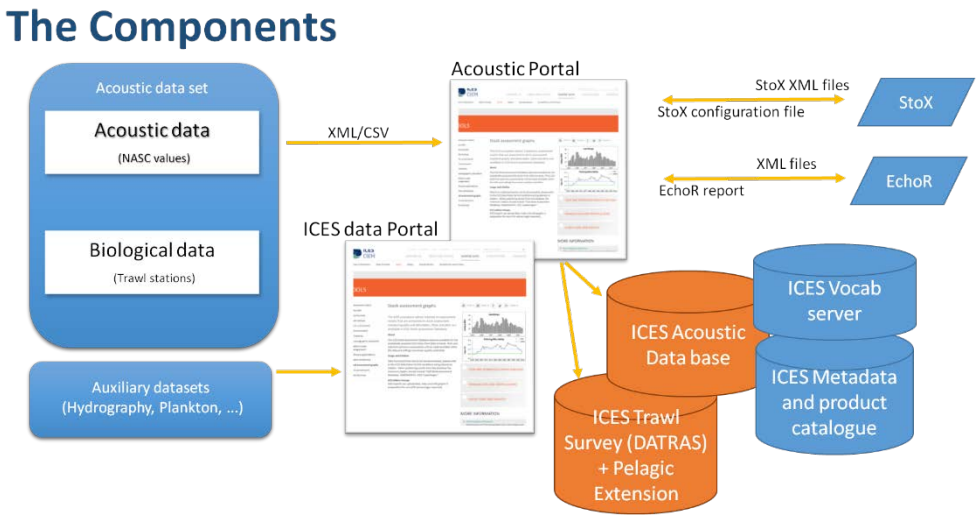


**Figure 1. The components of system, where interpreted acoustic data (acoustic data associated to an acoustic category) and the biological sampling data are uploaded to the ICES data portal. Data are stored in the ICES acoustic database and an extended version of the ICES Trawl survey database (DATRAS). The controlled vocabularies and product catalogue are stored in the ICES vocabulary and metadata and product catalogue databases, respectively. Output from the system is data files that can be directly interfaced with StoX and EchoR.**

# 3 Post-processing software

## 3.1 STOX

Espen Johnsen presented the StoX software (Figure 2), including the data requirements from the Data Centre. Information about the software can be found here: http://www.imr.no/forskning/prosjekter/stox/nb-no

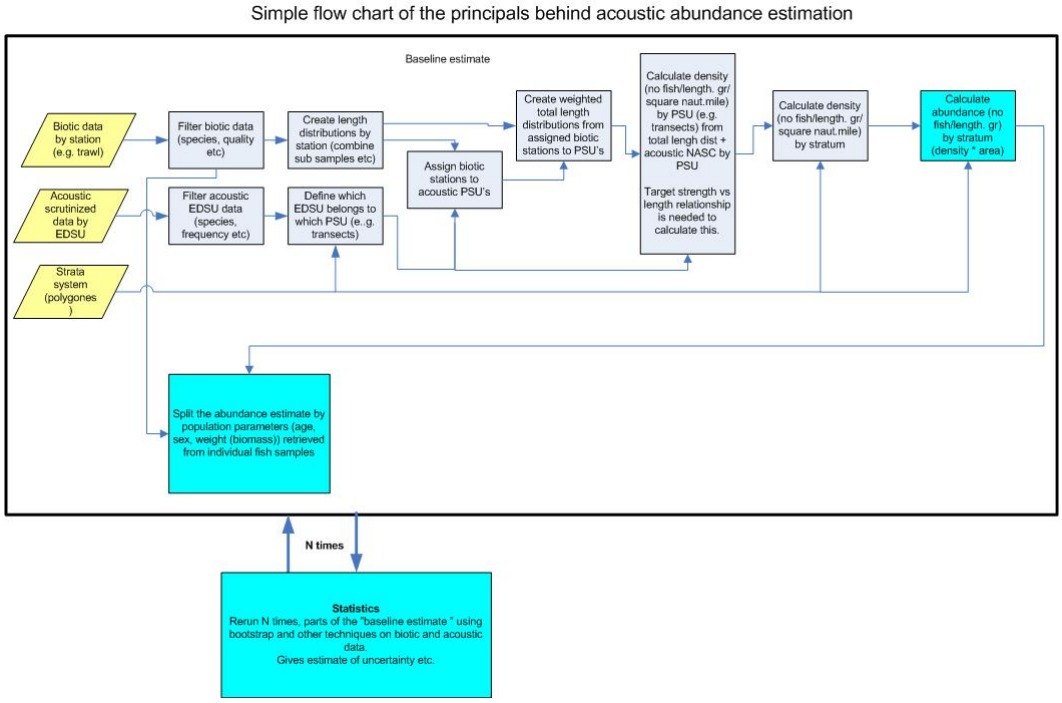


**Figure 2. The flowchart of acoustic data processing: From interpreted acoustic data, trawl sampling stations and strata system to abundance indices by population parameters.**

## 3.2 EchoR and EchoDB

Mathieu Doray presented the EchoDb and the EchoR library.

EchoR is a suite of R codes for handling preprocessed fisheries acoustics data collected during sea surveys and computing standard ec(h)osystemic indicators based on those fisheries acoustics data. These indicators include biomass estimates per fish species and elementary sampling distance units (ESDU), biomass-at-length estimates per fish species and ESDU, biomass-at-age estimates per fish species and ESDU, biomass estimates per fish species and post-stratification regions, synthetic spatial indicators and maps based on ESDU data, and methods for acoustic fish biomass assessment implemented.

Source codes can be found at:

https://forge.ifremer.fr/plugins/mediawiki/wiki/echor/index.php/Accueil

EchoR introduces the use of "echotypes", or acoustic categories, that allow to handle mixed pelagic species cases, typical of (sub)tropical pelagic ecosystems. This notion has been discussed extensively and an "acoustic category" added to the acoustic data table to ensure compatibility with EchoR data and procedures."

# 4 Data and data flow

## 4.1 The ICES acoustic database

The ICES Data Centre presented the first version of the acoustic data model. The version was based on the proposed format from the Workshop on evaluating current national acoustic abundance estimation methods for HERAS surveys (WKEVAL). A large part of the meeting was used to review and discuss each individual field, resulting in a revised data model for the ICES acoustic database.

All the details of the discussions are not presented, but the ones generated the most discussion are referred to below.

The position of the acoustic value needs to be documented in the new format, i.e. whether it is the start, middle or end of the acoustic mile. The controlled vocabulary from the ICES Acoustic metadata standard should be adopted.

There was a discussion on echotypes. The SAC at variable proposed by WKEVAL can only deal with a simple species mix whereas the echotype definition used in the EchoR package allows a more complex mix of stratum/species by survey. An additional dimension table is needed to make a lookup between the depth/species combination with the catch allocation in the haul. A subgroup looked into this during the workshop and concluded that the solution is to delete the "acousticcat" table and deal with a defined list of categories with associated aphiaID's in a controlled vocabulary. The categories need to be associated to a vocabulary and the content of this needs a revision. The revision should include its context, i.e. there should be clear categories and, e.g. labelling plankton if it is really 'everything else' should be avoided. The vocabulary needs to be defined by each survey expert group in their survey protocol.

It was also a recommendation from the group to adopt the ICES platform code system and coding for identifying research vessels/platforms, and the acoustics metadata standard should be updated accordingly:

http://vocab.ices.dk/Request/Login.aspx?ReturnUrl=%2frequest

The different surveys will be available through the ICES data portal. For the purpose of updating the survey list that includes the acoustic-trawl surveys, it is important that each of the survey groups generates shape files of the surveyed areas. This will in turn be available through the ICES geoportal and will make the data easier available.

ICES will develop a facility (web page and web service) to allow the submission of Acoustic Data. The file format is not yet agreed upon, but it should be a tagged file format that contains the fields specified in Annex 4. The major software companies that develop post-processing software will be contacted and asked if they could provide export functionality that holds this information.

In the case of XML a template will be released for the users to validate their data before submitting it to ICES. This is done by using an XSD (XML Schema) of the Acoustic format.

The data will be screened and uploaded together with the biological records, which will be accommodated in DATRAS database (see below).

The data output will be facilitated through the ICES Acoustic data website. The website should provide the user with basic functionality and the output format will be compatible with the acoustic post-processing software (StoX and EchoR).

## 4.2 Biological sampling and DATRAS

ICES hosts a dedicated database for fish trawl survey data – DATRAS (the DAtabase of TRAwl Surveys). The system has been developed to collate and document the survey data, assure data quality, standardize data formats and calculations. DATRAS gathers data for more than 15 surveys in the Northeast Atlantic and is addressed to find specific data.

Acoustic biological trawl sampling design is often close to the sampling performed in surveys contributing to DATRAS. Thus, expertise gather in data reporting of the trawl data can be re-used for the acoustic-trawl surveys. DATRAS data format can be a good starting point to finalize the best options for the acoustic trawl data. DATRAS stores survey-specific quality checks that could be beneficial to adapt for the acoustic trawl surveys as well.

The Biological sampling is composed by three types or record types:

1 ) The haul information (HH)

2 ) The Length based information (HL)

3 ) The biological information (CA)

This data content format was developed from the pelagic redfish surveys, which offers most suitable field definitions to be adopted by the acoustic trawl surveys.

Some of the notes to the format are:

**Gear**: The Gear Type Codes (Used in the DATRAS). There are no information about the platform that might be considered (catchability effects etc.)

**HaulDuration**: The definition needs to be expanded to cover pelagic hauls, minor revision.

**Day Night**: This filed is used to describe if the haul was performed during day or night and it should not be mandatory in the Acoustic Data.

**Shootlan/shootlat**: This field was not clear For the Acoustic data, this should be start fishing.

**HaulVal**: This field described if the haul is valid or not. In the trawl surveys a haul can be reported but not be valid for the calculations. We need a new HaulVal code for "Valid haul for adaptive sampling on acoutic refgistrations."

**Log**: This field that was proposed by the Acoustic group and it is sometimes used to link the acoustic data with the biological data. It should be added to the new biological format of the acoustic.

**Min and Max haul depth** need to be better defined for pelagic trawl: We need the headline vertical and the vertical opening. Shallowest depth of headline during trawling and deepest depth of the headline during trawling and the vertical opening.

#### Comments specific to the Haul Length record (HL)

**Length distribution**: individual length although recorded in pelagic surveys doesn't need to be in regional system of DATRAS.

**StomFullness**: this field should be removed

More info needed on *Vertebrae*

**Ichtyophonous** this field should be removed

**FAOCode**: This field should be removed as it is redundant; the standard WoRMS identifier (AphiaID) enables a lookup to the FAO ASFIS code.

The Biological sampling data will be accommodated in the ICES DATRAS database. ICES will develop a facility (web page and web service) to allow the submission pelagic hauls to DATRAS. The data should be delivered in the agreed format and ICES will provide a screen and upload facility of the format (Annex 4). In the case of XML a template will be released for the users to validate their data before submitting it to ICES. This is done by using an XSD (XML Schema).

The data output will together with the Acoustic data be available through the acoustic data website. The output format will be compatible with the Acoustic software (StoX and EchoR). If users are only interested in the biological data, the data will be also available for download through the DATRAS website.

### 4.3 Auxillary data

The auxiliary data contemplates the following data:

- Oceanographic data
- Plankton data
- Stomach content data

These proprietary data formats/flows that are already established in the ICES system should use the existing infrastructure, and this will not be a part of the acoustics DB as such.

The Oceanographic data. More information on how to submit the data can be found here: http://www.ices.dk/marine-data/data-portals/Pages/ocean.aspx

In the case of Plankton, ICES has a community dataset in the Environmental Database (DOME) and data can be submitted contacting ICES (dome.ices.dk)

The Stomach data have also a dedicated database and can be found here (http://www.ices.dk/marine-data/data-portals/Pages/Fish-stomach.aspx) and the submission of these data will be done upon request.

It is encouraged that the Acoustic community submit their auxiliary data. The data flow will be created to make sure these data are included in the input files for the Acoustic software.

Although the data will be in separate databases, all these data (if submitted) will be available for the acoustic community and the collation of the relevant auxiliary data with the acoustic data will be handled by the ICES Data Centre.

# 5 Post Workshop activities and timeline for next stage of development

After the meeting, the ICES Secretariat finalized the first version of the content fields proposal and provided a first version of the data model. This document was sent on review to the acoustic survey groups (WGACEGG[1], WGIDEEPS[2], WGBIFS[3], WGIPS[4]). The first version of the data model (AtlantOs milestone) is included in Annex 4 in this report.

Tim Ryan, the main author of the ICES metadata standard, has been contacted for some clarification regarding the range definitions in the metadata standard (Chair of the SCICOM/ACOM Steering Group on Integrated Ecosystem Observation and Monitoring (SSGIEOM)).

To allow a smooth transition to the new system, an application to transform the current data format will be developed by the ICES Data Centre. This will allow the Expert Groups to transform their current data format to be used by the post-processing software.

For the Baltic International Acoustic Survey (BIAS) there does not exist a common format for historical data, and BIAS will provide data in their agreed format. The ICES Data Centre will transform these data into the new standard. For this purpose, it is suggested that Espen Johnsen (IMR), Carlos Pinto (ICES Secretariat) and Mehdi Abbassi (ICES Secretariat) attend and provide support and guidance during the upcoming WGBIFS meeting on 30 March-03 April 2016 (Rostock, Germany) on how to implement the data flow: test files to be delivered beforehand.

A similar process will be set up for WGIPS, but in this case the data are more standardized and require less direct involvement from the developers, but an application to transform the data to the ICES data format will be provided by the ICES Data Centre.

The intention is that StoX and EchoR will adapt to the ICES data format, and this will be part of their contribution to the AtlantOS project. However, as an interim solution for WGIPS and WGBIFS in 2016, a conversion program will be generated to generate the current format to STOX – XML. Espen Johnsen will be contacted to ensure that we have a realistic timeline.

A similar process needs to be established for EchoDB, and a transport format to ECHODB csv currently needs to be established, but there might be a possibility to use web services for this purpose.

It is expected that the AtlantOS project will have regular (approximatively every 2 months) web calls to follow-up on actions/recommendations from WKIACTDB. The first internal milestone is the AtlantOS MSI 2.4.1, due December 2015, that includes the specification of the ICES acoustic database, and specification of missing functionality of the DATRAS, is given in Annex 4. The first deliverable is in March 2016 where an interface compatible with standard data processing software will be provided and the

[1] Working Group on Acoustic and Egg Surveys for Sardine and Anchovy in ICES Areas VII, VIII and IX (WGACEGG)
[2] Working Group on International Deep Pelagic Ecosystem Surveys (WGIDEEPS)
[3] Baltic International Fish Survey Working Group (WGBIFS)
[4] Working Group of International Pelagic Surveys (WGIPS)

databases to serve data product generation from pelagic fish surveys will be operational.

# 6 Annexes

## Annex 1: List of participants

| NAME | ADDRESS | E-MAIL |
|---|---|---|
| Mathieu Doray (remotely) | Ifremer<br>Centre Atlantique - Rue de l'Ile d'Yeu - BP 21105 - 44311 Nantes Cedex 03, France | Mathieu.Doray@ifremer.fr |
| Carlos Pinbto | ICES<br>H. C. Andersens Boulevard 44-46,<br>DK 1553 Copenhagen V, Denmark | carlos@ices.dk |
| Neil Holdsworth | ICES<br>H. C. Andersens Boulevard 44-46,<br>DK 1553 Copenhagen V, Denmark | neilh@ice.dk |
| Nils Olav Handegard | Institute of Marine Research IMR<br>PO Box 1870 Nordnes<br>5817 Bergen, Norway | nilsolav@IMR.no |
| Vaishav Soni | ICES<br>H. C. Andersens Boulevard 44-46,<br>DK 1553 Copenhagen V, Denmark | vaishav@ices.dk |
| Anna Osypchuk | ICES<br>H. C. Andersens Boulevard 44-46,<br>DK 1553 Copenhagen V, Denmark | anna.osypchuk@ices.dk |
| Mehdi Abassi | ICES<br>H. C. Andersens Boulevard 44-46,<br>DK 1553 Copenhagen V, Denmark | mehdi.abbasi@ices.dk |
| Benjamin Planque | IMR<br>9294 Tromsø<br>Sykehusveien 23,<br>9019 Tromsø, Norway | benjamin.planque@imr.no |
| Sascha Fässler | IMARES Wageningen<br>Haringkade 1<br>1976CP IJMUIDEN, Netherlands | sascha.fassler@wur.nl |
| Leon Smith | Faroe Marine Research Institute<br>Nóatún 1 Tórshavn,<br>FO 110 Faroe Islands | leonsmit@hav.fo |
| Eydna ì Homrum | Faroe Marine Research Institute<br>Nóatún 1 Tórshavn,<br>FO 110 Faroe Islands | eydnap@hav.fo |
| Olavi Kaljuste | SLU<br>Kustlaboratoriet, Skolgatan 6<br>742 42 ÖREGRUND, Sweden | olavi.kaljuste@slu.se |

## Annex 2: Agenda

### Workshop on the review of the ICES acoustic trawl survey database design (WKIACTDB)

01 October, 09:30 – 17:30

02 October, 09:00 – 16:00

ICES Secretariat, North Sea room

Copenhagen, Denmark

Chairs: Neil Holdsworth (DK) – Nils Olav Handegard (NO)

1 ) Opening and round table introductions

2 ) Intro to the aim of the workshop and overall activity in relation to the AtlantOS project (Nils Olav)

3 ) Intro to the overall dataflow and technical development of the Acoustic DB and related data flows (Neil)

4 ) **Acoustic data flow and specifications**;

- 4.1 ) STOX
- 4.2 ) EchoDB
- 4.3 ) Report back from WKEVAL
- 4.4 ) Specifications for data flow to ICES Acoustic DB (building on WKEVAL experience)
- 4.5 ) Specifications for data flow from ICES Acoustic DB

5 ) **Biological sampling data flow**

- 5.1 ) Introduction to the existing DATRAS trawl survey data system, flow of data and specification of data inputs
- 5.2 ) Specific adaptations for acoustic biological trawl data (also see biological sampling protocol flow chart in Annex of WKEVAL 2015 report)
- 5.3 ) Data flow to ICES
- 5.4 ) Data output from ICES

6 ) **Auxiliary data sampling**

- 6.1 ) Agreement on data flow/format i.e. CTD's, plankton sampling etc.

7 ) Compilation of input to specifications and identify further issues to resolve i.e. missing input from acoustic groups

8 ) Information on timeline for next stage of development

## Annex 3: Recommendations

| RECOMMENDATION | ADRESSED TO |
|---|---|
| 1. Recommendation to SSGIEOM to formally approach manufacturers to implement the ICES data format as a standard report. | SSGIEOM |
| 2. Review the acoustic metadata standard in relation to the proposed database design. | WGFAST |
| 3. Provide a controlled vocabulary for acoustic categories to be used by the ICES expert groups. | WGTC |
| 4. It would be better if the Surveyid was mandatory and quarter optinal for all surveys. | DIG/ DATRAS |
| 5. Provide shape files specifying the survyed area of the acoustic trawl-survey for the ICES geoportal, and provide the survey names/acronyms for inclusion in the SurveyID vocabulary. | WGACEGG, WGIDEEPS, WGBIFS, WGIPS |
| 6. Implement some changes to their format/definitions to align with the review in WKIACTDB | WGIDEEPS |

## Annex 4: AtlantOS deliverable

Atlantos WP2.4 Delivery 1 – Month 9 – Acoustic Specification delivered by ICES available on the ICES website under the AtlantOS project page.

Project page:

http://ices.dk/explore-us/projects/Pages/ATLANTOS.aspx

Acoustic Specification document:

http://ices.dk/marine-data/Documents/Specification%20of%20the%20acoustic%20database.pdf