

ICES WKDATR-NSCS REPORT 2018

ECOSYSTEM OBSERVATION STEERING GROUP

ICES CM 2018/EOSG:44

REF ACOM AND SCICOM

Report of the Workshop on DATRAS surveys in Greater North Sea and Celtic Sea (WKDATR-NSCS)

11 – 13 June 2018

ICES HQ, Copenhagen, Denmark



ICES
CIEM

International Council for
the Exploration of the Sea

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. 2018. Report of the Workshop on DATRAS surveys in Greater North Sea and Celtic Sea (WKDATR-NSCS). 11–13 June 2018, Copenhagen, Denmark. ICES CM 2018/EOSG:44. 24 pp. <https://doi.org/10.17895/ices.pub.8175>

The material in this report may be reused for non-commercial purposes using the recommended citation. ICES may only grant usage rights of information, data, images, graphs, etc. of which it has ownership. For other third-party material cited in this report, you must contact the original copyright holder for permission. For citation of datasets or use of data to be included in other databases, please refer to the latest ICES data policy on ICES website. All extracts must be acknowledged. For other reproduction requests, please contact 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.

© 2018 International Council for the Exploration of the Sea

Contents

Executive summary	1
1 Terms of reference	2
2 Overview	3
2.1 Background.....	3
3 Review of national submission issues.....	5
3.1 Current status.....	5
3.2 Belgium Beam Trawl Survey BEL-BTS	8
3.3 Norway	10
3.4 Irish Angler and Megrim Survey IE-IAMS	14
3.5 Irish Groundfish Survey IE-IGFS	15
3.6 UK-WEC-BTS, Western English Channel Beam Trawl Survey	16
3.7 UK Q1 Ecosystem survey	16
4 Common submission issues	17
4.1 General ICES & DGMARE Requirements.....	17
4.2 Reference field for CTD Data	17
4.3 Reference field for HH Data	18
4.4 Benthic Data	18
4.5 ALK borrowing.....	18
Annex 1: List of participants.....	19
Annex 2: Agenda.....	20
Annex 3: Recommendations	21
Annex 4: DATRAS Relational Database Model Schema and unique constraints of exchange data in HH HL and CA	22
Annex 5: The IBTS ALK substitution procedure.....	24

Executive summary

The Workshop on DATRAS surveys in Greater North Sea and Celtic Sea (WKDATR-NSCS) met in Copenhagen, 11–13 June 2018 and was chaired by Vaishav Soni, ICES Secretariat, and David Stokes from the Marine Institute, Ireland.

There were 7 attendees and 1 remote participant representing 4 countries (Norway, Belgium, Ireland and UK-England). The list of participants is provided in Annex 1.

The main goals of the workshop were to incorporate new trawl surveys, missing data submissions and missing data products into DATRAS, as well as to resolve some submission doubts and issues from the data submitters. Another goal was to identify and resolve data quality issues in DATRAS, proposing quality check procedures for missing data to be applied during uploading/reloading processes, as well as in already existing exchange data.

DATRAS (<http://datras.ices.dk>) is the ICES Database on Trawl Surveys. The database contains data of seventeen surveys, most of them coordinated by one of the three ICES trawl survey working groups, Baltic International Fish Survey Working Group (WGBIFS), International Bottom Trawl Survey Working Group (IBTSWG) and Working Group on Beam Trawl Surveys (WGBEAM). The fish stock assessment working groups rely on data products from DATRAS, such as CPUE per haul by length and age, age-length keys (ALK), etc. Most of these indices are derived from data collected under the DCF funding scheme, which require good quality error-free survey data submissions and outputs.

Primarily, the participants of the workshop worked towards identifying the issues regarding current missing time-series for exchange data, resolve them, and identify the hurdles. Participant data submitters worked closely with the DATRAS team of ICES Data Centre to resolve their issues, which allowed them to upload the data directly into DATRAS after applying the necessary changes to their data. Erroneous data and misinterpreted data have been corrected in this workshop. In addition, the workshop facilitated discussions about future calculation methods for length- and age-based products. There are still challenges in order to resolve all submission issues such as data errors, screening facility problems, how to create the exchange files or the need of additional training. WKDATR-NSCS worked on these submission issues in order to improve the submission and data gathering process.

1 Terms of reference

The **Workshop on DATRAS surveys- Greater North Sea Celtic Sea** (WKDATR-NSCS), Co-Chaired by David Stokes (Ireland), and Vaishav Soni, ICES, will be established and will meet at ICES Headquarters, Copenhagen, 11–13 June 2018 to:

- a) Incorporate new trawl surveys into DATRAS including checking procedures to be applied during uploading/reloading processes, targeting:
 - i) Review the reporting format
 - ii) Agree on errors and warnings that should apply to the new datasets
 - iii) Produce test files for uploading
- b) Incorporate missing data submissions to the existent surveys already hosted within DATRAS. The missing data can relate to:
 - i) Historical data (to achieve full time series)
 - ii) Biological data for all analysed species (length and age)
- c) Incorporate missing data products to existent surveys already hosted within DATRAS. The missing products can relate to CPUE products, indices, etc. The main objectives would be to:
 - i) Prepare data product calculation documentation

2 Overview

2.1 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.

In late 2017, a DG Mare special request¹ highlighted three stages of quality assurance for data collected under the DCF, and requested ICES to review and engage with a) development of the SmartDots platform for age reading data, b) specifications of the RDBES (Regional DataBase and Estimation System) for commercial catch sampling data, and c) inclusion of DCF-funded fisheries independent surveys/variables not yet held in DATRAS into such database. All three developments would enable ICES to reduce uncertainty and error, but also to more precisely quantify errors in input data to assessments.

In the light of the most recent errors in input data to the assessments from surveys which at present are not part of the DATRAS system, or where not all required variables are provided, ICES was requested to incorporate these surveys in DATRAS to increase options for quality control and transparency/traceability of assessment outcomes.

Two designated Workshops have been established to address this request: the Bay of Biscay and Iberian Coast (WKDATR-BoB), followed by the Greater North Sea and Celtic Sea workshop (WKDATR NSCS). These are intended to ensure experts have the time and dedicated technical support to work on identified surveys (based on the Liaison Meeting 2017 analysis of surveys). The objective being inclusion of quality assured data within DATRAS and output of appropriate products.

The surveys highlighted for inclusion in the workshop are given in table 1 and include surveys currently absent from DATRAS through to those with missing variable(s) or products.

Table 2.1.1: Targeted surveys for WKDATR-NSCS (from scientific justification for workshop ToRs).

Survey	Country
UK-WEC-BTS, Western English Channel Beam Trawl Survey	UK-England
Q1 South West Ecosystem Beam Trawl Survey	UK-England
Norwegian surveys	Norway
IE-IGFS	Ireland
IAMS	Ireland
BTS	Belgium, Germany
Survey	Country
UK-WEC-BTS, Western English Channel Beam Trawl Survey	UK-England
Q1 South West Ecosystem Beam Trawl Survey	UK-England
Norwegian surveys	Norway
IE-IGFS	Ireland
IAMS	Ireland
BTS	Belgium, Germany

¹(http://ices.dk/explore-us/Documents/Cooperation%20agreements/EU/2018_ICES_EC_AA_signed_WEB.pdf) and ICES Request code: 1711_DCF_quality

DATRAS is an online database of trawl surveys hosted by ICES. DATRAS stores data from internationally coordinated fish trawl surveys in ICES Area 27 in a standardized format and provides a data quality check upon data upload. The DATRAS portal offers online services for uploading and downloading of survey data. Details of the data exchange flat file (*.txt) format can be also found at the portal.

Upon successful submission, the exchange files are mapped and uploaded to the DATRAS SQL server relational database and are made publically available from DATRAS portal (see Annex 5 for schema and unique constraints of exchange data in HH HL and CA).

3 Review of national submission issues

3.1 Current status

The workshop opened with a round table discussion of the issue list circulated to the group prior to meeting (Table 3.1.1). This helped identify sub-groups for participants at a similar stage in the data upload process having common support requirements. A lack of information in areas of the table generally reflected the participant's lesser familiarity with the DATRAS upload process, vocabulary and formats.

In general the survey submission status fell into three categories.

- i) **Pre-submission** – surveys starting to map national data to the DATRAS exchange format and files not yet complete.
- ii) **New Submission** – survey is mapped, exchange file available, but support needed to address DATRAS Screening errors.
- iii) **Incomplete Submission** – time series of survey data held in DATRAS, but one or more variables or products incomplete. Support needed around updating of Exchange file(s) and/or agreement around calculation procedure for the relevant product.

In general terms,

UK fell under pre-submission,

The Irish survey IE-IAMS fell under New Submission. The Irish survey IE-IGFS were addressed as incomplete submissions.

Norwegian surveys were addressed as incomplete submissions.

Belgium BTS survey fell under pre-submission for historic data and some potential re-submission of existing data following additional quality control (QC) checks.

The workshop was initially structured then into a pre-submission sub-group looking at data formats, conversion and field mapping. The second sub-group dealt with outstanding submission errors or calculation of products.

In a number of cases, there was overlap and therefore, within the report the workshop activities are reviewed by country with common issues to several or all surveys dealt with separately.

The ICES Data Centre also gave an overview of DATRAS, exchange file formats and relevant vocabularies.

Table 3.1.1 Submission issues table circulated prior to the workshop.

Country	DATRAS Survey code	Time Series	Issue	Exchange Files available?	Exchange files in development?	Products available?	National Data Storage format	Comments
Ireland	IE-IGFS	2003 -2017	IE-IGFS data product development	Yes	NA	No	SQL Server	Need to agree requirements for general data products.
Ireland	IE-IAMS	2016-2018	IAMS submission set up	Yes	NA	No	SQL Server	Time series needs to be incorporated into DATRAS and data products defined.
Ireland	IE-IGFS	2003 -2017	Litter data	Yes	NA	No	SQL Server	This just needs some time allocation to upload and check the output.
Belgium	BEL-BTS	1992-2009	BTS historical data submission issue	No	Yes	No	Excel	Migrate data to SQL Server and check quality
Belgium	BEL-BTS	2011-2017	Litter data	No	Yes	No	Paper	Input paper records into csv with DATRAS format
Belgium	BEL-BTS	2010-2017	Data products based on BTS standard Swept Area routine	Yes	NA	No	Microsoft SQL Server	Generate required data products
Belgium	BEL-BTS	2011-2017	Litter data	No	Yes	No	Paper	Input paper records into csv with DATRAS format

England			Resolving issues with "new" beam trawl data for the ongoing CEFAS Q1 South West Ecosystem Survey (2006 onwards)					
England			CA and HL data format training					
Norway			NS-IDPS data	Yes				
Norway			Data discovery: other Norwegian surveys?					

3.2 Belgium Beam Trawl Survey BEL-BTS

ILVO has developed an application to automatically import historical length and biological data to the national SQL server. This procedure was discussed to see if the process could be supported during the meeting in order to speed up the submission of Belgian historical data to DATRAS. However, it was decided that it was more relevant for the ICES data centre to assist with the quality assurance of the data that goes into DATRAS. Consequently additional scripts for quality screening such as ALK outlier plots were developed. This was based on all data submitted for the southern part of the North Sea (also by other countries than Belgium). Belgium provided historical biological data for sole and plaice to check for errors with this tool.

Once the historical data is in the national database, Belgium will screen the data by year and upload it to DATRAS. They will start from 2009 and work backwards in time. The data that is already in DATRAS (2010-2017) will be rechecked for errors (especially in age data) and re-uploaded to DATRAS. The ICES Working Group on the Assessment of Demersal Stocks in the North Sea and Skagerrak (WGNSSK), the most important end-user of this data, will be informed once the updated data is available in DATRAS.

When exchange data time series correctly uploaded in to DATRAS from 2010 onwards then updated combined survey index for North Sea plaice can be calculated. The aim is also to develop a tuning index for North Sea sole based on the Belgian BTS data, which is to be evaluated for use in the future North Sea sole assessments.

A similar procedure will be carried out for the Belgian DYFS data. Starting with upload of the 2017 data and then working back in time. Both fish and shrimp length data will be submitted to DATRAS. From 2018 onwards, Belgium is planning to age sole and plaice on the DYFS, so they can start producing their own ALKs for survey index calculations.

Litter data submission

With support from the ICES Data Centre, the template for Litter submission was applied and tested for BTS data from 2017. There were some minor errors and warnings, which were all resolved. Now Belgium can start uploading all Litter data (2012-2017) that can then be used for assessments (for example the Working Group on Marine Litter, WGML assessment in the fall of 2018).

DYFS shrimp data

Belgian shrimp data was reviewed to identify what parameters, and in which format they should be uploaded to DATRAS. Based on how the data are recorded on the DYFS, there will be two size categories for shrimp: a large (Category 1) and small (Category 2) fraction.

- For each of these categories (fractions) the following parameters exist:
 - o TotalNo (total number of shrimp in the fraction)
 - o NoMeas (number of shrimp measured in the fraction)
 - o CatCatchWgt (total weight of shrimp in the fraction)
 - o SubWgt (weight of the subsample of shrimp that was measured)
 - o SubFactor (based on the weight proportion between Catcatchweight and weight of the subsample)
- Other parameters are:
 - o LngtClass

- HLNoAtLngt
- LngtCode will be "." because the shrimp are measured to mm
- SpecVal will always be 1
-

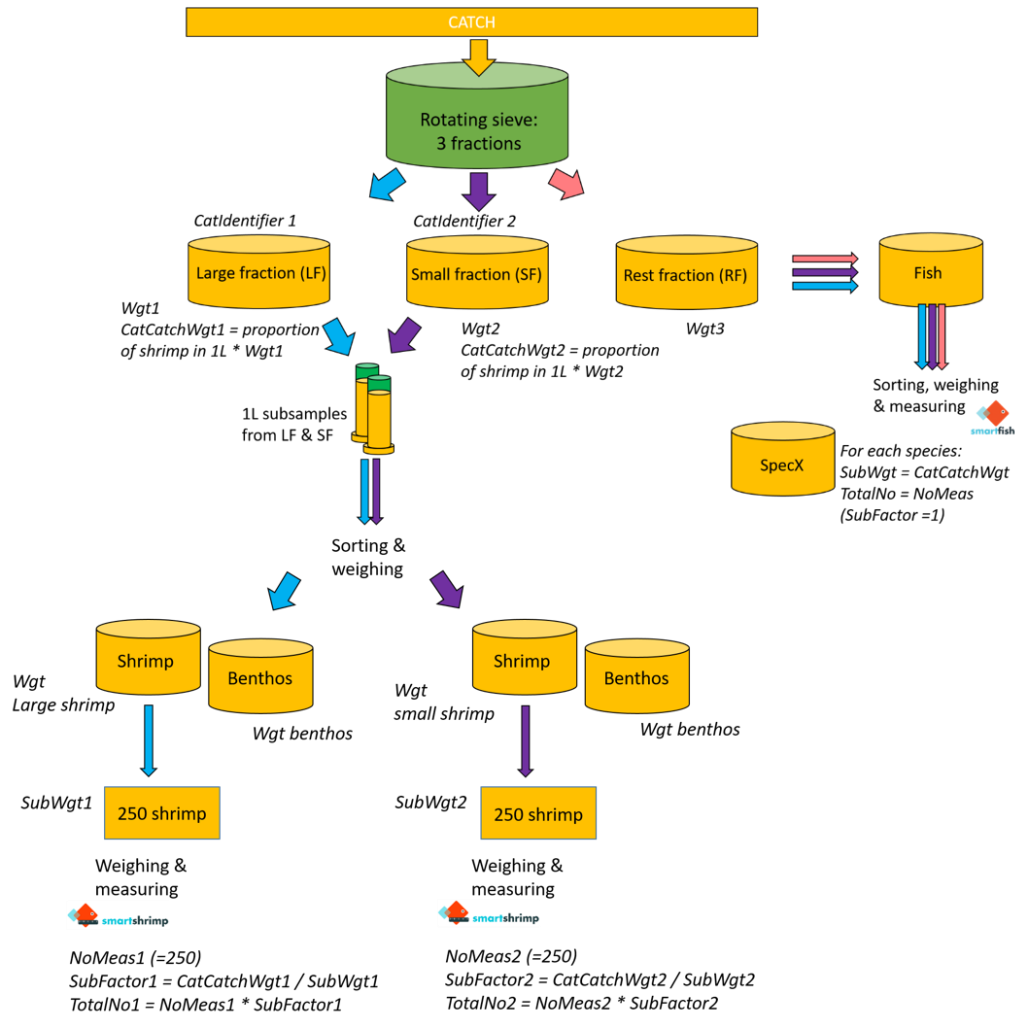


Figure 3.2.1. : Scheme of the catch processing on the Belgian DYFS with the related DATRAS parameters in italics.

It might be interesting for data users if other parameters such as “Length measurement type” and “Sample processing method” are added to the HL for inshore surveys. “Length measurement type” would clarify what part of the shrimp was measured (full body length, only carapax, etc.) and “Sample processing method” would deal with how the shrimp were processed before measuring (raw, cooked, preserved in formaldehyde, etc.). However, this is more an option to be discussed at the DATRAS governance group.

3.3 Norway

There were several data reporting issues covered on the first two days of the meeting. For example, the day and night field reported by Norway in the HH data raised a warning message because the Sunrise/Sunset algorithms of Norway and DATRAS slightly differ. This may be due to time zones in the survey area not adequately accounted for in the screening utility. The mismatch was discovered and discussed and the problem will be fixed prior to the 2019 data submission.

There was a suggestion for the CTD station number (additional field) in DATRAS data (NS-IBTS) to be the same CTD station number as in the oceanographic database, so that there is a 1-to-1 link between oceanographic and biological data. Other issues included;

- How are national data submitted to ICES, which format or station numbers are used, and whether the ICES Data Centre retains the national station numbers or use a different station.
- How do global databases pull from the ICES databases.
- There was a lengthy discussion on benthos data. The IBTSWG 2017 request was for guidance on whether benthos (non-mandatory species) should be uploaded for the NS-IBTS data (e.g., presence/absence, counts, counts and total weight, other?). It was also discussed that for representation of benthos data in exchange data a new field in HH or HL is needed.

In addition, borrowing of ALK indices was discussed. Different ways to present ALK data (when indices are downloaded) were reviewed, and it was agreed that ALKs for all 3 types of borrowing that were (potentially) used should be provided:

- final ALK used, with field stating which roundfish areas were used for each roundfish area (e.g., area 1 ALK used data from areas 1, 2, 4);
- raw ALK for each roundfish area;
- Illustration of borrowing of length data from surrounding areas within a given age.

Work was carried out on the conversion of Norwegian survey data to DATRAS format. Firstly R coding, and afterwards addressing pitfalls and internal/national data checks (without R syntax).

Several issues were clarified in relation to the Norwegian data uploads.

1. It was agreed that the Norwegian data submissions would move to raw data format (DataType=R); we are currently working on a conversion program to do this. ICES Data Centre agreed to assist with consistency checks between the DataType=C and DataType=R, so that errors are not introduced in e.g. total numbers, catch weights.
2. When DataType=R (raw) is adopted, data will be re-uploaded with SubFactor and SubWgt added to the HL data,
3. Clarified 'subsampling'. The Norwegian understanding had been that subsampling only happened when the total catch had not been sorted. It was clarified that the Norwegian method of measuring the catch, i.e., measuring 100 individuals of each species, where more than 100 (e.g. 135) were in the total catch, is also considered subsampling.

4. An error that occurs occasionally during Norwegian data collection is that total catch weight and total number is omitted/forgotten/accidentally deleted/entered incorrectly, but length measurements exist. These data do not fall under any of the currently valid SpecVal fields. It was clarified that to solve the issue either a new field/code could be created in DATRAS or data could be submitted as raw. There was a further suggestion from the ICES Data Centre to use SpecVal 10 in HL, and get the biological data (length-weight) to CA, with SpecVal = 10 weight information in CA are not mandatory. This suggestion needs to be tested in a sample with a similar problem.

On the last day of the meeting participants had a detailed discussion on subsampling. Norway currently uploads DataType= C, so there is not an issue with the non-reporting of subfactor (If DataType is C, CatIdentifier is always 1 and there is no subfactor or sub-weight). However, Norway would like to move to submitting only DataType=R.

Figure 3.3.1 illustrates how Norway samples their catch. The entire catch is always sorted. If two (or more) distinct, non-overlapping size distributions are seen (e.g. very small, large) in the catch, the species is sorted into different subsamples. The total weight of each subsample is recorded, but then Norway takes length measurements (or full biological samples) for a maximum of 100 individuals (50 in years prior to 2013). Norway then records the weight of those 100 measured fish and estimates the number of fish in the subsample.

There was a discussion regarding the non-overlapping length distributions. Since length-distributions in subsamples can overlap many countries will have separate subsamples specifically for the tails of the length distributions to ensure that very large or very small individuals have adequate samples and that their numbers are not over-inflated by being included with the rest of the catch (particularly when the length distribution has long tails). Norway will need to discuss this internally to determine if their sampling procedures (currently 1 per 1 cm except when IBTSWG has other specifications) needs to be reviewed.

Norway will do as specified in Figure 1 when submitting DataType=R:

1. In the HL records, CatIdentifier will be used for each sampled group within a species.
2. CatCatchWgt is the total weight of that sampled group.
3. SubWgt is the sample weight for that sampled group.
4. SubFactor is the raising factor to go from measured fish to total fish in the subsample.
5. HLNoAtLngt is number measured at length
6. $TotNo = \text{Sum}(HLNoAtLngt) * \text{SubFactor}$
7. NoMeas = number measured

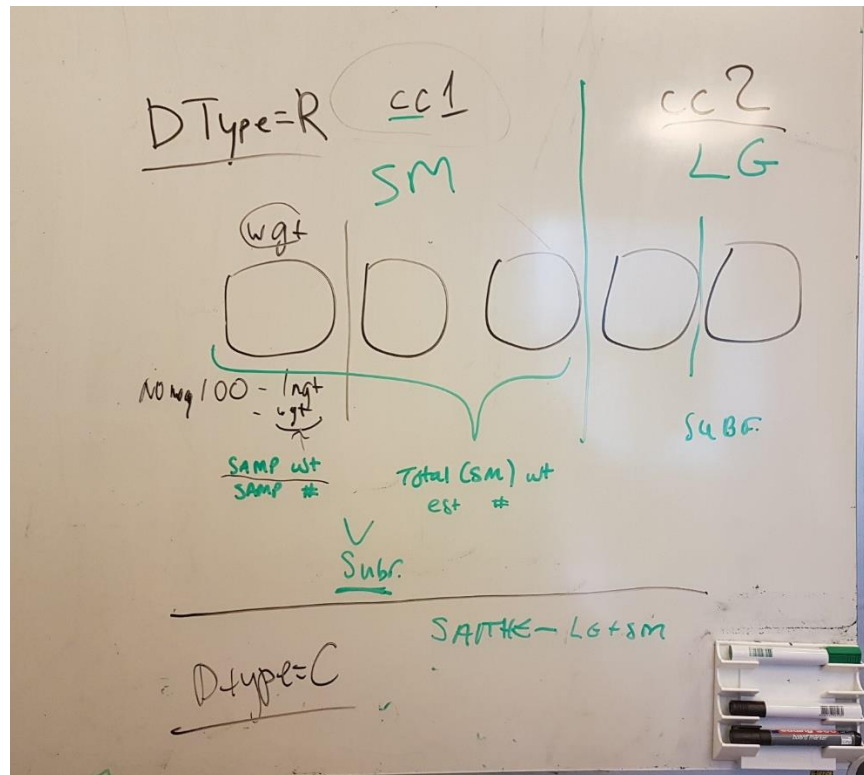


Figure 3.3.1. Breakdown of subsampling for Norwegian data and clarification on how to submit under Data Type R. Catch Component 1 = cc1; Small = SM; Large = LG. In this example 2 length groups are included in cc1 while a more common part of the length frequency distribution is allocated to a separate subfactor in cc2.

Several other issues were noted and solved during the meeting:

1. Ship codes. In the old dataset some ship codes were missing and are currently being made. It was ensured that old ship codes are linked to the IBTS data platform.
2. Issues with old (pre-2004) data that need to be resolved and re-uploaded:
 - a. No benthos or cephalopod species appear to have been recorded. To reflect this the BycSpecRecCode category must be checked and changed if it is found not to be correct.
 - b. Many species are identified in the data to species level, when they are actually species groups. The data will be reclassified to family level and re-uploaded.
 - c. Tows of not enough duration were included as valid and they need to be changed to invalid.
 - d. No data exist on haul positions for many of the early surveys.
3. Issues to be corrected with data (2004-present):
 - a. In 2014 Norway corrected a significant error in the data that affected most years (Q1 and Q3) but at the time and because of time constraints not all station parameters (e.g. (haul position, net opening, door spread, warp) were included in the re-upload. This needs to be corrected. To avoid potentially adding errors to the data or creating confusion about the number of re-uploads Norway will wait until the Data Type=R data

are uploaded to complete this task. The current timeline is to complete this task by end of 2019.

- b. There may be small changes in the total number of recorded fish species at a station (1-4 stations in certain quarters). This is because there is a small error in the current conversion program that removes records that do not have total/measured weight or number (i.e. where data appeared incomplete), but where CA records did exist. Dummy HL records will be made so that CA data can be uploaded.

New members of staff will be taking over the upload process. A plan was made to ensure that the upload process will be without problems and there will be a lengthy training process.

When uploading data for e.g. Sepiolidae, warnings are common for the length range (e.g. individuals are smaller than the minimum length allowed). The ICES data centre was asked to decrease the minimum size on the range to avoid these warnings since it is real data and to have individuals smaller than the current minimum range is very common. Further communication: as the minimum size check is generic for all surveys and species in DATRAS, and does not prevent data upload, no changes will be made to it, but the suggestion can be brought up to the attention of the WGDG. Maximum ranges per species can be revised and recommended by the respective survey groups.

3.4 Irish Angler and Megrim Survey IE-IAMS

The DATRAS Exchange File format for IE-IAMS was looked at in detail, identifying codes for reference fields and upper/lower values for numeric fields for each HH, HL and CA format.

The main issues addressed were the selection of a tickler code from the BTS for the Jackson trawl as well as the inclusion of a Codend Mesh Size code given there is no liner in the IAMS trawl.

The second important discussion was around products and index calculation. To avoid bias, the same stratification should be applied to all species because the density of stations is different for different strata. This implies that even though the used stratification might not be the optimum stratification for, say, plaice, you would still have to produce your estimates using a single stratification. In the IE-IGFS survey the density of stations is pretty much proportional to the area throughout so there is more flexibility.

On the second day of the workshop, a subgroup was formed between the ICES Data Centre and the Irish participant to examine the survey setup for IE-IAMS. General data fields (e.g. quarter, country, ship) and specific fields (e.g. gear, sweep length, door type, tickler, haul duration and distance) were discussed. Quality control ranges were established for each data field in the HH table. The spatial scale for input and data products was determined to be at the survey stratification level (not at the level of ICES divisions or rectangles). Shape files of survey strata were provided to the ICES Data Centre. Future data products and marine litter data were also discussed.

On the third day of the meeting, there was a round table discussion of progress by each country. David Stokes (Workshop co-chair) led a discussion of workshop recommendations e.g. minimum data products, links to the ICES TAF (Transparent Assessment Framework), by-catch species, mapping reference from EUROGOOS to DATRAS, input and output of survey data outside standard operations. In addition, a test Litter exchange file was produced for IE-IAMS and uploaded to DATRAS webpage for data screening.

3.5 Irish Groundfish Survey IE-IGFS

The outstanding data gathered from the IE-IGFS to be incorporated into DATRAS is the Litter data. There have been several uploads done manually by the ICES Data Centre, but upload through the DATRAS screening utility seems to have caused problems. There has been a recurrent issue with a small number of IE-IGFS station names that were converted to 'Date' format which has caused screening issues when matching Litter data to the relevant HH data. Despite the exchange file being manually checked against the download file, the screening error message have suggested a mismatch. The source of the problem is due to the conversion of exchange data into CSV format. Ireland has been aware of potential causes such as the use of national spreadsheets, but will review their processes and re-upload the remaining Litter data. Any ongoing anomalies will be reported back to the ICES Data Centre.

The second area of discussion concerned outputs rather than inputs since there are no products currently available for the IE-IGFS. The main reasons for this are:

1. **Confirmation of TS SubArea** – for the IGFS this is in line with the other western area surveys and it is provided as Strata. It is required to establish the vocabulary relation with SubArea fields.
2. **Confirmation of Gear Geometry equations** – simple algorithms to fill in missing data for gear geometry parameters have been provided for IBTS surveys. These are outlined in the recent Manual of the IBTS North Eastern Survey (SISP 15)² for the relevant surveys including IE-IGFS.

² ICES.2018. Manual of the IBTS North Eastern Atlantic Surveys. Series of ICES Survey Protocols SISP 15. 92 pp. <http://doi.org/10.17895/ices.pub.3519>

3.6 UK-WEC-BTS, Western English Channel Beam Trawl Survey

We had a remote WebEx session with Gary Burns (CEFAS) on data reporting plans and problems. The discussion covered some general and specific data mapping issues, but generally these seemed solvable and were in progress at CEFAS, and assistance from the ICES Data Centre is not required.

One point worth investigation from the ICES Data Centre side is the use of the 'seeded' otoliths. This is the case where a dummy record is produced for a fish below a defined length at which an otolith is not required (e.g. assumed juvenile 0-group).

There was also a request from CEFAS that extra tables and fields may be required to allow greater control over the CA area types. CEFAS will contact the ICES Data Centre with further information.

3.7 UK Q1 Ecosystem survey

See section 3.6 above.

4 Common submission issues

During the workshop a numbers of themes arose that were applicable to several if not all surveys and these are outlined below.

4.1 General ICES & DGMARE Requirements

Beyond submission of the basic exchange file format it is not clear what is commonly expected of surveys in terms of data products and outputs. Survey data in DATRAS is not generally included in ICES Data Calls so where data fails to meet the deadlines or expectations of users there is not a clear feedback loop to highlight issues. Extension of the DATRAS documentation or ICES data policy to cover some of this may be useful, even if it can't be definitive.

This is somewhat more complicated in the western area surveys where combined indices are not the norm and generally spatial methods and modelled ALK's for missing age data are used. Some of these indices are being reviewed by upcoming workshops (e.g. Workshop on evaluating survey information Celtic Sea gadoids, WKESIG), 2019) and benchmarks (e.g. WKCELTIC in 2020). Therefore, for assessment purposes, it may be more appropriate for standard outputs from DATRAS to be the inputs for index calculation to be used by assessment groups rather than output of indices per se. However, standard outputs for more general use should still be made available such biomass per haul/strata, mean weigh at age and so on rather than CPUE at age for a management area.

Furthermore, the actual data model (schema) behind DATRAS was discussed. An understanding of this can help users to follow how the various data fields are normalised and linked to each other, and therefore why certain functionality works and other aspects can be problematic. A schema was provided to the workshop (see Annex 5), but includes a lot of detail. A link to a scalable diagram on the DATRAS documentation page would be appreciated.

4.2 Reference field for CTD Data

Norway highlighted an issue in maintaining a reference between the biological data submitted to DATRAS and the oceanographic CTD data collected on the same IBTS surveys and submitted to EUROGOOS. Not all survey participants present are involved in this process so one of the chairs agreed to follow up with the relevant data centre representative after the workshop as they were not available at the time. Following the workshop David Stokes discussed the issue with Hjalte Parner from the ICES Data Centre. The suggestion was that there is a 4 digit alfa numeric station code in EUROGOOS that can be used, but rarely is. However, using the platform or ship code as part of a dynamic linkages/search algorithm including time and/or position as well is indeed already possible within the EUROGOOS portal. Requests for tailored solutions for standard searches for IBTS or other groups are also very welcome.

There is an argument that many CTD deployments are not spatially and/or temporally contiguous with the biological hauls during fisheries or other surveys. Therefore, defining at what point a CTD record is no longer temporally/spatially linked to a HH record is worth considering. This topic should be revisited by IBTS with a recommendation to define a 4 digit code and/or liaise with the ICES Data Centre to provide the relevant search tool(s) to retrieve the data requested (possibly based on a DATRAS HH input file over web services).

4.3 Reference field for HH Data

The recommendation 6 from the IBTSWG Report 2018 was discussed:

Add a field in HH records allowing to specify experimental tows, e.g. for trawl calibration or non-standard tow durations.

This was raised historically, but has come up again as there are a number of studies ongoing within IBTS currently that involve multiple surveys across areas and years. Comparative work can often be done as part of normal valid tows and therefore flagging it as anything other than valid (for routine assessment purposes) is not appropriate. However, identifying the relevant tows across ships and years can be slow and the Data Centre can assist the data users with this task. The proposal is for a clear reference code/field, visible in the ICES vocabulary pages for the HH exchange file format that could identify experimental tows. The code would be requested by the relevant expert group and used by data submitters to facilitate extracting and aggregating data sets together based on a particular study or issue. The same code could then be used to refer to the relevant study in the pertinent expert group report.

4.4 Benthic Data

An introductory discussion at the workshop related to the completeness of data sets, for example to include non-mandatory invertebrate bycatch species. This has been discussed on a number of occasions at expert group level, certainly within IBTSWG. The problem arises when there are no agreed quality standards for non-mandatory species. This is further compounded where these data are submitted by only a fraction of the surveys, producing a highly biased dataset for these taxa with likely inconsistencies in taxonomy, measurement data (e.g. weights, counts) as well as the obvious bias in spatial and temporal coverage.

While it was acknowledged that there is a growing demand for these data (e.g. OSPAR, MSFD), it was reiterated that the relevant expert groups such as Benthos Ecology Working group (BEWG) have been requested to provide advice on the issue (see IBTSWG2017 Report, Recommendation 3). Until expert advice, and probably resources, are available to support standardizing the taxonomy and measurement of the many non-mandatory species across the surveys delivering data primarily for stock assessment modelling it will likely prove difficult.

4.5 ALK borrowing

ALK allocation in DATRAS for missing ages in index calculation has been a topic of discussion amongst survey and assessment scientists for many years. The workshop reviewed different ways to present original and final ALK data when survey indices are downloaded. It was agreed that users should have ALKs for all 3 types of borrowing that could be (potentially) used:

- final ALK used, with field stating which roundfish areas were used for each roundfish area (e.g., area 1 ALK used data from areas 1, 2, 4);
- raw ALK for each roundfish area;
- Illustration of borrowing of length data from surrounding areas within a given age. See Annex 6

Annex 1: List of participants

Name	Institute	Country	Email
Dave Stokes (co-chair)	The Marine Institute, Galway	Ireland	david.stokes@marine.ie
Eoghan Kelly	The Marine Institute, Galway	Ireland	Eoghan.kelly@marine.ie
Anna Osypchuk	ICES Secretariat	Denmark	anna.osypchuk@ices.dk
Vaishav Soni (co-chair)	ICES Secretariat	Denmark	Vaishav@ices.dk
Trine Haugen	Institute of Marine Research	Norway	trine.haugen@hi.no
Loes Vandecasteele	ILVO	Belgium	loes.vandecasteele@ilvo.vlaanderen.be
Jennifer Devine	Institute of Marine Research	Ireland	jennifer.devine@imr.no
Gary Burt	CEFAS	UK-England	gary.burt@cefas.co.uk
Adriana Villamor	ICES secretariat	Denmark	Adriana.villamor@ices.dk

Annex 2: Agenda

Monday, 11/6				
10:00		Start, setting-up IT	Plenary	
10:30		Welcome and housekeeping	Introduction:	Dave,Vaishav
		Adoption of agenda		
		Introductory information and initial discussion on Country submission issue Personal goal and expectaiton		
11:00	COFFEE			
11:30		Data submission issue by countries Norway Belgium Ireland		Jennifer Loes Dave
13:00	LUNCH			
14:00		DATRAS Exchange data in details	Presentation	Anna,Vaishav
14:30		Data preparation	Subgroups	
		Preparation of Exchange file	Lead:	Anna Vaishav
16:00	COFFEE			
16:30		Preparation of Exchange file... Continue	Subgroups Lead:	As above
17:30		Adjourn		

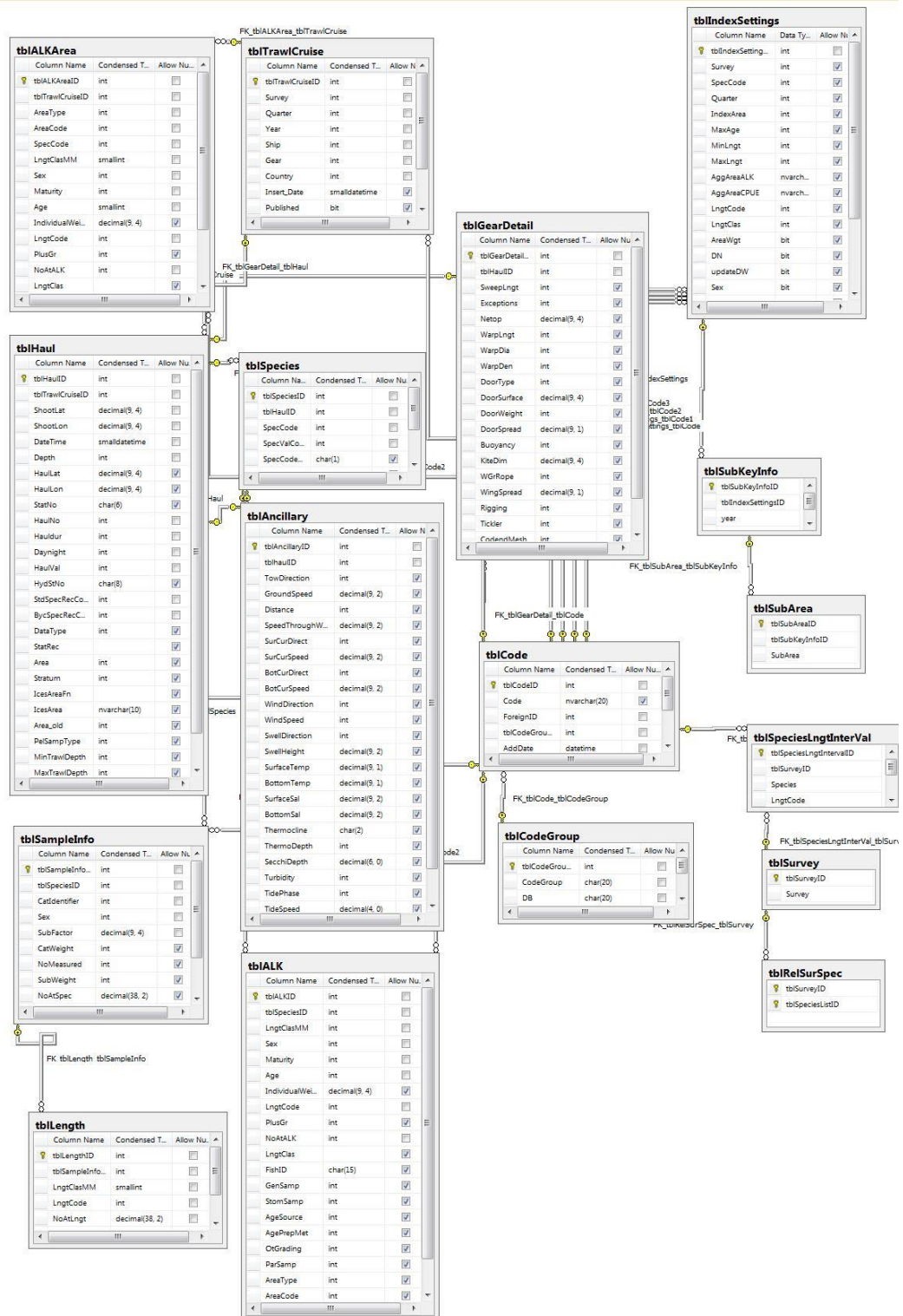
Monday, 12/6				
		Start, review	Plenary	
9:00		Progress summary	Tasks achieved and planned for sub groups	All participants
9:30		Topic Sub-groups		
		Introductory information and initial discussion on Country submission issue Personal goal and expectaiton		
11:00	COFFEE			
11:30		Data submission issue by countries Norway Belgium Ireland		Jennifer Loes Dave
13:00	LUNCH			
14:00		DATRAS Exchange data in de-tails	Presentation	Anna,Vaishav
14:30		Data preparation	Subgroups	
		Preparation of Exchange file	Lead:	Anna Vaishav
16:00	COFFEE			
16:30		Preparation of Exchange file... Continue	Subgroups Lead:	As above
17:30		Adjourn		

Annex 3: Recommendations

The general recommendations from the workshop are to implement the body of work highlighted by and for the national participants and ICES Data Centre and are therefore not addressed to other expert groups. There were a number of more general issues raised however these are summarised below including a suggested shortlist of relevant expert groups.

RECOMMENDATION	ADDRESSED TO
1. Create a list of minimum products for assessed species as output from a coordinated survey series. The product/format could be combined with the list of stocks using DATRAS data in their assessment as recommended by IBTS 2017.	IBTS, BTS, WGNSSK, WGCSE, ...
2. Highlight what outputs might be required from DATRAS for Western Area combined indices following Workshop/Benchmark review. Also other survey related data analysis such as calibration studies or non standard surveys where one vessel may fill in for another.	WGISDAA, WKESIG, WKUSER, ICES-DC
3. Agree on approach to by-catch species, in particular benthic animals (taxonomic level/grouping, species validity, count/weight/observation data?).	IBTS, BTS, BEWG
4. Reciprocal mapping reference from EUROGOOS db to DATRAS	IBTS, BTS, Data Centre
5. Inclusion of DATRAS surveys in official data calls	ACOM, ICES Data Centre
6. Additional fields in HH, HL and CA data need to be discussed and implemented in near future. These new fields need to be adopted by all surveys generic exchange data format for Trawl Survey data in DATRAS (e.g Length measurement type, Sample processing method)	DATRAS governance group, WGDGG

Annex 4: DATRAS Relational Database Model Schema and unique constrains of exchange data in HH HL and CA



	HH_Unique_Key		HL_Unique_Key		CA_Unique_Key
1	RecordType	1	RecordType	1	RecordType
2	Quarter	2	Quarter	2	Quarter
3	Country	3	Country	3	Country
4	Ship	4	Ship	4	Ship
5	Gear	5	Gear	5	Gear
6	SweepLngt	6	SweepLngt	6	SweepLngt
7	GearExp	7	GearExp	7	GearExp
8	DoorType	8	DoorType	8	DoorType
9	StNo	9	StNo	9	StNo
10	HaulNo	10	HaulNo	10	HaulNo
11	Year	11	Year	11	Year
12	Month	12	SpecCodeType	12	SpecCodeType
13	Day	13	SpecCode	13	SpecCode
14	TimeShot	14	SpecVal	14	AreaType
15	Stratum	15	Sex	15	AreaCode
16	HaulDur	16	TotalNo	16	LngtCode
17	DayNight	17	CatIdentifier	17	LngtClass
18	ShootLat	18	NoMeas	18	Sex
19	ShootLong	19	SubFactor	19	Maturity
20	HaulLat	20	SubWgt	20	PlusGr
21	HaulLong	21	CatCatchWgt	21	AgeRings
22	StatRec	22	LngtCode	22	CANoAtLngt
23	Depth	23	LngtClass	23	IndWgt
24	HaulVal	24	HLNoAtLngt		
25	HydroStNo				
26	StdSpecRecCode				
27	BycSpecRecCode				
28	DataType				
29	Netopening				
30	Rigging				
31	Tickler				
32	Distance				
33	WarpLngt				
34	Warpdia				
35	WarpDen				
36	DoorSurface				
37	DoorWgt				
38	DoorSpread				
39	WingSpread				
40	Buoyancy				
41	KiteDim				
42	WgtGroundRope				
43	TowDir				
44	GroundSpeed				
45	SpeedWater				
46	SurCurDir				
47	SurCurSpeed				
48	BotCurDir				
49	BotCurSpeed				
50	WindDir				
51	WindSpeed				
52	SwellDir				
53	SwellHeight				
54	SurTemp				
55	BotTemp				
56	SurSal				
57	BotSal				
58	ThermoCline				
59	ThClineDepth				

Annex 5: The IBTS ALK substitution procedure

This has in the past been done manually by Henrik Sparholt.

The procedure is as given below:

1. A table with no. of otoliths by species and Roundfish (RF) area is inspected to see which cells (i.e. species * RF) are missing or has below 25 otoliths. If zero otoliths → there will always be a substitution with data from other RFs, following the procedure given below.
2. The age data table mentioned above is compared with a table with no./hr of fish caught, by species and RF. If the number of otolith for a given cell is low, but the number caught is high special care has to taken to check whether the given otolith data are sufficiently good, i.e. spread out over length and ages and extensive enough. If not a substitution is made.
3. All ALKs are inspected manually and if there are only a few age groups represented or the length range is limited a substitution is made.
4. A substitution is actually a supplementation. Age data from neighbouring RF's are used. Often only one or a few with the largest set of age data are used.
5. For saithe all age data are merged and applied to all RFs.
6. For mackerel often the situation is the same as for saithe.
7. The manual inspection of data is also a quality control procedure where peculiarities are spotted. Often individual outliers or bulks of data not consistent with the rest of data are spotted. These data are then looked into and send back to the country of origin for checking. Manual inspection of the alk data is often a way of identifying missing data submission from a country, appearing as some RFs in the otolith table without age data.

The substitution can probably be simplified by deciding that a substitution for a given RF is made by age data from all neighbouring RFs. This will mean that:

RF 1	will be supplemented by data from	RF 2, 3
RF2	"	RF 1, 3, 4, 6, 7
RF3	"	RF 1, 2, 4
RF4	"	RF 2, 3, 5, 6
RF5	"	RF 4, 6
RF6	"	RF 2, 4, 5, 7
RF7	"	RF 2, 6
RF8	"	RF 7, 9
RF9	"	RF 8.

There are probably also some possibilities for using the age data table and the length data tables mentioned above to come up with algorithms like "if No/hr is larger than X% of total and number of age data points are less than Y then do this and that substitutions".